Whole genome sequencing of the fast-swimming Southern bluefin tuna (Thunnus maccoyii)

Xiaomeng Zhao1,2†, Yu Huang2†, Chao Bian2, Xinxin You2,3, Xinhui Zhang1,2,3, Jieming Chen2, Min Wang4, Cancan Hu3, Yun Xu2,3,4, Junmin Xu2,3,4 and Qiong Shi1,2,3*

1College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China, 2Shenzhen Key Lab of Marine Genomics, Guangdong Provincial Key Lab of Molecular Breeding in Marine Economic Animals, BGI Academy of Marine Sciences, BGI Marine, Shenzhen, China, 3Aquatic Breeding Center, BGI Marine, Shenzhen, China, 4BGI Zhenjiang Institute of Hydrobiology, Zhenjiang, China

The economically important Southern bluefin tuna (Thunnus maccoyii) is a world-famous fast-swimming fish, but its genomic information is limited. Here, we performed whole genome sequencing and assembled a draft genome for Southern bluefin tuna, aiming to generate useful genetic data for comparative functional prediction. The final genome assembly is 806.54 Mb, with scaffold and contig N50 values of 3.31 Mb and 67.38 kb, respectively. Genome completeness was evaluated to be 95.8%. The assembled genome contained 23,403 protein-coding genes and 236.1 Mb of repeat sequences (accounting for 29.27% of the entire assembly).

Comparative genomics analyses of this fast-swimming tuna revealed that it had more than twice as many hemoglobin genes (18) as other relatively slow-moving fishes (such as seahorse, sunfish, and tongue sole). These hemoglobin genes are mainly localized in two big clusters (termed as "MN" and "LA" respectively), which is consistent with other reported fishes. However, Thr39 of beta-hemoglobin in the MN cluster, conserved in other fishes, was mutated as cysteine in tunas including the Southern bluefin tuna. Since hemoglobins are reported to transport oxygen efficiently for aerobic respiration, our genomic data suggest that both high copy numbers of hemoglobin genes and an adjusted function of the beta-hemoglobin may support the fast-swimming activity of tunas. In summary, we produced a primary genome assembly and predicted hemoglobin-related roles for the fast-swimming Southern bluefin tuna.

KEYWORDS
southern bluefin tuna (Thunnus maccoyii), genome sequencing, assembly, hemoglobin, evolution
1 Introduction

As one migratory fish in the order of Scombriformes, Southern bluefin tuna (Thunnus maccoyii; TM) has been an economically important marine species due to its good meat quality. Its length and weight can reach up to 2.45 m and 260 kg (Miller, 1992). The worldwide tuna production continuously increased and reached the highest level to about 7.9 million tons in 2018. Because of overfishing and other human activities, the practical number of the Southern bluefin tuna has decreased significantly, and it has been considered as endangered in the IUCN Red List since 2021 (Collette et al., 2021).

Various tuna species are skilled at continuous and fast swimming (Magnuson, 1973; Blank et al., 2007b). For example, Atlantic bluefin tuna (T. thynnus) are reported to reach the highest speed of 54 km per hour (Wardle et al., 1989). Southern bluefin tuna, Atlantic bluefin tuna and Pacific bluefin tuna (T. orientalis) are collectively called as bluefin tunas. These bluefin tunas can migrate long distances over vast stretches and tolerate a wide range of temperatures. They can maintain body temperature at 10°C higher than the ambient waters (Blank and tolerates a wide range of temperatures. They can maintain body temperature at 10°C higher than the ambient waters (Blank et al., 2007a). They are warm-blooded (rarely among the abundant cold-blooded fishes), enabling them to adapt cold waters and to dive deeper (Bernal et al., 2017). More importantly, bluefin tunas have a high demand for oxygen to keep swimming continuously and at high speeds.

Hemoglobins (Hbs) transport oxygen from the lungs or gills to the rest parts of the body for aerobic respiration in a wide range of animals (Maton et al., 1993). They are usually tetramers with two α and two β subunits (denoted as αβ2), and each is associated with a heme group. These subunits are similar in size and structure, consisting of seven to eight α-helices. A total of 17 hemoglobin genes (hbgs) and one pseudogene, consisting of nine α-globins and eight β-globins, were identified in zebrafish (Opazo et al., 2013). hb genes in various teleost species are often distributed in two separated clusters, named as MN and LA respectively. MN refers to the left Methylpurine-DNA Glycosylase (mpg) and Nitrogen permease regulator-like 3 (nprl3) cluster, while LA refers to the right Leucine carboxyl methyltransferase 1 (lcm1) and Aquaporin-8 (aqp8) cluster (Hardison, 2008; Opazo et al., 2013).

The genome assemblies of several popular tuna species, such as Pacific bluefin tuna (Nakamura et al., 2013; Suda et al., 2019), Atlantic bluefin tuna (Puncher et al., 2018) and yellowfin tuna (Barth et al., 2017) have been published. However, the detailed genome information of Southern bluefin tuna is still limited except for one assembly uploaded in the public NCBI database (accession no. GCA_910596095.1), which can be used for genomics comparison. In previous studies, the hb gene clusters have been well studied in humans and several model fish species, while they remain largely unknown in bluefin tunas. Here, we performed whole genome sequencing of Southern bluefin tuna to provide a valuable genetic resource for identification and characterization of hb genes in this fast-swimming tuna species.

2 Materials and methods

2.1 Sample collection, library construction, and genome sequencing

We collected muscle samples and then extracted genomic DNAs from a Southern bluefin tuna (Figure 1A), which was captured from the Great Australian Bight and cultured around the Port Lincoln of Australia. The extracted genomic DNAs were used to construct five libraries following the Illumina protocols, including three short-insert libraries (270 bp, 500 bp, and 800 bp), and two long-insert libraries (2 kb and 5 kb). Subsequently, based on the routine shotgun sequencing strategy (You et al., 2014), whole genome was sequenced on Hiseq 2500 and Hiseq X-ten platforms (Illumina, San Diego, CA, United States). Raw reads were filtered by SOAP filter (v2.2) (Luo et al., 2015) before assembly.

2.2 Genome size estimation and de novo genome assembly

We conducted a 17-mer distribution analysis to estimate the target genome size based on the short-insert libraries (Liu et al., 2013; Vultur et al., 2017). We calculated the genome size (G) using the following formula as reported before (You et al., 2014): G = k_num/peak_depth (see more explanations in the legend of Figure 1B).

We employed Platanus v1.2.4 (Tokyo Institute of Technology, Tokyo, Japan) with an optimized parameter “-k 29” to obtain a De Bruijn assembly by using Illumina paired-end reads (Kajitani et al., 2014). We generated a total of 1,564,055 contigs (Supplementary Table S5), and then we constructed scaffolds and filled the gaps of intra-scaffolds using the reads of short-insert sizes (270 bp, 500 bp, and 800 bp) with Platanus v1.2.4. To further estimate the completeness of our assembly, the BUSCO (Benchmarking Universal Single-Copy Orthologs; version 1.22) software was applied based on the actinopterygiiodb9 database.

2.3 Repeat sequence analysis

To analyze the assembled genome for repeat sequences, we first employed Tandem Repeats Finder (version 4.07) to search tandem repeats (Benson, 1999). Secondly, we employed RepeatMasker (version 4.0.6) and RepeatProteinMask (version 4.0.6, an updated software in the RepeatMasker package) to...
detect known transposable elements (TEs) based on the public Repbase TE library (release 21.01) (Jurka et al., 2005; Tarailo-Graovac and Chen, 2009; Bao et al., 2015). Thirdly, we applied RepeatModeler (version 1.0.8) and LTR_FINDER (version 1.0.6) with default parameters (Xu and Wang, 2007; Abrusan et al., 2009) to generate a de novo repeat library. Finally, we employed RepeatMasker to identify and classify homologous repeats against this de novo repeat library.

2.4 Genome annotation and evaluation

We utilized three different approaches to annotate detailed structures of the predicted genes in our assembled genome, including de novo prediction, homology-based prediction, and transcriptome-based annotation (You et al., 2014). For the de novo prediction, we employed AUGUSTUS (version 3.2.1) and GENSCAN (version 1.0) to identify protein-coding genes with the repeat-masked genome (Burge and Karlin, 1997; Stanke et al., 2006). For the homology-based prediction, we aligned the homologous proteins of seven representative fish species, including zebrafish (Danio rerio), threespine stickleback (Gasterosteus aculeatus), Nile tilapia (Oreochromis niloticus), medaka (Oryzias latipes), Japanese pufferfish (Takifugu rubripes), green spotted pufferfish (Tetraodon nigroviridis), and Atlantic cod (Gadus morhua) (Ensembl 83 release), to the repeat-masked genome using tblastn (Blastall version 2.2.26) with an e-value ≤ 1e-5 (Mount, 2007; Cunningham et al., 2015). Subsequently, Solar (version 0.9.6) and GeneWise (version 2.4.1) were executed to define the potential gene structures for all alignments (Birney et al., 2004; Li et al., 2010). Next, we employed TopHat (version 2.0.13) to map RNA reads and ensemble genomes, and then we applied Cufflinks (version 2.2.1) to assemble and merge transcripts using the accepted hits of TopHat (Trapnell et al., 2009; Trapnell et al., 2010). Finally, we combined the aforementioned three datasets to obtain a consistent and comprehensive gene set by GLEAN (version 1.0) (Elsik et al., 2007).

The predicted coding proteins of the Southern bluefin tuna were aligned against KEGG (Kyoto Encyclopedia of Genes and Genomes), SwissProt and TrEMBL databases for prediction of functions and pathways by using BLASTP (Kanehisa and Goto, 2000; Boeckmann et al., 2003). Subsequently, we applied InterProScan (version 5.16-55.0) to identify functional motifs and domains through the Pfam, PRINTS, ProDom and SMART databases (Attwood, 2002; Letunic et al., 2004; Bru et al., 2005; Hunter et al., 2009; Finn et al., 2014).

2.5 Genomics comparison of different specimens of southern bluefin tuna

There is a chromosome-level genome assembly of Southern bluefin tuna in NCBI (GCA_910596095.1) without publication of related paper; in order to compare our genome assembly with this public version, we reordered our scaffolds into pseudochromosomes by RaGOO (Alonge et al., 2019) using the public version as the reference. Then, genome-wide alignments were performed using minmap (Li, 2016), and the best homology segments were selected for visualization by SyRI (Goel et al., 2019). The repeat content and gene annotation of the public assembly were also retrieved from.
NCBI. Numbers of genes and exons of each gene were counted, and length distribution of the CDS, exons and introns were plotted by the routine R package.

### 2.6 Identification of hemoglobin genes

We used zebrafish hemoglobin protein sequences, downloaded from the Ensemble database (Supplementary Table S1), as the queries to extract \( h\beta \) genes in the genome assemblies of Southern bluefin tuna (this study), Pacific bluefin tuna, swordfish, sailfish, Atlantic cod, tiger tail seahorse, greater amberjack, yellowtail amberjack, tongue sole, large yellow croaker, European seabass, zebrafish, and ocean sunfish (Supplementary Table S2). We constructed a local database for each fish, and performed tblastn (v2.2.28) to align these sequences with an e-value cutoff of \( 10^{-5} \) for localizing the orthologs of these queries (Mount, 2007). Subsequently, we employed Exonerate (v2.2.0) to evaluate the alignment results within the entire encoding regions (Slater and Birney, 2005).

To further characterize hemoglobin structures, we downloaded sea lamprey, spotted gar and human hemoglobin amino acid sequences (Supplementary Table S1) from the public Ensemble database. In addition, we used sequences of \( h\beta \) genes and 12 neighboring genes from the zebrafish genome to determine syntenic correlations among Southern bluefin tuna and other examined fishes. To evaluate the conservation of \( h\beta \) genes, we used five genes upstream plus two genes downstream of the MN cluster, and three genes upstream plus two genes downstream in the LA cluster for a synteny analysis (Opazo et al., 2013), which included \( mgrp1b \) (Mahogunin ring finger 1b), \( aanat2 \) (Arylalkylamine N-acetyltransferase 2), \( rhbdf1a \) (Rhomboid family member 1a), \( mpg \), \( npr3 \), \( kanl2 \) (KN motif and ankyrin repeat domain-containing protein 2), \( dock6 \) (Dedicator of cytokinesis 6), \( mgrp1a \), \( fosl1b \) (Forkhead box j1b), \( rhbdf1b \), \( aqp8 \), \( lmt1 \), and \( arhgap17a \) (Rho GTPase-activating protein 17a) (see detailed accession numbers in Supplementary Table S1). They are conserved in teleosts, thus we used their sequences to search for corresponding syntenic locations. We obtained the related genome data from NCBI as mentioned above. Subsequently, the routine strategy of protein sequences aligned to nucleotides was employed to examine these extracted synteny genes in various fish genomes by using tblastn (Blastall version 2.2.26) with an e-value \( \leq 1e-5 \) (Mount, 2007). Next, Solar (version 0.9.6) was executed to define the potential gene structures for all alignments (Li et al., 2010). Finally, we draw the figures with GSDS (Gene structure display server v2.0; Hu et al., 2014).

### 2.7 Phylogeny of hemoglobin genes

These \( h\beta \) genes in Southern bluefin tuna were then deduced as protein sequences for phylogenetic analyses. Then, we predicted their best nucleotide substitution model using IQ-TREE (version 1.6.10) under the Bayes information criterion (BIC) (Nguyen et al., 2015; Kalyaanamoorthy et al., 2017). Parameters within the best nucleotide substitution model of WAG + G (bluefin tunas) was applied into PhyML (version 3.0) to construct phylogenetic trees with the method of maximum likelihood (ML) and 4,000 replicates for evaluation of their branch supports (Guindon et al., 2010). We selected the \( h\beta \) genes in Southern bluefin tuna, Pacific bluefin tuna, swordfish, greater amberjack, and spotted gar, for the phylogenetic analysis using jModeltest2 under AIC with the best nucleotide substitution model of GTR + I + G.

### 3 Results

#### 3.1 Summary of the genome assembly and annotation

| Genome assembly | Data |
|-----------------|------|
| Contig N50 size (kb) | 67.38 |
| Scaffold N50 size (Mb) | 3.31 |
| Assembled genome size (Mb) | 806.54 |
| Genome coverage (x) | 168.37 |
| The longest scaffold (bp) | 14,783,706 |
| Genome annotation | Data |
| Protein-coding gene number | 23,403 |
| Annotated functional gene number | 22,485 (96.08%) |
| Unannotated functional gene number | 918 (3.92%) |
| Repeat sequences | 29.27% |

We obtained 173.19 Gb of raw reads and 135.80 Gb of clean reads for genome assembly of the Southern bluefin tuna (Supplementary Table S3). The estimated genome size of Southern bluefin tuna was 852 Mb by the k-mer analysis (Figure 1B; Supplementary Table S4), and the final assembly was 806.54 Mb in length, with contig and scaffold N50 values of 67.38 kb and 3.31 Mb, respectively (Table 1; Supplementary Table S5). The
BUSCO analysis shows that the assembled genome contains 4,390 (95.8%) complete and 105 (2.3%) duplicated BUSCOs, suggesting that the current assembly was qualified for downstream analyses.

We then annotated repeat elements and protein-coding genes in the assembled genome. Repeat sequences accounted for 29.27% (236.1 Mb) of the assembled genome, including 9.87% DNA, 5.04% LINE, 0.28% SINE, 1.65% LTR (Supplementary Tables S6, S7). A total of 23,403 protein-coding genes were annotated, with an average of 9.82 exons and 1800-bp coding sequences per gene (Supplementary Table S8). Based on the public KEGG, SwissProt and TrEMBL databases, 96.08% (22,485) of the predicted genes were assigned to at least one function (Supplementary Table S9). To estimate the quality of our annotated genes, we determined a total of 91.7% complete BUSCOs.

3.2 High-throughput identification of hemoglobin genes

We selected hb genes with the main aim to study their potential roles for continuous swimming with a high speed. A total of 18 hb genes distributing on two scaffolds were identified from the assembled genome, including nine α-hemoglobin genes (hba) and nine β-hemoglobin genes (hbb) (Figure 2A; Table 2). All these hb genes were composed of three exons, which are classic for hb genes (Figures 2B,C). Their total length ranged from 687 bp to 1,846 bp (143–147 amino acids), although the lengths of introns were variable.

Almost all hb genes (except for two) were localized in the MN cluster (scaffold10), displaying in both forward and reverse orientations (upper Figure 2A). They were named from left (close to nprl3) to right sequentially as follows: hba1, a1, a2, b2, a3, b3, a4, b4, a5, b5, a6, b6, a7, b7, b8, and a8. The remaining two hemoglobin genes, named hba9 and hbb9, were identified in the LA cluster (scaffold90840) (lower Figure 2A). Neighboring genes to both MN and LA clusters were also annotated (Figure 2A).

The phylogenetic tree shows that the two types of Southern bluefin tuna hb genes formed two independent groups as expected, i.e., the clades hba and hbb (Figure 2D). Interestingly, hba2, a3, a4, and a5 had similar gene structures with similar lengths of exons and introns (Figure 2B), and they...
were close in evolution (top panel in Figure 2D). Similar phenomenon was also found for hba1, a6, a7, and a8 (Figure 2D). Such a topology supports a common evolutionary origin of the two types of hb genes, and each type underwent expansion through a series of gene duplications and independent evolution (Storz, 2016; Lei et al., 2021).

### 3.3 Genomic comparisons of hemoglobin genes in representative teleost species

To investigate hb genes among various vertebrates, a collinearity analysis of them as well as their neighboring genes was performed using the genomes of sea lamprey, human, spotted gar and 11 teleosts (Table 2). We chose these representatives because 1) their genome assemblies are of high quality; 2) some are model species; 3) they are representatives through the fish tree of life; 4) they are fast-swimming as tunas or slow-swimming in contrast. The conserved genes located in upstream and downstream regions of hbs were used as markers to distinguish MN and LA clusters (Figure 3). Our results showed that the genomic positions of teleost hb gene clusters were highly conserved but the gene number and order in the clusters varied between species, especially in the MN cluster. All fish MN clusters contained the five genes (marn1b, annat2, rhbdfl1a, mpg, and nprl3) upstream and two genes (kank2 and dock6) downstream to the hemoglobin duplicates (Figure 3B). However, LA clusters were relatively diverse among various fishes. Most fishes possessed two genes (aqp8 and lcmt1) upstream and two genes (rhbdf1b and foxj1b) downstream to the hemoglobin family, with exceptions for ocean sunfish, tongue sole and tiger tail seahorse (Figure 3C). We cannot detect a complete hba9 sequence in the LA cluster of zebra fish, therefore only 15 hb genes of zebra fish were presented in Figure 3.

After identification of hb genes in all examined fishes, we found that the fast-swimming tunas and billfishes contained more hb genes than other fishes, suggesting adaptation by obvious expansion of the hb gene family in these lineages (17–18 copies; see Table 2; Figure 3). In contrast, those languid species (such as seahorses, tongue sole and sunfish) had much fewer hemoglobin gene copies (only 4–8; see more details in Table 2; Figure 3). These interesting findings suggest that the copy number variation of hb genes may be involved in the fish swimming activity; it seems that the more hb genes a fish contains, the faster it can swim.

### 3.4 Phylogenetic relationships among representative teleost fishes

To investigate the evolutionary history of fish hb genes, we constructed an unrooted phylogenetic tree using the coding sequences from five teleosts and spotted gar (Figure 4). These hb genes were grouped into two clades of a (with blue branches) and β (with red branches) with seven hb genes of spotted gar as the ancestors (Feng et al., 2014), and clustered into discrete MN and LA clades (Figure 4). The hemoglobin protein in sea lamprey

---

**TABLE 2 Copy numbers of hemoglobin (hb) genes in the examined genomes.**

| Scientific name          | Common name         | hba | hbb | MN cluster | LA cluster | Total | Note* |
|--------------------------|---------------------|-----|-----|------------|------------|-------|-------|
| *Thunnus maccocyii*      | Southern bluefin tuna | 9   | 9   | 16         | 2          | 18    | Fast  |
| *Thunnus orientalis*     | Pacific bluefin tuna | 9   | 7   | 12         | 4          | 16    | Fast  |
| *Xiphias gladius*        | Swordfish           | 8   | 8   | 14         | 2          | 16    | Fast  |
| *Danio rerio*            | Zebrafish           | 8   | 8   | 13         | 3          | 16    | Ref.  |
| *Seriola dumerili*       | Greater amberjack   | 6   | 7   | 11         | 2          | 13    | Fast  |
| *Larimichthys crocea*    | Large yellow croaker | 8   | 5   | 11         | 2          | 13    | Fast  |
| *Seriola aurorivittata*  | Yellowtail amberjack| 5   | 7   | 10         | 2          | 12    | Fast  |
| *Dicentrarchus labrax*   | European seabass    | 6   | 5   | 8          | 3          | 11    | Fast  |
| *Gadus morhua*           | Atlantic cod        | 4   | 5   | 4          | 5          | 9     | Slow  |
| *Mola mola*              | Ocean sunfish       | 5   | 3   | 4          | 4          | 8     | Slow  |
| *Cynoglossus semilaevis* | Tongue sole         | 4   | 3   | 2          | 5          | 7     | Slow  |
| *Hippocampus comes*      | Tiger tail seahorse | 3   | 1   | 2          | 2          | 4     | Slow  |
| *Homo sapiens*           | Human               | 5   | 5   | 5          | 5          | 10    | Ref.  |
| *Lepisosteus osculatus*  | Spotted gar         | 3   | 4   | —          | —          | 7     | Ref.  |
| *Petromyzon marinus*     | Sea lamprey         | —   | —   | 8          | 2          | 10    | Ref.  |

*Fast: fast-swimming fishes; Slow: slow-swimming fishes; Ref.: reference species.
is a monomeric molecule (Hendrickson and Warner, 1971; Schwarze et al., 2014), but in jawed vertebrates it is a tetrameric molecule with two α and two β subunits. It is believed that the proto hb gene was duplicated to form α and β combined hemoglobins (Goodman et al., 1975; Hoffmann et al., 2012; Storz et al., 2013). The phylogenetic tree of hbb genes were grouped into four well-supported subclades, and those genes associated with the LA and MN clusters were resolved as paraphyletic (Figure 4). The genealogical relationships among the orthologous sequences were consistent with known organismal relationships.

All hbb genes derived from LA clusters were clustered in one clade (Figure 4). The LA hbb subclades possessed orthologous sequences derived from each species. The relationship in the MN cluster was more complicated, since the cluster was divided into three subclades. The hbb genes of tunas [T. maccoyii (TM) and T. orientalis (TO)] formed an independent clade (MN clades 3), while in the other teleost species they were in each of the three MN clades. In this study, the MN cluster of both TM and TO possessed a more extensive and varied composition of hb genes when compared with other examined fishes. Similar to the hbb genes, hba genes were arranged into three subclades, with genes from the MN cluster forming a paraphyletic group relative to those from the LA cluster. The hba genes derived from TM and TO in the MN cluster formed the sister group with the other species in MN clade 2 (Figure 4).

### 3.5 Sequence alignment of fish hemoglobin proteins

Complete amino acid sequences of hb genes from representative teleosts and human were aligned for comparing the gene structures and conservation. Protein sequence alignments showed that both α (Supplementary Figure S4) and β (Figure 5) hemoglobins were highly conserved, although the two types varied slightly in several ways. First, the β-hemoglobins contained eight α-helixes (named as A – H), with a length of 147 or 148 amino acid (aa) long, while the α-hemoglobins were 141–144 aa in length, losing the α-helix D. Second, histidines in helixes E and F, the putative binding sites to heme (Bolton and Perutz, 1970; Blouquit et al., 1984; Frischknecht et al., 1996), were completely conserved in all examined β-hemoglobins (blue arrows in Figure 5), whereas
in one zebrfish a-hemoglobin (hba8), the histidine in helix E was mutated to glutamine. Last but most importantly, we observed an interesting substitution of threonine (Thr, T) to cysteine (Cys, C) at the position 39 of β-hemoglobin in the MN cluster, and such a point mutation (Figure 5; Supplementary Figure S5) was found in all reported tuna genomes (T. maccoyii, T. orientalis, T. albacares; Supplementary Table S10). Previous reports have proved that this conserved site was strongly related to the oxygen affinity (Blouquit et al., 1984; Frischknecht et al., 1996), which inspires us to speculate that this linages-specific mutation was potentially conducive to oxygen transport for their high speed and continue swimming in tunas.

4 Discussion

4.1 Comparison of reported tuna genome assemblies

There are eight species in the genus Thunnus (including T. alalunga, T. albacares, T. atlanticus, T. maccoyii, T. obesus, T. orientalis, T. thynnus, and T. tonggol), and four of them had been sequenced and assembled (Supplementary Table S10). This study assembled an 806-Mb draft genome for Southern bluefin tuna, which is an indispensable supplement to the comparative genomic study of this important genus.
Here, we curated all available tuna genome data and made some comparisons. The sizes of these tuna genomes were mainly between 720 Mb and 830 Mb (Supplementary Table S10), with a similar BUSCO completeness, with an exception for *T. thynnus* (Puncher et al., 2018). In fact, the reported genome size of *T. thynnus* reached 944 Mb, but the downloaded genome from NCBI was only 648 Mb with 36.9% complete BUSCOs, which may influence the quality of this genome assembly for comparative analysis. That is likely why some *hb* genes and their flanking genes were not detectable in one cluster of *T. thynnus*. The GC content of the four tuna genomes were similar (Supplementary Table S10), supporting relative reliability of our genome assembly.

Based on the public genome data of a different specimen of Southern bluefin tuna from the NCBI, we anchored our scaffold-level assembly into 24 pseudochromosomes too. A total of 791.3 Mb was anchored, corresponding to 98.1% of our assembled genome with a high coverage (93.92%) of the public assembly (Supplementary Table S12). There were 716.59-Mb syntenic regions between the two genome versions. Subsequently, we performed a collinearity analysis of these two assemblies, which shows perfect one-to-one synteny relationship between each pair of chromosomes, with exceptions of a few translocations, inversions, and duplications (Supplementary Figure S1). After compared the two genome assemblies of Southern bluefin tuna, we found that the distributions of exon length, intron length, exon number and CDS length were similar (Supplementary Figure S2). We also compared the repeat contents and gene sets of the two assemblies. It seems that our genome assembly has a slightly higher repeat content (29.27%) than the public version (26.07%), but both assemblies have an approximate number of predicted genes (23,403 vs. 24,659), and similar distributions of exon number, exon length, intron length and CDS length (Supplementary Figure S2). All these results suggest that our present assembly is somehow comparable to the public chromosome-level version.

### 4.2 Copy number variation of fish hemoglobin genes

The *hb* gene repertoire of Southern bluefin tuna consisted of nine *hba* and nine *hbb* genes. With these 18 genes in total, the Southern bluefin tuna had the largest number of *hb* genes among all examined diploid fishes so far (Table 2). According to previous report, tunas and billfishes can reach speeds of 50 km per hour (Wardle et al., 1989; Wu et al., 2021). Tongue sole is a benthic fish and rarely swims speedily (Chen et al., 2014).
Sunfish can swim at continuous speeds of 1.4 km/h–2.5 km/h (Pope et al., 2010). The swim speed of Atlantic cod is at 0.72 km/h–1.44 km/h (Blaikie and Kerr, 1996). Without demands of speed, the more languid seahorses stay still for most of the time (Blake, 1976). Billfishes, maintaining continuous swimming activities and high speed, and similar in functional morphology to tunas (Wu et al., 2021), have a relatively high copy number of \( hb \) genes as well. Our comparative genomics analysis indicated that this fast-swimming tuna and other active fishes (such as Pacific bluefin tuna, swordfish, and greater amberjack) had more \( hb \) genes than those relatively languid fishes (such as seahorse, sunfish, and tongue sole; see more details in Table 2; Figure 3).

Unlike most mammals, fishes have multiple \( hb \) genes and their copy number variation has been discussed in several studies. These previous reports suggest that variations in \( hb \) copy number may be related to physiological differences in blood oxygen transport and aerobic energy metabolism (Hoffmann and Storz, 2007; Rutjes et al., 2007). As previously discussed, the diversity of hemoglobin in fishes is related to their adaptation to different environmental conditions or habitats (Verde et al., 2006). It has been proposed that less \( hb \) gene copies in polar fishes were related to their sluggish style of life and lower metabolism (Verde et al., 2006; Verde et al., 2007; Dettaï et al., 2008). In a previous report of channel catfish, the embryonic \( hb \) genes showed importance in coping with the low oxygen conditions under heat stress (Feng et al., 2014). In contrast, fishes with higher metabolism and diverse environmental conditions, such as Atlantic salmon, tend to have more \( hb \) copies (Quinn et al., 2010). Tunas always have high rates of energy turnover and metabolism to maintain fast and continuous swimming in an open pelagic environment (Brill, 1996; Korsmeyer et al., 1996; Jensen, 2001), thus more oxygen is absolutely necessary for the high-efficiency metabolism. Improved repertoire of more \( hb \) genes in these fast-swimming fishes may be beneficial for oxygen transportation and metabolism (Hoffmann et al., 2010; Opazo et al., 2013).

### 4.3 Expansion of hemoglobin genes in the tuna lineage

To figure out how the \( hb \) gene family is expanded in Southern bluefin tuna, we performed subsequent synteny and phylogenetic analyses. Our data show that almost all tuna \( hb \) genes (16 out of 18) were localized in the same MN cluster, indicating that the \( hb \) expansion in tuna is a result of emerging \( hb \) genes in the MN instead of the LA cluster. Specially, those genes in the MN cluster encoding \( \alpha \) subunits were clustered together with the \( \beta \)-hemoglobin genes, appearing as pairs of one \( \alpha \) plus one \( \beta \)-globin genes, such as \( hba2-hbb2 \) and \( hba3-hbb3 \) (Figure 2A).

Such an alternate arrangement of \( hb \) genes was reported in some other fish species, which is benefit for efficient oxygen transcription (Flint et al., 2001; Maruyama et al., 2004; Hardison, 2008; Negrisolo et al., 2008; Quinn et al., 2010; Opazo et al., 2013; Feng et al., 2014; Storz et al., 2020; Lei et al., 2021).

The expansion of \( hb \) genes in the MN cluster was not only valid in the Southern bluefin tuna, but also in other tunas such as the yellowfin tuna and Pacific bluefin tuna (Supplementary Figure S3). However, the LA clusters among the tunas were relatively variable. In addition to the four \( hb \) genes, the T. maccocoiy1 and T. alabcarae assemblies had extra copies of \( aqp8-hba-hbb \) repeats downstream, which has not been identified in any other fishes before. The public version assembly probably recovered a full LA cluster, we therefore added the \( aqp8-hba-hbb \) repeats in Figures 2, 3 with the public version as the reference.

Tandem duplicates often have similar coding sequences owing to interparalog gene conversion. A balanced and synchronous expression of \( \alpha \) and \( \beta \) subunits is reported to be good for efficient production of hemoglobin tetramers (Chan et al., 1997; Miyata and Aoki, 1997; Chu et al., 2006; Weatherall and Clegg, 2008) with the same transcriptional polarity, thereby increasing the capacity to transport oxygen. Expansion of the \( \alpha-\beta \) hemoglobin gene pairs is supposed to adapt for high speeds and continues swimming for various environments in tunas. Such expansion of tuna \( hb \) genes in the MN cluster was further supported by the predicted phylogeny (Figure 4), which also provides insights into the origin of those newly emerged \( hb \) genes.

### 4.4 Mutation of \( \beta39 \)Thr-Cys in tunas

We observed an interesting substitution of Thr to Cys at the position 39 of \( \beta \)-hemoglobins in MN clusters of all reported tuna genomes (Figure 5; Supplementary Figure S5). As reported before, the highly conserved residue Thr-39 is often contacted to the heme at the vinyl side chain of the heme ring, thus it is essential for hemoglobin function (Bolton and Perutz, 1970; Frischknecht et al., 1996). Other studies proved that Thr-a38 is also invariant in mammals (Dickerson and Geis, 1983), and it is important for the oxygen binding property (Hashimoto et al., 1993).

In human, there are three known types of mutations at the position 39 in \( \beta \)-hemoglobins, including Hb Hazebrouk (Thr–Pro) and Hb Grove City (Thr–Ser) that may decrease oxygen affinity and cause hemolytic anemia (Blouquit et al., 1984; Taleriaco et al., 2013), and Hb Hinwil (Thr–Asn) that in contrast can elevate oxygen affinity but cause a remarkable reduction of cooperativity (Frischknecht et al., 1996). We analyzed the 12 species in Table 2 (see Supplementary Figures S5, S6) and selected more eleven fishes for further analysis, which represented nine orders and near the tunas in the tree of life (Supplementary Table S11). We confirmed that the mutation
happened only in the MN cluster of all tunas and one possibly in the LA cluster of Boleophthalmus pectinirostris [BP, one of three selected mudskippers; (You et al., 2014)]. Whether it is true in BP and other mudskippers are worthy of more investigations.

Analysis of the coding genes of hemoglobin in various fishes showed that this substitution was ACT/ACC/ACG/ACA to TGC/TGT (Supplementary Figure S6). In the MN cluster of tunas, the codons for Cys are TGC (27/29) and TGT (2/29). While in other fishes and spotted gar, the codons for Thr are ACT (48/67), ACC (14/67), ACG (3/67) and ACA (2/67) (Supplementary Table S11). In the LA cluster of fishes, the codons for Thr are ACT (37/41), ACC (3/41), and TGT (1/41, only in mudskippers). When this site mutates through the path of Thr- > Ser- > Cys, the first position takes A- > T, the second goes by C- > G, and the third position retains C or T. Because we could not detect the codons for Ser (TCT/TCC/TCG/TCA) in all examined fishes at this site, we are not sure that disadvantageous mutations might have occurred. In some previous reports, cysteine residues of human hemoglobins mainly function for local nitric oxide regulation as needed (Pawloski et al., 2001; Feng et al., 2014). Since beta39Thr- > Cys substitution is not observed in other selected fast-swimming fishes (eg, swordfish), it may not be critical for the fast-swimming ability, while it is still a unique position in tunas. This cysteine substitution at the position 39 in tunas may therefore adjust functions of the hemoglobins, which possibly benefits for aerobic respiration and supports the fast-swimming activities of tunas.

5 Conclusion

We performed whole genome sequencing and assembled a draft genome for Southern bluefin tuna, based on which a total of 18 hemoglobin genes were identified. In detail, there were eight hba and eight hbb genes in the MN cluster and one hba and one hbb in the LA cluster. Genomic comparison analysis shows that this fast-swimming tuna had the largest hemoglobin gene number among all examined diploid fishes. The phylogenetic tree supports a common origin of the two types of hemoglobin genes, and each type expanded through gene duplication, followed by independent evolution. In addition, the β39 threonine was mutated to cysteine in the Southern bluefin tuna genome and other tunas, which may help to transport more oxygens for fast and continuous swimming. In summary, we produced a draft genome assembly and predicted improved hemoglobin-related roles for the fast-swimming activities in Southern bluefin tuna.

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 below: the data of genome assembly used in this work were deposited at NCBI under the bioproject PRJNA843977 and biosample SAMN28775172. Raw sequencing reads were deposited at China National Genebank (CNGB) under the project CNP0000961 with accession numbers CNX0447365–0447369.

Ethics statement

The animal study was reviewed and approved by The Institutional Review Board of BGI.

Author contributions

Conceptualization, QS and JX; methodology, XmZ and YH; software, XhZ and CB; data curation, JC, MW, and CH; writing—original draft preparation, XmZ; writing—review and editing, QS, YH, and XY; visualization, XmZ; supervision, QS; funding acquisition, QS and CB. All authors have read and agreed to publish this manuscript.

Funding

This research was financially supported by Shenzhen Science and Technology Innovation Program for International Cooperation (No. GJHZ20190819152407214) and Natural Science Foundation for Fundamental Research in Shenzhen (No. JCYJ20190812105801661).

Conflict of interest

Authors XmZ, YH, CB, XY, XhZ, JC, MW, CH, YY, JX and QS were employed by company BGI Group.

Publisher’s note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1020017/full#supplementary-material
References

Abrusan, G., Grundmann, N., DeMester, L., and Makalowski, W. (2009). TkClass—a tool for automated classification of unknown eukaryotic transposable elements. Bioinformatics 25 (10), 1329–1330. http://dx.doi.org/10.1093/bioinformatics/btp684

Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlacek, F. J., et al. (2019). RaGTOO: Fast and accurate reference-guided scaffolding of draft genomes. Genome Biol. 20 (1), 224. http://dx.doi.org/10.1186/s13059-019-1819-6

Atwood, T. K. (2002). The PRINTS database: A resource for identification of protein families. Brief. Bioinform. 3 (3), 252–263. http://dx.doi.org/10.1093/bib/bbf067

Bao, W., Kojima, K. K., and Kohany, O. (2015). Rebase Update, a database of repetitive elements in eukaryotic genomes. Mob. DNA 6, 11. http://dx.doi.org/10.1186/s13001-015-0041-9

Barth, J. M. I., Damerau, M., Matschner, M., Jentoft, S., and Hanel, R. (2017). Genomic differentiation and demographic histories of atlantic and indo-pacific yellowfin tuna (Thunnus albacares) populations. Genome Biol. Evol. 9 (4), 1084–1098. http://dx.doi.org/10.1093/gbe/evx067

Beitz, E. (2000). TEXshade: Shading and labeling of multiple sequence alignments using LATEX2 epsilon. Genome Biol. 16.2.135

Cunningham, F., Amode, M. R., Barrett, D., Beal, K., Billis, K., Brent, S., et al. (2015). Ensembl 2015. Nucleic Acids Res. 43, D662–D669. http://dx.doi.org/10.1093/nar/gku1010

Dettai, A., di Prisco, G., Lacouttre, G., Parisi, E., and Verde, C. (2008). Inferring evolution of fish proteins: The globin case study. Methods Enzymol. 456, 539–570. http://dx.doi.org/10.1016/j.9678-0983(06)303-3

Dickerson, R. E., and Geis, I. (1983). Hemoglobin: Structure, function, evolution, and pathology. Menlo Park, CA, USA: Benjamin/Cummings Pub Co.

Eluk, C. G., Mackey, A. J., Reese, J. T., Mishina, N. V., Roos, D. S., and Weinstock, G. M. (2007). Creating a honey bee consensus gene set. Genome Biol. 8 (1), R13. http://dx.doi.org/10.1186-7-1-r13

Feng, J., Liu, S., Wang, X., Wang, R., Zhang, J., Jiang, Y., et al. (2014). Channel catfish hemoglobin genes: Identification, phylogenetic and systemic analysis, and specific induction in response to heat stress. Comp. Biochem. Physiol. Part D. Genomics Proteomics 9, 11–22. http://dx.doi.org/10.1016/j.cbpd.2013.11.002

Finn, R. D., Bateman, A., Clements, I., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: The protein families database families. Nucleic Acids Res. 42, D222–D230. http://dx.doi.org/10.1093/nar/gkt1223

Flint, J., Tufarelli, C., Peden, J., Clark, K., Daniels, R. J., Hardison, R., et al. (2001). Comparative genome analysis delimits a chromosomal domain and identifies key regulatory elements in the alpha globin cluster. Hum. Mol. Genet. 10 (4), 371–382. http://dx.doi.org/10.1093/hmg/1.4.371

Goel, M., Sun, H., Jiao, W. B., and Schneeberger, K. (2019). SryR: Finding genomic rearrangements and local sequence differences from whole-genome assemblies. Genome Biol. 20 (1), 277. http://dx.doi.org/10.1186/s13059-019-1911-0

Goodman, M., Moore, G. W., and Matsuda, G. (1975). Darwinian evolution in the genealogy of haemoglobins. Nature 253 (5493), 683–688. http://dx.doi.org/10.1038/253603a0

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 PHYLML. Syst. Biol. 59 (3), 307–321. http://dx.doi.org/10.1093/sysbio/sys010

Hardison, R. C. (2012). Evolution of hemoglobin and its genes. Cold Spring Harb. Perspect. Med. 2 (12), a011627. http://dx.doi.org/10.1101/cshperspect.a011627

Hardison, R. C. (2008). Globin genes on the move. J. Biol. 7 (9), 35. http://dx.doi.org/10.1186/ jbiol92

Hashimoto, M., Ishimori, K., Imai, K., Miyazaki, G., Morimoto, H., Wada, Y., et al. (1993). Site-directed mutagenesis in hemoglobin: Functional and structural study of the intersubunit hydrogen bond of threonine-38(C3beta)alpha at the alpha 1-beta 2 interface in human hemoglobin. Biochemistry 32 (49), 16888–16895. http://dx.doi.org/10.1021/bi00212a038

Hendrickson, W. A. L., and Warner, E. (1971). Structure of lamprey haemoglobin. Nat. New Biol. 232 (33), 197–203. http://dx.doi.org/10.1038/newbio232197a0

Hoffmann, F. G., and Storz, J. F. (2007). The alphadelta-globin gene originated via duplication of an embryonic alpha-like globin gene in the ancestor of tetrapod vertebrates. Mol. Biol. Evol. 24 (9), 1982–1990. http://dx.doi.org/10.1093/molbev/msm127

Hoffmann, F. G., Opazo, J. C., and Storz, J. F. (2010). Gene cooption and convergent evolution of oxygen transport hemoglobins in jawed and jawless vertebrates. Proc. Natl. Acad. Sci. U. S. A. 107 (32), 14274–14279. http://dx.doi.org/10.1073/pnas.1006756107

Hoffmann, F. G., Opazo, J. C., and Storz, J. F. (2012). Whole-gene duplications spurred the functional diversification of the globin gene superfamily in vertebrates. Mol. Biol. Evol. 29 (1), 303–312. http://dx.doi.org/10.1093/molbev/msr207

Hu, B., Jin, J., Guo, A.-Y., Zhang, H., Luo, J., and Gao, G. (2014). Gods 2.0: An upgraded gene feature visualization server. Bioinformatics 31 (8), 1296–1297. http://dx.doi.org/10.1093/bioinformatics/btu187

Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Bults, D., et al. (2009). InterPro: the integrative protein signature database. Nucleic Acids Res. 37, D211–D215. http://dx.doi.org/10.1093/nar/gkn785

Jensen, F. B. (2001). Hydrogen ion binding properties of tuna haemoglobins. Nat. New Biol. 232 (33), 197–203. http://dx.doi.org/10.1038/newbio232197a0

Kajita, J., Kapitanov, V. V., Pavlicek, A., Kloumovsky, P., Kobany, O., and Walchli, C. (2005). Rebase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. 110 (1-4), 462–467. http://dx.doi.org/10.1159/000084979
Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Ogura, Y., Okuno, M., et al. (2014). Efficient de novo assembly of highly heterogeneous genomes from whole-genome shotgun short reads. *Genome Res.* 24 (8), 1384–1395. doi: 10.1101/gr.170720.113

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jerison, L. S. I. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14 (6), 587–589. doi: 10.1038/nmeth.4285

Kanehsa, M., and Goto, S. (2000). Kegg: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28 (1), 27–30. doi: 10.1093/nar/gan28

Korsmeyer, K., Dewar, H., Lai, N., and Graham, J. (1996). The aerobic capacity of shrimps: Effect of feeding and handling stress on muscle oxygen consumption. *J. Fish. Biol.* 49 (1), 129–137. doi: 10.1111/j.1095-8641.1996.tb00095.x

Lei, Y., Yang, L., Jiang, H., Chen, J., Sun, N., Lv, W., et al. (2021). Recent genome duplications facilitate the phenotypic diversity of Hb repertoire in the Cyprinidae. *Sci. China. Life Sci.* 64 (7), 1149–1164. doi: 10.1007/s11427-020-1809-0

Liu, L., Fan, W., Tian, G., Zhu, H., Li, C., Li, J., et al. (2010). The sequence and de novo assembly of the giant genome of a giant predator. *Nature* 463 (7297), 311–317. doi: 10.1038/nature08696

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., et al. (2015). Erratum: SOAPdenovo2. An empirically improved memory-efficient short-read de novo assembler. *Genome Biol. 4*, 30. doi: 10.1186/s13037-015-0069-2

Miyata, M., and Aoki, T. (1997). Head-to-head linkage of carp alpha- and beta-globin genes. *Biochim. Biophys. Acta* 1376 (1), 17–24. doi: 10.1016/S0005-2760(97)00043-1

Nawar, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Mol. Biol. Evol.* 34 (3), 620–638. doi: 10.1093/molbev/msx069

Quinn, N. L., Borovec, K. A., Lubieniecki, K. P., Chow, W., Davidson, E. A., Phillips, R. B., et al. (2010). Genomic organization and evolution of the Atlantic salmon hemoglobin repertoire. *BMC Genomics* 11, 539. doi: 10.1186/1471-2164-11-53

Rutjes, H., Nieven, M. C., Weber, R. E., Witte, F., and Van den Thillart, G. (2007). Multiple strategies of Lake Victoria cichlids to cope with life-long hypoxia include hemoglobin switching. *Ann. Physiol. Regul. Integ. Comp. Physiol.* 293 (3), R1376–R1383. doi: 10.1016/j.apregu.2006.05.006

Schwarze, K., Campbell, K. L., Hankeln, T., Storz, J. F., Hoffmann, F. G., and Burmester, T. (2014). Theoglobin gene repertoire of lampreys: Convergent evolution of hemoglobin and myoglobin in jawed and jawless vertebrates. *Mol. Biol. Evol.* 31 (10), 2708–2721. doi: 10.1093/molbev/mzu116

Slater, G. S., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6, 31. doi: 10.1186/1471-2105-6-31

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

Storz, J. F. (2016). Gene duplication and evolutionary innovations in hemoglobin- oxygen transport. *Physiology* 31 (3), 223–232. doi: 10.1152/physiol.00060.2015

Storz, J. F., Natarajan, C., Groudeff, M. K., Vandeweghe, M., Hoffmann, F. G., You, X., et al. (2020). Oxygenation properties of hemoglobins and the evolutionary origins of isoform multiplicity in an amphibious, air-breathing fish, the blue-spotted mudskipper (*Boleophthalmus pectinirostris*). *J. Exp. Biol.* 223 (2), 3217. doi: 10.1242/jeb.217307

Tanaka, K., Ichiki, K., Watanabe, S., and Sasa, M. (2021). Recent genome duplications provide insights into the terrestrial adaptation of amphibious fishes. *Nature* 589, 611–615. doi: 10.1038/s41586-021-03799-3

Taliercio, R. M., Ashton, R. W., Horwitz, L., Swanson, K. C., Wendl, P. C., Hoyer, J. D., et al. (2013). HG barley gis. *db (Thunus orintalsis) reference genome and development of male-specific DNA markers. Sci. Rep.* 3 (1), 14450. doi: 10.1038/srep014450

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis version 5.0 for bigger datasets. *Bioinformatics* 27 (21), 2422–2423. doi: 10.1093/bioinformatics/btr467

Weatherall, D. J., and Clegg, J. B. (2008). "Human haemoglobin," in *Human biology and health*. Editors O. Gon and P. C. Heemstra (Grahamstown J.L.B.: Smith Institute of Ichthyology). 301–315.

Xu, Z., and Wang, H. (2007). LTR_FINDER: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 31 (1), 140–153. doi: 10.1093/molbev/msm121

Xu, Z., and Wang, H. (2007). LTR_FINDER: An efficient tool for the prediction of full-lengthLTR retrotransposons. *Nucleic Acids Res.* 35, W265–W268. doi: 10.1093/nar/gkm286

You, X., Bian, C., Zan, Q., Xu, X., Liu, X., Chen, J., et al. (2014). Mudskipper genomes provide insights into the terrestrial adaptation of amphibious fishes. *Nat. Commun.* 5, 5994. doi: 10.1038/ncomms6094