ZR | Zoological Research

Article

Open Access

Chromosome-level genome assembly of the glass catfish (*Kryptopterus vitreolus*) reveals molecular clues to its transparent phenotype

Chao Bian[1,2,3,*], Rui-Han Li[3,4], Zhi-Qiang Ruan[3], Wei-Ting Chen[5], Yu Huang[3], Li-Yue Liu[6], Hong-Ling Zhou[7], Cheong-Meng Chong[8], Xi-Dong Mu[2,*], Qiong Shi[1,3,9,*]

[1] *Laboratory of Aquatic Genomics, College of Life Sciences and Oceanography, Shenzhen University, Shenzhen, Guangdong 518057, China*

[2] *Key Laboratory of Prevention and Control for Aquatic Invasive Alien Species, Ministry of Agriculture and Rural Affairs, Guangdong Modern. Recreational Fisheries Engineering Technology Center, Pearl River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Guangzhou, Guangdong 510380, China*

[3] *Shenzhen Key Lab of Marine Genomics, Guangdong Provincial Key Lab of Molecular Breeding in Marine Economic Animals, BGI Academy of Marine Sciences, BGI Marine, Shenzhen, Guangdong 518081, China*

[4] *CAS Key Laboratory of Mountain Ecological Restoration and Bioresource Utilization & Ecological Restoration and Biodiversity Conservation Key Laboratory of Sichuan Province, Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu, Sichuan 610041, China*

[5] *School of Life Sciences, Jiaying University, Meizhou, Guangdong 514015, China*

[6] *China Zebrafish Resource Center, National Aquatic Biological Resource Center, State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Science, Wuhan, Hubei 430072, China*

[7] *Shenzhen 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, Guangdong 518116, China*

[8] *State Key Laboratory of Quality Research in Chinese Medicine, Institute of Chinese Medical Sciences, University of Macau, Macao, Macau 999078, China*

[9] *Center for Aquatic Genomics, College of Life Sciences, Neijiang Normal University, Neijiang, Sichuan 641100, China*

ABSTRACT

Glass catfish (*Kryptopterus vitreolus*) are notable in the aquarium trade for their highly transparent body pattern. This transparency is due to the loss of most reflective iridophores and light-absorbing melanophores in the main body, although certain black and silver pigments remain in the face and head. To date, however, the molecular mechanisms underlying this transparent phenotype remain largely unknown. To explore the genetic basis of this transparency, we constructed a chromosome-level haplotypic genome assembly for the glass catfish, encompassing 32 chromosomes and 23 344 protein-coding genes, using PacBio and Hi-C sequencing technologies and standard assembly and annotation pipelines. Analysis revealed a premature stop codon in the putative albinism-related *tyrp1b* gene, encoding tyrosinase-related protein 1, rendering it a nonfunctional pseudogene. Notably, a synteny comparison with over 30 other fish species identified the loss of the endothelin-3 (*edn3b*) gene in the glass catfish genome. To investigate the role of *edn3b*, we generated *edn3b*$^{-/-}$ mutant zebrafish, which exhibited a remarkable reduction in black pigments in body surface stripes compared to wild-type zebrafish. These findings indicate that *edn3b* loss contributes to the transparent phenotype of the glass catfish. Our high-quality

chromosome-scale genome assembly and identification of key genes provide important molecular insights into the transparent phenotype of glass catfish. These findings not only enhance our understanding of the molecular mechanisms underlying transparency in glass catfish, but also offer a valuable genetic resource for further research on pigmentation in various animal species.

Keywords: Glass catfish; Whole-genome sequencing; *edn3b*$^{-/-}$ mutant; Transparent phenotype

INTRODUCTION

Kryptopterus vitreolus (NCBI Taxonomy ID: 2012678) is a freshwater fish species in the family Siluridae, order Siluriformes. Commonly referred to as the "glass catfish" or "ghost catfish", this species is distinguished by its highly transparent body, which exposes its internal organs and intermuscular bones (Figure 1A). The fish retains minimal pigmentation, with small amounts of black, silver, and yellow pigments concentrated in a narrow zone along the dorsal midline and head. This transparency provides a significant adaptive advantage by rendering the fish virtually invisible to predators (Ng & Kottelat, 2013).

Glass catfish are native to river drainages in the peninsular and southeastern waters of Thailand (Ng & Kottelat, 2013). Its natural transparency has made the species a popular choice for aquariums and an important research model for pigmentation (Hayashi & Fujii, 1994, 2001), cell transplantation and isolation (Han et al., 2011), DNA immunization (Dijkstra et al., 2001), and vascular system studies (Dahl Ejby Jensen et al., 2009; Rummer et al., 2014). Nevertheless, the species has often been misidentified as either *K. minor* or, more frequently, *K. bicirrhis*. Ng & Kottelat (2013) resolved this long-standing issue by classifying the glass catfish as a distinct species, *K. vitreolus*. Their study also redefined the genus *Kryptopterus*, emphasizing the unique transparency and small size of the glass catfish compared to the larger and more opaque *K. bicirrhis* (Ng & Kottelat, 2013).

To the best of our knowledge, no genome sequences of the glass catfish have been published. Therefore, developing genomic resources for this species is essential to provide molecular insights into its transparent phenotype. In this study, we sequenced the genome and transcriptome of the glass catfish for the first time. We identified the endothelin-3 (*edn3b*) gene and conducted knockout experiments to investigate its role in transparency. The phenotype of *edn3b*$^{-/-}$ mutants was consistent with previous findings (Spiewak et al., 2018). These data provide valuable insights for comparative genomic analyses aimed at predicting evolutionary convergence in transparent fish species, including *casper* zebrafish (Bian et al., 2020) and icefish (Liu et al., 2017; Zhang et al., 2020). This research may help inform molecular breeding programs for the development of ornamental fish varieties with transparent body patterns.
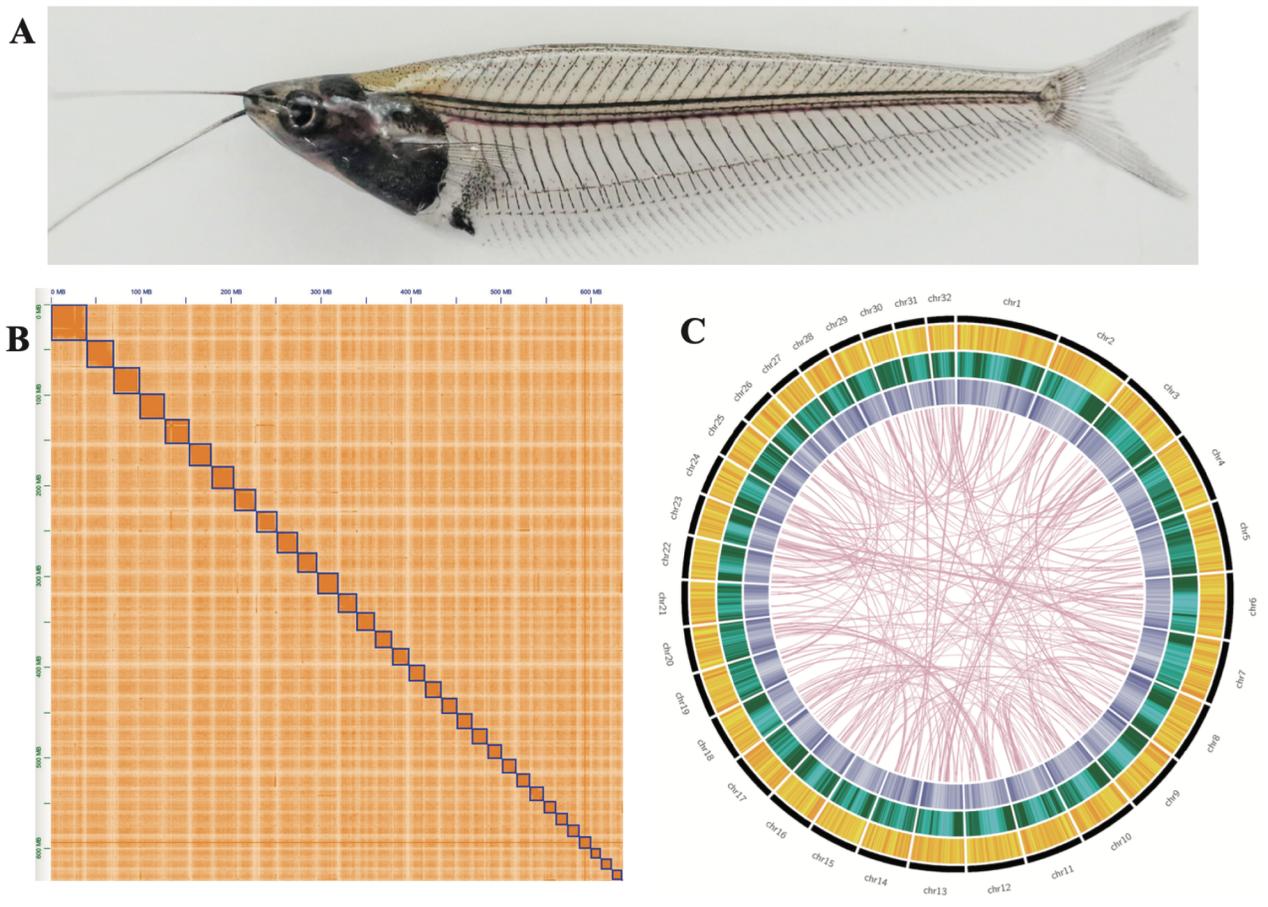
Figure 1 Overview of sequenced glass catfish

A: Image of glass catfish. B: Hi-C heatmap of chromosomal interactions. C: Circos plot of genome. Data from outside to inside include 32 anchored chromosomes, gene density, repeats, GC content, and synteny blocks.

MATERIALS AND METHODS

Sample preparation, ethics statement, and whole-genome sequencing
A female glass catfish (Figure 1A) (2.0 g) was collected from a local market in Guangzhou City, Guangdong Province, China. Genomic DNA was extracted from the muscle tissue using a genomic DNA isolation kit (Qiagen, Germany). All animal experiments were approved by the Institutional Review Board on Bioethics and Biosafety of BGI (approval ID: FT18134).

A comprehensive strategy for whole-genome shotgun sequencing was implemented. Seven Illumina paired-end libraries were constructed, including three short-insert libraries (270, 500, and 800 bp) and four long-insert libraries (2, 5, 10, and 20 kb), for sequencing using the Illumina Hiseq Xten platform (Illumina, USA). Low-quality raw reads (with more than 10 Ns or over 20% low-quality bases), polymerase chain reaction (PCR)-duplicated reads, and adaptor sequences were filtered to obtain clean data using SOAPnuke v.1.5 (optimized parameters: filter -1 fq1 -2 fq2 -M 1 -A 0.5 -l 5 -q 0.5 -n 0.05 -d -S -O 10 -P 0.1 -Q 1 -C cleanFq1 -D cleanFq2) (Chen et al., 2018).

For PacBio sequencing, high-quality genomic DNA was extracted using AMPure PB magnetic beads (Pacific Biosciences, USA) for subsequent library construction with the SMRTbell template prep kit 2.1 (Pacific Biosciences, USA). Long reads were sequenced on a PacBio Sequel platform (Pacific Biosciences, USA).

High-throughput chromosome conformation capture (Hi-C) experiments were conducted to generate Hi-C reads to anchoring chromosomes of the assembled glass catfish genome. DNA extraction, chromatin digestion, and proximity-ligation experiments were carried out in accordance with previous research (Zhu et al., 2017). The restriction enzyme MboI was used to digest DNA, and the constructed Hi-C library was sequenced on the Illumina Xten platform (Illumina, USA) with 2×150 bp paired-end reads.

Genome size estimation by a *k*-mer analysis
To predict the genome size of the glass catfish, routine *k*-mer analysis of frequency distribution was performed (You et al., 2014). Approximately 30-fold clean reads were collected for 17-mer analysis using kmerfreq v.1.0 (Wang et al., 2020) with optimized parameters (kmerfreq -k 17 -l fq.list >17mer.freq 2>17mer.log).

Genome assembly, chromosome construction, and completeness evaluation
We applied an integrated approach utilizing two sequencing technologies: PacBio long reads for *de novo* genome assembly and clean Illumina short reads for correcting sequencing errors in the PacBio reads. Specifically, Platanus v.1.2.4 (Kajitani et al., 2014) was employed to assemble the paired-end Illumina reads into preliminary contigs. These contigs, along with the corrected PacBio reads, were used as input data for DBG2OLC (Ye et al., 2016) to assemble super contigs with optimized parameters (DBG2OLC LD 0 Contigs contig.fa k 17 KmerCovTh 5 MinOverlap 10 AdaptiveTh 0.001 f super_contig.fa). Illumina paired-end reads from short-insert libraries were aligned to the super contigs using BWA-MEM (Li, 2014). Pilon v.1.24 (Walker et al., 2014) was applied to polish the super contigs based on the BWA alignments. The polished assembly was then elongated with PacBio reads using SSPACE-LongReads v.1.1 (Boetzer & Pirovano, 2014).

The completeness of the glass catfish genome assembly was elevated using Benchmarking Universal Single Copy Ortholog (BUSCO v.5.2.2) (Simão et al., 2015) with default parameters (-l actinopterygii_odb10 -m genome -c 3).

High-quality Hi-C reads were subsequently employed to construct a chromosome-level genome assembly. Putative chromosomes were clustered, ordered, and assigned orientations using Juicer v.1.5 (Durand et al., 2016) and 3D-DNA with optimized parameters (-m haploid -s 4 -c 30) (Dudchenko et al., 2017). Interaction frequency analysis was conducted through a routine segmentation-allowed method, in which all scaffolds were clustered and reassembled using the clean Hi-C data. Based on cross-linking strength, the arrangements and orientations of the initial chromosomal scaffolds were anchored. Overlaps among the initial chromosomal scaffolds were merged and linked, resulting in the final putative chromosomes.

Annotation of repeat sequences and gene structures
Tandem Repeat Finder v.4.07 (Benson, 1999) with core parameters "Match=2, Mismatch=7, Delta=7, PM=80, PI=10, Minscore=50, and MaxPerid=2000" was utilized to identify tandem repeats. A *de novo* repeat library was constructed using RepeatModeler v.1.0.8 (Chen, 2004) and LTR_harvest (Ellinghaus et al., 2008), which was subsequently used to identify repeat sequences with RepeatMasker (Tarailo-Graovac & Chen, 2009). Transposable elements (TEs) were identified using the RepBase library (RepBase v.21.01) (Jurka et al., 2005) with RepeatModeler (Chen, 2004).

Gene structures were annotated using homology-based and transcriptome-derived predictions. For homology-based annotation, protein sequences from seven representative fish species, including zebrafish (*Danio rerio*), threespine stickleback (*Gasterosteus aculeatus*), barramundi (*Lates calcarifer*), spotted gar (*Lepisosteus oculatus*), Japanese medaka (*Oryzias latipes*), Japanese puffer (*Takifugu rubripes*), and green spotted puffer (*Tetraodon nigroviridis*), were downloaded from Ensemble (release 87) for subsequent alignment to the assembled glass catfish genome using TBLASTn (Kent, 2002) with an e-value≤1e–5. Genewise v.2.2.0 (Birney et al., 2004) was then employed to predict gene structures in the alignments.

For transcriptome-based annotation, HISAT2 v.2.21 (Kim et al., 2015) was used to map the transcriptome sequencing (RNA-seq) reads onto the assembled glass catfish genome. Alignments were sorted using SAMtools (Li et al., 2009) and potential gene structures were predicted using Cufflink (Trapnell et al., 2013). The results from both annotation strategies were merged with MAKER (Cantarel et al., 2008). Functional annotation was performed by aligning all deduced protein sequences from the gene set against four public databases, including TrEMBL (Boeckmann et al., 2003), Swiss-Prot (Boeckmann et al., 2003), KEGG (Kanehisa et al., 2017), and InterPro (Hunter et al., 2009), using BLASTp (McGinnis & Madden, 2004) with an e-value≤1e-5.

Gene family clustering and phylogenetic analysis
Genome and protein sequences of 12 representative species, including Chinese large-mouth catfish (*Silurus meridionalis*), black bullhead catfish (*Ameiurus melas*), Asian red-tailed catfish (*Hemibagrus wyckioides*), channel catfish (*Ictalurus punctatus*), blue catfish (*I. furcatus)*, striped catfish (*Pangasianodon hypophthalmus*), yellow catfish (*Tachysurus fulvidraco*), North African catfish (*Clarias gariepinus*), shark

catfish (*Pangasius djambal*), armorhead catfish (*Cranoglanis bouderius*), striped eel catfish (*Plotosus lineatus*), zebrafish, Asian arowana (*Scleropages formosus*), and Japanese medaka, were downloaded from the NCBI database. BLASTp (McGinnis & Madden, 2004) (e-value≤1e-5) was utilized to align all protein sequences, including the protein set of the glass catfish. Gene families were then clustered using OrthoMCL v.2.0.9 (Fischer et al., 2011) with default parameters. Subsequently, coding sequence (CDS) regions of all single-copy genes in each species were concatenated and subjected to multiple sequence alignment using MUSCLE v.3.8.31 (Edgar, 2004). Finally, PhyML v.3.0 (Guindon et al., 2010) was applied to construct a phylogenetic tree with the maximum-likelihood method.

Divergence time analysis was performed using MCMCtree in the PAML package (Yang & Rannala, 2006). Putative calibration times were adopted from TIMETREE (Kumar et al., 2017). Expanded and contracted gene families were predicted using CAFE v.4.2.1 (De Bie et al., 2006), which defined gene family expansion and contraction by comparing gene family sizes between ancestors and each extant species through a random birth and death process model to identify gene gain and loss along each lineage. Gene families were excluded they contained ≥200 genes in one species or ≤2 genes in all other species. The expanded or contracted gene families were separately enriched by mapping to the Gene Ontology (GO) (Gene Ontology Consortium, 2004) and KEGG pathway (Kanehisa & Goto, 2000) databases.

Reconstruction of ancestral chromosomes from representative catfish species

The reconstruction of ancestral chromosomes was based on 13 pairs of potential ancestral chromosomes (Chra~m), predicted as the original chromosome karyotype of ancestral teleosts from previous studies (Bian et al., 2016; Kasahara et al., 2007; Lu et al., 2022). Protein sets from 10 representative catfish species with diverse chromosome numbers were collected for this reconstruction. The protein set of each species was aligned to the protein set of the putative ancestral teleost using BLASTp (e-value≤1e-10). Best-hit alignments were then identified, and alignments with fewer than 20 hits per chromosome were filtered out. Finally, chromosome fissions, fusions, and translocations were identified in each catfish species and the karyotype of ancestral catfish was predicted. All karyotypes were displayed using SVG in Perl.

Prediction of genes related to melanophore and iridophore development

Genes involved in melanophore and iridophore development were predicted by consulting various published studies on melanophore and iridophore synthesis pathways (Lin and Fisher, 2007; Sprague et al., 2003). Relevant genes and corresponding protein sequences for zebrafish and medaka were downloaded from the Ensembl and NCBI databases. These protein sequences were aligned to the assembled glass catfish genome using tBLASTn v.2.2.28 (Mount, 2007) with an e-value threshold of 10^{-5}. Alignments were further filtered and processed using a Perl script to retain best-hit alignments with putative target genes showing over 50% identity and alignment ratio. GeneWise v.2.2.0 (Birney et al., 2004) was then employed to predict target genes among the best-hit alignments from the glass catfish genome. In cases where gene loss was identified, corresponding tBLASTn alignments

were manually validated.

Gene knockout with CRISPR-Cas9

The coding sequence of the zebrafish *edn3b* gene (Ensembl ID: ENSDARG00000086669), which contains four exons and three introns, was downloaded from Ensembl. A gRNA targeting the second exon was designed with the sequence GGAACCACCAGAGAAGAC, confirmed to be unique in the zebrafish genome for subsequent amplification from the wild-type zebrafish. This gRNA, along with *Cas9* mRNA, was synthesized *in vitro* for co-injection into one-cell stage zebrafish embryos. PCR was conducted to screen the P0 embryonic population for targeted mutations. Mosaic P0 zebrafish were crossed with wild-type zebrafish to obtain the F1 generation, which was screened to verify multiple mutation types. Selected F1 individuals were cultured separately according to their mutation types. Once the F1 generation reached sexually maturity, the F2 generation were produced by inbreeding within mutation types. Recovered mosaics (*edn3b*$^{-/-}$) were identified by sequencing the genomic DNA of examined fish, yielding a mutation ratio of 30.3% (33/109). The primers used were F-CATTTTCACCGTGGATACATC GTTTT and P-TTTGGGCATCATCTGGATCAACACCC.

RESULTS

Summary of genome assembly and annotation

Approximately 269.4 Gb of Illumina raw reads from seven libraries were generated. After a series of filtering steps, 182.7 Gb of clean reads were obtained (Supplementary Table S1). In addition, 15.1 Gb of raw reads were produced using the PacBio platform (Supplementary Table S2). The glass catfish genome was predicted to be 642.7 Mb in size based on Illumina short reads (from the 500 bp library) and 17-mer frequency distribution analysis (Genome Size=k-mer_num/k-mer_depth; see Supplementary Table S3). The core peak of the k-mer depth was 49, with a total 17-mer number of 31 494 249 372 (Supplementary Figure S1).

After assembling the combined Illumina and PacBio reads, a *de novo* genome assembly of 652.4 Mb was obtained, with a contig N50 of 1.2 Mb (Supplementary Table S4). The BUSCO results showed 90.3% completeness (Supplementary Table S5), demonstrating that the glass catfish genome assembly covered most genetic regions with relatively high completeness.

For chromosome construction, 146.6 Gb of raw reads from the Hi-C library were sequenced. After filtering, 137.3 Gb of clean data were generated for the HiC-Pro pipeline. Finally, 634.7 Mb of contigs were anchored to 32 chromosomes (Chr; Figure 1B; Supplementary Table S6), accounting for 97.1% of the total assembled genome. The scaffold N50 value of this chromosome-level genome assembly was 22.5 Mb (Figure 1C).

For further annotations, the integrated repeat sequences were predicted to cover approximately 35.2% of the *de novo* genome assembly, totaling 223.4 Mb (Supplementary Table S7). Among these, DNA repeat sequences were the predominant type, accounting for 25.5% of the total repeats. Long interspersed nuclear element (LINE), short terminal repeated element (SINE), and long terminal repeat (LTR) sequences accounted for 5.2%, 7.1%, and 7.0% of the total repeats, respectively. For the prediction of gene structures, seven homolog-based annotation results (16 442–33 943

genes) and transcriptome annotation results (21 375 genes; Supplementary Table S8) were merged into a non-redundant set of 23 344 genes. These genes contained an average of 9.7 exons, each with an average length of approximately 175.0 bp. Within the final gene set of the glass catfish, 22 537 genes (96.5%, Supplementary Table S9) exhibited annotated functions with at least one hit from the searched databases, including InterPro, KEGG, SwissProt, and TrEMBL.

Evolutionary characterization of glass catfish

To determine the evolutionary relationship among glass catfish and other catfish species, a phylogenetic tree was constructed using 12 representative species. A total of 25 404 orthologous gene families were identified from 15 representative species, including glass catfish (Supplementary Table S10), with 2 904 single-copy orthologs genes used to establish the phylogenetic topology and divergence time tree (Figure 2).

Based on the maximum-likelihood phylogenetic tree, the 12 representative catfish species diverged from zebrafish approximately 173.5 million years ago (Ma), which is later than the divergence time reported in the blue catfish genome study (Wang et al., 2022). These species were divided into two main clades (Figure 2). The glass catfish showed the closest relationship to the Chinese large-mouth catfish, belonging to the family of Siluridae, with a predicted divergence time of approximately 42.3 Ma.

Furthermore, 277 expanded and 1 034 contracted gene families were identified in the glass catfish genome (Figure 2). GO and KEGG enrichment analyses were performed on these gene families. The top enriched GO terms in both expanded and contracted gene families included "cellular process", "biological regulation", and "cell" (Supplementary Figure S2). Compared to other catfish species, molecular cellular process genes have undergone significant changes in the glass catfish during evolution. The top enriched KEGG pathways in both

the expanded and contracted gene families were "Transport and catabolism" and "Cell growth and death" (Supplementary Figure S3).

**Reconstruction of ancestral karyotype for various catfish species**

Catfish (Siluriformes) constitute a large and diverse order of teleosts, comprising 36 families primarily found in freshwater habitats (Ferraris, 2007). Despite limited information on the chromosomal evolution of catfish, the chromosome karyotypes of various catfish species exhibit a range of diploid numbers ($2n$) from $2n=40$ to $2n=62$ (de MC Sassi et al., 2020; Legrande, 1981). Genome and protein sequences from 10 representative catfish species were collected to reconstruct their ancestral karyotype. Over 10 000 best-hit gene pairs were identified between each catfish and the teleost ancestor protein set (Supplementary Table S11). The reconstruction yielded 29 ancestral chromosomes (Figure 3A) with nine major rearrangements, including five fissions, three fusions, and a complex rearrangement involving both fissions and translocations (the translocations of three original chromosomes to five rearranged chromosomes are marked in red boxes in Figure 3A). Additionally, some lineage-specific small-scale translocations were identified in the catfish ancestor. These observations suggest that the catfish ancestor may have experienced relatively complex inter-chromosome rearrangements after diverging from the teleost ancestor. Some of these rearrangements were similar to those observed in zebrafish (Bian et al., 2016), implying that both Siluriformes and Cypriniformes may share similar chromosomal rearrangements.

Chromosome numbers of the 10 representative catfish species with chromosome-level genomes ranged from 26 to 32 (see Supplementary Table S12). Detailed rearrangements for each species are marked by arrows in Figure 3C–L. Two different translocations were observed in the five fish species
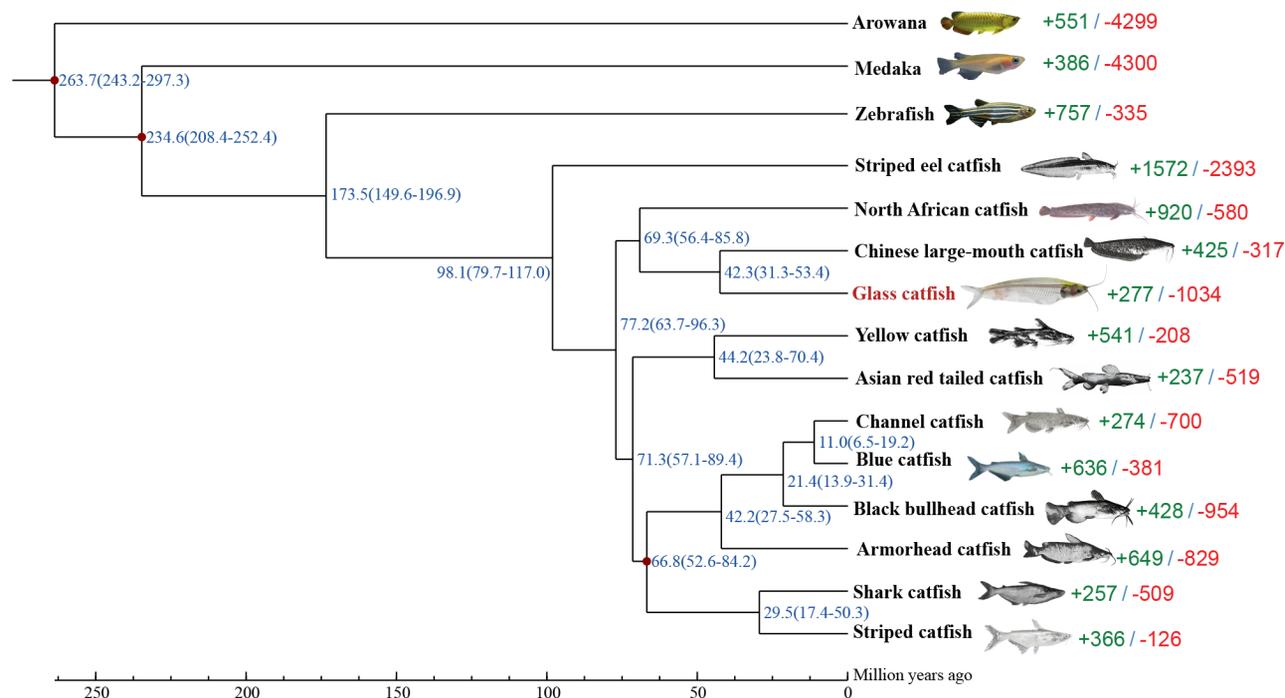
Figure 2 Phylogenetic tree of glass catfish and 14 representative fish species

Divergence times were estimated (million years ago, Ma) and are marked in blue. Expanded and contracted gene families are in green and red, respectively. Red points represent putative calibration times.
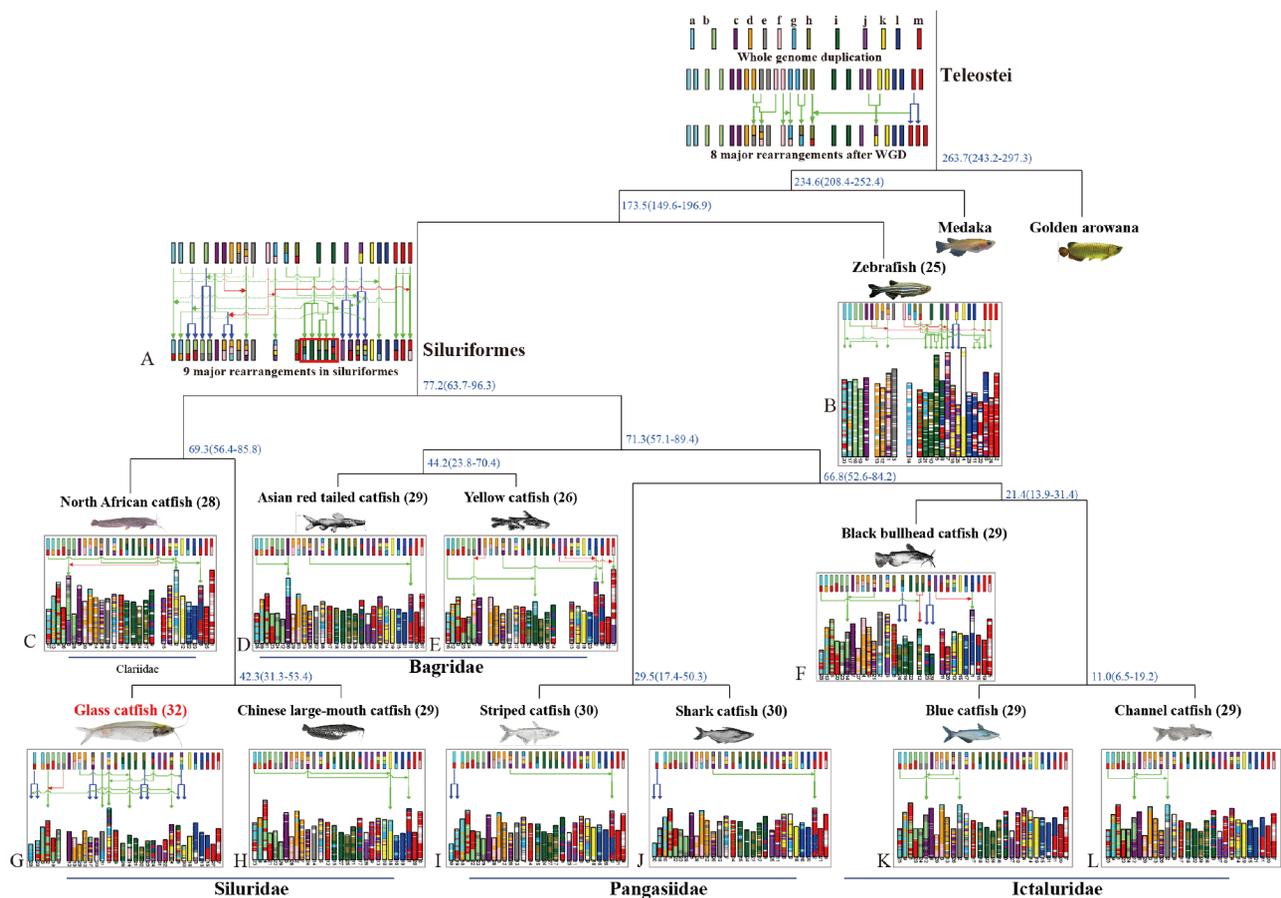
Figure 3 Ancestral karyotype and chromosomal evolution of various catfish species

Thirteen color bars represent ancestor chromosomes a–m. Blue, red, and green arrows represent fission, fusion, and translocation events, respectively. A: Predicted ancestral karyotype of Siluriformes. B: Chromosome karyotypes of zebrafish. C–L: Chromosome karyotypes of 10 examined catfish species.

with 29 chromosomes (Figure 3D, F, H, K, L), which were closest to the ancestral catfish karyotype, although the black bullhead catfish exhibited two extra fissions and two fusions (Figure 3F).

The glass catfish contained the highest haplotypic chromosome number (32) among the representative catfish species. Notably, compared to other closely related species, the glass catfish has undergone specific chromosomal rearrangements, including three fissions, one fusion, and five translocations (Figure 3G). Additionally, a common fission and a similar translocation were also observed in the glass catfish, Chinese large-mouth catfish, and two Pangasiidae catfish species (Figure 3G–J).

Species from the orders Siluriformes and Pangasiidae appear to have originated with an ancestral haplotypic chromosome number of 29. These species then underwent extensive chromosomal rearrangements during evolution. However, some sublineages developed 30 chromosomes through fissions, while others evolved reduced chromosome karyotypes by continuous chromosomal fusions. These findings are consistent with previous reports (De Oliveira et al., 2009; Legrande, 1981). Moreover, three active chromosomal segments were identified as the most unstable, including the light blue segment of Chr30, light purple segment of Chr25, and gray & dark purple segment of Chr2 in striped catfish (see Figure 3I). Two Bagridae catfish species (yellow and North African catfish) with fewer chromosomes (26 and 28, respectively) were observed to have undergone one or

three fusions (Figure 3C, E). These large-scale chromosomal arrangements contribute to the high diversity in chromosome karyotypes observed among various catfish species.

Molecular clues to transparent phenotype of glass catfish and morphological changes in *edn3b* knockout zebrafish mutants

After collecting nucleotide and protein sequences related to melanophore and iridophore development in zebrafish and medaka (Supplementary Table S13), we identified their corresponding orthologous genes in the glass catfish genome. Detailed comparisons of the melanin synthesis pathway revealed that the overall gene copy numbers in the glass catfish were similar to those in zebrafish and medaka (Supplementary Tables S14, 15). However, the *tyrp1b* gene (encoding tyrosinase-related protein 1b) in the glass catfish contained a premature stop codon in its first exon (Figure 4), likely resulting in the gene becoming a nonfunctional pseudogene.

As the silver pigment synthesis pathway remains unclear, we compiled a list of iridophore development-related genes from several previous studies (Lin and Fisher, 2007; Sprague et al., 2003). Interestingly, the glass catfish genome contained most of these important genes, similar to zebrafish and medaka, except for *edn3b*. Genomic synteny comparisons demonstrated the loss of *edn3b* in glass catfish (top in Figure 5). Consequently, we generated a homozygous zebrafish mutant (*edn3b⁻/⁻*) using CRISPR/Cas9 knockout. The *edn3b⁻/⁻* mutant exhibited a remarkable reduction of
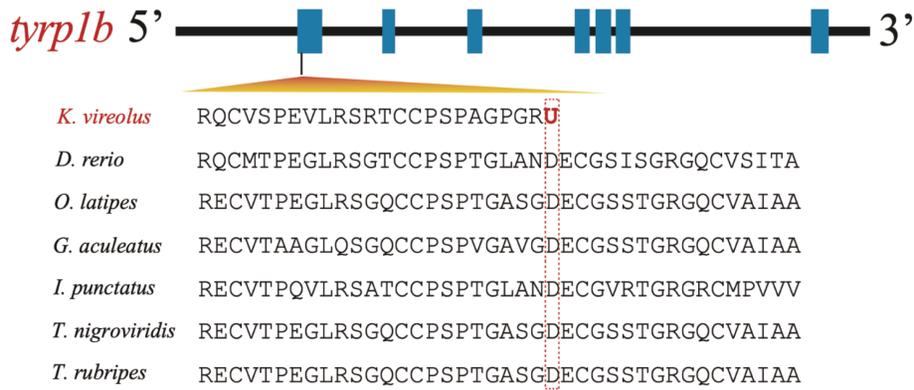
Figure 4 Multiple sequence alignments of *tyrp1b* in seven representative fish species

Nonsense mutation causing a premature stop codon in the first exon of the glass catfish *tyrp1b* is highlighted in brown.
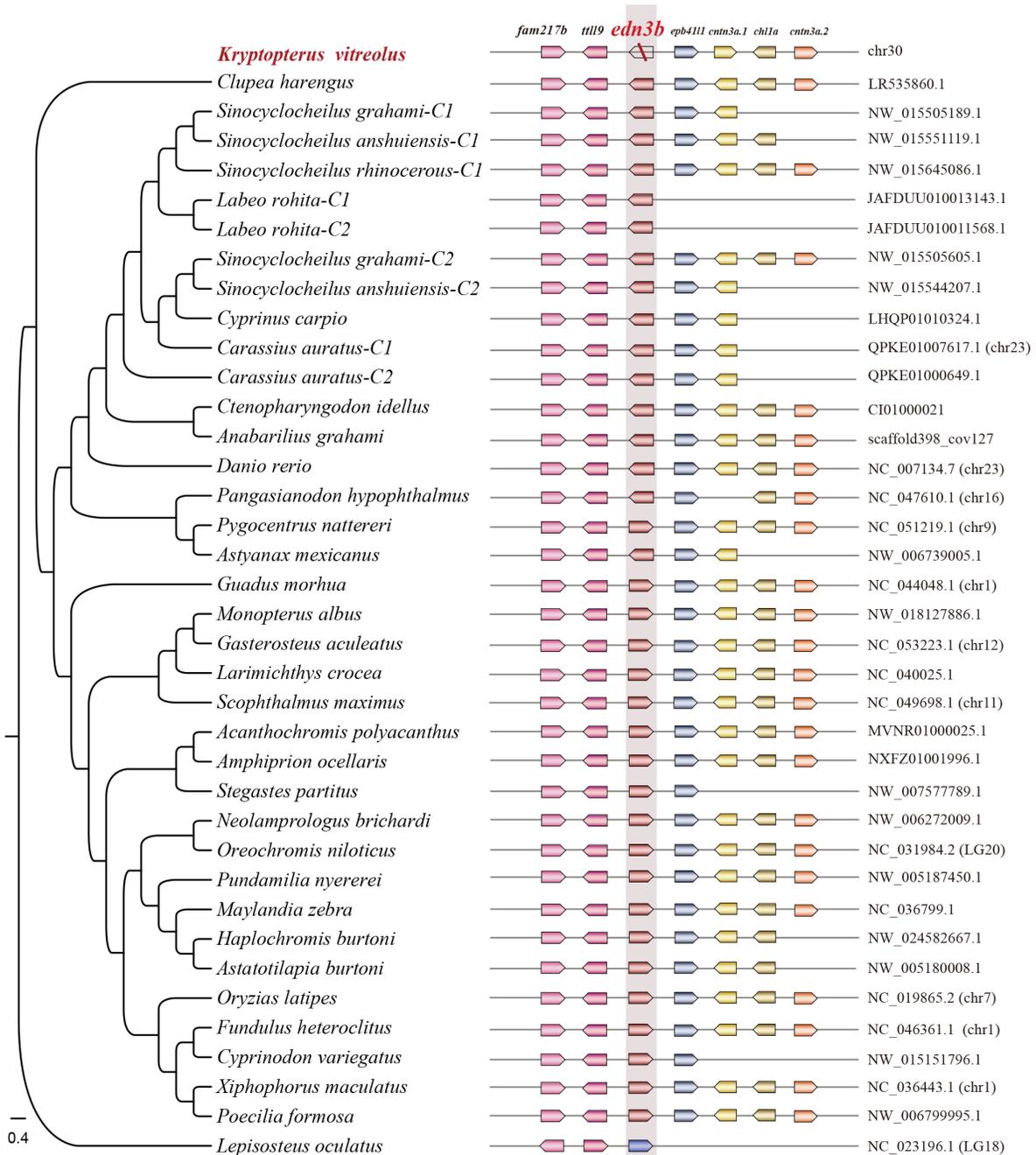
Figure 5 Gene tree (left) and synteny comparison of *edn3b* genes (right) in 33 representative species

First line of gene synteny block (right panel) illustrates loss of *edn3b* in glass catfish genome.

melanin and silver pigment in its stripes (Figure 6A) compared to wild-type zebrafish (Figure 6B). The stripe patterns in the mutants were notably different, breaking into circular spots or disappearing entirely (Figure 6A).

DISCUSSION

The most striking and unique feature of the glass catfish is its transparent coloration (Figure 1A). This species possesses small black eyes and two long barbels extending outward, with internal organs and even the spinal cord visible through its skin (Ng & Kottelat, 2013). The transparent phenotype facilitates predator avoidance (Ng & Kottelat, 2013). While the glass catfish has lost most black and silver pigmentation on its trunk, it has retained sparse pigmentation on its head. Interestingly, genomic comparisons and gene identification revealed that the glass catfish genome contains the majority of known black pigment-related genes.

However, the critical *tyrp1b* gene in the glass catfish harbors a nonsense mutation in its first exon (Figure 4B), resulting in a loss-of-function pseudogene. The *tyrp1b* gene, expressed in melanophores, plays a crucial role in melanocyte maintenance and proliferation (Braasch et al., 2009; Krauss et al., 2014b). Black eumelanin formation relies on the presence of *tyrp1*, a function likely conserved from the common ancestor of bony vertebrates (Braasch et al., 2009). In addition, *tyrp1a* exhibits significantly lower transcription levels in transparent muscle compared to colored muscle. Double knockdown of *tyrp1a* and *tyrp1b* in zebrafish mutants is reported to produce hypo-pigmented melanophores and brown coloration instead of black eumelanin (Braasch et al., 2009). Therefore, the loss-of-function mutation in *tyrp1b* may be a significant factor contributing to the melanin loss phenotype in glass catfish.

To further investigate the transparent phenotype of the glass catfish, we demonstrated loss of the *edn3b* gene (Figure 5) through synteny block analysis of glass catfish, zebrafish, medaka, and over 30 other species. The *edn3b* gene, identified as a ligand for the endothelin receptor signaling in larval iridophores, significantly reduces iridophore number when knocked down in zebrafish (Krauss et al., 2014a). Our *edn3b*$^{-/-}$ knockout mutant exhibited melanophore deficiencies, presenting with a spotted pattern rather than the wild-type stripes (Figure 6A). These findings confirmed that the *edn3b* gene plays a role in altering cellular coloration and stripe formation. Thus, loss of *edn3b* likely contributes to the transparent phenotype observed in glass catfish.

Our high-quality genome assembly and gene annotation of the glass catfish provide a valuable genetic resource for studying this transparent fish. We constructed a chromosome-level haplotypic genome assembly with 32 chromosomes and 23 344 protein-coding genes. A premature stop codon in the *tyrp1b* gene, potentially resulting in a nonfunctional pseudogene, was identified in the glass catfish genome. Additionally, genomic comparisons with over 30 other fish species confirmed the loss of the *edn3b* gene in the glass catfish. The generated *edn3b* knockout zebrafish mutant exhibited a marked reduction in black pigment in skin stripes, further supporting the role of *edn3b* in pigmentation. These findings not only enhance our understanding of the transcriptional regulation of body coloration and the transparent phenotype in the glass catfish but also provide a valuable genetic resource for in-depth investigations into pigmentation across various animal species.
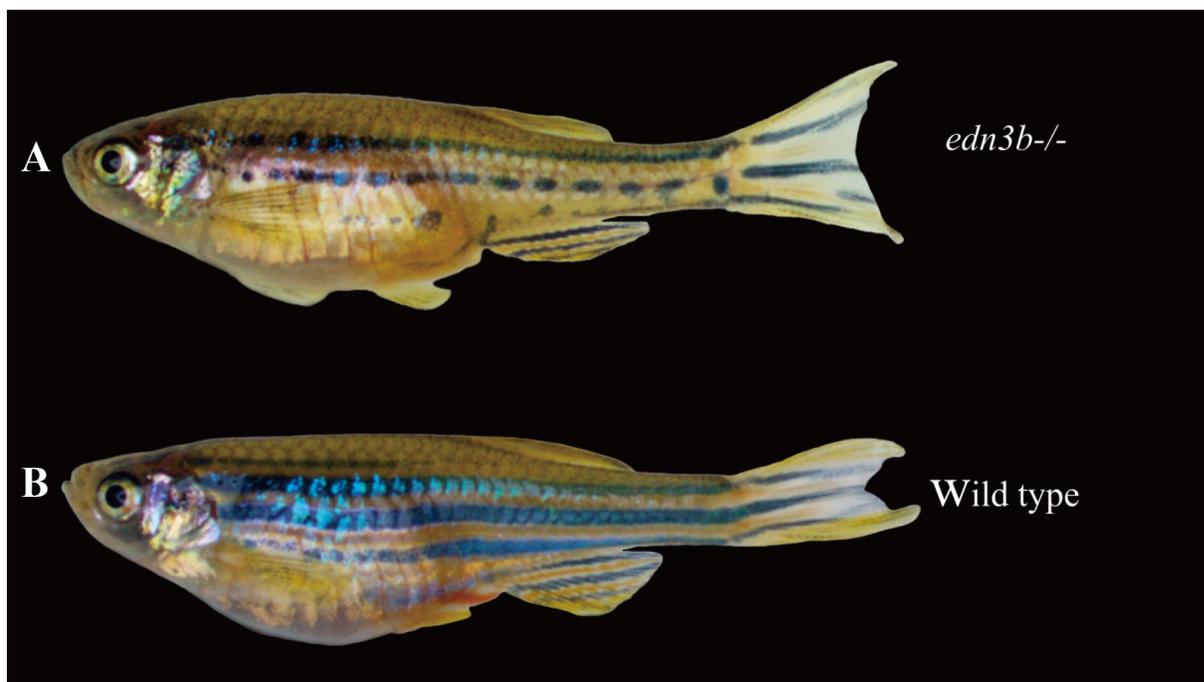
Figure 6 Phenotype of *edn3b*$^{-/-}$ zebrafish mutant

A: *edn3b*$^{-/-}$ mutant exhibits two or three scattered stripes with normal peritoneal iridophores. Bottom stripes show melanophore deficiencies, presenting with a spotted pattern. B: Wild-type zebrafish exhibits three or four melanophore stripes with sparse iridophores and light inter-stripes with abundant iridophores.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

Q.S., C.B., and X.D.M. conceived the project; C.B., R.H.L., and Y.H. performed data analysis; X.D.M. and Z.Q.R. collected samples; Z.Q.R., L.Y.L., and W.T.C. conducted the knockout and validation experiments. C.B., R.H.L., and Q.S. wrote the manuscript. Q.S., C.M.C., W.T.C., H.L.Z., and X.D.M. revised the manuscript. All authors read and approved the final version of the manuscript.

REFERENCES

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, **27**(2): 573–580.

Bian C, Chen WT, Ruan ZQ, et al. 2020. Genome and transcriptome sequencing of *casper* and *roy* zebrafish mutants provides novel genetic clues for iridophore loss. *International Journal of Molecular Sciences*, **21**(7): 2385.

Bian C, Hu YC, Ravi V, et al. 2016. The Asian arowana (*Scleropages formosus*) genome provides new insights into the evolution of an early lineage of teleosts. *Scientific Reports*, **6**: 24501.

Birney E, Clamp M, Durbin R. 2004. Genewise and genomewise. *Genome Research*, **14**(5): 988–995.

Boeckmann B, Bairoch A, Apweiler R, et al. 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Research*, **31**(1): 365–370.

Boetzer M, Pirovano W. 2014. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics*, **15**: 211.

Braasch I, Liedtke D, Volff JN, et al. 2009. Pigmentary function and evolution of *tyrp1* gene duplicates in fish. *Pigment Cell & Melanoma Research*, **22**(6): 839–850.

Cantarel BL, Korf I, Robb SM, et al. 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research*, **18**(1): 188–196.

Chen NS. 2004. Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*, doi: 10.1002/0471250953.bi0410s05.

Chen YX, Chen YS, Shi CM, et al. 2018. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *GigaScience*, **7**(1): gix120.

Dahl Ejby Jensen L, Cao RH, Hedlund EM, et al. 2009. Nitric oxide permits hypoxia-induced lymphatic perfusion by controlling arterial-lymphatic conduits in zebrafish and glass catfish. *Proceedings of the National Academy of Sciences of the United States of America*, **106**(43): 18408–18413.

De Bie T, Cristianini N, Demuth JP, et al. 2006. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics*, **22**(10): 1269–1271.

de MC Sassi F, Deon GA, Moreira-Filho O, et al. 2020. Multiple sex chromosomes and evolutionary relationships in amazonian catfishes: the outstanding model of the genus *harttia* (Siluriformes: Loricariidae). *Genes*, **11**(10): 1179.

De Oliveira RR, Feldberg E, Dos Anjos MB, et al. 2009. Mechanisms of chromosomal evolution and its possible relation to natural history characteristics in *Ancistrus* catfishes (Siluriformes: Loricariidae). *Journal of Fish Biology*, **75**(9): 2209–2225.

Dijkstra JM, Okamoto H, Ototake M, et al. 2001. Luciferase expression 2 years after DNA injection in glass catfish (*Kryptopterus bicirrhus*). *Fish & Shellfish Immunology*, **11**(2): 199–202.

Dudchenko O, Batra SS, Omer AD, et al. 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*, **356**(6333): 92–95.

Durand NC, Shamim MS, Machol I, et al. 2016. Juicer provides a one-click system for analyzing loop-resolution hi-c experiments. *Cell Systems*, **3**(1): 95–98.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**(5): 1792–1797.

Ellinghaus D, Kurtz S, Willhoeft U. 2008. *LTRharvest*, an efficient and flexible software for *de novo* detection of LTR retrotransposons. *BMC Bioinformatics*, **9**: 18.

Ferraris Jr CJ. 2007. Checklist of catfishes, recent and fossil (Osteichthyes: Siluriformes), and catalogue of siluriform primary types. *Zootaxa*, **1418**(1): 1–628.

Fischer S, Brunk BP, Chen F, et al. 2011. Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. *Current Protocols in Bioinformatics*, doi: 10.1002/0471250953.bi0612s35.

Gene Ontology Consortium. 2004. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research*, **32**(suppl_1): D258–D261.

Guindon S, Dufayard JF, Lefort V, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**(3): 307–321.

Han JE, Choresca CH, Koo OJ, et al. 2011. Establishment of glass catfish (*Kryptopterus bicirrhis*) fin-derived cells. *Cell Biology International Reports*, **18**(1): e00008.

Hayashi H, Fujii R. 1994. Pharmacological profiles of the subtypes of muscarinic cholinoceptors that mediate aggregation of pigment in the melanophores of two species of catfish. *Pigment Cell Research*, **7**(3): 175–183.

Hayashi H, Fujii R. 2001. Possible involvement of nitric oxide in signaling pigment dispersion in teleostean melanophores. *Zoological Science*, **18**(9): 1207–1215.

Hunter S, Apweiler R, Attwood TK, et al. 2009. InterPro: the integrative protein signature database. *Nucleic Acids Research*, **37**(suppl_1): D211–D215.

Jurka J, Kapitonov VV, Pavlicek A, et al. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research*, **110**(1-4): 462–467.

Kajitani R, Toshimoto K, Noguchi H, et al. 2014. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Research*, **24**(8): 1384–1395.

Kanehisa M, Furumichi M, Tanabe M, et al. 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, **45**(D1): D353–D361.

Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, **28**(1): 27–30.

Kasahara M, Naruse K, Sasaki S, et al. 2007. The medaka draft genome and insights into vertebrate genome evolution. *Nature*, **447**(7145): 714–719.

Kent WJ. 2002. BLAT—the BLAST-like alignment tool. *Genome Research*, **12**(4): 656–664.

Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, **12**(4): 357–360.

Krauss J, Frohnhöfer HG, Walderich B, et al. 2014a. Endothelin signalling in iridophore development and stripe pattern formation of zebrafish. *Biology*

Open, **3**(6): 503–509.

Krauss J, Geiger-Rudolph S, Koch I, et al. 2014b. A dominant mutation in *tyrp1A* leads to melanophore death in zebrafish. *Pigment Cell & Melanoma Research*, **27**(5): 827–830.

Kumar S, Stecher G, Suleski M, et al. 2017. TimeTree: a resource for timelines, timetrees, and divergence times. *Molecular Biology and Evolution*, **34**(7): 1812–1819.

LeGrande WH. 1981. Chromosomal evolution in North American catfishes (Siluriformes: Ictaluridae) with particular emphasis on the madtoms. *Noturus. Copeia*, **1981**(1): 33–52.

Li H. 2014. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics*, **30**(20): 2843–2851.

Li H, Handsaker B, Wysoker A, et al. 2009. The sequence alignment/map format and samtools. *Bioinformatics*, **25**(16): 2078–2079.

Lin JY, Fisher DE. 2007. Melanocyte biology and skin pigmentation. *Nature*, **445**(7130): 843–850.

Liu K, Xu DP, Li J, et al. 2017. Whole genome sequencing of Chinese clearhead icefish. *Protosalanx hyalocranius. Gigascience*, **6**(4): giw012.

Lu YS, Li RH, Xia LQ, et al. 2022. A chromosome-level genome assembly of the jade perch (*Scortum barcoo*). *Scientific Data*, **9**(1): 408.

McGinnis S, Madden TL. 2004. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Research*, **32**(suppl_2): W20–W25.

Mount DW. 2007. Using the basic local alignment search tool (BLAST). *CSH Protocols*, **2007**: pdb. top17.

Ng HH, Kottelat M. 2013. After eighty years of misidentification, a name for the glass catfish (Teleostei: Siluridae). *Zootaxa*, **3630**(2): 308–316.

Rummer JL, Wang S, Steffensen JF, et al. 2014. Function and control of the fish secondary vascular system, a contrast to mammalian lymphatic systems. *Journal of Experimental Biology*, **217**(Pt 5): 751–757.

Simão FA, Waterhouse RM, Ioannidis P, et al. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**(19): 3210–3212.

Spiewak JE, Bain EJ, Liu J, et al. 2018. Evolution of endothelin signaling and diversification of adult pigment pattern in *Danio* fishes. PLoS Genetics, **14**(9): e1007538.

Sprague J, Clements D, Conlin T, et al. 2003. The zebrafish information network (ZFIN): the zebrafish model organism database. *Nucleic Acids Research*, **31**(1): 241–243.

Tarailo-Graovac M, Chen NS. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*, doi: 10.1002/0471250953.bi0410s25.

Trapnell C, Hendrickson DG, Sauvageau M, et al. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology*, **31**(1): 46–53.

Walker BJ, Abeel T, Shea T, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One, **9**(11): e112963.

Wang HC, Liu B, Zhang Y, et al. 2020. Estimation of genome size using k-mer frequencies from corrected long reads. *arXiv*, doi: https://doi.org/10.48550/arXiv.2003.11817.

Wang HL, Su BF, Butts IAE, et al. 2022. Chromosome-level assembly and annotation of the blue catfish *Ictalurus furcatus*, an aquaculture species for hybrid catfish reproduction, epigenetics, and heterosis studies. *Gigascience*, **11**: giac070.

Yang ZH, Rannala B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution*, **23**(1): 212–226.

Ye CX, Hill CM, Wu SG, et al. 2016. DBG2OLC: efficient assembly of large genomes using long erroneous reads of the third generation sequencing technologies. *Scientific Reports*, **6**: 31900.

You XX, Bian C, Zan QJ, et al. 2014. Mudskipper genomes provide insights into the terrestrial adaptation of amphibious fishes. *Nature Communications*, **5**: 5594.

Zhang J, Qi JW, Shi FL, et al. 2020. Insights into the evolution of neoteny from the genome of the Asian icefish *protosalanx chinensis. iScience*, **23**(7): 101267.

Zhu WS, Hu B, Becker C, et al. 2017. Altered chromatin compaction and histone methylation drive non-additive gene expression in an interspecific *Arabidopsis* hybrid. *Genome Biology*, **18**(1): 157.