Mu et al. *Zool. Res.* 2026, 47(6): 2054−2066
https://doi.org/10.24272/j.issn.2095-8137.2026.191
https://cstr.cn/32248.14.zr.2026.191

ZR | Zoological Research

Article Open Access

Osteoglossid pangenome reveals ancient teleost evolution and population dynamics

Xi-Dong Mu[1,#], Ru-Jing-Wen Huan[2,#], Ye-Xin Yang[1,3], Yuan-Yuan Wang[1], Huan-Lei Li[4], Yi Liu[1,*], Chao Bian[2,*]

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

[2] *College of Life Sciences and Oceanography*, *Shenzhen University*, *Shenzhen*, *Guangdong* 518060, *China*

[3] *Key Laboratory of Aquatic Animal Immune Technology of Guangdong Province*, *Guangzhou*, *Guangdong* 510380, *China*

[4] *Heze Yellow River Beach Agricultural Development Co.*, *Ltd*, *Heze*, *Shandong* 274000, *China*

ABSTRACT

Osteoglossidae, one of the most ancient freshwater fish lineages exhibiting transcontinental disjunction, offers important insights into early teleost evolution and historical biogeography. Using integrated sequencing strategies, we report two telomere-to-telomere (T2T) chromosome-level genome assemblies for pirarucu (*Arapaima gigas*) and Australian bonytongue (*Scleropages jardinii*) and a near T2T genome assembly for black arowana (*Osteoglossum ferreirai*). The BUSCO completeness was >97% for all the assemblies, markedly surpassing most of the previous osteoglossid assemblies. Using these reference genomes, we reconstructed a time-calibrated phylogeny, estimated major divergence times and evaluated vicariance hypotheses related to Gondwana. We found that the divergence between *Osteoglossum* and *Scleropages* is estimated at 52.6 million years ago (Ma), supporting the hypothesis that their vicariant events may have resulted from the separation of Australia from the Antarctica–South America landmass. In addition, we reconstructed and analyzed the ancestral karyotypes of *Scleropages* and *Osteoglossum*, revealing five and ten major chromosomal rearrangements, respectively. We further systematically reconstructed and parsed the historical population dynamics across most of the Osteoglossidae species, demonstrating dramatic declines in effective population sizes for most Osteoglossidae species during the Last Glacial Period. Together, our findings offer a comprehensive genomic perspective on the evolutionary history, genome architecture, and historical biogeography of Osteoglossidae while also providing essential resources for future conservation and research.

Keywords: Osteoglossidae; T2T genomes; Species separation; Ancestor chromosomes; Population dynamics

INTRODUCTION

Osteoglossidae represents one of the earliest divergent teleost lineages, offering a valuable opportunity to investigate deep-time teleost evolution, genome architectures, and historical biogeography. These ancient freshwater fishes exhibit a striking intercontinental distribution across South America, Africa, Southeast Asia, and Australia, making them a classic model for testing the hypotheses of Gondwanan vicariance (Betancur-R et al., 2017; Hilton & Lavoué, 2018; Schoch et al., 2020).

Among the extant osteoglossids, three genera—*Arapaima* (e.g., *Arapaima gigas,* pirarucu) and *Osteoglossum* (e.g., *Osteoglossum ferreirai,* black arowana) in South America and *Scleropages* (e.g., *Scleropages jardinii,* Australian bonytongue) in northern Australia—offer detailed insights into dissecting parallel and divergent evolutionary trajectories. These species share physiological adaptations such as facultative air breathing and tolerance to hypoxia (de Oliveira et al., 2019; Queiroz & Camargo, 2008); however, they occupy ecologically and geographically distinct habitats, ranging from the oxygen-poor floodplain lakes of the Amazon (pirarucu) to acidic blackwater environments (black arowana) and clear, fast-flowing escarpment rivers in northern Australia (Australian bonytongue) (de Bello Cioffi et al., 2019; de Oliveira et al., 2019; de Souza et al., 2019; Escobar Lizarazo et al., 2013;

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2026 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

Received: 02 April 2026; Accepted: 27 May 2026; Online: 10 June 2026

Foundation items: This study was supported by the Guangdong Provincial Special Fund for Modern Agriculture Industry Technology Innovation Teams (2024CXTD26), China-ASEAN Maritime Cooperation Fund (CAMC-2018F), National Freshwater Genetic Resource Center (FGRC18537) and the Guangdong Basic and Applied Basic Research Foundation (no.2026A1515011346)

[#]Authors contributed equally to this work

*Corresponding authors, E-mail: bianchao@szu.edu.cn; seetohrb@126.com

Hrbek et al., 2005; Stewart, 2013). Despite their phylogenetic proximity, they exhibit marked differences in chromosome karyotype (2*n*=48–56), suggesting that chromosomal rearrangements played uncharacterized roles in speciation (de Bello Cioffi et al., 2019; de Oliveira et al., 2019; de Souza et al., 2019).

Although the osteoglossid phylogeny has been broadly established, the timing and mechanism of their global radiation remain controversial. For example, estimates of the divergence time between *Scleropages* and *Osteoglossum* range from 31.3 to 80.0 million years (Ma) (Capobianco & Friedman, 2024; de Bello Cioffi et al., 2019; Lavoué, 2016; Peterson et al., 2022). While some previous studies supported a Gondwanan origin and vicariance-driven divergence (Hao et al., 2020; Kumazawa & Nishida, 2000; Mu et al., 2014; Sparks & Smith, 2005), others pointed to long-distance marine dispersal as a more plausible explanation (Capobianco & Friedman, 2019, 2024; de Bello Cioffi et al., 2019; Lavoué, 2015, 2016; Peterson et al., 2022).

Recent breakthroughs in long-read sequencing technologies have enabled telomere-to-telomere (T2T) genome assembly in teleosts, providing unprecedented resolution for fish genomics. The T2T genome of the large yellow croaker (*Larimichthys crocea*) has characterized its centromeric architecture and adaptive divergence (Cui et al., 2025), while the gapless T2T assembly of the grass carp (*Ctenopharyngodon idella*) elucidates its potential for genetic improvement (Liu et al., 2025). Furthermore, T2T-level assemblies of black and golden koi carp variants reveal critical chromosomal rearrangements and evolutionary mechanisms within Cyprinidae (Bian et al., 2025). By resolving previously inaccessible genomic regions—including centromeres, telomeres, and complex repeats—T2T genome assemblies can provide highly accurate references for further analyses, such as divergence time estimation and genomic feature analysis.

Moreover, osteoglossid fishes face growing conservation threats from human activities, particularly commercial harvesting and the ornamental trade. Once central to Amazonian fisheries, the population of pirarucu has declined, leading to genetic drift and differentiation (Araripe et al., 2013; Hrbek et al., 2005). Similarly, black arowana, despite its IUCN Least Concern status (this status may require reassessment), is classified as Endangered (EN) in the Red Book of Colombian Freshwater Fishes (de Souza et al., 2019; Salvador et al., 2023). Furthermore, the black arowana likely requires a more elevated conservation status because of its restricted distribution in specific water types (de Souza et al., 2019).

Here, we present three high-quality genome assemblies for black arowana (*O. ferreirai*), pirarucu (*A. gigas*), and Australian bonytongue (*S. jardinii*). Notably, the assemblies of pirarucu and Australian bonytongue achieved T2T level, whereas the black arowana assembly attained near-T2T level. These genomic resources allow us to reconstruct ancestral chromosome karyotypes, quantify lineage-specific genome expansions and contractions, and infer demographic histories in relation to paleogeography and climatic oscillations. Our results offer a comprehensive evolutionary perspective on one of the most ancient fish families, shedding light on the genetic underpinnings of their biogeographic diversification and providing a solid foundation for future conservation and functional studies.

MATERIALS AND METHODS

Sample collection and DNA extraction

All procedures for handling and treatment of experimental fishes during this study were approved by the Laboratory Animal Ethics Committee of the Pearl River Fisheries Research Institute (PRFRI), Chinese Academy of Fishery Sciences (CAFS), China. All animal experiments were carried out in accordance with the Guide for the PRFRI in Guangzhou, China. The black arowana, pirarucu and Australian bonytongue samples used in this study were obtained from the National Freshwater Genetic Resource Center, Guangzhou, China. Their muscle tissues were separately dissected for genomic DNA (gDNA) extraction using the TIANamp Genomic DNA Kit (TIANGEN, China) according to the manufacturer's protocol. The concentration and size of the DNA were assessed using a Qubit 3.0 fluorometer (Thermo Fisher Scientific, USA) and 1% agarose gel electrophoresis.

Genomic and transcriptome library construction and sequencing

Genomic DNA from black arowana, pirarucu and Australian bonytongue was used to construct sequencing libraries for MGI, PacBio HiFi, Oxford Nanopore (ONT), and Hi-C technologies. MGI libraries with an insert size of 500 bp were prepared using the MGIEasy UDB Universal Library Prep Set following the manufacturer's instructions and sequenced on the DNBSEQ-T7 platform (MGI Tech Co. Ltd., China). PacBio HiFi libraries were generated using the SMRTbell Express Template Prep Kit 2.0 following the PacBio standard protocol and sequenced on the PacBio Sequel II system (Pacific Biosciences, USA) (Rhoads & Au, 2015). Circular Consensus Sequencing analysis was performed using SMRT Link v.9.0 (Pacific Biosciences) (Chin et al., 2013) to generate high-fidelity consensus reads. ONT libraries were constructed using an SQK-ULK001 kit according to the instructions from the manufacturer and sequenced on a PromethION flow cell (Oxford Nanopore Technologies, UK). These ONT reads were corrected by NECAT v.200221 software with default parameters (Chen et al., 2021). Hi-C libraries were prepared using a GrandOmics Hi-C Kit and *DpnII* enzyme according to the manufacturer's standard protocol (GrandOmics, China) and sequenced on an Illumina NovaSeq platform.

Total RNA was extracted from muscle tissues of three Osteoglossidae species using a TRIZOL Kit (Invitrogen, USA) following the manufacturer's instructions. The integrity and quality of the extracted RNA were assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies, USA), and only samples demonstrating RNA Integrity Number (RIN) values >7.0 were retained for downstream library preparation. cDNA libraries were constructed using DNA nanoballs (DNBs) according to the DNBSEQ platform manufacturer's protocol (MGI, China). These libraries were subsequently sequenced on the MGISEQ-2000 platform (MGI, China) using a paired-end 150 bp sequencing strategy.

Genome size prediction, assembly, and quality evaluation

The sizes of the genomes of black arowana, pirarucu and Australian bonytongue were estimated using 17-mer and 21-mer frequency analyses of MGI sequencing data (Marçais & Kingsford, 2011). Three initial genome assemblies of black arowana, pirarucu and Australian bonytongue were assembled using Hifiasm v.0.19.8 (parameters: -t 16 --n-hap 4

--hg-size 1520m) (Ruan & Li, 2020) with PacBio HiFi and ONT reads. Hi-C data were aligned to the above-generated contigs using Bowtie 2 v.2.2.5 (parameters: --very-sensitive -L 30 --score-min L, -0.6, -0.2 --end-to-end --reorder) (Langmead & Salzberg, 2012). Chromosome sequences were subsequently constructed using YaHs v.1.0 (Zhou et al., 2023) and 3D-DNA v.170123 (parameters: -m haploid -r 2) (Dudchenko et al., 2017). To correct misassemblies and reduce redundancies, we made additional adjustments using Juicebox (v.1.11.08) (Durand et al., 2016). To achieve T2T-level genome assemblies, we employed LR_GapCloser v.1.0 (parameters: -t 35 -m 1000000 -v 10000) (Xu et al., 2019) and TGS-GapCloser v.1.0.1 (parameters: -min_match 2000) (Xu et al., 2020) for gap closure. PacBio HiFi and Oxford Nanopore Technologies (ONT) reads were also aligned against three genome assemblies using Minimap2 (v.2.24) (Li, 2018) to improve structural accuracy. Centromeric and telomeric regions were identified using QuarTeT software (v.1.6) (Lin et al., 2023). Assembly completeness results from the above three genome assemblies were evaluated using Benchmarking Universal Single-Copy Orthologs (BUSCO v.5.4.3) (Waterhouse et al., 2018). Genome quality values (QVs) were calculated with Merqury at a k-mer depth threshold of 17 (Rhie et al., 2020).

Repeat element annotation

Repetitive elements in three genomes were annotated using an integrated approach by combining *de novo* prediction and homology-based identification. *De novo* repeat detection was performed using RepeatModeler v.1.0.8 (Abrusán et al., 2009) and LTR_Finder v.1.0.6 (Xu & Wang, 2007) to comprehensively identify diverse repeat types. Custom repeat libraries were subsequently constructed by merging *de novo* predicted repeats with reference sequences from Repbase TE v.21.01 (Jurka et al., 2005) through RepeatMasker v.4.0.6 (Tarailo-Graovac & Chen, 2009). Tandem repeats were annotated using Tandem Repeats Finder (v.4.09) (Benson, 1999) with the following parameters: 2 7 7 80 10 50 2000 -d -h. The refined repeat libraries were subsequently applied to annotate transposable elements using both RepeatProteinMask v.4.1.0 and RepeatMasker v.4.1.0 with default parameters (Tarailo-Graovac & Chen, 2009).

For the insertion time analysis of long terminal repeat (LTR) retrotransposons, we first selected pairs of LTRs with complete 5' and 3' ends from the above LTR-FINDER results. Self-alignment of these LTR pairs was performed using MUSCLE (v.3.8.31) (Edgar, 2004). The nucleotide divergence (K) for each aligned pair was subsequently calculated using the Kimura two-parameter model from the EMBOSS suite (Rice et al., 2000). Finally, the insertion time (T) was determined according to the formula T=K/2r, applying a mutation rate (r) of 3.51×10^{-9} substitutions per site per year (Eyre-Walker, 2000).

Gene prediction and functional annotation

Protein-coding genes were annotated through the integration of homology-based prediction and transcriptome-based prediction. For homology-based prediction, protein sequences from four phylogenetically representative species, namely, *Danio rerio* (zebrafish), *Oryzias latipes* (medaka), *Heterotis niloticus* (African arowana), and *Scleropages formosus* (Asian arowana), were downloaded from NCBI and aligned to the assembled genomes using TBLASTN (e-value 10^{-5}) (Mount, 2007). Gene structures were subsequently predicted from these alignments using GeneWise v.2.2.0 (Birney et al., 2004) with optimized parameters (parameters: --blast_eval $1e^{-5}$, --align_rate 0.5, --extend_len 500). Transcriptome reads were aligned to the genomes via HISAT2 (Kim et al., 2019), followed by transcript assembly using Cufflinks v.2.2.1 (Pollier et al., 2013) to generate transcriptome-derived gene models. These homology- and transcriptome-based predictions were merged through MAKER (parameters: max_dna_len=300,000, min_contig=500, pred_flank=500, AED_threshold=1) (Cantarel et al., 2008) to produce nonredundant gene sets. The final protein-coding genes were functionally annotated against the SwissProt, TrEMBL, Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO), and InterPro databases (Ashburner et al., 2000; Boeckmann et al., 2003; Hunter et al., 2009; Kulikova et al., 2004; Ogata et al., 1999).

Gene families, phylogenetic tree construction, and divergence time estimation

The construction of single-copy orthologous gene families enabled phylogenetic analysis and estimation of divergence times. Chromosome-level genomes of four osteoglossiform species were obtained from NCBI and our previous studies, including *Gnathonemus petersii (*elephantnose fish), African arowana, *Osteoglossum bicirrhosum* (silver arowana), and golden arowana. We performed cross-species protein sequence alignments using BLASTp (McGinnis & Madden, 2004). The alignment results were clustered into homologous gene families using OrthoMCL (Fischer et al., 2011). All single-copy orthologous gene families were subjected to multiple sequence alignments with MUSCLE v.3.8.31 (Edgar, 2004), after which the sequences of the conserved regions were filtered using Gblocks (Castresana, 2000). Finally, this dataset was used to construct a phylogenetic tree using the maximum likelihood method via PhyML v.4.9 (Guindon et al., 2010). The MCMCtree software of PAML v.4.9 (Yang, 2007) was used to estimate the divergence times. Here, we used the fossil † *Scleropages sinensis* (47.8–56 Ma) to calibrate our phylogenetic tree (Capobianco & Friedman, 2024; Zhang & Wilson, 2017).

Gene family expansion and contraction were assessed using CAFE v.4.2.1 (De Bie et al., 2006). This approach compared gene family sizes between ancestral and target species, applying a random birth-death model to detect gene gain and loss along lineages. Gene families were excluded from the analysis if they contained more than 200 genes in one species or fewer than 2 genes across all other species.

Construction of ancestral chromosomes

According to previous studies, the pre-WGD (whole-genome duplication) teleost ancestral karyotype consisted of 13 chromosomes (a–m). After eight major chromosomal rearrangements, the teleost ancestor possessed 24 chromosomes following the WGD event (Kasahara et al., 2007; Bian et al., 2016; Lu et al., 2022; Bian et al., 2024a; Bian et al., 2024b). On this basis, the protein dataset of each species was aligned by BLASTP (McGinnis & Madden, 2004) with an E-value threshold of $1e^{-10}$ against the predicted teleost ancestor protein sequences. After the reciprocal best hits were obtained, aligned regions with <20 hits shared color-coded segments per chromosome were filtered out. Chromosome rearrangement events (fissions, fusions, translocations) were inferred from the retained synteny blocks. Finally, perl-based SVG figures were used to show these conserved syntenic regions on the reconstructed osteoglossomorph ancestor

karyotype, using corresponding color lumps for orthologous chromosomal segments.

Analysis of historical population dynamics

The pairwise sequentially markovian coalescent (PSMC) model was used to estimate the most recent common ancestor of Osteoglossiformes at heterozygous sites (Li & Durbin, 2011). For each species, MGI reads were aligned to each genome using BWA v.0.7.17 (Li & Durbin, 2009) to generate diploid genome heterozygosity information. Heterozygous sites were identified using SAMtools. The whole-genome diploid consensus sequences were generated using SAMtools and BCFtools (Li, 2011) on the basis of the parameter C50 -d8000 -uf. The program fq2psmcfa was used to transform the consensus sequence into a FASTA-like format, where the characters indicate heterozygous positions in consecutive bins of 100 bp. The distribution of the time to the most recent common ancestor was estimated using an assumed mutation rate (μ) and the observed generation time (also referred to as the age at sexual maturity) to reconstruct the historical population dynamics of each species. The mutation rate (μ=5.55×10^{-9} per year; Wang & Obbard, 2023) was used, and different generations were used, including 4 years for Asian arowana (three color variants; Yue et al., 2020) and 2 years for African bonytongue (Adekeye, 1993), 5 years for pirarucu (Torati et al., 2019) and Australian bonytongue (Merrick et al., 1983), and 2 years for silver arowana and black arowana (de Souza et al., 2019). The historical population results of each Osteoglossidae species were subsequently visualized using gnuplot v.4.4 (Janert, 2009).

RESULTS

Summary of three novel T2T genome assemblies

High-quality genomic sequencing data were generated for black arowana, pirarucu and Australian bonytongue using a combination of MGI short-read, PacBio HiFi long-read, ONT ultralong read, and Hi-C technologies. For these speciesspecies, there were 51.4, 56.9, and 50.2 Gb of MGI reads; 84.2, 133.0, and 101.1 Gb of PacBio HiFi reads; and 36.1, 26.4, and 35.5 Gb of ONT ultralong reads were generated, respectively. Hi-C sequencing yielded 120.0, 165.5, and 128.7 Gb of data for black arowana, pirarucu and Australian bonytongue, respectively, enabling chromosome-level scaffolding (Table 1; Supplementary Tables S1–S3).

Through *k*-mer analysis of the MGI data, we estimated that the genome sizes of black arowana, pirarucu and Australian bonytongue were 791.2, 678.2 and 703.7 Mb, respectively (Supplementary Figure S1; Supplementary Table S4). By integrating the sequencing reads above with Hi-C technology, we achieved a near-T2T genome assembly for black arowana and T2T assemblies for pirarucu and Australian bonytongue. The black arowana genome assembly was ultimately anchored onto 27 chromosomes (Figure 1A), with contig N50 and scaffold N50 values of 26.9 and 27.7 Mb, respectively (see detailed length of each chromosome in Supplementary Table S5). The pirarucu and Australian bonytongue genome assemblies were resolved into 28 and 24 chromosomes (Figure 1B, C), with contig N50 values reaching 27.0 and 33.1 Mb, respectively (see detailed length of each chromosome in Supplementary Tables S6 and S7). BUSCO assessments revealed 97.0%, 97.7%, and 97.1% completeness for the black arowana, pirarucu and Australian bonytongue assemblies, respectively. Genome quality assessment using Merqury showed high consensus accuracy for all three assemblies, with QV values of 54.6 for black arowana, 57.0 for pirarucu, and 60.3 for Australian bonytongue. These comprehensive evaluations confirm the high quality and completeness of all three genome assemblies.

Repeat element and gene annotation data

Repetitive elements accounted for 42.1%, 19.4%, and 28.9% of the black arowana, pirarucu, and Australian bonytongue genomes, respectively (Supplementary Table S8). Protein-coding genes were predicted through homology-based and transcriptome-supported methods, yielding 28 110, 22 745, and 29 963 genes for black arowana, pirarucu, and Australian bonytongue, with average transcript lengths of 15.5, 17.2 and 15.5 kb, respectively (Supplementary Tables S9–S11). Functional annotations assigned roles to 93.8%, 95.6%, and 94.6% of genes using the SwissProt, KEGG, TrEMBL, GO, and InterPro databases. BUSCO assessments of protein-coding genes revealed 93.3%, 92.4%, and 91.2% completeness, respectively (Supplementary Table S12).

Table 1 Statistics of the genome assemblies for three Osteoglossidae species and other published genome assemblies

Parameter	black arowana	pirarucu	Australian bonytongue	golden arowana	African arowana	silver arowana
MGI reads (Gb)	51.4	56.9	50.2	/	/	/
HiFi reads (Gb)	84.2	133.0	101.1	27.4 (PacBio)	/	75.6 (PacBio)
ONT reads (Gb)	36.1	26.4	35.5	/	10.2 (Nanopore)	/
Hi-C reads (Gb)	120.0	165.5	128.7	80.7	21.2	173.4
Genome size (Mb)	941.9	676.7	789.6	781.5	650.4	739.8
Contig N50 (Mb)	26.9	27.0	33.1	2.7	2.3	12.6
Scaffold N50 (Mb)	27.7	/	/	29.8	32.4	27.1
Gap number	2	0	0	795	608	503
GC rate	44.2%	43.0%	42.8%	40%	43.0%	45.1%
Chr number	27	28	24	25	20	28
BUSCO	97.0%	97.7%	97.1%	96.5%	97.6%	96.7%
Repeat ratio	42.1%	19.4%	28.9%	27.8%	18.7%	25.7%
Gene number	28 110	22 745	29 963	25 328	24 146	30 221
Average gene length (bp)	15 381.6	17 163.8	15 494.6	12 358.4	14 911.2	13 824.8
Functional gene number	26 354	21 741	28 354	22 250	21 621	27 242
Protein BUSCO	93.3%	92.4%	91.2%	/	94.4%	87.3%

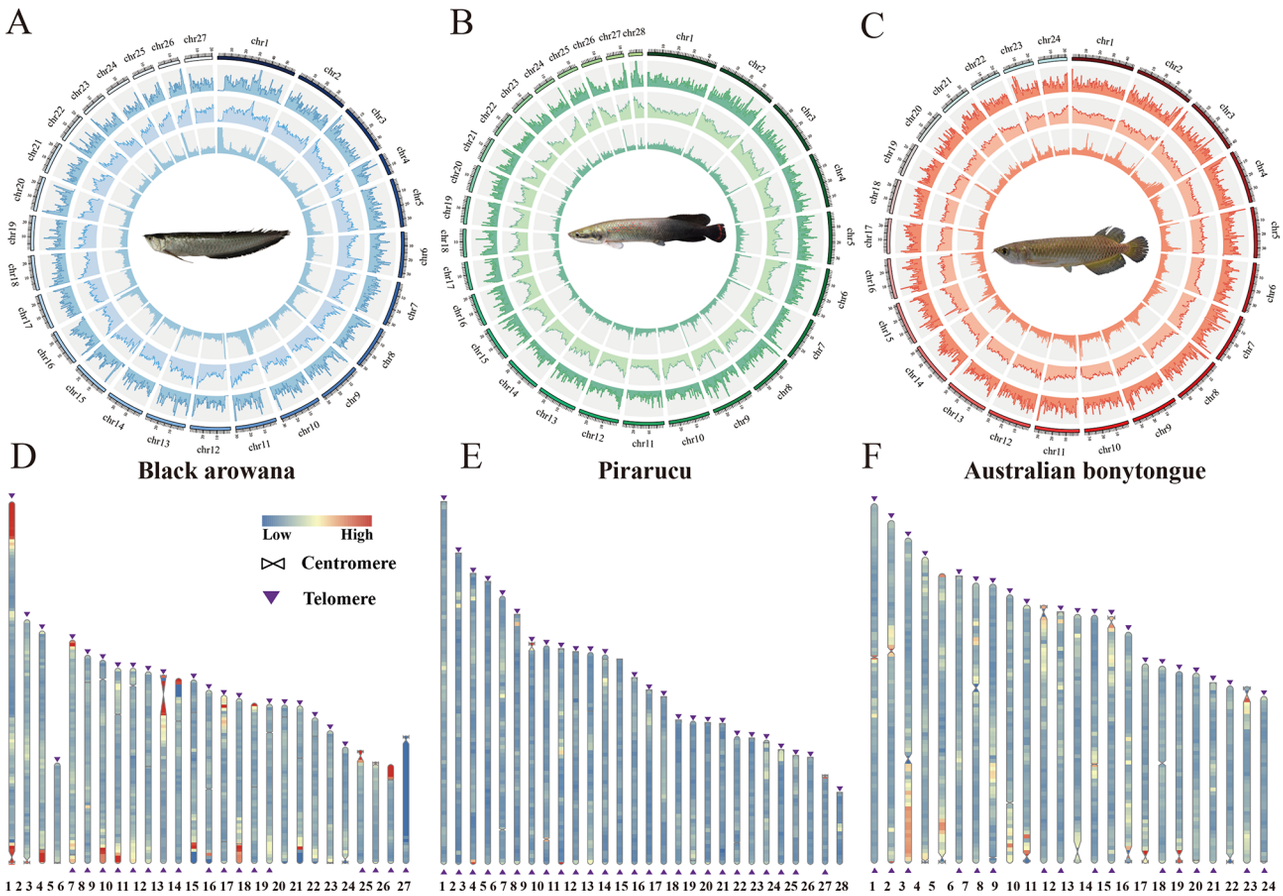


Figure 1 Genome assemblies of black arowana, pirarucu and Australian bonytongue (from left to right)

A–C: Circos plots of these genome assemblies. Data from the outside to the inside include the length and number of each chromosome, gene density distribution (bin=0.1 Mb), GC content (bin=0.1 Mb), repetitive sequence density (bin=0.1 Mb), and the fish photograph (in the center). D–F: Localization of telomeres and centromeres and the distribution density of repetitive sequences in these assembled genomes.

Divergence genome sizes in Osteoglossidae species driven by TEs

The black arowana genome was markedly larger than those of all the other Osteoglossidae species analyzed, ranging from 19.3% to 44.8% greater in size, primarily because of its higher transposable elements (TE) and intron contents (Figure 2A). As major components of vertebrate genomes, TEs play a fundamental role in shaping genome architecture through episodic amplification. Comparative genomic analysis revealed extensive expansion of both class II DNA transposons and class I retrotransposons, including long interspersed nuclear elements (LINEs), long terminal repeats (LTRs), and short interspersed nuclear elements (SINEs) in black arowana, whose LINEs were nearly twice as long as those in related genomes (Figure 2A). Conversely, these repeat elements account for a relatively smaller proportion of the genomes in the other two annotated species, representing 19.4% of the pirarucu genome (52.4 Mb of DNA transposons, 25.2 Mb of LINEs, 21.9 Mb of LTRs, and 11.9 Mb of SINEs) and 28.9% of the Australian bonytongue genome (76.3 Mb of DNA transposons, 51.2 Mb of LINEs, 34.2 Mb of LTRs, and 25.4 Mb of SINEs). Kimura distance-based divergence profiling revealed a recent TE burst event in Osteoglossidae, dominated by specific DNA transposon and LINE families. These elements, including DNA/TcMar-Tc1, DNA/PIF-Harbinger, LINE/L1, and LINE/CR1, were evenly distributed across 27 chromosomes of the black arowana genome (Figure 2B).

Phylogenetics, divergence time and gene family

Using genome-wide single-copy orthologs, we reconstructed a time-calibrated phylogeny to infer the divergence times among osteoglossid lineages (Figure 3A). The resulting phylogenetic tree is consistent with the topology of previous trees based on the mitochondrial genomes of black arowana from our own previous study (Mu et al., 2014) and Asian arowana (Gao & Zhang, 2017), confirming similar evolutionary relationships. In this phylogenetic tree, the divergence time between African arowana and pirarucu was 68.0 Ma (48.3–98.1). The divergence between the Asian–Australian clade *Scleropages* and South American *Osteoglossum* was estimated at 52.6 Ma (48.1–56.1). Within *Scleropages*, the divergence between the Australian bonytongue and the Asian arowanas (including golden, red, and green varieties) was estimated at 38.2 Ma (20.2–50.8 Ma).

To further characterize genome evolution, we performed a comparative analysis of gene family evolution across all osteoglossid species (Figure 3A). Among them, black arowana exhibited the most pronounced expansion, with 1 123 expanded and 2 352 contracted gene families, whereas silver arowana showed 508 expansions and 1 347 contractions. The pirarucu showed 850 expanded and 1 113 contracted families. The African arowana presented 453 expanded and 2 272 contracted gene families. Among the Asian lineages, the golden arowana displayed 622 expansions and 3 092 contractions; the green arowana had 344 expansions and 968 contractions, and the red arowana

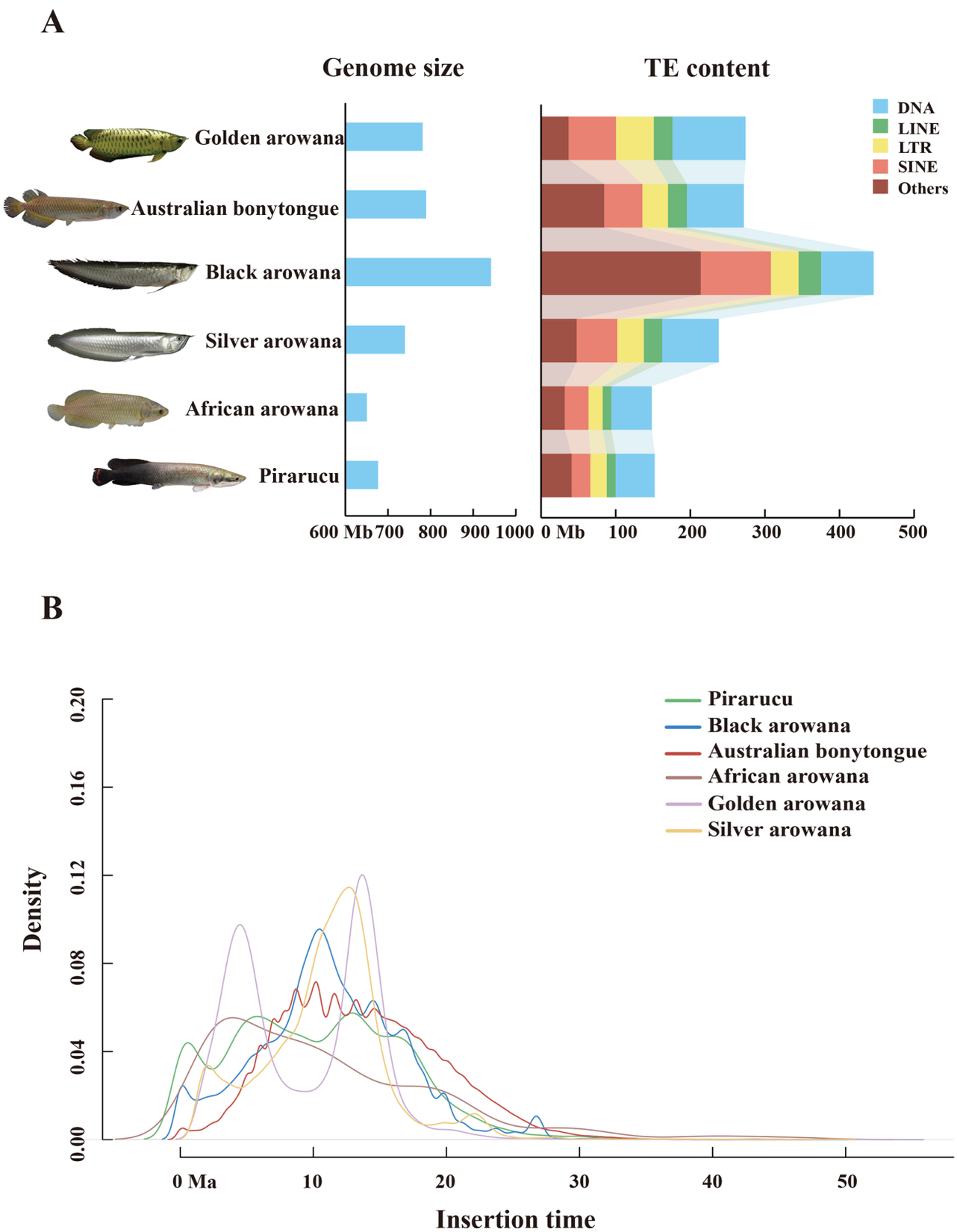


Figure 2 Divergence genome sizes in six Osteoglossidae species driven by TEs

A: Genome size and TE content statistics of these Osteoglossidae species. The categories of TEs included DNA, LINEs, LTRs, SINEs, and other genomic regions in each species. B: Density plot of the timing of LTR insertions among these Osteoglossidae species

showed 263 expansions and 1 228 contractions. These results indicate substantial lineage-specific genomic reshaping within Osteoglossidae. These differences are likely closely associated with species-specific habitats and adaptive evolution. The pronounced gene family expansions and contractions observed in the black arowana may reflect

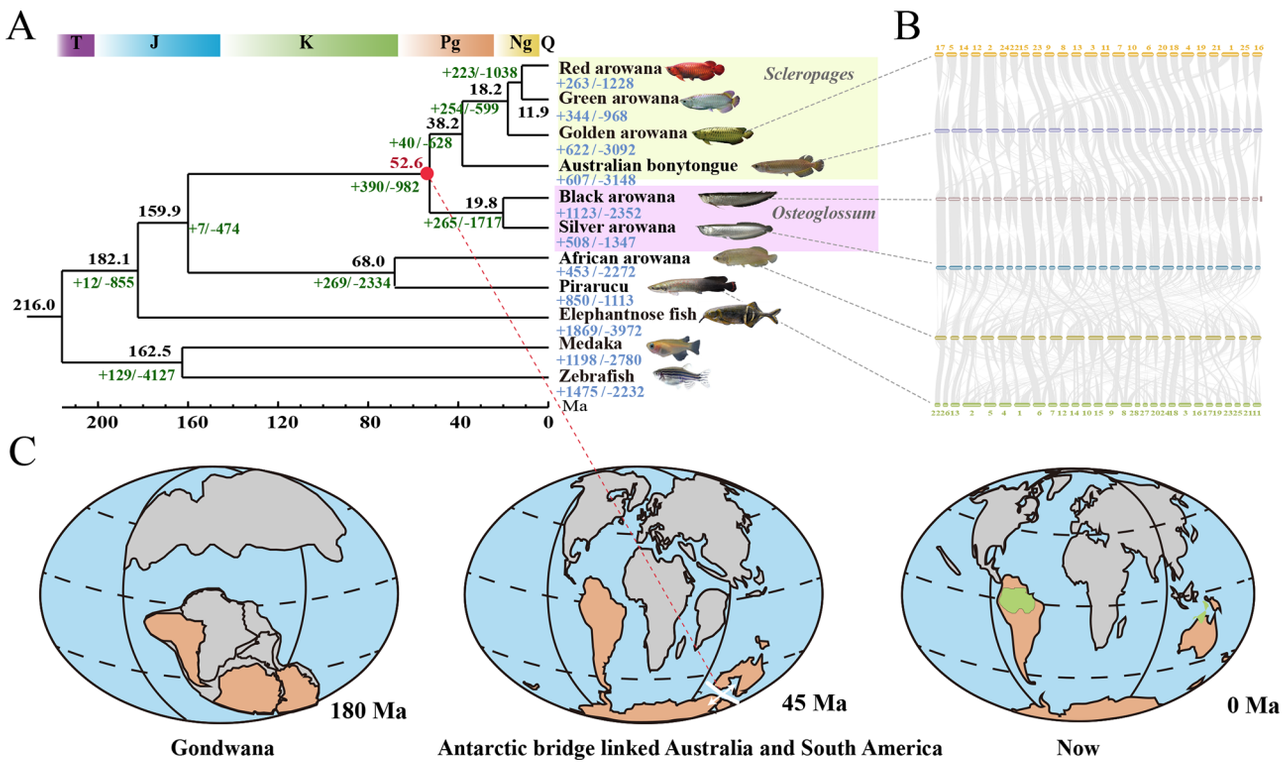


Figure 3 Phylogeny, gene family and biogeography analyses of Osteoglossidae

A: Time-calibrated phylogeny and estimated divergence time of Osteoglossidae. The results of the gene family analyses are indicated by blue numbers, with "+" representing expansions and "−" representing contractions. B: Comparative genomic synteny analysis among eight Osteoglossidae species. C: Paleogeographic maps depicting the configuration of Earth's continents at three critical time points (180, 45, and 0 Ma). The separation of Australia from the South America–Antarctic landmass is estimated to have occurred ~45 Ma. This timing overlaps with the estimated divergence time (52.6 MaMa) between the *Scleropages* and *Osteoglossum*.

adaptive evolutionary processes driven by its unique acidic and hypoxic aquatic environments. In addition, the synteny analysis revealed signals of chromosomal rearrangements among osteoglossid species, including potential chromosome fusion events and gene translocations (Figure 3B).

Ancestral chromosomes of Osteoglossiformes

We karyotyped and analyzed the elephantnose fish (*Gnathonemus petersii, Mormyridae,* Osteoglossiformes) and representative species of Osteoglossidae (Figure 4C–I). We reconstructed the ancestral karyotype of *Scleropages* with 25 chromosomes, involving five major chromosomal rearrangements—specifically, four translocations and one fission (Figure 4A). Similarly, the ancestral *Osteoglossum* karyotype was inferred to have 28 chromosomes with ten major rearrangements, including six translocations and four fissions (Figure 4B). Compared with the elephantnose fish (*n*=25), Osteoglossidae clearly underwent more complex chromosomal restructuring after diverging from the last common ancestor of teleosts. Furthermore, we observed one shared chromosomal fission event and one lineage-specific fusion event in both the elephantnose fish and the *Scleropages* species (golden arowana, *n*=24; Australian bonytongue, *n*=25) (Figure 4C, F, G). Conserved fission was absent in *Osteoglossum* species (black arowana, *n*=27; silver arowana, *n*=28) (Figure 4H, I). Thus, we propose that the common ancestor of Mormyridae and Osteoglossidae experienced one instance of chromosomal fission prior to their divergence, while fusion occurred after speciation. Additionally, the ancestral *Osteoglossum* lineage underwent four unique chromosomal fissions before diverging from the *Scleropages* ancestor.

Historical population dynamics

Using the PSMC model (Li & Durbin, 2011), we analyzed changes in the effective population size (N_e) across eight species of osteoglossid fishes. We observed that the Last Glacial Period (LGP, ~110–12 kya) significantly affected the populations of nearly all the Osteoglossidae species (Figure 5). All analyzed Osteoglossidae species (except Asian arowana and African arowana) exhibited substantial declines in effective population size coinciding with the LGP. This reduction occurred almost synchronously with falling sea levels and temperatures.

In South America, black arowana and silver arowana experienced marked population declines during the LGP, with N_e reaching the lowest levels toward the end of this period (Figure 5H, I). This trend of population reduction is consistent with the findings of previous studies (de Souza et al., 2019). By contrast, pirarucu exhibited an initial population expansion during the early LGP, which persisted for approximately the first half of the glacial period before undergoing a sustained decline (Figure 5F). Australian bonytongue likewise displayed a sustained decline throughout the LGP (Figure 5E), mirroring the demographic trajectories observed in black arowana and silver arowana. In contrast, the African arowana population exhibited a modest increase in N_e during the same period, with the population size gradually increasing throughout the LGP (Figure 5G). With respect to the Asian arowana, including its three geographic variants (golden, red, and green), N_e remained relatively stable throughout the LGP (Figure 5B–D).

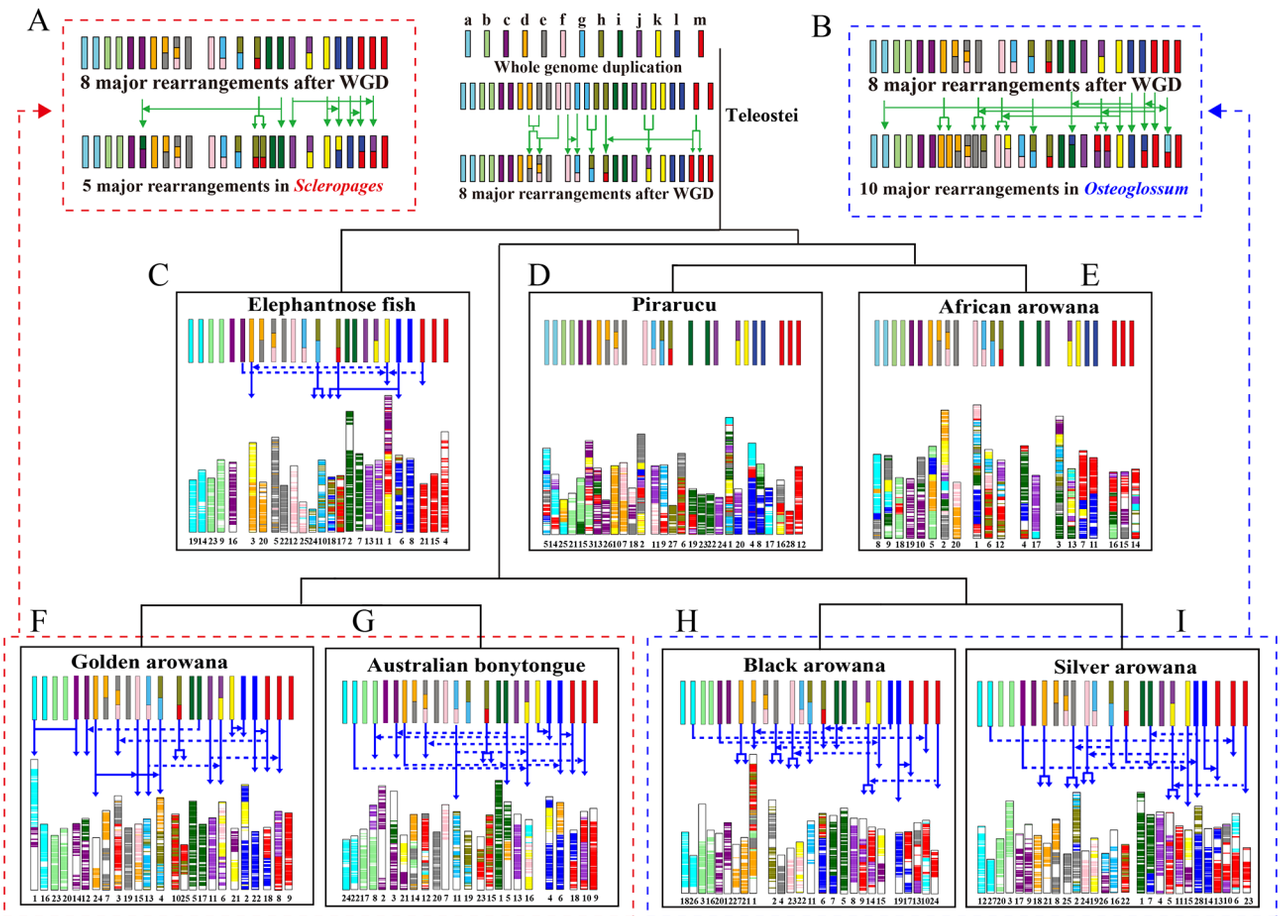


Figure 4 The ancestral karyotype of *Osteoglossum* and *Scleropages*.

Thirteen color bars represent teleost ancestor chromosomes a–m. A: Chromosomal rearrangements in the ancestral karyotype of the *Scleropages* lineage. B: Chromosomal rearrangements in the ancestral karyotype of the *Osteoglossum* lineage. C–I: Chromosomal karyotype of elephantnose fish and six Osteoglossidae species: elephantnose fish (C), pirarucu (D), African arowana (E), golden arowana (F), Australian bonytongue (G), black arowana (H) and silver arowana (I).

DISCUSSION

Large improvement in osteoglossid genome assemblies

By integrating MGI short-read, PacBio HiFi, ONT ultralong read, and Hi-C assembly technologies, this study presents high-quality chromosome-level genome assemblies for three osteoglossid species (pirarucu, black arowana, and Australian bonytongue). Compared with existing genomic data, our assemblies demonstrate significant improvements in both genome integrity and continuity (Table 1). The contig N50 values of pirarucu, black arowana, and Australian bonytongue are 27.0, 26.9 and 33.1 Mb, respectively, significantly surpassing those in previously sequenced assemblies, such as the African arowana with a contig N50 of 2.3 Mb (Hao et al., 2020) and pirarucu with a contig N50 of 0.05 Mb (Vialle et al., 2018). Notably, our genome assemblies of pirarucu and Australian bonytongue have reached the T2T level. To our knowledge, these represent the first T2T chromosome-level assemblies reported for osteoglossid species. These genome sequences provide important genetic resources for further research.

Biogeographical analysis of the osteoglossid

Extant osteoglossid fishes, as freshwater species distributed across South America, Africa, Asia, and Australia, have long been the subject of debate regarding whether their intercontinental distribution is linked to the breakup of the Gondwana supercontinent. Three main hypotheses have been proposed: 1) the divergence between the African arowana and the pirarucu was caused by continental separation between Africa and South America; 2) the divergence between the South American genus *Osteoglossum* and the Asian–Australian genus *Scleropages* resulted from the rifting of Australia from the Antarctica–South America landmass; and 3) the “Indian biotic ferry” hypothesis, which proposes that the ancestor of Southeast Asian *Scleropages* was transported to Asia via the Indian plate after its northward drift and collision with Eurasia (de Bello Cioffi et al., 2019; Hilton & Lavoué, 2018).

Our molecular divergence time results provide a framework for evaluating these hypotheses. According to our results, the divergence between African arowana and pirarucu occurred approximately 68.0 Ma, which is substantially younger than the initial separation of Africa and South America (~110 Ma) (McIntyre et al., 2017); thus, we reject the hypothesis that their divergence was driven by continental isolation. The divergence between *Osteoglossum* and *Scleropages* is estimated at 52.6 Ma, overlapping with the estimated period of connection between Australia and Antarctica–South America (62.5–43.5 Ma) (Figure 3A, C) (McIntyre et al., 2017), which supports the second hypothesis. Their common ancestor likely dispersed across these connected landmasses via Antarctic

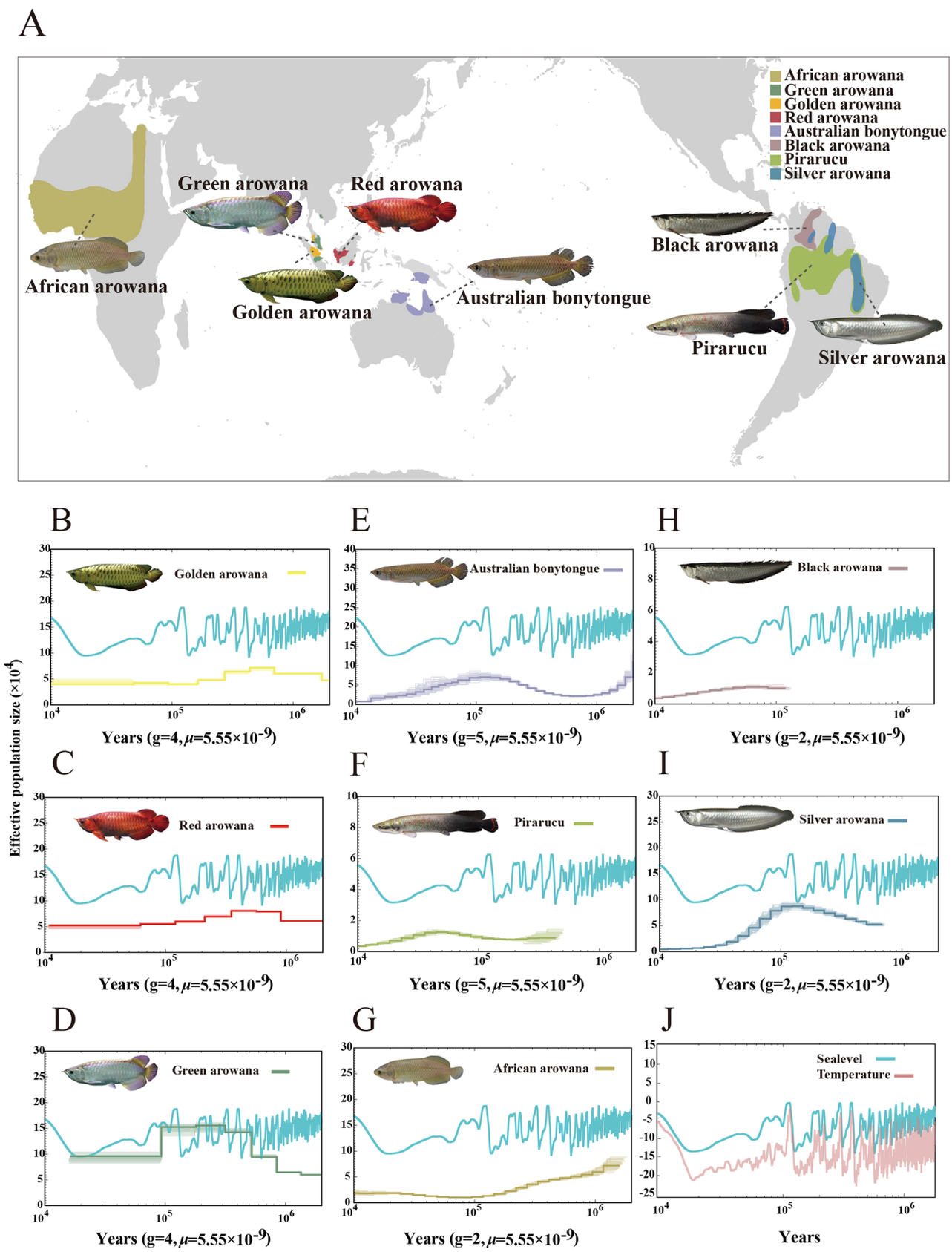


Figure 5 Distributions and demographic trajectories of Osteoglossidae species

A: Current distribution ranges of eight Osteoglossidae species (Asian arowana includes three color variants). B–I: Demographic trajectories of eight Osteoglossidae species obtained using the PSMC method. The x-axis corresponds to the time before it is present in years on a log scale, whereas the y-axis corresponds to the effective population size (*Ne*). J: The plot shows the sea level and temperature with the same x-axis as before.

land bridges (van den Ende et al., 2017) and was later isolated because of the separation of Australia. In contrast, the divergence between the Australian bonytongue and Southeast Asian *Scleropages* species is estimated at 38.2 Ma, which is much later than the separation of India and Australia (~120 Ma) (McIntyre et al., 2017), thus strongly refuting the "Indian biotic ferry" hypothesis. Instead, it strongly supports marine dispersal events across the Wallace Line.

Historical population dynamics of osteoglossid species

Climate and hydrological fluctuations are among the key drivers of evolutionary dynamics and population structure, with particularly strong impacts on tropical freshwater fishes because of their ecological dependence on stable aquatic environments (Ivory et al., 2016). Our results reveal distinct population histories among osteoglossid species that coincide with major climatic events, especially during the Late Pleistocene. A shared signal among most taxa—especially South American and Australian lineages—is a marked population decline during the LGP, suggesting that global cooling and aridification exerted severe ecological pressures on these tropical freshwater lineages.

The dramatic decline in population sizes observed in South American and Australian osteoglossid species may be attributed to widespread temperature decline and aridification in some areas (particularly in the Amazon region) during the LGP, which likely led to the contraction and fragmentation of freshwater habitats (Nixon et al., 2025; Wang et al., 2017), limiting gene flow and reducing population viability. During the same period, there may be two reasons for the increasing trend of the African arowana population. Their omnivorous diet likely conferred an adaptive advantage (Adite et al., 2005), enhancing their survival during glacial periods. Additionally, the Nile River and West African drainage systems likely maintained intermittent hydrological connectivity during this period, potentially providing more aquatic habitats for this species (Osborne et al., 2008; Pachur & Kröpelin, 1987). Previous research has demonstrated that forest mosaics prevailed over savanna corridors in Southeast Asia throughout the LGM period (Hamilton et al., 2024). These forest mosaics likely sustained stable freshwater ecosystems, creating a critical foundation for the survival of Asian arowana populations. Furthermore, the complexity of interconnected river-lake networks in the region may have provided refugia that ensured population stability for Asian arowana during glacial cycles. This potential climate resilience due to geographic location suggests that the current endangered status of Asian arowana may be caused largely by anthropogenic pressure, especially overexploitation and habitat destruction, rather than past climate change.

Surprisingly, most Osteoglossidae species did not recover demographically during the late LGP (except for the African arowana), despite rising sea levels and temperatures. This contrast suggests that African arowana were more resilient to glacial environmental stressors than their other counterparts, indicating greater ecological adaptability. While glacier-driven habitat fragmentation imposed major constraints, such abiotic factors alone cannot fully explain the long-term demographic stagnation observed. Nonclimatic processes, such as ecological niche displacement and the emergence of novel competitors, may further limit population recovery in South America and Australia. In contrast, the African arowana's omnivorous diet likely reflects a flexible survival strategy, allowing it to exploit diverse resources and avoid direct competition. These findings highlight that population resilience depends not only on environmental recovery but also on species ecological flexibility and interactions within shifting biotic communities over evolutionary timescales.

We discussed several hypotheses regarding the cross-continental distribution of Osteoglossidae, including the possibility that the divergence between *Osteoglossum* and *Scleropages* may have been driven by geographic isolation following the breakup of Gondwana, and further analyzed the historical population dynamics of these species at specific periods in relation to paleogeographic and climatic changes.

Elucidating the molecular mechanisms underlying adaptive divergence across varying environments—such as differences in temperature, diet, water pH, body color, and morphology—requires extensive sampling for population resequencing and single-cell transcriptome sequencing. Therefore, this falls beyond the scope of the current study. Future research can integrate multi-omics, such as resequencing, transcriptomics, and methylation data to gain a deeper understanding of the molecular basis of their adaptation in Osteoglossidae species.

CONCLUSION

Our research generated three chromosome-level genome assemblies, two of which—pirarucu and Australian bonytongue—achieved T2T completeness, whereas the black arowana assembly reached the near-T2T level. These assemblies represent the most contiguous and complete genomic resources currently available for Osteoglossidae, with BUSCO scores exceeding 97% and contig N50 values ranging from 26.9 to 33.1 Mb. Our phylogenomic analyses resolve long-standing biogeographic debates: the divergence between *Scleropages* and *Osteoglossum* (~52.6 Ma) aligns with the separation of Australia from the South America–Antarctica landmass.

We also reconstructed the ancestral karyotypes of *Scleropages* and *Osteoglossum*, which underwent five and ten major chromosomal rearrangements, respectively. Population history analyses further revealed that most osteoglossid species experienced significant declines in effective population size during the LGP. In contrast, the African arowana exhibited greater ecological adaptability and demographic resilience during the same period. The lack of postglacial recovery in most species suggests that nonclimatic factors may further constrain population resilience, underscoring the urgency of conservation for many Osteoglossidae species. Overall, our study provides a comprehensive framework for understanding the evolutionary history, biogeography, and population dynamics of Osteoglossidae.

DATA AVAILABILITY

All of these assembled genomes and raw reads are available under NCBI BioProject: PRJNA1331523 for pirarucu; PRJNA1328470 for black arowana; PRJNA1328591 for Australian bonytongue. The assembled genome sequences generated in this study have been deposited in Science Data Bank (DOI: 10.6084/m9.figshare.32309736). The detailed annotation files including coding sequences, protein sequences and gene location were deposited in Figshare with linkage: https://doi.org/10.6084/m9.figshare.32309736.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

X.M., C.B., designed research. Y.L., Y.Y., H.L. and Y.W. provided and collected samples. C.B., R.H. analyzed data. C.B., R.H., Y.L. and X.M. wrote the paper with input from all authors. C.B., and X.M. revised the article. All authors read and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

We thank the National Freshwater Genetic Resource Center, Guangzhou, China, for providing the samples of three Osteoglossid species (including black arowana, pirarucu and Australian bonytongue) in this study.

REFERENCES

Abrusán G, Grundmann N, DeMester L, et al. 2009. TEclass—a tool for automated classification of unknown eukaryotic transposable elements. *Bioinformatics*, **25**(10): 1329–1330.

Adekeye AO. 1993. Age and growth of *Heterotis niloticus* around Pategi in the middle River Niger, Kwara State. *In: Eyo AA, Balogun AM, eds. Proceedings of the 10th Annual Conference of the Fisheries Society of Nigeria (FISON). Lagos: Fisheries Society of Nigeria*, 152–158.

Adite A, Winemiller KO, Fiogbe ED. 2005. Ontogenetic, seasonal, and spatial variation in the diet of *Heterotis niloticus* (Osteoglossiformes: Osteoglossidae) in the Sô River and Lake Hlan, Benin, West Africa. *Environmental Biology of Fishes*, **73**(4): 367–378.

Araripe J, do Rêgo PS, Queiroz H, et al. 2013. Dispersal capacity and genetic structure of *Arapaima gigas* on different geographic scales using microsatellite markers. *PLoS One*, **8**(1): e54470.

Ashburner M, Ball CA, Blake JA, et al. 2000. Gene ontology: tool for the unification of biology. *Nature Genetics*, **25**(1): 25–29.

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

Betancur-R R, Wiley EO, Arratia G, et al. 2017. Phylogenetic classification of bony fishes. *BMC Evolutionary Biology*, **17**(1): 162.

Bian C, Hu YC, Ravi V, et al. 2016. The Asian arowana (*Scleropages formosus*) genome provides new insights into the evolution of an early lineage of teleosts. *Scientific Reports*, **6**(1): 24501.

Bian C, Huan RJW, Shi Q. 2025. Telomere-to-telomere chromosome-scale genome assemblies of black and golden koi carp variants support construction of an ancient karyotype of Cypriniformes. *GigaScience*, **14**: giaf073.

Bian C, Huang Y, Li RH, et al. 2024a. Genomics comparisons of three chromosome-level mudskipper genome assemblies reveal molecular clues for water-to-land evolution and adaptation. *Journal of Advanced Research*, **58**: 93–104.

Bian C, Li RH, Ruan ZQ, et al. 2024b. Chromosome-level genome assembly of the glass catfish (*Kryptopterus vitreolus*) reveals molecular clues to its transparent phenotype. *Zoological Research*, **45**(5): 1027–1036. *Gigascience*, 14: giaf073.

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

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

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

Capobianco A, Friedman M. 2019. Vicariance and dispersal in southern hemisphere freshwater fish clades: a palaeontological perspective. *Biological Reviews*, **94**(2): 662–699.

Capobianco A, Friedman M. 2024. Fossils indicate marine dispersal in osteoglossid fishes, a classic example of continental vicariance. *Proceedings of the Royal Society B: Biological Sciences*, **291**(2028): 20241293.

2018. *Biological reviews of the Cambridge Philosophical Society*, **94**(2): 662-699.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, **17**(4): 540–552.

Chen Y, Nie F, Xie SQ, et al. 2021. Efficient assembly of nanopore reads via highly accurateand intact error correction. *Nature Communications*, **12**(1): 60.

Chin CS, Alexander DH, Marks P, et al. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods*, **10**(6): 563–569.

Cui Y, Yuan YB, Wang B, et al. 2025. T2T genomes unveil centromere architecture and adaptive divergence in large yellow croaker (*Larimichthys crocea*). *Advanced Science*, **12**(43): e06374.

de Bello Cioffi M, Ráb P, Ezaz T, et al. 2019. Deciphering the evolutionary history of arowana fishes (Teleostei, Osteoglossiformes, Osteoglossidae): insight from comparative cytogenomics. *International Journal of Molecular Sciences*, **20**(17): 4296.

2025. T2T Genomes Unveil Centromere Architecture and Adaptive Divergence in Large Yellow Croaker (Larimichthys crocea).

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

de Oliveira EA, Bertollo LAC, Rab P, et al. 2019. Cytogenetics, genomics and biodiversity of the South American and African Arapaimidae fish family (Teleostei, Osteoglossiformes). *PLoS One*, **14**(3): e0214225.

de Souza FHS, Perez MF, Bertollo LAC, et al. 2019. Interspecific genetic differences and historical demography in South American arowanas (Osteoglossiformes, Osteoglossidae, *Osteoglossum*). *Genes*, **10**(9): 693.

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

Durand NC, Robinson JT, Shamim MS, et al. 2016. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Systems*, **3**(1): 99–101.

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

Escobar Lizarazo MD, Farias IP, Taphorn DC, et al. 2013. Molecular diagnosis of the arowanas *Osteoglossum ferreirai* Kanazawa, 1966 and *O. bicirrhossum* (Cuvier, 1829) from the Orinoco and Amazon River basins. *Neotropical Ichthyology*, **11**(2): 335–340.

Eyre-Walker A. 2000. Fundamentals of molecular evolution (2nd edn). *Heredity*, **84**(6): 735.

Fischer S, Brunk BP, Chen F, et al. 2011. Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. *Current Protocols in Bioinformatics*, **35**(1): 6.12. 1–6.12. 19.

Gao BJ, Zhang J. 2017. The complete mitochondrial genome of *Scleropages formosus* (Osteoglossiformes: Osteoglossidae) and phylogenetic studies of Osteoglossiformes. *Conservation Genetics Resources*, **9**(2): 265–268.

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

Hamilton R, Amano N, Bradshaw CJA, et al. 2024. Forest mosaics, not savanna corridors, dominated in Southeast Asia during the Last Glacial Maximum. *Proceedings of the National Academy of Sciences of the United States of America*, **121**(1): e2311280120.

Hao SJ, Han K, Meng LF, et al. 2020. African arowana genome provides insights on ancient teleost evolution. *iScience*, **23**(11): 101662.

Hilton EJ, Lavoué S. 2018. A review of the systematic biology of fossil and living bony-tongue fishes, Osteoglossomorpha (Actinopterygii: Teleostei). *Neotropical Ichthyology*, **16**(3): e180031.

Hrbek T, Farias IP, Crossa M, et al. 2005. Population genetic analysis of *Arapaima gigas*, one of the largest freshwater fishes of the Amazon basin: implications for its conservation. *Animal Conservation*, **8**(3): 297–308. *Animal Conservation*, **8**: 297-308.

Hunter S, Apweiler R, Attwood TK, et al. 2009. InterPro: the integrative protein signature database. *Nucleic Acids Research*, **37**(S1): D211–D215.

Ivory SJ, Blome MW, King JW, et al. 2016. Environmental change explains cichlid adaptive radiation at Lake Malawi over the past 1.2 million years. *Proceedings of the National Academy of Sciences of the United States of America*, **113**(42): 11895–11900.

Janert PK. 2009. Gnuplot in Action: Understanding Data with Graphs.

Jiangyong Z, Wilson M Lewis Street Greenwich: Manning Publications Co.

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

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

Kim D, Paggi JM, Park C, et al. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, **37**(8): 907–915.

Kulikova T, Aldebert P, Althorpe N, et al. 2004. The EMBL nucleotide sequence database. *Nucleic Acids Research*, **32**(Suppl 1): D27–D30.

Kumazawa Y, Nishida M. 2000. Molecular phylogeny of osteoglossoids: a new model for gondwanian origin and plate tectonic transportation of the Asian arowana. *Molecular Biology and Evolution*, **17**(12): 1869–1878.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**(4): 357–359.

Lavoué S. 2015. Testing a time hypothesis in the biogeography of the arowana genus *Scleropages* (Osteoglossidae). *Journal of Biogeography*, **42**(12): 2427–2439.

Lavoué S. 2016. Was Gondwanan breakup the cause of the intercontinental distribution of Osteoglossiformes? A time-calibrated phylogenetic test combining molecular, morphological, and paleontological evidence. *Molecular Phylogenetics and Evolution*, **99**: 34–43.

Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, **27**(21): 2987–2993.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, **34**(18): 3094–3100.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**(14): 1754–1760.

Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature*, **475**(7357): 493–496.

Lin YZ, Ye C, Li XZ, et al. 2023. quarTeT: a telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. *Horticulture Research*, **10**(8): uhad127.

Liu F, Li Y, Wang GS, et al. 2025. The telomere-to-telomere gapless genome of grass carp provides insights for genetic improvement. *GigaScience*, **14**: giaf059.

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

Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics*, **27**(6): 764–770.

McGinnis S, Madden TL. 2004. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Research*, **32**(S2): W20–W25.

McIntyre SRN, Lineweaver CH, Groves CP, et al. 2017. Global biogeography since Pangaea. *Proceedings of the Royal Society B: Biological Sciences, Proceedings of the Royal Society B*, **284**(1856): 20170716.

Merrick JR, Midgley SH, Rimmer MA. 1983. Preliminary studies on growth of two spotted barramundi species, *Scleropages leichardti* and *S. jardini* (Pisces: Osteoglossidae). *Aquaculture*, **32**(1-2): 195–199.

Mount DW. 2007. Using the basic local alignment search tool (BLAST). *Cold Spring Harbor Protocols*, **2007(7)**: pdb. top17.

Mu XD, Liu Y, Wang XJ, et al. 2014. Characterization of the mitochondrial genome and phylogeny of the black arowana (*Osteoglossum ferreirai*). *Biologia*, **69**(9): 1222–1230.

Nixon FC, Tyler JJ, Klaebe RM, et al. 2025. Temperature and hydrological change in central Australia through the last glacial cycle: insights from clumped isotope analysis of lacustrine molluscs. *Quaternary Science Reviews*, **350**: 109157.

Ogata H, Goto S, Sato K, et al. 1999. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, **27**(1): 29–34.

Osborne AH, Vance D, Rohling EJ, et al. 2008. A humid corridor across the Sahara for the migration of early modern humans out of Africa 120, 000 years ago. *Proceedings of the National Academy of Sciences of the United States of America*, **105**(43): 16444–16447.

Pachur HJ, Kröpelin S. 1987. Wadi Howar: paleoclimatic evidence from an extinct river system in the southeastern Sahara. *Science*, **237**(4812): 298–300.

Peterson RD, Sullivan JP, Hopkins CD, et al. 2022. Phylogenomics of bony-tongue fishes (osteoglossomorpha) shed light on the craniofacial evolution and biogeography of the weakly electric clade (mormyridae). *Systematic Biology*, **71**(5): 1032–1044.

Pollier J, Rombauts S, Goossens A. 2013. Analysis of RNA-Seq data with TopHat and cufflinks for genome-wide expression analysis of jasmonate-treated plants and plant cultures. *In*: Goossens A, Pauwels L. Jasmonate Signaling. Totowa: Humana, 305–315.

Queiroz HL, Camargo M, eds. . 2008. Biologia, Conservação e Manejo dos Aruanãs na Amazônia Brasileira. *Tefé: Instituto de Desenvolvimento Sustentável Mamirauá*.

Rhie A, Walenz BP, Koren S, et al. 2020. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology*, **21**(1): 245.

Rhoads A, Au KF. 2015. PacBio sequencing and its applications. *Genomics, Proteomics & Bioinformatics*, **13**(5): 278–289.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European molecular biology open software suite. *Trends in Genetics*, **16**(6): 276–277.

Ruan J, Li H. 2020. Fast and accurate long-read assembly with wtdbg2. *Nature Methods*, **17**(2): 155–158.

Salvador GN, Reis, et al. 2023. *Osteoglossum ferreirai*. The IUCN Red List of Threatened Species 2023: e. T167687A160183419. https://doi.org/10.2305/IUCN.UK.2023-1.RLTS.T167687A160183419.en.

Schoch CL, Ciufo S, Domrachev M, et al. 2020. NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database*, **2020**: baaa062.

Sparks JS, Smith WL. 2005. Freshwater fishes, dispersal ability, and nonevidence: "Gondwana Life Rafts" to the rescue. *Systematic Biology*, **54**(1): 158–165.

Stewart DJ. 2013. Re-description of *Arapaima agassizii* (Valenciennes), a rare fish from brazil (Osteoglossomorpha: Osteoglossidae). *Copeia*, **2013**(1): 38–51.

Tarailo-Graovac M, Chen NS. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*, **25**(1): 4.10. 1-4.10. 14.

Torati LS, Taggart JB, Varela ES, et al. 2019. Genetic diversity and structure in *Arapaima gigas* populations from Amazon and Araguaia-Tocantins river basins. *BMC Genetics*, **20**(1): 13.

van den Ende C, White LT, van Welzen PC. 2017. The existence and break-up of the Antarctic land bridge as indicated by both amphi-Pacific distributions and tectonics. *Gondwana Research*, **44**: 219–227.

Vialle RA, de Souza JES, de Paiva Lopes K, et al. 2018. Whole genome sequencing of the Pirarucu (*Arapaima gigas*) supports independent emergence of major teleost clades. *Genome Biology and Evolution*, **10**(9): 2366–2379.

Wang XF, Edwards RL, Auler AS, et al. 2017. Hydroclimate changes across the Amazon lowlands over the past 45, 000 years. *Nature*, **541**(7636): 204–207.

Wang YG, Obbard DJ. 2023. Experimental estimates of germline mutation rate in eukaryotes: a phylogenetic meta-analysis. *Evolution Letters*, **7**(4): 216–226.

Waterhouse RM, Seppey M, Simão FA, et al. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution*, **35**(3): 543–548.

Xidong M, Liu Y, Wang X, et al. *Biologia*, **69**: 1222–1230.

Xu GC, Xu TJ, Zhu R, et al. 2019. LR_Gapcloser: a tiling path-based gap closer that uses long reads to complete genome assembly. *GigaScience*, **8**(1): giy157.

Xu MY, Guo LD, Gu SQ, et al. 2020. TGS-GapCloser: a fast and accurate gap closer for large genomes with low coverage of error-prone long reads. *GigaScience*, **9**(9): giaa094.

Xu Z, Wang H. 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research*, **35**(S2): W265–W268.

Yang ZH. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**(8): 1586–1591.

Yue GH, Chang A, Yuzer A, et al. 2020. Current knowledge on the biology and aquaculture of the endangered Asian Arowana. *Reviews in Fisheries Science & Aquaculture*, **28**(2): 193–210.

Zhang JY, Wilson MVH. 2017. First complete fossil Scleropages (Osteoglossomorpha). *Vertebrata Palasiatica*, **55**(1): 1–23.

Zhou CX, McCarthy SA, Durbin R. 2023. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics*, **39**(1): btac808.