Whole-Genome Sequencing of *Sinocyclocheilus maitianheensis* Reveals Phylogenetic Evolution and Immunological Variances in Various *Sinocyclocheilus* Fishes

Ruihan Li[1,2†], Xiaoai Wang[3†], Chao Bian[1,2†], Zijian Gao[1,2], Yuanwei Zhang[3], Wansheng Jiang[4], Mo Wang[5], Xinxin You[1,2], Le Cheng[6], Xiaofu Pan[3], Junxing Yang[3*] and Qiong Shi[1,2*]

[1]College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China, [2]Shenzhen Key Lab of Marine Genomics, Guangdong Provincial Key Lab of Molecular Breeding in Marine Economic Animals, BGI Academy of Marine Sciences, BGI Marine, BGI, Shenzhen, China, [3]State Key Laboratory of Genetic Resources and Evolution, The Innovative Academy of Seed Design, Yunnan Key Laboratory of Plateau Fish Breeding, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China, [4]Hunan Engineering Laboratory for Chinese Giant Salamander's Resource Protection and Comprehensive Utilization, and Key Laboratory of Hunan Forest and Chemical Industry Engineering, Jishou University, Zhangjiajie, China, [5]Key Laboratory for Conserving Wildlife with Small Populations in Yunnan, Faculty of Biodiversity Conservation, Southwest Forestry University, Kunming, China, [6]BGI-Yunnan, Kunming, China

An adult *Sinocyclocheilus maitianheensis*, a surface-dwelling golden-line barbel fish, was collected from Maitian river (Kunming City, Yunnan Province, China) for whole-genome sequencing, assembly, and annotation. We obtained a genome assembly of 1.7 Gb with a scaffold N50 of 1.4 Mb and a contig N50 of 24.7 kb. A total of 39,977 protein-coding genes were annotated. Based on a comparative phylogenetic analysis of five *Sinocyclocheilus* species and other five representative vertebrates with published genome sequences, we found that *S. maitianheensis* is close to *Sinocyclocheilus anophthalmus* (a cave-restricted species with similar locality). Moreover, the assembled genomes of *S. maitianheensis* and other four *Sinocyclocheilus* counterparts were used for a fourfold degenerative third-codon transversion (4dTv) analysis. The recent whole-genome duplication (WGD) event was therefore estimated to occur about 18.1 million years ago. Our results also revealed a decreased tendency of copy number in many important genes related to immunity and apoptosis in cave-restricted *Sinocyclocheilus* species. In summary, we report the first genome assembly of *S. maitianheensis*, which provides a valuable genetic resource for comparative studies on cavefish biology, species protection, and practical aquaculture of this potentially economical fish.

Keywords: *Sinocyclocheilus maitianheensis*, whole genome sequencing, assembly, annotation, phylogeny, immunity, cave adaptation

Abbreviations: *ask1*, apoptosis signal-regulating kinase 1, also known as mitogen-activated protein kinase 5 (*map3k5*); *bcl-2*, B-cell lymphoma 2; *bcl2l1*, *bcl-2*-like protein 1; *cd40*, tumor necrosis factor receptor superfamily member 5; *cd401*, tumor necrosis factor ligand superfamily member 5; *daxx*, death domain-associated protein 6; *fadd*, fas-associated protein with death domain; *fas*, tumor necrosis factor receptor superfamily member 6; *fasl*, tumor necrosis factor ligand superfamily member 6; *mkk4*, mitogen-activated protein kinase kinase 4, also named *map2k4*; *mkk6*, mitogen-activated protein kinase kinase 6, also named *map2k6*; *tab1*, mitogen-activated protein kinase 7-interacting protein 1; *tak1*, mitogen-activated protein kinase kinase kinase 7, also named *map3k7*; *tnfa*, Tumor necrosis factor ligand superfamily member 1, also named *tnf*.

INTRODUCTION

Cave-restricted animals live in dark subterranean environments. They have evolved over time to adapt to the cave environments through various trait changes in morphology, behavior, and physiology (Jeffery, 2001). Cavefishes have degraded eyes, less body pigments, lower immune activities, and decrease in circadian rhythms in comparison to surface-dwelling fishes (Jeffery, 2009; Qiu et al., 2016; Yang et al., 2016; Krishnan and Rohner, 2017). As a compensation, the nonvisual sensory system of cavefishes is usually enhanced, such as development of extra taste buds and increased vibration attraction behavior (Yoshizawa et al., 2010; Yang et al., 2016). These facts of cavefishes are extremely interesting and worth to be explored with more investigations.

Previously, we have proved that cavefishes have fewer copies of major histocompatibility complex–related gene families (genes of cell surface proteins essential for acquired immune system) than surface-dwelling and semi–cave-dwelling counterparts, possibly suggesting relatively lower immune activities in cavefishes (Qiu et al., 2016), which may be a specific strategy for cave adaptation. Cave-restricted Mexican tetra (*Astyanax mexicanus*) also shows a big increase in appetite, but its fatty liver did not affect this fish's health (Aspiras et al., 2015), implying that there may be some other immune-related molecular mechanisms for fighting inflammation in cavefishes (Xiong et al., 2019). A recent study (Peuß et al., 2020) proposed that organisms in various environments have developed differential immune strategies with innate immune degradation and T-cell overexpression in cavefishes. However, many of these putative molecular mechanisms are not fully understood.

S. maitianheensis lives originally in the surface of Maitian river in Kunming City, Yunnan Province, China (**Supplementary Figure S1**). Some *Sinocyclocheilus* fishes are also residents in caves. *S. maitianheensis* can therefore be used as a control for comparative studies on cave adaptation. Meanwhile, as a genus of state second-class protection in China, *Sinocyclocheilus* has been propagated with a series of artificial breeding for protection from extinction (Yin et al., 2021). Various *Sinocyclocheilus* species likely shared tetraploid origin and have 96 chromosomes that are twice of most teleosts (Heng et al., 2002). The genome diversity makes fishes in this genus as good models for studying cave adaptation and phylogenetic evolution. Although there are some reports on morphological and mitochondrial genomic evolution of various *Sinocyclocheilus* species (Wu et al., 2010; He et al., 2012; Chen Y.-Y. et al., 2018), whole genome–based comparative studies are rare, except for three representative *Sinocyclocheilus* fishes that we published before (Yang et al., 2016).

Here, we performed whole-genome sequencing, assembly, and annotation of *S. maitianheensis* and subsequently conducted comparative genomic analysis and immune-gene inquiry with four other *Sinocyclocheilus* counterparts (including surface-dwelling *Sinocyclocheilus grahami,* semi–cave-dwelling *Sinocyclocheilus rhinocerous*, and cave-restricted *Sinocyclocheilus anophthalmus* and *Sinocyclocheilus anshuiensis*; their genome sequences are publicly available).

Our main purpose is to provide a genetic resource for in-depth studies on cave adaption and cavefish biology. Our study can also contribute to the species protection and exploitation of potentially economical value for *S. maitianheensis*.

MATERIALS AND METHODS

Sampling, Library Constructing, and Genome Sequencing

An adult *S. maitianheensis* was collected from Maitian river in Kunming City, Yunnan Province, China, for genome and transcriptome sequencing. Genomic DNAs were extracted from muscle sample. Seven Illumina paired-end sequencing libraries (with insert sizes of 270 bp, 500 bp, 800 bp, 2 kb, 5 kb, 10 kb, and 20 kb, respectively) were constructed for a routine shotgun whole-genome sequencing in an Illumina HiSeq 2,500 platform (San Diego, CA, United States). SOAPfilter v2.2 (Li R. et al., 2009) (-z -p -g 1 -f -o clean -M 2 -f 0) was used to filter reads. Duplicate reads from polymerase chain reactions, those reads with 10 or more nonsequencing bases (Ns), adapter sequences, and bases with low quality were removed.

Total RNAs were extracted from muscle, skin, eye, liver, heart, and brain for construction of individual cDNA library. cDNAs were then sequenced in an Illumina HiSeq X platform and filtered by SOAPnuke v1.0 (Chen Y. et al., 2018) with optimized parameters [-l 10 (default: 5) -q 0.2 (default: 0.5) -n 0.05 -c 0 -Q 2 (default: 1)]. Reads with nonsequenced (N) base ratio of more than 5% or low-quality base (base quality ≤10) ratio of greater than 20% were discarded to generate a new set of higher-quality reads for subsequent transcriptome-based annotation.

Genome Survey, *de Novo* Assembly, and Assessment

Genome size was estimated via the routine 17-mer frequency distribution analysis with the following formula: genome size $= K_{num}/K_{depth}$, where k_{num} is the number of k-mers obtained from reads, and K_{depth} is the expected depth of k-mer at a maximum frequency (Song et al., 2016). Two Illumina short-insert libraries (500 and 800 bp) were used for this 17-mer analysis.

The genome assembly strategy includes three steps. First, SOAPdenovo2 v2.04.4 (Luo et al., 2012) was applied to produce primary and final scaffolds with the following parameters: pregraph -K 27 -p 16 -d 1; contig -M 1; scaff -F -b 1.5 -p 16. Contigs and primary scaffolds were generated by using filtered reads from short-insert libraries (200, 500, and 800 bp), and the final scaffolds were constructed by mapping long-insert libraries (2, 5, 10, and 20 kb) onto the primary scaffolds. Second, gaps in scaffolds were then filled in two rounds using paired-end reads from the three short-insert libraries (270, 500, and 800 bp) via GapCloser v1.12 and v1.10 (Li R. et al., 2009) (-t 8 -l 150 and -t 25 -p 25, respectively). Finally, SSPACE V2.0 (Boetzer and Pirovano, 2014) (-k 5 -T 25 -g 2) was used to further extend and fill up both contigs and scaffolds. Completeness assessment of the final genome assembly was

performed by BUSCO v5.2.2 (Simão et al., 2015; Manni et al., 2021) (e-value ≤1e-3) with the popular actinopterygii_odb10 database.

Repeat Annotation

Repeat sequence annotation is composed of three routine methods, including *de novo* annotation, homology prediction, and tandem repeat prediction. First, we used RepeatModeller v1.04 (Chen, 2004) and LTR-FINDER v1.0.6 (Xu and Wang, 2007) to construct a local *de novo* repeat reference, and then our assembled genome was aligned to this reference library by RepeatMasker v4.06 (Chen, 2004). In addition, RepeatMasker v4.06 (Chen, 2004) and RepeatProteinMask v4.06 (Chen, 2004) were applied for homology prediction after identification of transposable elements based on RepBase (Jurka et al., 2005). Moreover, Tandem Repeat Finder v4.09 (Benson, 1999) was separately used to predict comprehensive tandem repeats in our pipeline as previously reported (Liu et al., 2019; Zhao et al., 2021). Finally, these results from the aforementioned three methods are integrated by our in-house perl scripts (https://github.com/liruihanguo/Repeats_integration). These scripts separately classified each type of repeat, integrated all repeats, and then removed those overlaps to obtain a nonredundant repeat set.

Gene Structure and Function Annotations

Two different methods were used for gene annotation to generate a total gene set, including homology annotation and transcriptome-based annotation (Bian et al., 2019). For the homology annotation, we downloaded protein sequences of four representative vertebrates from NCBI (Benson et al., 2006), including zebrafish (*Danio rerio*), Japanese medaka (*Oryzias latipes*), and two *Sinocyclocheilus* fishes (*S. anshuiensis* and *S. rhinocerous*), to align them against our *S. maitianheensis* genome assembly by TBLASTn (e-value ≤1e-5) (Gertz et al., 2006). Each gene structure was predicted by GeneWise v2.4.2 (Birney et al., 2004). For the transcriptome-based annotation, Tophat v2.0.13 (Trapnell et al., 2009) was utilized to obtain potential genes by mapping transcriptome reads onto our assembled genome. Subsequently, Cufflink v2.2.1 (Trapnell et al., 2013) was applied to predict the structures of potential genes on the alignments sorted by samtools v1.1 (Li H. et al., 2009). Lastly, the final consensus gene set was integrated by MAKER v2.31.8 (Cantarel et al., 2008).

All these predicted genes were aligned onto several public databases, including Interpro (Hunter et al., 2009), KEGG (Kanehisa et al., 2017), TrEMBL (Boeckmann et al., 2003), and Swiss-Prot (Boeckmann et al., 2003), using BLASTp (McGinnis and Madden, 2004) (e-value ≤1e-5) to perform function annotation. These results were then assessed by comparing coding sequence (CDS) length, intron length, gene length, exon length, and exon number distributions with the four closely related *Sinocyclocheilus* species, including *S. anshuiensis* (Yang et al., 2016) (SAMN03320099. WGS_v1.1 in NCBI), *S. rhinocerous* (Yang et al., 2016) (SAMN03320098_v1.1), *S. grahami* (Yang et al., 2016)

(SAMN03320097. WGS_v1.1), and *S. anophthalmus* [genome assembly was deposited at NCBI under accession no. PRJNA669129 (GCA_018155175.1)]. This unpublished genome of *S. anophthalmus* was sequenced by us on both Illumina Hiseq2500 and PacBio Sequel platforms using muscle genomic DNAs, and the final assembly of 1.9 Gb (with a contig N50 of 229.8 kb, a scaffold N50 of 309.9 kb, and prediction of 49,865 protein-coding genes) was assembled by combining the corrected long PacBio reads and the primary assembly from short Illumina reads by DBG2OLC v1.1 (Ye et al., 2016). We also assessed the completeness of these protein-coding gene sets by BUSCO v5.2.2 (Manni et al., 2021).

Orthogroup Cluster

The protein sequences of *S. maitianheensis* and other ten representative species were used for clustering orthogroups and phylogenetic analyses. These vertebrates include four *Sinocyclocheilus* species (*S. anophthalmus*, *S. grahami*, *S. anshuiensis*, and *S. rhinocerous*), common carp (*Cyprinus carpio*, GCF_000951,615.1 in NCBI), zebrafish (GCF_000002035.6), Japanese medaka (GCF_002234675.1), and Asian arowana (*Scleropages formosus*, GCF_900964775.1)*,* as well as the outgroup of human (*Homo sapiens*, GCF_000001405.39) and mouse (*Mus musculus*, GCF_000001635.27). We used BLASTp (McGinnis and Madden, 2004) (e-value ≤1e-5) to align these protein sequences with each other and OrthoMCL v2.0.92 (Fischer et al., 2011) with default parameters to identify orthologous genes and construct orthogroups.

Phylogenetic and Divergence Time Analyses

Single-copy orthogroups were aligned using MUSCLE v3.8.31 (Edgar, 2004). Subsequently, conserved regions were obtained by Gblocks (Castresana, 2000), and the CDS regions of all single-copy genes from each species were connected to form a supergene for extraction of the 4d sites. We also constructed a phylogenetic tree by using PhyML v3.0 with the maximum likelihood method (Guindon et al., 2010). MCMCtree in the PAML package (Yang and Rannala, 2006) was used to estimate the divergence time of five *Sinocyclocheilus* fishes and other species by three calibration time points of fossil records (Benton and Donoghue, 2007), including 61.5–100.5 Mya for *H. sapiens* and *M. musculus*, 159.9–165.2 Mya for *D. rerio* and *O. latipes*, and 416.1–421.8 Mya for *D. rerio* and *H. sapiens*.

4dTv Analysis to Identify Specific Whole-Genome Duplication in *Sinocyclocheilus* Fishes

To estimate the *Sinocyclocheilus* specific whole-genome duplication (WGD) event, we performed a fourfold degenerate third-codon transversion (4dTv) analysis by comparing five *Sinocyclocheilus* genomes with zebrafish and common carp genome assemblies. The WGD periods were calculated by using the following formula [The recognized

time of 3R WGD (~320 Mya)/the 4dTv peak values of 3R WGD in *Sinocyclocheilus* (0.65–0.75)] * the 4dTv peak values of lineage-specific WGD of *Sinocyclocheilus* (0.04–0.05).

An all-to-all alignments of protein sequences from these seven genomes were applied by using BLASTp (McGinnis and Madden, 2004) with an e-value of 1e-5. Syntenic blocks between species were identified using i-ADHoRe 3.0 (Proost et al., 2012) with default parameters, and then homologous proteins were obtained. Subsequently, homologous pairs were aligned using MUSCLE (Edgar, 2004), after we retrieved these homologous protein sequences and converted them to nucleotide sequences. Lastly, we calculated and corrected the 4dTv values for each gene pair by using the HKY model in PAML package (Yang and Rannala, 2006).

Identification of Immune Genes and Pseudogenes in P38 and Mitochondrial Pathways

Fifteen apoptosis-related genes were identified in five *Sinocyclocheilus* genomes and other five representative vertebrate genomes with high quality to investigate the gene copy number in P38 and mitochondrial pathway. These other five vertebrates include common carp, zebrafish, Japanese medaka, Asian arowana, and Mexican tetra (*A. mexicanus*, GCF_000372,685.1 in NCBI). Each genome sequence was used to construct a standard aligned database in the first place. Protein sequences of *tak1, tab1, ask1, fas, fasl, fadd, tnfa, cd40, cd40l, daxx, mkk4a, mkk4b, mkk6, bcl-2a,* and *bcl2l1* of zebrafish were downloaded from public uniport databases (**Supplementary Table S1**) as the queries. These protein sequences were then aligned onto above 10 genomes by TBLASTn (e-value ≤1e-5). The alignments with aligned ratio of less than 0.5, sequences similarity of less than 50%, and redundant data were filtered out to obtain the final hit alignments. Subsequently, target apoptosis-related genes were predicted by GeneWise v2.4.1 (Birney et al., 2004) from these 10 vertebrate genomes.

We performed a multiple-sequence alignment using the Muscle module in MEGA v7.0 (Kumar et al., 2016) to identify pseudogenes in each species indicated above. The whole open reading frames were performed with codon-based alignment to identify potential pseudogenes with irregular shifts of premature stop codon(s), codon frameshifts, or missing exon regions.

RESULTS AND DISCUSSION

Summary of the Genome Assembly and Assessment

We generated a total of 236.7-Gb raw reads, among them approximately 179.5 Gb of clean reads were obtained after removal of low-quality data. For the transcriptome sequencing, a total of 39.3-Gb raw reads were generated. The genome size of *S. maitianheensis* was estimated to be about 1.8 Gb by the routine 17-mer frequency distribution analysis, because the k_{num} is 40,525,178,512, and the K_{depth} is 23 (see **Supplementary Figure S2**).

After *de novo* assembly and gap closing, the final genome assembly was 1.7 Gb in total length, with a scaffold N50 of

TABLE 1 | Statistics of the genome assembly for *S. maitianheensis*.

Genome assembly	Parameter
Contig N50 (kb)	24.7
Contig number (>100 bp)	212,978
Scaffold N50 (Mb)	1.4
Scaffold number (>100 bp)	106,737
Total length (Gb)	1.7
Genome coverage (×)	139.2
The longest scaffold (Mb)	12.5
Genome annotation	**Parameter**
Protein-coding gene number	39,977
Mean transcript length (bp)	15,881
Mean exons per gene	8.6
Mean exon length (bp)	188.3
Mean intron length (bp)	1,728

1.4 Mb, a contig N50 of 24.7 kb, and the GC content of 37.6% (**Tables 1** and **Supplementary Table S2**). The final assembled genome accounted for 94.4% of the estimated genome size (1.8 Gb). For assessment of our genome assembly, we searched a total of 3,640 BUSCO (Benchmarking Universal Single-Copy Orthologs) groups and determined that 3,536 (97.1%) were complete (with 1,643 single-copy BUSCOs and 1,893 duplicated BUSCOs), suggesting a high completeness of our genome assembly for *S. maitianheensis.*

We compared the genome assembly and annotation results for the five examined *Sinocyclocheilus* species and observed that their genome sizes are at a narrow range of 1.7–1.9 Gb (**Supplementary Table S3**). The largest genome is 1.9 Gb for *S. anophthalmus* (GCA_018155175.1 in NCBI) whose assembly was generated with both Illumina (next-generation) and PacBio (third-generation) sequencing reads, whereas others were sequenced only by an Illumina platform with relatively lower scaffold N50 values. GC contents of the five examined *Sinocyclocheilus* species ranged from 37.2 to 38.6% (**Supplementary Table S3**).

Genome Annotation

For repeat annotation, we predicted 664,837,705-bp repeat elements (accounting for 39.4% of the genome assembly; **Supplementary Table S4**). Among them, 455.9 Mb of DNA repeat elements, 135.2 Mb of long terminal repeats (LTRs), 104.9 Mb of long interspersed nuclear elements (LINE), and 9.1 Mb of short interspersed nuclear elements (SINE) were detected (**Supplementary Table S5**). After homology and transcriptome-based annotations, our MAKER results integrated a total of 39,977 protein-coding genes with an average length of 15.9 kb (**Supplementary Table S6**).

For function annotation, 38,677 genes were aligned onto four public databases (Interpro, KEGG, TrEMBL, and Swiss-Prot) with function assignments, which account for 96.8% of the annotated genes (**Supplementary Table S7**). The distributions of CDS, intron, gene, and exon length in *S. maitianheensis* were generally similar to those in other four examined *Sinocyclocheilus* fishes (**Supplementary Figure S3**), which suggested a good reliability of the genome annotation for *S. maitianheensis.* However, there are still some distinctions among them, such as the higher percentages of CDS and genes at length
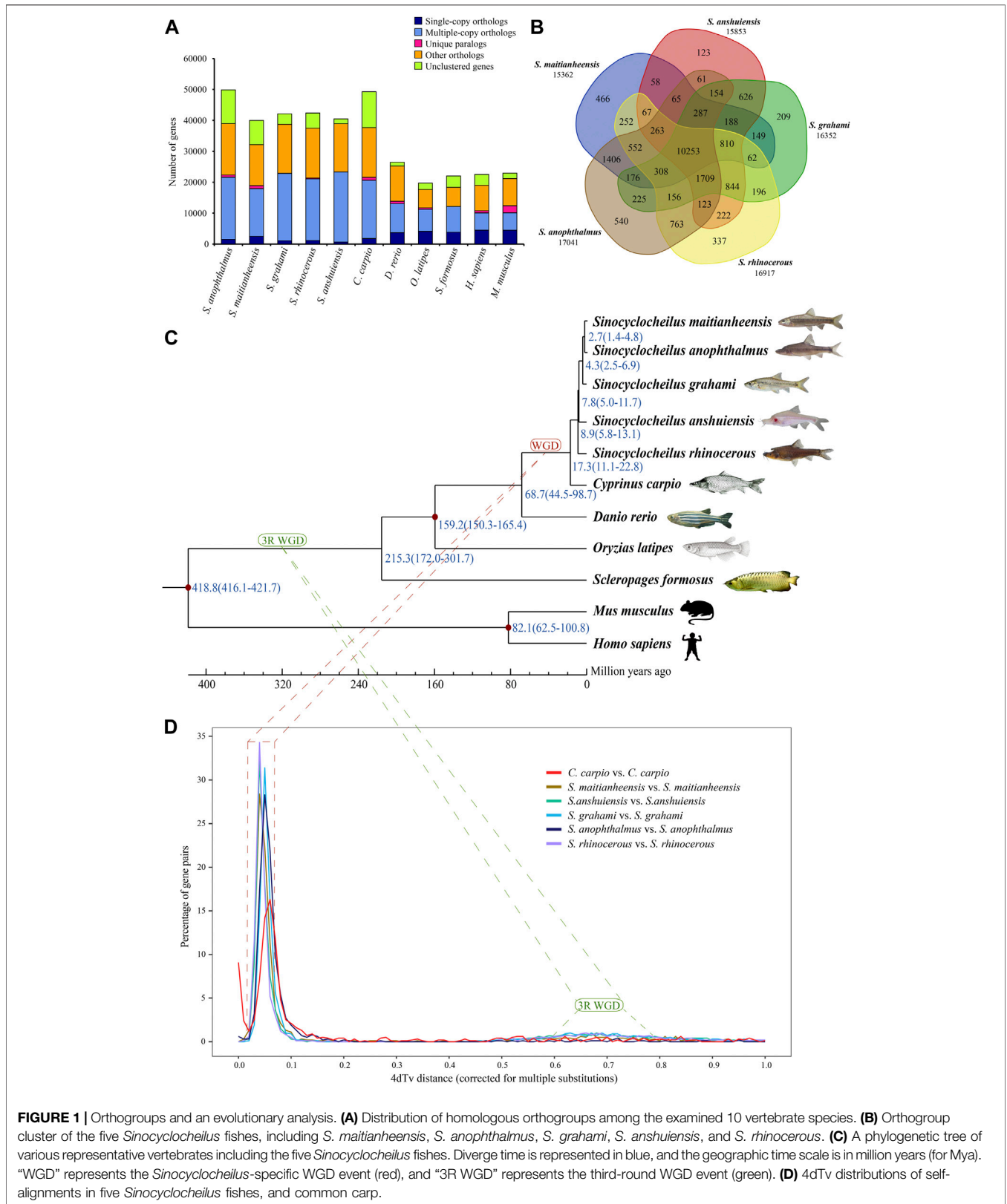
FIGURE 1 | Orthogroups and an evolutionary analysis. **(A)** Distribution of homologous orthogroups among the examined 10 vertebrate species. **(B)** Orthogroup cluster of the five *Sinocyclocheilus* fishes, including *S. maitianheensis*, *S. anophthalmus*, *S. grahami*, *S. anshuiensis*, and *S. rhinocerous*. **(C)** A phylogenetic tree of various representative vertebrates including the five *Sinocyclocheilus* fishes. Diverge time is represented in blue, and the geographic time scale is in million years (for Mya). "WGD" represents the *Sinocyclocheilus*-specific WGD event (red), and "3R WGD" represents the third-round WGD event (green). **(D)** 4dTv distributions of self-alignments in five *Sinocyclocheilus* fishes, and common carp.

approximately 1,000 bp in *S. maitianheensis*. Moreover, our annotation results reveal that *S. maitianheensis* has less protein-coding gene number (39,977) than the other four *Sinocyclocheilus* counterparts (49,865 for *S. anophthalmus*, 42,109 for *S. grahami*, 40,470 for *S. anshuiensis*, and 42,377 for *S. rhinocerous*). The assessment of protein-coding genes among the five fishes showed that there were more missing BUSCOs in *S. maitianheensis* (**Supplementary Table S8**), which may lead to a lower amount of gene number. However, this divergence of gene number may be caused in part by different annotation methods.

Summary of the Gene Orthogroups

A total of 26,875 orthogroups were predicted for the five examined *Sinocyclocheilus* fishes and six other representative vertebrate species. For *S. maitianheensis*, 32,150 protein-coding genes were clustered into 15,617 orthogroups, although 7,827 genes were not clustered; the numbers of multiple-copy, single-copy orthologs, and unique paralogs were 15,400, 2,508 and 981, respectively (**Figure 1A** and **Supplementary Table S9**). We summarized the numbers of orthogroups shared with each other between the five *Sinocyclocheilus* in **Figure 1B**. There are 10,253 common orthogroups among these *Sinocyclocheilus* species.

Phylogenetic Position of *S. maitianheensis*

A total of 191 common single-copy orthogroups among all the examined species were used for construction of a phylogenetic tree. A progressive evolutionary relationship in the five *Sinocyclocheilus* species revealed a new phylogeny based on both genome and transcriptome data (**Figure 1C**). This topology is consistent with the phylogenetic tree based on genomes (Yang et al., 2016) and mitochondrial genes (Zhang and Wang, 2018), but it is different from a previous report of a division into two branches, in which *S. maitianheensis*, *S. anophthalmus*, and *S. anophthalmus* are located in one branch, whereas *S. anshuiensis* and *S. rhinocerous* are located in another branch (Mao et al., 2021) based on the morphological trait of eyes. Therefore, our latest topology provides novel insights into detailed phylogeny for the five *Sinocyclocheilus* species at a genome level.

S. maitianheensis and *S. anophthalmus* diverged at approximately 2.7 (1.4–4.8) Mya, and the divergence time periods with *S. grahami, S. anshuiensis*, and *S. rhinocerous* were 4.3 (2.5–6.9), 7.8 (5.0–11.7), and 8.9 (5.8–13.1) Mya, respectively (**Figure 1C**). Surface-dwelling *S. maitianheensis* has the closest relationship with cave-restricted *S. anophthalmus*; however, there are huge differences in morphological traits between them. Interestingly, the geographical positions of both habitats are close, located in the same Yiliang County (Kunming City, Yunnan Province, China; **Supplementary Figure S1**). *S. anophthalmus* resides in a dark environment of several caves in Jiuxiang Town (Zhao and Zhang, 2009), whereas *S. maitianheensis* lives in the surface of Maitian river. Divergence time of the two *Sinocyclocheilus* species was relatively late, approximately 2.7 (1.4–4.8) Mya. Their separation may be due to a geographical isolation generated by the continuous uplift of the Yunnan–Guizhou Plateau after Himalayan orogeny (50–40 Mya) (Yin and Harrison, 2000),

and some of their ancestors swam down along the Maitian river into these surrounding caves.

Estimation of the Lineage-specific WGD in *Sinocyclocheilus*

Cyprinidae experienced a recent genome-wide duplication event (Larhammar and Risinger, 1994; David et al., 2003) after the third-round WGD (3R WGD, also known as teleost-specific WGD) (Jaillon et al., 2004). We performed a 4dTv analysis to estimate the timing of this recent lineage-specific WGD in *Sinocyclocheilus*. Self-alignments with paralogous gene pairs of *S. maitianheensis*, *S. anshuiensis*, *S. rhinocerous*, *S. anophthalmus*, *S. grahami*, and common carp displayed distinct peaks (**Figure 1D**). Their 4dTv distances were calculated to be 0.04, 0.04, 0.04, 0.05, 0.05, and 0.06, respectively. The peak values of the five *Sinocyclocheilus* fishes and common carp are very close (0.04–0.06), implying that these fishes might share the recent genome-wide duplication event.

In order to compare the recent specific WGD between *Sinocyclocheilus* fishes and common carp, we performed a 4dTv analysis on 13,579 orthologous gene pairs between *S. maitianheensis* and common carp genomes (**Supplementary Figure S4**). The peak values in each group of *C. carpio–C. carpio*, *S. maitianheensis–C. carpio*, and *S. maitianheensis–S. maitianheensis* were estimated to be 0.06, 0.04, and 0.04, respectively. The nearly overlapping of peaks indicated that *S. maitianheensis* and common carp might have undergone the recent specific WGD together. Based on the aforementioned 4dTv analyses and the construction of phylogenetic tree, we predicted that the carp-specific WGD occurred ~18.1 Mya, before the evolutionary separation of *Sinocyclocheilus* fishes and common carp (~17.3 Mya; **Figure 1D**). This estimate is little earlier before the specific WGD of common carp based on the divergence time of transposable elements (Xu et al., 2019).

The time estimation of the latest WGD in *Cyprinidae* is contentious, from ~8.2 to ~16 Mya (Larhammar and Risinger, 1994; David et al., 2003; Xu et al., 2014), although these studies almost focused on the common carp. However, a recent study of common carp defined a rough time range (9.7–23 Mya) and further predicted this time point to about 12.4 Mya (Xu et al., 2019). Our result of ~18.4 Mya in the present study is within this time range, and the 4dTv analysis of *Sinocyclocheilus* fishes provides more evidence for the timing of the latest genome duplication in *Cyprinidae*.

Copy Number Variations of Several Immune Genes in P38 and Mitochondrial Pathways

S. anophthalmus has evolved a series of traits to adapt to the caved environment. It had developed huge differences from its sister species (*S. maitianheensis*), such as loss of eyes and the semitransparent body (Zhao and Zhang, 2009). Therefore, compared to *S. maitianheensis*, *S. anophthalmus* is an independent cave species with many valuable traits that are waiting for in-depth explorations.

To investigate potential immunological variances in cave-restricted *Sinocyclocheilus* species, we identified the copy number

of 15 immune genes within P38 and mitochondrial pathways in the five examined *Sinocyclocheilus* genomes (see detailed statistics in **Supplementary Table S10**). These genes include *tak1*, *tab1*, *ask1*, *fas*, *fasl*, *fadd*, *tnfa*, *cd40*, *cd40l*, *daxx*, *mkk4a*, *mkk4b*, *mkk6*, *bcl-2a*, and *bcl2l1* (**Supplementary Figure S5**). In general, five *Sinocyclocheilus* species and common carp usually own twofold copies compared to the other diploid teleosts. Interestingly, we found copies of some genes in the cave-restricted fishes. For *S. anophthalmus*, a copy of *ask1* (an apoptosis signal-regulating kinase) (Patel et al., 2019) was predicted as a pseudogene; only one copy of *bcl2l1* (encoding apoptotic regulators in BCL-2 family) (Warren et al., 2019) was retained in *S. anophthalmus* compared with *S. maitianheensis* that contained two copies; one more copy of *bcl-2a* and no copy of *mkk4a* were also observed in *S. anophthalmus* genome. In addition, among most of these genes we studied, another cave-restricted fish Mexican tetra (*A. mexicanus*) had fewer gene copy numbers than the other examined fishes. These variances in gene copy number imply that the apoptotic activity might have been decreased in cave-restricted fishes, which is consistent with our previous report of relatively lower immunity in cavefishes (Qiu et al., 2016). However, apoptotic activity is regulated by many factors, and more investigations should be done for in-depth verification.

In addition, we performed high-throughput identification of antimicrobial peptides (AMPs) (Yi et al., 2017; Mwangi et al., 2019). A total of 379, 551, 522, 545, and 552 putative AMP sequences were identified in *S. maitianheensis*, *S. rhinocerous*, *S. anshuiensis*, *S. anophthalmus*, and *S. grahami* genomes, respectively (**Supplementary Table S11**). Thrombin, histone, lectin, chemokine, scolopendin, and ubiquitin are the most abundant AMPs in the five examined *Sinocyclocheilus* species. The lowest number of AMP sequences in *S. maitianheensis* genome may be related to its least protein-coding genes in the five *Sinocyclocheilus* fishes.

CONCLUSION

In summary, we reported the first genome assembly of *S. maitianheensis*, which provides a valuable genetic resource for comparative studies on cavefish biology, species protection and practical aquaculture of this potentially economical teleost fish. This genome assembly also supplies essential genomic data for in-depth genetic analysis. Based on these genomic data, we observed a close relationship between *S. maitianheensis* and *S. anophthalmus*. Some variations of gene copy number in the immune system might indicate the variation in immunity and apoptosis in cave-restricted *Sinocyclocheilus* species.

VALUE OF THE DATA

This is the first draft genome of a representative surface-dwelling Chinese golden-line barbel fish, *Sinocyclocheilus maitianheensis*. The final assembly was 1.7 Gb with a scaffold N50 of 1.4 Mb and a contig N50 of 24.7 kb.

The phylogenetic tree revealed that *S. maitianheensis* is close to *S. anophthalmus* (a cave-restricted species with similar

locality). The divergence time between the two relatives is about 2.7 million years ago (Mya).

The 4dTv analysis demonstrated that the recent carp-specific WGD event occurred approximately 18.1 Mya.

A decrease in the copy number of many important immunological genes was observed in cave-restricted *Sinocyclocheilus* species.

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

All animal experiments in this study were performed in accordance with the guidelines of the Animal Ethics Committee and were approved by the Institutional Review Board on Bioethics and Biosafety of BGI, China (No. FT18134). Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

Conceptualization, JY and QS; data analysis, RL and ZG; samples collection and assisted data analysis, YZ, WJ, and MW; data curation, XY, LC, and XP; writing-original draft preparation, RL and ZG; writing-review and editing, QS, CB, and XW; supervision, QS and JY; funding acquisition, JY and QS. All authors have read and approved the published version of the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.736500/full#supplementary-material

REFERENCES

Aspiras, A. C., Rohner, N., Martineau, B., Borowsky, R. L., and Tabin, C. J. (2015). Melanocortin 4 Receptor Mutations Contribute to the Adaptation of Cavefish to Nutrient-Poor Conditions. *Proc. Natl. Acad. Sci. USA* 112 (31), 9668–9673. doi:10.1073/pnas.1510802112

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Wheeler, D. L. (2006). *Genbank. Nucleic Acids Res.* 34, D16–D20. doi:10.1093/nar/gkj157Database issue

Benson, G. (1999). Tandem Repeats Finder: a Program to Analyze DNA Sequences. *Nucleic Acids Res.* 27 (2), 573–580. doi:10.1093/nar/27.2.573

Benton, M. J., and Donoghue, P. C. J. (2007). Paleontological Evidence to Date the Tree of Life. *Mol. Biol. Evol.* 24 (1), 26–53. doi:10.1093/molbev/msl150

Bian, C., Huang, Y., Li, J., You, X., Yi, Y., Ge, W., et al. (2019). Divergence, Evolution and Adaptation in ray-finned Fish Genomes. *Sci. China Life Sci.* 62 (8), 1003–1018. doi:10.1007/s11427-018-9499-5

Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and Genomewise. *Genome Res.* 14 (5), 988–995. doi:10.1101/gr.1865504

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., et al. (2003). The SWISS-PROT Protein Knowledgebase and its Supplement TrEMBL in 2003. *Nucleic Acids Res.* 31 (1), 365–370. doi:10.1093/nar/gkg095

Boetzer, M., and Pirovano, W. (2014). SSPACE-LongRead: Scaffolding Bacterial Draft Genomes Using Long Read Sequence Information. *BMC Bioinformatics* 15, 211. doi:10.1186/1471-2105-15-211

Cantarel, B. L., Korf, I., Robb, S. M. C., Parra, G., Ross, E., Moore, B., et al. (2008). MAKER: an Easy-To-Use Annotation Pipeline Designed for Emerging Model Organism Genomes. *Genome Res.* 18 (1), 188–196. doi:10.1101/gr.6743907

Castresana, J. (2000). Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Mol. Biol. Evol.* 17 (4), 540–552. doi:10.1093/oxfordjournals.molbev.a026334

Chen, N. (2004). Using Repeat Masker to Identify Repetitive Elements in Genomic Sequences. *Curr. Protoc. Bioinformatics* 5. , 2004 Chapter 4Unit 4.10. doi:10.1002/0471250953.bi0410s05

Chen, Y.-Y., Li, R., Li, C.-Q., Li, W.-X., Yang, H.-F., Xiao, H., et al. (2018b). Testing the Validity of Two Putative Sympatric Species from Sinocyclocheilus (Cypriniformes: Cyprinidae) Based on Mitochondrial Cytochrome B Sequences. *Zootaxa* 4476 (1), 130–140. doi:10.11646/zootaxa.4476.1.12

Chen, Y., Chen, Y., Shi, C., Huang, Z., Zhang, Y., Li, S., et al. (2018a). SOAPnuke: a MapReduce Acceleration-Supported Software for Integrated Quality Control and Preprocessing of High-Throughput Sequencing Data. *Gigascience* 7 (1), 1–6. doi:10.1093/gigascience/gix120

David, L., Blum, S., Feldman, M. W., Lavi, U., and Hillel, J. (2003). Recent Duplication of the Common Carp (*Cyprinus carpio* L.) Genome as Revealed by Analyses of Microsatellite Loci. *Mol. Biol. Evol.* 20 (9), 1425–1434. doi:10.1093/molbev/msg173

Edgar, R. C. (2004). MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput. *Nucleic Acids Res.* 32 (5), 1792–1797. doi:10.1093/nar/gkh340

Fischer, S., Brunk, B. P., Chen, F., Gao, X., Harb, O. S., Iodice, J. B., et al. (2011). Using OrthoMCL to Assign Proteins to OrthoMCL-DB Groups or to Cluster Proteomes into New Ortholog Groups. *Curr. Protoc. Bioinformatics Chapter* 6, 11–19. Unit 6.12. doi:10.1002/0471250953.bi0612s35

Gertz, E. M., Yu, Y.-K., Agarwala, R., Schäffer, A. A., and Altschul, S. F. (2006). Composition-based Statistics and Translated Nucleotide Searches: Improving the TBLASTN Module of BLAST. *BMC Biol.* 4, 41. doi:10.1186/1741-7007-4-41

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst. Biol* 59 (3), 307–321. doi:10.1093/sysbio/syq010

He, S., Liang, X.-F., Chu, W.-Y., and Chen, D.-X. (2012). Complete Mitochondrial Genome of the Blind Cave barbelSinocyclocheilus furcodorsalis(Cypriniformes: Cyprinidae). *Mitochondrial DNA* 23 (6), 429–431. doi:10.3109/19401736.2012.710216

Heng, X., Rendong, Z., Jianguo, F., Ming, O., Weixian, L., Shanyuan, C., et al. (2002). Nuclear DNA Content and Ploidy of Seventeen Species of Fishes in Sinocyclocheilus. *Dong Wu Xue Yan jiu= Zoolog. Res.* 23 (3), 195–199.

Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., et al. (2009). InterPro: the Integrative Protein Signature Database. *Nucleic Acids Res.* 37 (Database issue), D211–D215. doi:10.1093/nar/gkn785

Jaillon, O., Aury, J.-M., Brunet, F., Petit, J.-L., Stange-Thomann, N., Mauceli, E., et al. (2004). Genome Duplication in the Teleost Fish *Tetraodon nigroviridis* Reveals the Early Vertebrate Proto-Karyotype. *Nature* 431 (7011), 946–957. doi:10.1038/nature03025

Jeffery, W. R. (2001). Cavefish as a Model System in Evolutionary Developmental Biology. *Dev. Biol.* 231 (1), 1–12. doi:10.1006/dbio.2000.0121

Jeffery, W. R. (2009). Regressive Evolution inAstyanaxCavefish. *Annu. Rev. Genet.* 43, 25–47. doi:10.1146/annurev-genet-102108-134216

Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase Update, a Database of Eukaryotic Repetitive Elements. *Cytogenet. Genome Res.* 110 (1-4), 462–467. doi:10.1159/000084979

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: New Perspectives on Genomes, Pathways, Diseases and Drugs. *Nucleic Acids Res.* 45 (D1), D353–d361. doi:10.1093/nar/gkw1092

Krishnan, J., and Rohner, N. (2017). Cavefish and the Basis for Eye Loss. *Phil. Trans. R. Soc. B.* 372 (1713), 20150487. doi:10.1098/rstb.2015.0487

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 33 (7), 1870–1874. doi:10.1093/molbev/msw054

Larhammar, D., and Risinger, C. (1994). Molecular Genetic Aspects of Tetraploidy in the Common Carp *Cyprinus carpio*. *Mol. Phylogenet. Evol.* 3 (1), 59–68. doi:10.1006/mpev.1994.1007

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009a). The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* 25 (16), 2078–2079. doi:10.1093/bioinformatics/btp352

Li, R., Yu, C., Li, Y., Lam, T.-W., Yiu, S.-M., Kristiansen, K., et al. (2009b). SOAP2: an Improved Ultrafast Tool for Short Read Alignment. *Bioinformatics* 25 (15), 1966–1967. doi:10.1093/bioinformatics/btp336

Liu, H.-P., Xiao, S.-J., Wu, N., Wang, D., Liu, Y.-C., Zhou, C.-W., et al. (2019). The Sequence and De Novo Assembly of Oxygymnocypris Stewartii Genome. *Sci. Data* 6 (1), 190009. doi:10.1038/sdata.2019.9

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., et al. (2012). SOAPdenovo2: an Empirically Improved Memory-Efficient Short-Read De Novo Assembler. *GigaSc.i* 1 (1), 18. doi:10.1186/2047-217x-1-18

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., and Zdobnov, E. M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol. Biol. Evol.* 28, msab199. doi:10.1093/molbev/msab199

Mao, T.-R., Liu, Y.-W., Meegaskumbura, M., Yang, J., Ellepola, G., Senevirathne, G., et al. (2021). Evolution in Sinocyclocheilus Cavefish Is Marked by Rate Shifts, Reversals, and Origin of Novel Traits. *BMC Ecol. Evo* 21 (1), 45. doi:10.1186/s12862-021-01776-y

McGinnis, S., and Madden, T. L. (2004). BLAST: at the Core of a Powerful and Diverse Set of Sequence Analysis Tools. *Nucleic Acids Res.* 32, W20–W25. Web Server issue. doi:10.1093/nar/gkh435

Mwangi, J., Hao, X., Lai, R., and Zhang, Z.-Y. (2019). Antimicrobial Peptides: new hope in the War against Multidrug Resistance. *Zool. Res.* 40 (6), 488–505. doi:10.24272/j.issn.2095-8137.2019.062

Patel, P., Naik, M. U., Golla, K., Shaik, N. F., and Naik, U. P. (2019). Calcium-induced Dissociation of CIB1 from ASK1 Regulates Agonist-Induced Activation of the P38 MAPK Pathway in Platelets. *Biochem. J.* 476 (19), 2835–2850. doi:10.1042/bcj20190410

Peuß, R., Box, A. C., Chen, S., Wang, Y., Tsuchiya, D., Persons, J. L., et al. (2020). Adaptation to Low Parasite Abundance Affects Immune Investment and Immunopathological Responses of Cavefish. *Nat. Ecol. Evol.* 4 (10), 1416–1430. doi:10.1038/s41559-020-1234-2

Proost, S., Fostier, J., De Witte, D., Dhoedt, B., Demeester, P., Van de Peer, Y., et al. (2012). I-ADHoRe 3.0-fast and Sensitive Detection of Genomic Homology in Extremely Large Data Sets. *Nucleic Acids Res.* 40 (2), e11. doi:10.1093/nar/gkr955

Qiu, Y., Yang, J., Jiang, W., Chen, X., Bian, C., and Shi, Q. (2016). A Genomic Survey on the Immune Differences amongSinocyclocheilusfishes. *Communicative Integr. Biol.* 9 (6), e1255833. doi:10.1080/19420889.2016.1255833

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 (19), 3210–3212. doi:10.1093/bioinformatics/btv351

Song, L., Bian, C., Luo, Y., Wang, L., You, X., Li, J., et al. (2016). Draft Genome of the Chinese Mitten Crab, Eriocheir Sinensis. *GigaSci.* 5, 5. doi:10.1186/s13742-016-0112-y

Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L., and Pachter, L. (2013). Differential Analysis of Gene Regulation at Transcript Resolution with RNA-Seq. *Nat. Biotechnol.* 31 (1), 46–53. doi:10.1038/nbt.2450

Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: Discovering Splice Junctions with RNA-Seq. *Bioinformatics* 25 (9), 1105–1111. doi:10.1093/bioinformatics/btp120

Warren, C. F. A., Wong-Brown, M. W., and Bowden, N. A. (2019). BCL-2 Family Isoforms in Apoptosis and Cancer. *Cell Death Dis.* 10 (3), 177. doi:10.1038/s41419-019-1407-6

Wu, X., Wang, L., Chen, S., Zan, R., Xiao, H., and Zhang, Y.-p. (2010). The Complete Mitochondrial Genomes of Two Species from Sinocyclocheilus (Cypriniformes: Cyprinidae) and a Phylogenetic Analysis within Cyprininae. *Mol. Biol. Rep.* 37 (5), 2163–2171. doi:10.1007/s11033-009-9689-x

Xiong, J. B., Nie, L., and Chen, J. (2019). Current Understanding on the Roles of Gut Microbiota in Fish Disease and Immunity. *Zool. Res.* 40 (2), 70–76. doi:10.24272/j.issn.2095-8137.2018.069

Xu, P., Xu, J., Liu, G., Chen, L., Zhou, Z., Peng, W., et al. (2019). The Allotetraploid Origin and Asymmetrical Genome Evolution of the Common Carp *Cyprinus carpio*. *Nat. Commun.* 10 (1), 4625. doi:10.1038/s41467-019-12644-1

Xu, P., Zhang, X., Wang, X., Li, J., Liu, G., Kuang, Y., et al. (2014). Genome Sequence and Genetic Diversity of the Common Carp, *Cyprinus carpio*. *Nat. Genet.* 46 (11), 1212–1219. doi:10.1038/ng.3098

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. Web Server issue. doi:10.1093/nar/gkm286

Yang, J., Chen, X., Bai, J., Fang, D., Qiu, Y., Jiang, W., et al. (2016). The Sinocyclocheilus Cavefish Genome Provides Insights into Cave Adaptation. *BMC Biol.* 14, 1. doi:10.1186/s12915-015-0223-4

Yang, Z., and Rannala, B. (2006). Bayesian Estimation of Species Divergence Times under a Molecular Clock Using Multiple Fossil Calibrations with Soft Bounds. *Mol. Biol. Evol.* 23 (1), 212–226. doi:10.1093/molbev/msj024

Ye, C., Hill, C. M., Wu, S., Ruan, J., and Ma, Z. (2016). DBG2OLC: Efficient Assembly of Large Genomes Using Long Erroneous Reads of the Third Generation Sequencing Technologies. *Sci. Rep.* 6, 31900. doi:10.1038/srep31900

Yi, Y., You, X., Bian, C., Chen, S., Lv, Z., Qiu, L., et al. (2017). High-Throughput Identification of Antimicrobial Peptides from Amphibious Mudskippers. *Mar. Drugs* 15 (11), 364. doi:10.3390/md15110364

Yin, A., and Harrison, T. M. (2000). Geologic Evolution of the Himalayan-Tibetan Orogen. *Annu. Rev. Earth Planet. Sci.* 28 (1), 211–280. doi:10.1146/annurev.earth.28.1.211

Yin, Y. H., Zhang, X. H., Wang, X. A., Li, R. H., Zhang, Y. W., Shan, X. X., et al. (2021). Construction of a Chromosome-Level Genome Assembly for Genome-wide Identification of Growth-Related Quantitative Trait Loci in *Sinocyclocheilus grahami* (Cypriniformes, Cyprinidae). *Zool Res.* 42 (3), 262–266. doi:10.24272/j.issn.2095-8137.2020.321

Yoshizawa, M., Gorički, Š., Soares, D., and Jeffery, W. R. (2010). Evolution of a Behavioral Shift Mediated by Superficial Neuromasts Helps Cavefish Find Food in Darkness. *Curr. Biol.* 20 (18), 1631–1636. doi:10.1016/j.cub.2010.07.017

Zhang, R., and Wang, X. (2018). Characterization and Phylogenetic Analysis of the Complete Mitogenome of a Rare Cavefish, Sinocyclocheilus Multipunctatus (Cypriniformes: Cyprinidae). *Genes Genom* 40 (10), 1033–1040. doi:10.1007/s13258-018-0711-3

Zhao, N., Guo, H., Jia, L., Guo, B., Zheng, D., Liu, S., et al. (2021). Genome Assembly and Annotation at the Chromosomal Level of First Pleuronectidae: Verasper Variegatus Provides a Basis for Phylogenetic Study of Pleuronectiformes. *Genomics* 113 (2), 717–726. doi:10.1016/j.ygeno.2021.01.024

Zhao, Y., and Zhang, C. (2009). Threatened Fishes of the World: Sinocyclocheilus Anophthalmus (Chen and Chu, 1988) (Cyprinidae). *Environ. Biol. Fish.* 86 (1), 163. doi:10.1007/s10641-008-9361-7