ARTICLE

OPEN ACCESS

www.the-innovation.org

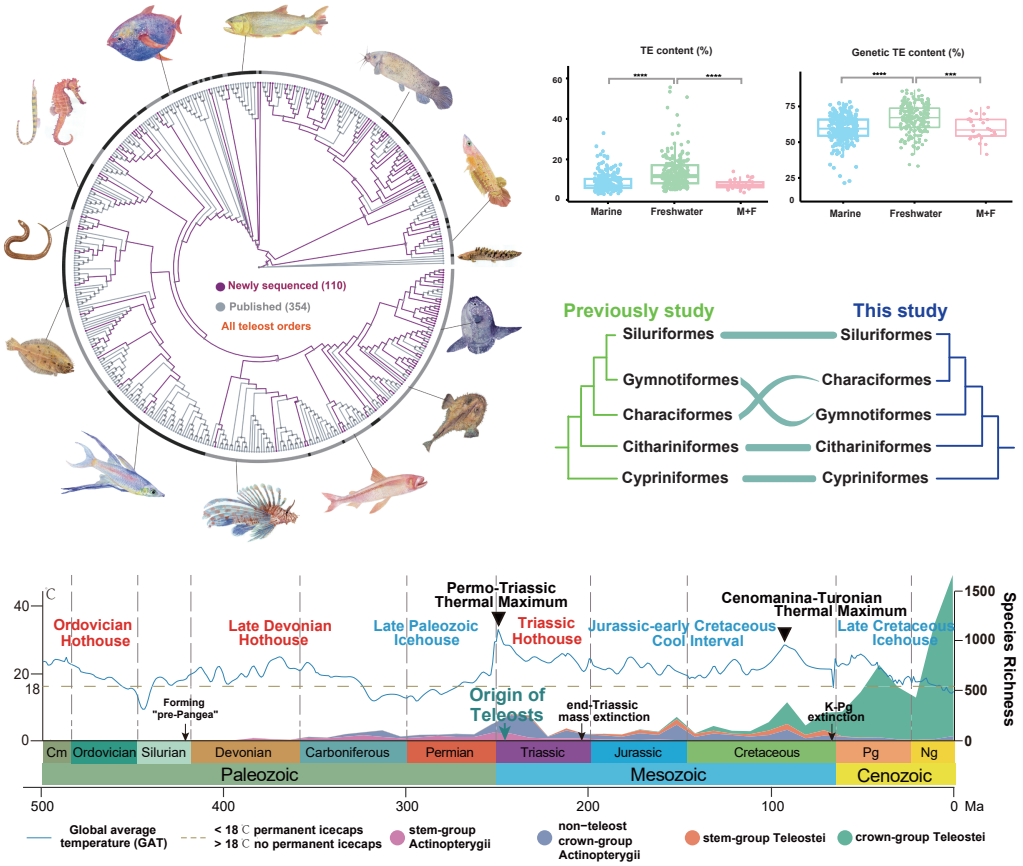
A genomic compendium of hundreds of teleost fishes reveals their evolutionary landscape

Yue Song,[1,2,3,17,28] Zengbao Yuan,[1,27,28] Chengchi Fang,[3,28] Suyu Zhang,[1,28] Liandong Yang,[3,28] Mingliang Hu,[4,28] Inge Seim,[5,28] Shanshan Liu,[1,28] Xiaolin Tian,[6] Cheng Wang,[3] Yaolei Zhang,[1,2] Zhaohui Pan,[7] Qingming Qu,[8] Hongyue Liu,[6] Yuanning Li,[6] Luyuan Pan,[9] Chenglong Zhu,[4] Hengjia Yang,[1] Xiao Chen,[10] Mengqi Zhang,[1] Gang Hou,[11] Meiru Liu,[1] Jiahao Wang,[1] Qun Liu,[1,17] Xiaoni Gan,[3] Honghui Zeng,[3] Wenjie Xu,[4] Chenguang Feng,[4] Mengjun Wang,[7,25] Zhuocheng Zhou,[12] Song He,[13] Chenglong Liu,[1] Mengjun Yu,[1] Hanbo Li,[1,14] Jian Liang,[15] He Zhang,[1] Yongxin Li,[4] Simon Ming-Yuen Lee,[16] Yonghua Sun,[9] Qiang Qiu,[4] Xin Liu,[2] Karsten Kristiansen,[17] Wen Wang,[4] Jian Wang,[2] Min Zhu,[7,26] Huanming Yang,[2,18,19,26] Songlin Chen,[20] Jianfang Gui,[3] Yiyu Chen,[3] Guojie Zhang,[21,22,*] Xun Xu,[1,2,23,*] Kun Wang,[4,*] Guangyi Fan,[1,2,16,23,*] and Shunping He[3,24,25,*]

*Correspondence: guojiezhang@zju.edu.cn (G.Z.); xuxun@genomics.cn (X.X.); wangkun@nwpu.edu.cn (K.W.); fanguangyi@genomics.cn (G.F.); clad@ihb.ac.cn (S.H.)

Received: September 2, 2025; Accepted: November 2, 2025; https://doi.org/10.1016/j.xinn.2025.101177

GRAPHICAL ABSTRACT



PUBLIC SUMMARY

- A 464 whole-genome alignment resource including 110 *de novo* assembled genomes.
- Phylogenetic analysis reveals controversies in the topology.
- Differences in repetitive elements are associated with the habitats of teleost fishes.
- Deep phylogenomics identifies evolutionary constraints and genetics innovation.

OPEN ACCESS

ARTICLE

The Innovation

A genomic compendium of hundreds of teleost fishes reveals their evolutionary landscape

Yue Song,[1,2,3,17,28] Zengbao Yuan,[1,27,28] Chengchi Fang,[3,28] Suyu Zhang,[1,28] Liandong Yang,[3,28] Mingliang Hu,[4,28] Inge Seim,[5,28] Shanshan Liu,[1,28] Xiaolin Tian,[6] Cheng Wang,[3] Yaolei Zhang,[1,2] Zhaohui Pan,[7] Qingming Qu,[8] Hongyue Liu,[6] Yuanning Li,[6] Luyuan Pan,[9] Chenglong Zhu,[4] Hengjia Yang,[1] Xiao Chen,[10] Mengqi Zhang,[1] Gang Hou,[11] Meiru Liu,[1] Jiahao Wang,[1] Qun Liu,[1,17] Xiaoni Gan,[3] Honghui Zeng,[3] Wenjie Xu,[4] Chenguang Feng,[4] Mengjun Wang,[7,25] Zhuocheng Zhou,[12] Song He,[13] Chenglong Liu,[1] Mengjun Yu,[1] Hanbo Li,[1,14] Jian Liang,[15] He Zhang,[1] Yongxin Li,[4] Simon Ming-Yuen Lee,[16] Yonghua Sun,[9] Qiang Qiu,[4] Xin Liu,[2] Karsten Kristiansen,[17] Wen Wang,[4] Jian Wang,[2] Min Zhu,[7,26] Huanming Yang,[2,18,19,26] Songlin Chen,[20] Jianfang Gui,[3] Yiyu Chen,[3] Guojie Zhang,[21,22,*] Xun Xu,[1,2,23,*] Kun Wang,[4,*] Guangyi Fan,[1,2,16,23,*] and Shunping He[3,24,25,*]

[1]BGI Research, Qingdao 266555, China

[2]BGI Research, Shenzhen 518083, China

[3]Key Laboratory of Breeding Biotechnology and Sustainable Aquaculture, Chinese Academy of Sciences, Wuhan 430072, China

[4]Shaanxi Key Laboratory of Qinling Ecological Intelligent Monitoring and Protection, School of Ecology and Environment, Northwestern Polytechnical University, Xi'an 710129, China

[5]Marine Mammal and Marine Bioacoustics Laboratory, Institute of Deep-Sea Science and Engineering, Chinese Academy of Sciences, Sanya 572000, China

[6]Institute of Marine Science and Technology, Shandong University, Qingdao 266237, China

[7]Key CAS Laboratory of Vertebrate Evolution and Human Origins, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences (CAS), Beijing 100044, China

[8]State Key Laboratory of Cellular Stress Biology, School of Life Sciences, Xiamen University, Xiamen 361102, China

[9]National Aquatic Biological Resource Center, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China

[10]South China Agricultural University, Guangzhou 510642, China

[11]College of Fisheries, Guangdong Ocean University, Zhanjiang 524088, China

[12]Professional Committee of Native Aquatic Organisms and Water Ecosystem of China Fisheries Association, Beijing, China

[13]Red Sea Research Center, Division of Biological and Environmental Science and Engineering, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

[14]Lars Bolund Institute of Regenerative Medicine, Qingdao-Europe Advanced Institute for Life Sciences, BGI Research, Qingdao 266555, China

[15]State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University, Xining 810016, China

[16]Department of Food Science and Nutrition, and PolyU-BGI Joint Research Centre for Genomics and Synthetic Biology in Global Ocean Resources, and State Key Laboratory of Chemical Biology and Drug Discovery, The Hong Kong Polytechnic University, Hong Kong 999077, China

[17]Laboratory of Integrative Biomedicine, Department of Biology, University of Copenhagen, 2100 Copenhagen, Denmark

[18]James D. Watson Institute of Genome Sciences, Hangzhou 310029, China

[19]Institute of Basic Medicine and Cancer (IBMC), Chinese Academy of Sciences, The Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital), Hangzhou 310022, China

[20]State Key Laboratory of Mariculture Biobreeding and Sustainable Goods, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071, China

[21]Center of Evolutionary & Organismal Biology, and Women's Hospital, Zhejiang University School of Medicine, Hangzhou 310058, China

[22]Liangzhu Laboratory, Zhejiang University Medical Center, Hangzhou 311121, China

[23]State Key Laboratory of Genome and Multi-omics Technologies, BGI Research, Shenzhen 518083, China

[24]Institute of Deep-Sea Science and Engineering, Chinese Academy of Sciences, Sanya 572000, China

[25]Center for Excellence in Animal Evolution and Genetics, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650201, China

[26]University of Chinese Academy of Sciences, Beijing 100049, China

[27]College of Fisheries, Southwest University, Chongqing 400715, China

[28]These authors contributed equally

*Correspondence: guojiezhang@zju.edu.cn (G.Z.); xuxun@genomics.cn (X.X.); wangkun@nwpu.edu.cn (K.W.); fanguangyi@genomics.cn (G.F.); clad@ihb.ac.cn (S.H.)

Received: September 2, 2025; Accepted: November 2, 2025; https://doi.org/10.1016/j.xinn.2025.101177

Citation: Song Y., Yuan Z., Fang C., et al., (2026). A genomic compendium of hundreds of teleost fishes reveals their evolutionary landscape. The Innovation **7(3)**, 101177.

The remarkable morphological diversity and species abundance of teleost fishes offer a valuable resource for understanding vertebrate evolution. In phase I of the Fish10K project, genomes of 110 teleost species were sequenced and assembled, filling gaps in 3 previously unrepresented orders, and integrated with existing data to generate a 464 species whole-genome alignment spanning all teleost orders—the largest such resource beyond mammals and birds. Comparative analyses reveal distinctive genomic features, including progressive genome compaction with shortened intron lengths relative to non-teleost ray-finned fishes. Analysis of the transposable element (TE) landscape suggests a potential association between TE expansion in teleost genomes and different habitats, as well as the uniqueness of teleosts' DNA-dominated transposon composition among vertebrates. Genome-wide phylogenetic analyses refute the widely accepted monophyly of "Siluriphysi" hypothesis and support the hypothesis of a single origin of electroreception followed by secondary loss in Characiformes. A refined evolutionary timeline of teleosts by whole-genome alignment resource placed teleosts at ~253 million years ago, predating the Permian-Triassic extinction, and delineates three diversification phases punctuated by mass extinctions, challenging continuous post-Cretaceous-Palaeogene acceleration models. This study establishes a large-scale genomic database and a foundational whole-genome alignment resource, advancing insights into the landscape of teleost genomic architecture and macroevolution.

INTRODUCTION

Teleost fishes, comprising almost more than 30,000 extant species, are the most abundant vertebrate group. After more than 250 million years of evolution, teleosts inhabit almost all waters on Earth—including the deep sea, polar regions, and plateaus—and are outstanding representatives of vertebrate adaptive evolution. Since the first teleost draft genome (*Fugu rubripes*) was published in 2003,[1] an increasing number of genomes have been published, shedding light on the genetic underpinnings of unique traits across various lineages. However, the lack of a robust, whole-genome comparative context has been a critical bottleneck, preventing a systematic investigation into the evolutionary landscape.

With advances in sequencing technology, several large-scale genome consortia with a core mission of producing genomes for thousands of species have been launched, revolutionizing our understanding of these lineages. We present here a foundational resource from phase I of the 10,000 Fish Genomes (Fish10K) Project.[2] One hundred and ten newly sequenced teleost genomes are reported here, filling in the gaps for three orders. A combination of short-read

and long-read sequencing technologies are applied, enabling the sequencing of rare and difficult-to-access species such as *Alepocephalus agassizii* and *Polyacanthonotus rissoanus* from the deep sea.[3] By integrating these with publicly available data, we have constructed the 464-way whole-genome alignment of teleosts, a resource that constitutes the first of its kind for fishes and brings the scope of genomic study for this group on par with other vertebrate lineages.

This comprehensive alignment allowed for the reconstruction of a robust phylogenetic tree, resolving several long-standing evolutionary debates. Our analysis also establishes a more precise evolutionary timeline, linking major teleost diversification events to global extinction events. Furthermore, we characterized evolutionary constraints and accelerations in the genome of teleosts at single-base resolution and systematically identified genetic innovation elements. Ultimately, this study provides both a foundational resource for the research community and novel insights into the genetic drivers of the incredible success of teleosts. Also, this matrix of genomic level in large scale provides us a chance to search the special genetic elements for important clades.

MATERIALS AND METHODS

Genome collection and library construction for sequencing and assembly

We collected a total of 464 teleost fish genomes spanning 44 taxa (orders or families), 3 of which had not been sequenced before. Among these genomes, 110 are newly assembled, with an average scaffold N50 size of 4.22 Mb (Figure S1A) and an average BUSCO gene completeness of 86.38% (Table S2). The new assemblies from this project are enumerated in Table S2 (listed under the "Newly sequenced" heading). The newly sequenced genomes show similar (or better) contig N50 continuity (see Figure S1B) and BUSCO completeness (see Figure S1C) to most of the previously published teleost fish genomes. Most of these genomes obtained from this study were sequenced at BGI, where stLFR technology was used for library construction, and then assembled with Supernova 2.1[4] using default parameters.

In addition, we classified these 464 species into "freshwater" and "marine" categories based on the salinity flags provided by FishBase (www.fishbase.org): fishes listed as occurring exclusively in "fresh water" were assigned to the freshwater group, whereas those recorded only in "marine" were treated as marine; any taxa flagged as brackish or euryhaline were excluded to avoid confounding effects. Detailed descriptions of these procedures are available in the supplemental information.

Cactus whole-genome alignment

The reference-free whole-genome alignment for 467 (including 3 outgroup species, coelacanth, bichir, and spotted gar) species was executed with Cactus (v.2.1.1)[5] on a cluster of integrated HPC systems. The guide species tree required for Cactus was generated by combining the maximum-likelihood BUSCO gene trees with ASTRAL-III software.[6] Each of these gene trees was constructed using RAxML (v.8.2.12),[7] with 100 independent tree searches and the "PROTGAMMAAUTO" substitution model (for further details, refer to the supplemental information). To obtain a final MAF file, we converted the HAL format alignment file using a parallelized version of the command "hal2maf" with the parameter "–onlyOrthologs –noAncestors" and "*Danio rerio*" as the reference.

Selection analysis on whole-genome alignments

Neutral model construction. To measure evolutionary conservation or acceleration, we first needed a neutral model to define the baseline rate of evolution for DNA that is not under selective pressure. Following previous studies,[8,9] we used ancestral repetitive elements as our source for these neutrally evolving sequences. To minimize computational effort, we used a dataset of 368 teleost species covering all 44 taxonomic units. We extracted the common ancestral sequences for these species and identified the repeated elements using RepeatMasker (v.3.3.0).[10] From these neutral regions, we randomly sampled 500,000 bases and used their corresponding MAF alignments as input to the phyloFit program (PHAST v.1.556) to generate our neutral model.

Calculation of conservation and acceleration scores. Based on the previously obtained neutral model, we ran PhyloP to estimate conservation and acceleration scores. PhyloP scores represent log-coded *p* values for acceleration, and we converted these scores into *p* values that were then converted into *q* values using the FDR correction method. Any base with a *q* value less than 0.05 was considered significantly conserved or accelerated.

Construction of concatenation-based and coalescent-based phylogenies

For concatenation-based analyses using a single model, we used the GTR+F+R7 and JTT+R6 models for whole-genome alignment supermatrix and protein supermatrix, respectively, because it was the best-fitting model for most of whole-genome loci or gene trees. All phylogenetic analyses were independently performed using IQ-TREE, v.2.3.6[11] with 1,000 ultrafast bootstrap replicates and "-bnni" parameter.

For coalescent-based analysis, individual loci (from the whole-genome alignment dataset) or gene trees (from BUSCO genes) were inferred using IQ-TREE, v.2.3.6, with an automatic detection for the best-fitting model with "-MFP" option using ModelFinder[12] under the Bayesian information criterion. For each locus or gene tree, the topological robustness of each tree was evaluated by 1,000 ultrafast bootstrap replicates and "-bnni" parameter.

Given the comprehensive nature of the whole-genome dataset, which includes a larger amount of genetic information compared with other datasets, the whole-genome ASTRAL-III (v.5.7.3)[6] analysis appears in our study as the main tree to provide a more robust and accurate resolution for these nodes. The topological robustness was evaluated using the local posterior probability (LPP).

Divergence time estimation

Divergence times were estimated using MCMCtree from the PAML package,[13] which required an input phylogeny. The input phylogeny was created by pruning, using ETE3,[14] a phylogenetic tree originally constructed from whole-genome information of 467 fishes. This pruning process reduced the tree to 47 taxa (Table S16), ensuring representation from all teleost fish orders (families) included in the original dataset. The MCMCtree analysis was calibrated using 19 soft-bound fossil calibrations (Table S17) that were carefully reviewed and adjusted according to our fossil record and the International Chronostratigraphic Chart. We applied the HKY85 model of sequence evolution and executed 4 independent runs with a burn-in phase of 100,000 iterations, sampling 10,000,000 times in each run. Divergence times were visualized with FigTree software (http://tree.bio.ed.ac.uk/software/figtree/) and compared using R's ggplot2 package.[15] The results are presented in Figure S5 (the order-level phylogenetic tree). Additional details are provided in the supplemental information.

Diversity trajectory of fishes using fossil sampling data

A total of 13,194 fossil records of Actinopterygii were utilized, with 9,841 specifically pertaining to Teleostei (Table S15). These data were sourced from the DeepBone database (accessible at https://www.deepbone.org). We focused our analysis on the genus and species levels, as they offer the most granular taxonomic information with adequate fossil evidence. Data cleaning involved the removal of species with ambiguous names and fossil records dating beyond the Silurian period due to their uncertainty. Rigorous manual verification ensured data integrity, and any updates were reflected on the DeepBone website.

Subsequently, the refined dataset was categorized into four categories: stem-group Actinopterygii, non-teleost crown-group Actinopterygii, stem-group Teleostei, and crown-group Teleostei. The non-teleost crown-group Actinopterygii includes data from Acipenseriformes/Chondrostei, Holostei, Scanilepiformes, and the stem-group Neopterygii. A subsample dataset was then generated by randomly selecting 90% of the data from each category.

We determined the species' origination and extinction time frames to calculate their existence intervals and mid-ages. The time frame from 500 million years ago (Ma) to the present was partitioned into 10-million-year intervals, and each fossil was assigned to the appropriate interval based on its mid-age. Using the divDyn package (v.0.8.2) in R, we computed the genus/species richness for each major category, considering their taxonomic units and designated time intervals. The resulting diversity curves were then plotted.

RESULTS AND DISCUSSION

A comprehensive comparative genomics resource covering all extant teleost orders

In phase I of the Fish10K project, we strategically sequenced 110 new genomes, which represent 59.09% of extant teleost orders across a total of 44 taxa (37 orders and 7 unclassified families,[16] Tables S1 and S2), filling gaps for 3 previously un-sequenced orders and alleviates genomic scarcity in key clades, thereby advancing comprehensive coverage of teleost diversity (Figure 1A). These genomes were sequenced using a combination of single-tube long-fragment read (stLFR)[17] and long-read sequencing technologies[18] (Table S2). The average scaffold N50, contig N50, and median BUSCO completeness of these newly assembled genomes were 4.22 Mb, 288 kb, and 95.17%, respectively (Table S2). These newly sequenced genomes are comparable in continuity and completeness to most published fish genomes (Figures S1A–S1C). Furthermore, integrating our dataset with 354 publicly available teleost genomes resulted in a comprehensive dataset of 464 teleost species, providing broad representation across all extant teleost orders

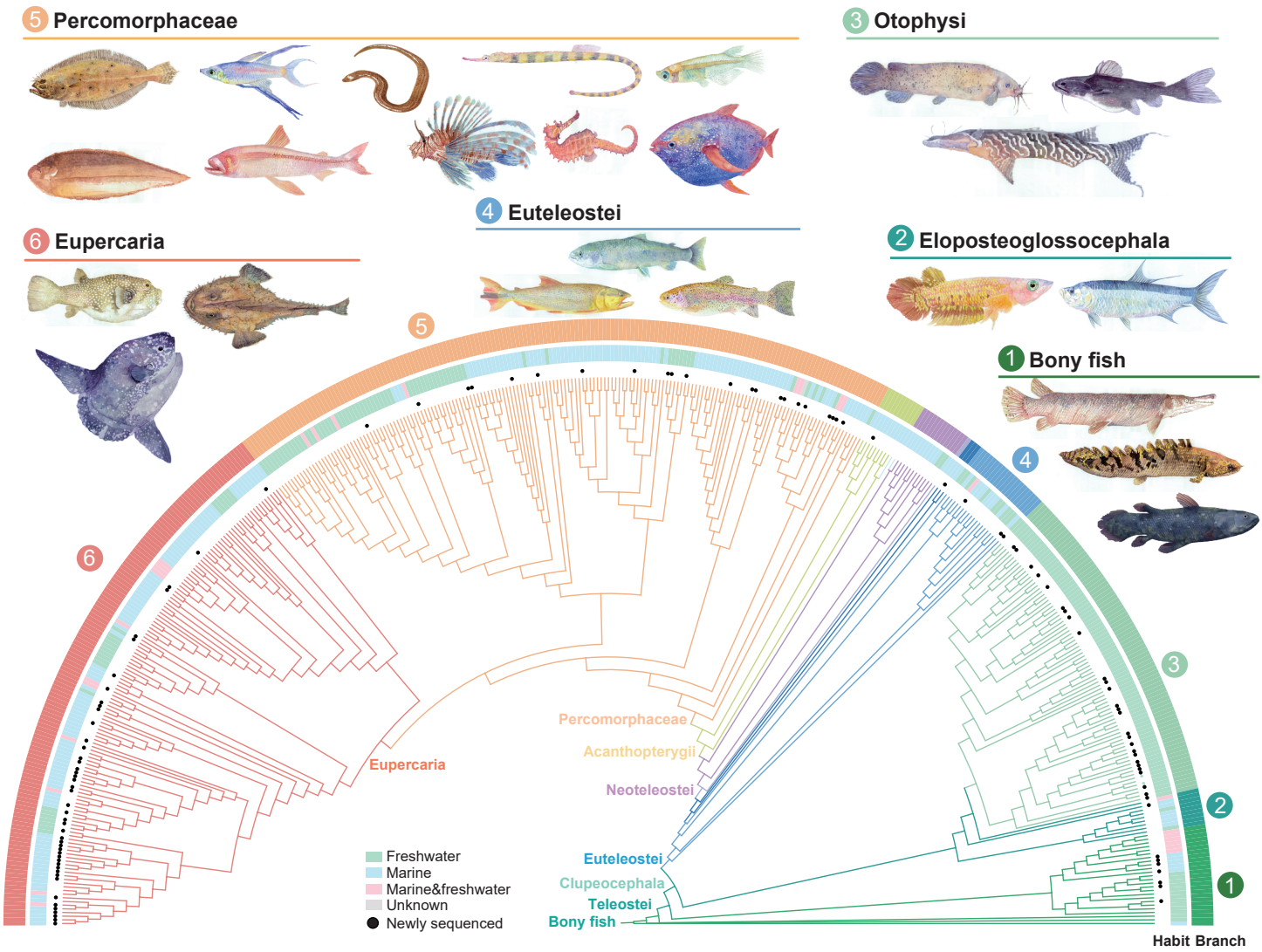


Figure 1. Overview of the teleost orders in the Fish10K phase I compendium A phylogeny of 464 teleost species, reconstructed from a large-scale, reference-free whole-genome alignment built with Cactus. The 110 newly sequenced species are marked with black circles.

(Table S2). A comparison of the genomic features of these 464 teleost species with all published non-teleost ray-finned fish genomes revealed a gradual reduction in genome size during the evolution of ray-finned fishes (Figure S1D). Specifically, intron length showed a progressive shortening from the non-teleost ray-finned fish to teleost. Notably, teleost exhibited significantly shorter introns compared with other ray-finned fish groups, while exon length remained relatively conserved across lineages (Figure S1E; Table S3). These findings provide important insights into the evolutionary patterns of genome architecture in ray-finned fishes.

DNA transposons as the main component of transposable elements in the genome of teleosts

We systematically characterized the distribution of transposable element (TE) content in teleost fish (Figure S1F). We further investigated the distribution patterns of TEs in. The results revealed that the proportion of various TE components in teleosts was significantly lower than in early ray-finned fish, which may explain the genomic contraction observed in teleosts (Figure S1G). In addition, we further investigated the distribution patterns of TEs in different orders of teleost (Figure S1H; Table S4).

Among the three main branches,[19] Notacanthiformes exhibited the highest TE content (11.64%) within Eloposteoglossocephala. The Cypriniformes (21.84%) exhibit the highest transposon content (Figure S1H), representing a major branch within freshwater fishes. To further explore the association between TEs and diverse aquatic habitats, we divided the 464 teleost species into marine and freshwater groups. Our findings revealed that freshwater fish exhibit significantly higher overall transposon content than marine fish (Figure 2A), with notably increased DNA transposon (Figure S1I). We further investigated the positional preferences of TEs across different environmental fish genomes, revealing that the proportion of genes containing TE insertions is significantly higher in freshwater fish than in other groups (Figure 2B). Comparative analysis of genomic features across different vertebrate lineages revealed that DNA transposons constitute the most abundant transposon category in teleost genomes. In contrast, LINE transposons dominate in cartilaginous fish, amphibians, reptiles, and mammals, exhibiting a LINE/DNA transposon ratio distinct from that observed in teleosts (Figures 2C and S1J). Previous studies suggested that TE expansions correlate with adaptive evolution in fishes. For example, a DNA transposon insertion underlies the recurrent gold-dark color polymorphism in Midas cichlids[20]; Antarctic notothenioids recruited LTR/LINE bursts for antifreeze-glycoprotein regulation[21]; and lineage-specific LTR expansions also coincide with alkaline-lake specialization in *Leuciscus waleckii*.[22] Comparative studies of multiple genomes further indicate that the transition from aquatic to terrestrial environments was accompanied by an expansion of the LINE-CR1 superfamily and a higher proportion of DNA transposons in freshwater fish compared with marine fish.[23,24] By extending our analysis to 464 teleost species, we find that (1) the content of TEs in genomes shows a potential association with the survival of teleost in different environments and (2) the TE landscape of teleost is fundamentally different from the "LINE-dominant" paradigm that prevails in chondrichthyans, reptiles, and

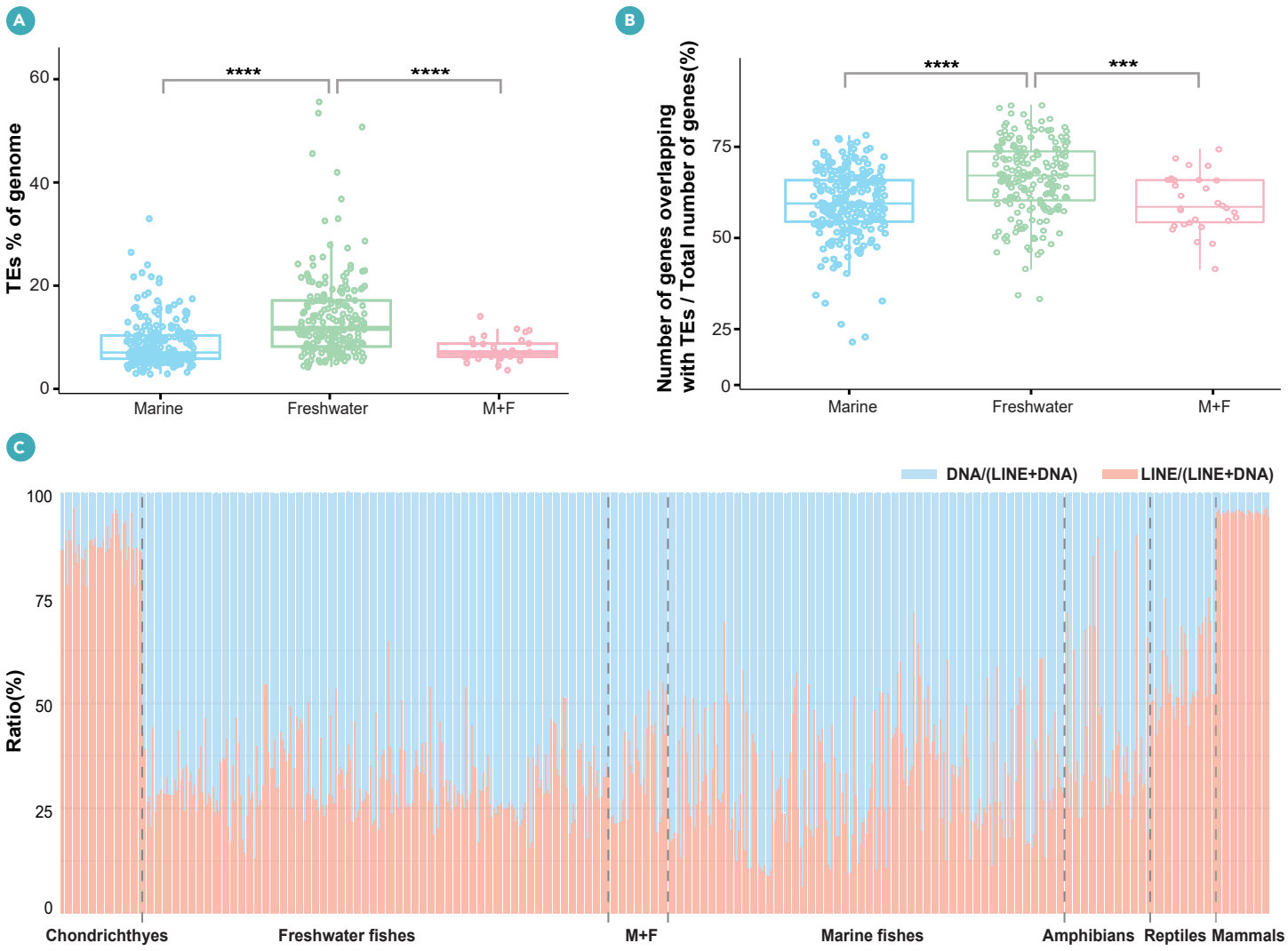


Figure 2. The TE composition of teleost fish (A) Comparison of total TEs content in the genome of teleosts under different environments. (B) Ratio of TEs overlapped with gene region in the genome of teleosts under different environments (marine, freshwater, marine + freshwater). (C) Comparison of TE composition in different branches of vertebrates.

mammals (Table S4).[23–25] Although these findings suggest a potential link between TE activity and environmental adaptation as well as evolution across different vertebrate lineages, further comparative and phylogenetic analyses are required to validate this hypothesis.

Evolutionary constraint and acceleration sites in the teleost genome

With Progressive Cactus[5] and 464 teleost genomes, we generated the largest fish whole-genome alignment to date—spanning ~300 My and all 44 orders—and the third largest for any vertebrate clade after mammals[26] and birds[8] (Figure 3A). This resource now furnishes the first genome-wide panorama of teleost genomic architecture, underscoring their unique evolutionary path relative to other vertebrate lineages.

Based on our whole-genome alignment, we first identified 0.11 Gb of ancestral sequences in the teleost lineages that was lower than in mammals and birds (Figure S2A). Long-term evolutionary conserved sequences are typically subject to purifying selection and tend to retain conserved functions.[27] Genome-wide alignments have proved effective for identifying conserved genomic regions in birds,[8] mammals,[9] and primates,[28] thereby elucidating the evolutionary significance of nucleotide changes. To identify and quantify such conserved regions in fish, we performed the first large-scale, single-base-resolution screen for constrained elements. We detected over 30 Mb of constrained sites, corresponding to ~2% of the zebrafish genome (FDR < 5%), a proportion lower than that reported for mammals and birds (Figure 3B). Previous studies have shown that the evolutionary rates on microchromosomes and sex chromosomes differ from those on autosomes in vertebrates.[29,30] To quantify the extent of selective constraint across the fish genomes, we calculated the proportion of constrained sites on each chromosome of zebrafish. Results indicate that the proportion of individual constraint sites on different chromosomes in zebrafish is lower than that of acceleration sites (Figure 3C). Chromosomes 4 and 18 exhibit the lowest and highest proportions of constraint sites, respectively, with a 2-fold difference between them (Figures 3D and S2B). This pronounced heterogeneity implies that different chromosomes experience distinct selection.

To further explore the functional characteristics and variability of genomic elements in teleosts, we constructed the contiguous accelerated elements (CAEs) and contiguous constrained elements (CCEs) by linking adjacent accelerated and constrained sites, respectively. Ultimately, we identified 24,462 CCEs (Table S5) and 787 CAEs (Table S6), and the length and number of CCEs are significantly higher than those of CAEs (Figure 3E). Previous studies have shown that, although genes are commonly considered the most strongly constrained sequences, most constrained and conserved regions are located in non-coding.[9] Our findings also support the idea that the majority of constrained areas are located in non-coding region (Figure 3F). Furthermore, studies in both mammals and birds have shown that highly conserved elements are enriched for transcriptional regulation and organ development pathways.[9,29] In the functional enrichment analysis of CCEs in teleosts, we identified a consistent pattern: functions are primarily associated with organ development (Figure 3G). This suggests that constrained regions in vertebrates may often perform similar functions during long-term evolution. We also examined the biological

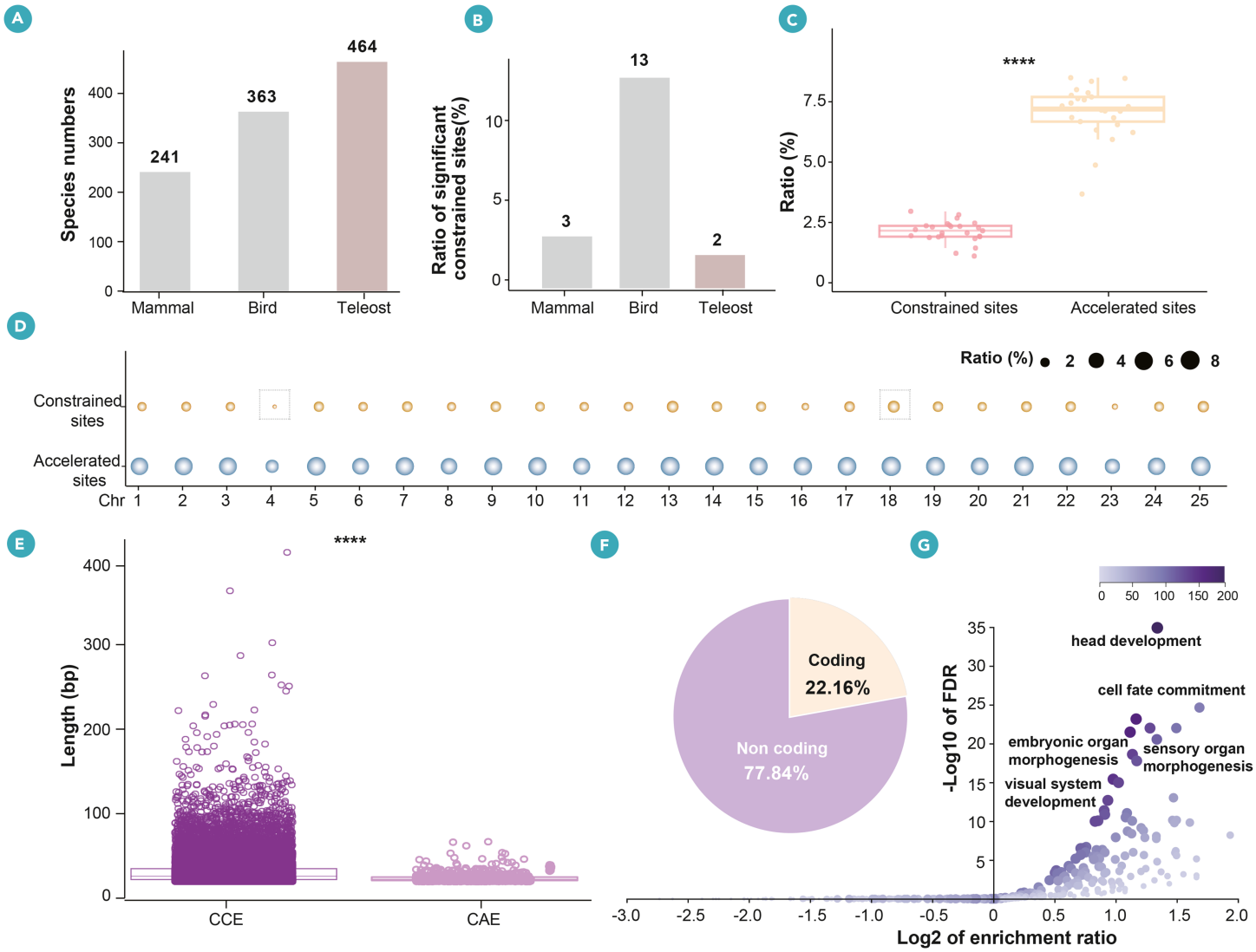


Figure 3. Evolutionary constraint and acceleration in teleost genomes (A) Species number used in different large-scale genome sequencing projects, including birds, mammals, and fishes. (B) Proportion of constrained sites in teleost genomes versus other vertebrate lineages. (C) The proportion of accelerated sites is significantly higher than that of constrained sites. (D) Proportions of constrained and accelerated sites across the 25 zebrafish chromosomes. Bubble size corresponds to the proportion. Dashed boxes highlight the chromosomes with the highest and lowest ratios. (E) Contiguous constrained elements (CCEs) are significantly longer than contiguous accelerated elements (CAEs). (F) Genomic locations of CCEs enriched in non-coding regions. (G) Functional enrichment summary. CCEs are associated with developmental processes.

functions underlying gene acceleration. Previous studies indicate that these accelerated genes are linked to mammalian immunity, skin development, and certain sensory capabilities.[9] Functional analysis of genes harboring accelerated regions in teleosts revealed a preference for immune functions (Figure S2C), consistent with patterns observed in other vertebrate lineages.

Teleost-specific conserved elements drive adaptive evolution post teleost genome duplication

Highly conserved elements (HCEs) are important genomic innovations in the process of macroevolution and are crucial for the adaptive evolution of species.[28–30] Whole-genome-wide comparisons have been instrumental in identifying such elements across vertebrates.[23,31–33] A notable example includes bird-specific conserved elements associated with feather and wing development.[29] By using whole-genome alignments across 464 fishes, we identified 28,878 teleost highly conserved elements (THCEs) (Figure 4A; Table S7). Comparative analysis with four Holostei species genomes (*Amia calva*, *Lepisosteus oculatus*, *Atractosteus spatula*, and *Lepisosteus osseus*) identified 1,689 teleost-specific HCEs (TSHCEs) conserved in >85% of teleosts (Figure S3A; Table S8). Notably, in contrast to other vertebrates where conserved elements are predominantly non-coding,[8,26] teleost TSHCEs exhibited an exon bias with 60% located in exonic regions (Figures S3B–S3D). The remaining non-coding TSHCEs showed substantial overlap with ATAC-seq peaks but limited overlap with TEs (Figure 4B), suggesting that they may retain regulatory potential (Figure 4C).

Previous studies have indicated that conserved elements play pivotal roles in shaping specific traits, often being under strong evolutionary constraint and involved in crucial regulatory functions.[28,29] By integrating zebrafish multi-organ transcriptomes and ZFIN phenotypic data, we explored the role of these conserved elements in teleost evolution. Functional analysis of these conserved elements revealed association with genes expressed in teleost organs, including the brain, pectoral fins, heart, and gills (Figure 4D). Function enrichment further highlighted their preferential involvement in nervous system development and fin morphogenesis (Figure 4E; Table S9). Teleost fishes exhibit extensive morphological and ecological diversification, driven by a long history of adaptive radiation. These novel morphological features, such as the refined fin structures enhancing swimming abilities and the evolution of the gas bladder, have played a pivotal role in seizing new ecological niches and displaying phenotypic diversity.[23,31–35] The emergence of lineage-specific elements has been proposed as a key driver of morphological diversification during vertebrate evolution.[29,36–38] Our findings suggest that some TSHCEs may be linked to genes regulating teleost organ development and contributed to the evolution of lineage-specific morphological traits in teleosts.

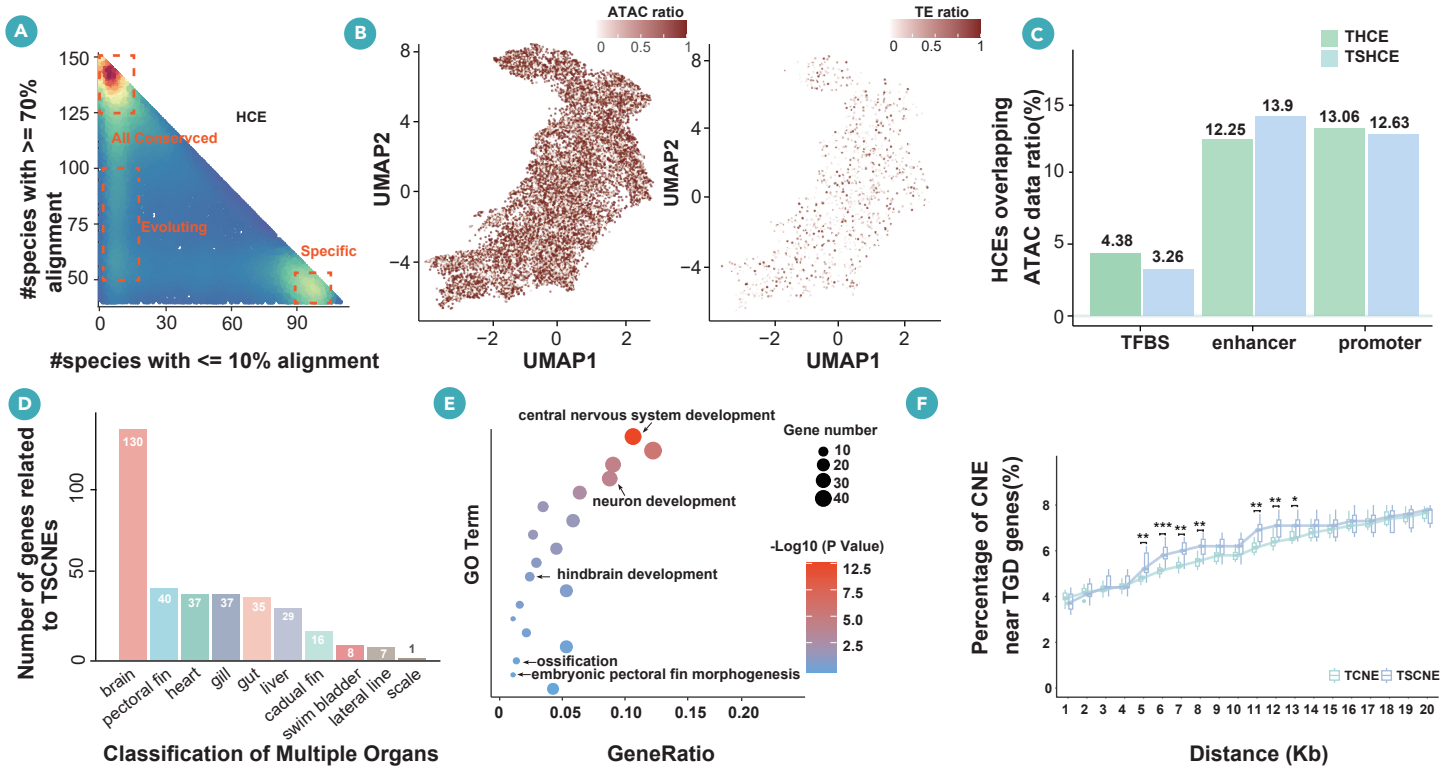


Figure 4. Tracing the evolutionary history and impact of TGD (A) Distribution of THCEs (teleost highly conserved elements) based on their alignment within respective genomes. The *y* axis denotes the number of species in which ≥70% of nucleotides in a THCE align. The *x* axis denotes the number of species in which ≤10% of a THCE align. Three groups corresponding to dense regions in the heatmap are highlighted. (B) The overlap between TSHCEs and ATAC-seq data and between TSHCEs and TEs, respectively. (C) Percentage of overlap between conserved elements and different types of regulatory elements. (D) Functional statistics of genes associated with teleost species conversed non-coding elements (TSCNEs), derived by integrating transcriptome and knockout experiment results from the ZFIN database (https://zfin.org/). (E) Presentation of the GO (Gene Ontology) enrichment results for genes located in proximity to teleost-specific non-coding conserved elements. (F) An analysis of the impacts of TGD (teleost-specific whole-genome duplication) events on teleost-specific conserved non-coding elements (TSCNEs) and teleost conserved non-coding elements (TCNEs).

The teleost-specific genome duplication (TGD) event[36–41] provided an evolutionary context for investigating the origins of these conserved elements. We extracted conserved non-coding elements (CNEs) from THCEs and TSHCEs respectively to detect the impact of TGD events on candidate regulatory elements. Notably, teleost-specific CNEs showed significant enrichment ($p < 0.05$) within 15 kb of TGD-derived genes compared with non-teleost-specific CNEs (Figures 4F and S3E), indicating that TGD not only duplicated protein-coding genes but also potentially facilitated the emergence of novel regulatory elements. This coordinated expansion of both coding and regulatory components demonstrates how whole-genome duplication could drive evolutionary innovation through simultaneous reorganization of gene and regulatory repertoires.

To explore the regulatory potential of TSHCEs, we examined a TSHCE segment located upstream of the *rfx4* gene (Figure 5A; Table S10). *rfx4* controls dorsal and ventral midline formation in the neural tube, and its loss-of-function mutants in zebrafish exhibit caudal neural defects, impaired swimming, and swim bladder inflation failure.[43] Those roles in posterior neural development and motor coordination suggest that *rfx4* may contribute to the evolution of axial patterning and swimming control in teleosts.[44] Hi-C data analysis revealed that this *rfx4*-TSHCE segment interacts strongly with the *rfx4* gene, implying a regulatory relationship (Figure 5B). Furthermore, ChIP-seq data identified enrichment of regulatory chromatin marks at this locus (Figure 5C). To validate this functionally, we constructed dual-luciferase reporter plasmids based on pGL3-Basic: a control group (pGL3 vector), and two experiment groups (pGL3–871bp and pGL3–871bp+TSHCE). The upstream region along (pGL3–871bp construct) exhibited luciferase activity comparable with the control, whereas inclusion of the TSHCE (pGL3–871bp+TSHCE) significantly increased reporter activity by approximately 4.2-fold (two-sample Student's *t* test, $p < 0.05$) (Figure 5D). Together, these results indicate that the *rfx4*-TSHCE could act as a *cis*-regulatory element of *rfx4*. It may have fine-tuned *rfx4* expression during posterior neural development and motor coordination, potentially contributing to teleost-specific adaptations in axial patterning and swimming efficiency, including modular axial skeletons, reconfigured myomeres for enhanced propulsion, and homocercal caudal fins.[44,45]

Whole-genome phylogenomics challenges the monophyly of "Siluriphysi"

The evolutionary history of teleosts, characterized by rapid radiation events and long-time divergences, requires phylogenomic approaches to overcome the resolution limitations inherent to traditional molecular markers.[8] Using whole-genome alignments from 464 species (comprising 3.18 Mb of orthologous sequences partitioned into 3,185 1-kb windows and 1,000 BUSCO genes, Figure S4), we reconstructed a robust phylogeny covering all ordinal-level lineages. Both concatenation and coalescent-based methods yielded highly supported trees (average bootstrap value = 88.27%) (Figures S5A–S5D) and high topological consistency of Robinson-Foulds distances (Figures S5E and S5F; Table S11) across all methods. While 89% of internodes exhibited concordance, persistent conflicts remained within certain lineages (e.g., Otophysi; Figures 6A, 6B, and S5A–S5D; Table S12). For these contentious nodes, we employed the ASTRAL coalescent approach to reduce the effects of incomplete lineage sorting.[8,26]

Our phylogenetic analyses provide novel insights into the relationships among Gymnotiformes, Siluriformes, and Characiformes (Figures 6A–6C). Although the protein-coding sequence analyses support the monophyly of Siluriformes and Gymnotiformes relationships (Table S12; Node A in Figure 6 A), whole-genome coalescent analyses strongly suggest an alternative topology placing Gymnotiformes as sister to Characiformes-Siluriformes clade (bootstrap [BS] = 100%, local posterior probability [LPP] = 1.0; Figures 6A and 6B), thereby challenging the current consensus on Otophysi relationships (Figure 6B; Table S13). Polytomy

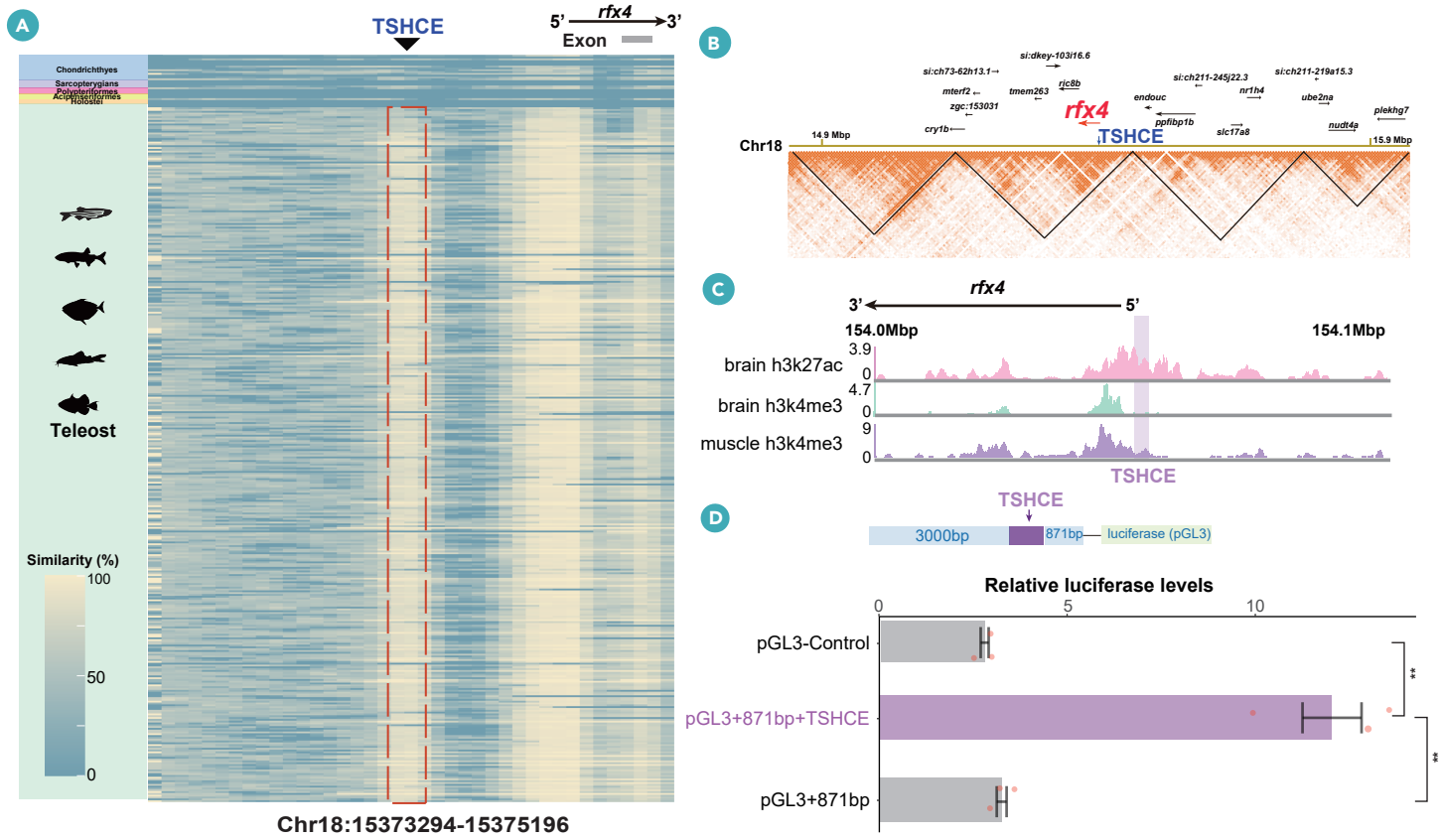


Figure 5. Functional validation of TSHCEs (A) A case showing the location of TSHCE and how well it is conserved in teleost fishes. (B) Hi-C data demonstrating that TSHCE and *rfx4* genes are located in the same structural domain (black triangles). (C) ChIP-seq data showing overlap between different peaks and TSHCE, with data from the DANIO-CODE[42] project of brain and muscle data. (D) To validate this functionally, we constructed dual-luciferase reporter plasmids based on pGL3-Basic: a control group (pGL3 vector), and two experiment groups (pGL3–871bp and pGL3–871bp+TSHCE). The upstream region along (pGL3–871bp construct) exhibited luciferase activity comparable with the control, whereas inclusion of the TSHCE (pGL3–871bp+TSHCE) significantly increased reporter activity by approximately 4.2-fold (two-sample Student's *t* test, $p < 0.05$).

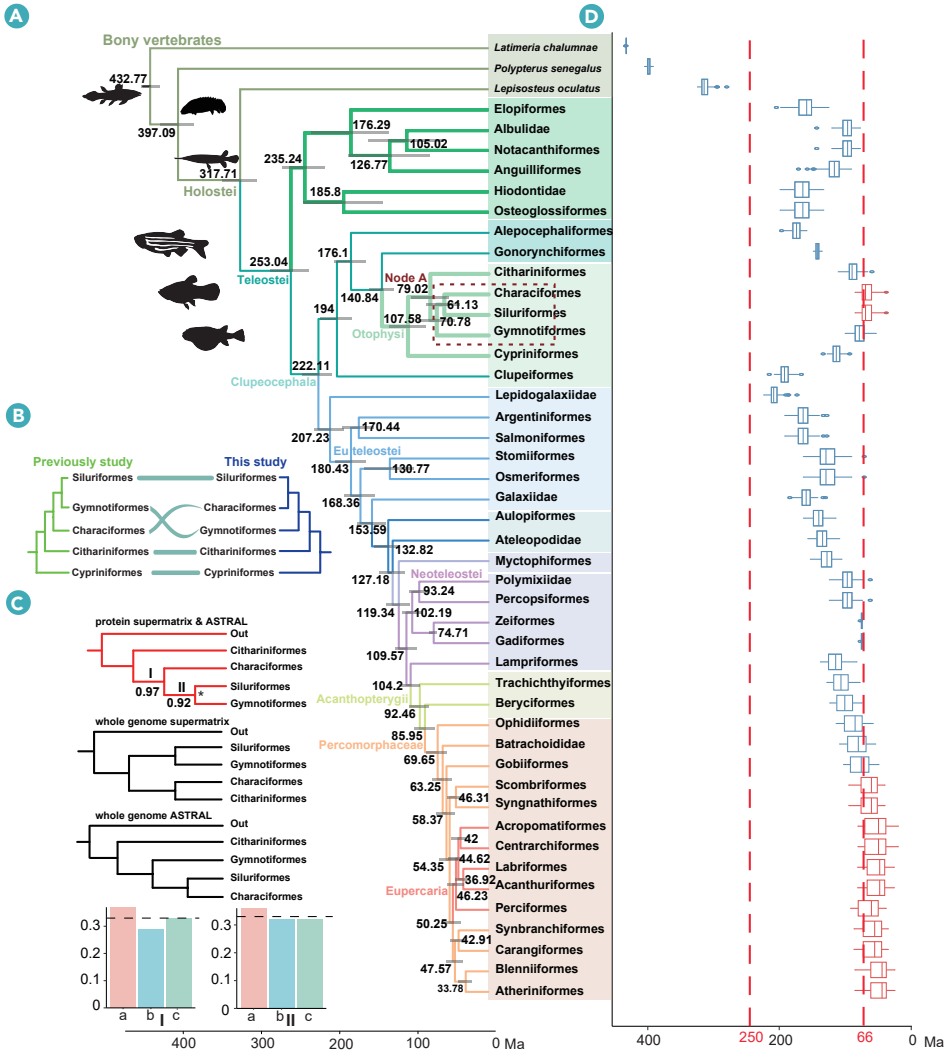
tests further support a potential trifurcation among these orders (Figure 6C). In addition, microsynteny analyses among Otophysi species, based on 14,294 orthologous genes located in conserved microsynteny blocks (zebrafish as the reference; Table S14) revealed that, while 37.12% of gene trees supported monophyly of "Siluriphysi," approximately 60% contradicted this hypothesis (Figure S6). Quartet support tests further confirmed phylogenetic instability within this clade. These conflicting signals both in protein-based and genome-scale datasets challenge the currently prevail view that Siluriphysi constitutes a monophyletic group[45–47] (Figure 6B; Table S13). We also collected mitochondrial genome data from 506 teleost species, covering all currently recognized orders. Using 12 mitochondrial protein-coding genes, we reconstructed a robust mitochondrial phylogeny. Notably, the mitochondrial tree strongly supports our previously conclusions, particularly regarding the challenges of monophyly of Siluriphysi (Figure S10).

Within the Otocephala clade, the monophyly hypothesis of Siluriphysi (comprising Gymnotiformes and Siluriformes), originally proposed based on morphological characteristics including the electroreception system,[45–47] has long been widely accepted. Although numerous molecular phylogenetic studies have demonstrated frequent discordance between molecular data and morphological expectations, the monophyly of Siluriphysi has generally been maintained in traditional interpretations, from early single-locus analyses[48] to recent multiple genomic-scale loci,[49–56] primarily to conform to morphological predictions. Nonetheless, some molecular phylogenetic studies (e.g., Figures S2, S4, and S5 in Hughes et al.'s study, and in Near et al.'s and Melo et al.'s studies)[54–56] failed to support Siluriphysi monophyly. The limited molecular markers employed in these studies resulted in inconsistent topological structures, precluding definitive challenges to the Siluriphysi hypothesis. Dornburg and Near[57] recently identified this issue as one of the unresolved questions in teleost evolution. Our whole-genome alignments data yielded a maximally supported topology (Figure 6) that provides conclusive evidence for the hypothesis proposed by Dornburg and Near[57] and Near and Thacker.[16] This result presents a reappraisal of how electroreception arose within Otophysi: either it evolved twice independently in catfish and knifefish (convergent-evolution hypothesis) or was secondarily lost in characiforms (ancestral-loss hypothesis). Beyond underlining the need for denser taxon sampling, the study demonstrates that only comprehensive genomic datasets can untangle the deepest branches of the teleost tree.

Genome-wide analyses uncover a novel model of teleost radiation dynamics

Our study leveraged the largest teleost whole-genome alignment to date (3.18 Mb of orthologous sequences, Figure S4) to precisely reconstruct the teleost evolutionary timeline. Integrating paleontological and molecular clock data, our analysis indicates that teleosts originated ~253 Ma, preceding the Permian-Triassic extinction (251.9 Ma) (Figures 6A and S7). By the time of Gondwana's breakup (~180 Ma), teleosts had already diversified into three primary clades: Osteoglossomorpha, Elopomorpha, and Clupeocephala (Figure 6A). Key mass extinctions—particularly the end-Permian (~250 Ma) and K-Pg (~66 Ma) events—stimulated teleost radiation by vacating ecological niches and reducing competition from non-teleost lineages (Figure 6D). The precise molecular clock and lineage-through-time (LTT) analyses reveal three evolutionary phases in crown teleosts: (1) rapid diversification (origin to ~170 Ma), (2) a period of declining diversification (~170–66 Ma), and (3) post-K-Pg phase beginning with a brief stabilization, followed by a transient resurgence in diversification before decreasing to modern levels (Figures 7A, 7B, and S7). These major inflection points coincide with two key mass extinction events, the end-Permian (~250 Ma) and K-Pg (~66 Ma) (Figures 7C and S8).

Additionally, our results challenge previous hypotheses suggesting a continuous increase in diversification rates following the K-Pg extinction.[58–61] In contrast, both our LTT analysis (Figure 7A) and fossil evidence (fossil data from 13,194 Actinopterygii records, including 9,841 Teleostei. Table S15) revealed stable net diversification rates (rate through time) for crown teleosts



After the Permian-Triassic extinction, life recovered and diversified as Pangaea's formation enabled

The coastlines underwent continuous expansion, providing various habitats for the spiny-finned fishes.

Figure 6. A reconstructed robust phylogeny covering all ordinal-level lineages by using whole-genome alignment dataset (A) The time-calibrated phylogeny of teleosts inferred from whole-genome alignments. Blue bars at nodes indicate the 95% confidence intervals for divergence times. (B) Competing phylogenetic hypotheses for the relationships within Ostariophysi based on different datasets. The whole-genome phylogenomics challenges the monophyly of "Siluriphysi." (C) Polytomy tests and the count of loci trees also support a potential trifurcation among these orders. (D) Estimated divergence times for all extant teleost orders. Blue boxes indicate an origin before the K-Pg extinction (66 Ma); red boxes indicate an origin after the extinction. (E) Paleomaps illustrating major geological events corresponding with the divergence of major teleost lineages.

throughout the Paleogene to early Neogene (60–23 Ma) with only minor fluctuations (Figures 7B and S8). This pattern aligns with established phylogenetic models[62] and supports our three-phase evolutionary framework. Except for the ecological opportunities created by mass extinction events (Figure 7B), our refined evolutionary timeline reveals that teleost groups responded differently to temperature fluctuations: modern crown teleosts experienced markedly suppressed diversification during episodes of global warming (Figures 7C and S9).

CONCLUSION

The large-scale genomic dataset of teleost fish constructed in this study provides robust support for systematic analysis of their evolutionary origins. Through systematic analysis of 464 teleost genomes, we discovered that DNA transposons have replaced LINE elements as the predominant components in teleost genomes. This finding represents a significant divergence from the traditional LINE-dominant pattern observed across most vertebrate lineages. This novel feature of teleost genomes warrants further investigation into its potential association with the adaptive radiation of teleosts. Furthermore, based on large-scale whole-genome alignment, we have for the first time identified genome-wide constraint sites in teleost fish at single-base resolution and systematically characterized genome sequences specifically generated in teleosts. This provides a critical data foundation for understanding genomic evolutionary pressures in teleosts. We also constructed a comprehensive teleost phylogenetic tree and confirmed a new phylogenetic relationship between the Gymnotiformes and Siluriformes, challenging the monophyly of the Siluriphysi clade. Furthermore, we systematically evaluated the evolutionary dynamics of teleost radiation through molecular clock estimation and fossil data.

The Fish10K project has produced a comprehensive genomic foundation to systematically reconstruct teleost evolutionary history and decipher the genomic basis of their phenotypic diversity. These resources provide enduring value by enabling both species-specific studies in areas such as trait evolution and conservation genomics, as well as cross-species comparisons to uncover fundamental patterns of vertebrate genome evolution and adaptive radiation.

RESOURCE AVAILABILITY

Materials availability

This study did not generate new unique materials or reagents. All experimental animal treatments in this study have been verified and identified by taxonomic experts and museum taxonomists according to the guidelines approved by the instituted Review Board of Bioethics and Biosafety (BGI-IRB, ethical permit ID: BGI-IRB A20007-T1).

Data and code availability

- For data availability, all the data produced in this study are stored in publicly available databases.
- The newly assembled genomes of 110 fishes from this study have been submitted to the CNSA of CNGBdb (https://db.cngb.org/cnsa/): CNP0004403.
- We also established a BioProject on NCBI specifically for depositing the fish genomes generated from the Fish10K project.
- The 110 newly assembled genomes generated from this research are also stored in the NCBI database (https://www.ncbi.nlm.nih.gov/): PRJNA1209848.
- The whole-genome HAL format alignment files for 467 fish species (including 3 non-teleost fishes as the outgroup) have also been submitted to the CNSA database of CNGBdb: ftp://ftp.cngb.org/pub/CNSA/data5/CNP0004661/Other/467-Bony-fish.hal and can be accessed using accession number CNP0004661; alternatively, users who wish to use this dataset can download it with the following command: "wget -c ftp://ftp.cngb.org/pub/CNSA/data5/CNP0004661/Other/467-Bony-fish.hal".
- The Newick format tree file for all 467 fish species (including 3 non-teleost fishes as the outgroup) can be found in the "Data S1–S4."

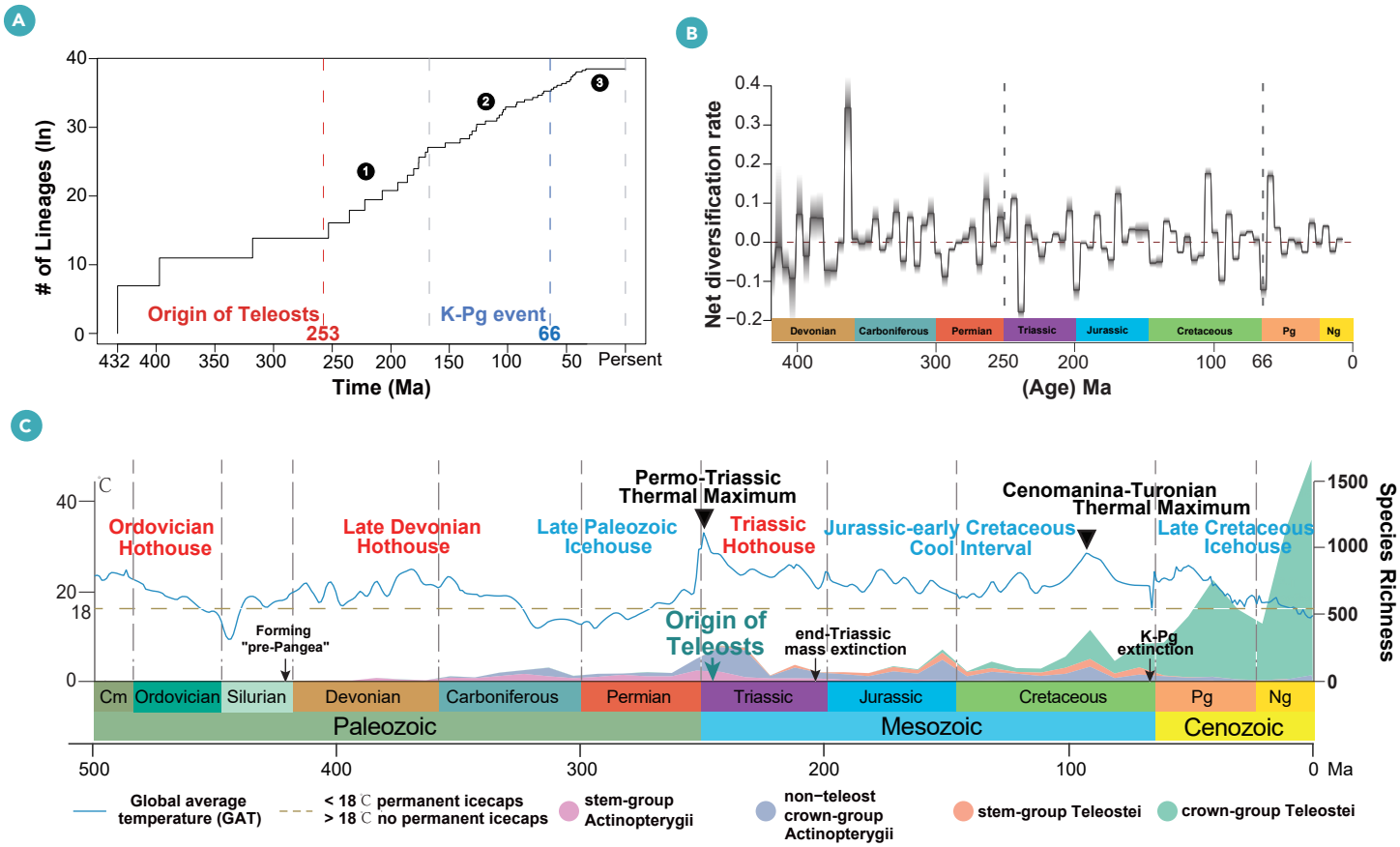


Figure 7. The distribution of teleost phyloteny and correlation between teleost diversification, geological events, and mass extinctions (A) Lineage-through-time (LTT) plot for crown teleosts. The analysis reveals three distinct phases in diversification dynamics. (B) Net diversification rates (speciation minus extinction) of teleosts through time. The solid line is the mean posterior rate; the shaded area is the 95% confidence interval. Rates below zero indicate declining diversity. (C) Changes in the species richness of Actinopterygii based on the fossil record. Data for (B) and (C) were sourced from the DeepBone database (https://www.deepbone.cn/home/).

- The database used for TE annotation in this study is available online (http://www.repeatmasker.org).
- The protein sequences of the four fish species used for homology annotation are sourced from the RefSeq database of NCBI and can be publicly accessed on NCBI via the following accession numbers: *Danio rerio* (GCF_000002035.6), *Lepisosteus oculatus* (GCF_040954835.1), *Takifugu rubripes* (GCF_901000725.2), and *Rhincodon typus* (GCF_021869965.1).
- For code availability, scripts used in this study, such as phylogenetic and gene annotation pipelines, are available on the GitHub page of Fish10K (https://github.com/BGI-Qingdao/fish10k).
- The scripts for running the Cactus software are based on the usage guidelines of Cactus; the program used for the "hal2maf" conversion can be found in the section "comparative genomic analysis" from Fish10K GitHub repository (https://github.com/BGI-Qingdao/fish10k/).

FUNDING AND ACKNOWLEDGMENTS

We would like to express our gratitude to Professor Thomas J. Near for providing critical samples that were essential to this study. We also extend our sincere thanks to Ms. Xunwei Xie, Ms. Linglu Li, Ms. Liyue Liu, Ms. Yuejia Sun, Dr. Yaqing Wang, Dr. Fang Zhou, Dr. Guangxin Wang, and Dr. Feng Xiong for their great help in the experiment design and cell culture. This research was supported by The General Program (Key Program, Major Research Plan) of National Natural Science Foundation of China nos. 32170439 (to G.Y.F.), 32122021 (to K.W.), 32293252, and 32170438 (to C.C.F.); the Thousands Marine Species Genome Sequencing Project of Qingdao Free Trade Zone Management Committee (to G.Y.F.); Shandong Province Key Research and Development Plan no. 2023CXPT057 (to G.Y.F.); as well as the Specially appointed Professor Program of Jiangsu Province, the Jiangsu Foreign Expert Bureau, and the Jiangsu Provincial Department of Technology grant JSSCTD202142 (to I.S.). This research also received technical support from the China National Gene Bank and supported by the High-performance Computing Platform of YZBSTCACC.

AUTHOR CONTRIBUTIONS

S.H., G.F., X.X., G.Z., K.W., Huanming Yang, and W.W. conceived and designed the study. L.Y., Hengjia Yang, Mengqi Zhang, S.L., Z.Z., G.H., S.H., J.L., C.L., and X.C. performed sample collection and sequencing. Y. Song, S.Z., Z.Y., Y.Z., J.W., M.L., M.H., M.Y., and Q.L. performed assembly and annotation. Y. Song, S.Z., Hongyue Liu, X.T., Yuanning Li, and Chenguang Feng performed phylogenetic analysis. Y. Song, Z.P., M.W., Qingming Qu, and Min Zhu collected and organized fossil record data as well as analyzing fossil biota abundance. Y. Song, Z.Y., and C.Z. performed highly conserved element identification. Chengchi Fang, C.W., L.P., and Y. Sun performed the functional analysis of highly conserved elements. L.Y., and L.P. performed *in situ* hybridization and CRISPR-Cas9-mediated experiments. I.S., K.W., Qiang Qiu, W.W., G.Z., S.H., W.X., X.G., Honghui Zeng, Hanbo Li, He Zhang, Yongxin Li, Simon Ming-Yuen Lee, X.L., K.K., J.W., S.C., J.G., and Y.C. participated in discussions and provided suggestions. Y. Song, I.S., K.W., G.F., G.Z., and S.H. wrote and revised the manuscript with the input from all co-authors. All authors contributed to the manuscript and approved the final version.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

It can be found online at https://doi.org/10.1016/j.xinn.2025.101177.

REFERENCES

1. Aparicio, S., Chapman, J., Stupka, E. et al. (2002). Whole-Genome Shotgun Assembly and Analysis of the Genome of Fugu rubripes. *Science* **297**:1301–1310. DOI:10.1126/science.1072104
2. Fan, G., Song, Y., Yang, L. et al. (2020). Initial data release and announcement of the 10,000 Fish Genomes Project (Fish10K). *GigaScience* **9**:giaa080. DOI:10.1093/gigascience/giaa080
3. Bass, A., Compagno, L. and Heemstra, P. (1986). Smiths' Sea Fishes (Springer-Verlag Berlin)

4. Weisenfeld, N.I., Kumar, V., Shah, P. et al. (2017). Direct determination of diploid genome sequences. *Genome Res.* **27**:757–767. DOI:10.1101/gr.214874.116
5. Armstrong, J., Hickey, G., Diekhans, M. et al. (2020). Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587**:246–251. DOI:10.1038/s41586-020-2871-y
6. Zhang, C., Rabiee, M., Sayyari, E. et al. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinf.* **19**:153. DOI:10.1186/s12859-018-2129-y
7. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**:1312–1313. DOI:10.1093/bioinformatics/btu033
8. Feng, S., Stiller, J., Deng, Y. et al. (2020). Dense sampling of bird diversity increases power of comparative genomics. *Nature* **587**:252–257. DOI:10.1038/s41586-020-2873-9
9. Christmas, M.J., Kaplow, I.M., Genereux, D.P. et al. (2023). Evolutionary constraint and innovation across hundreds of placental mammals. *Science* **380**:eabn3943. DOI:10.1126/science.abn3943
10. Tarailo-Graovac, M. and Chen, N. (2009). Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. *Curr. Protoc. Bioinform.* **25**:4.10.1–4.10.14. DOI:10.1002/0471250953.bi0410s25
11. Minh, B.Q., Schmidt, H.A., Chernomor, O. et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**:1530–1534. DOI:10.1093/molbev/msaa015
12. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F. et al. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**:587–589. DOI:10.1038/nmeth.4285
13. Dos Reis, M., Álvarez-Carretero, S. and Yang, Z. (2017). MCMCTree tutorials. http://abacus.gene.ucl.ac.uk/software/MCMCtree.Tutorials.pdf
14. Huerta-Cepas, J., Serra, F. and Bork, P. (2016). ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol. Biol. Evol.* **33**:1635–1638. DOI:10.1093/molbev/msw046
15. Wilkinson, L. (2011). ggplot2: Elegant Graphics for Data Analysis by WICKHAM, H (Oxford University Press)
16. Near, T.J. and Thacker, C.E. (2024). Phylogenetic classification of living and fossil ray-finned fishes (Actinopterygii). *Bull. - Peabody Mus. Nat. Hist.* **65**:3–302. DOI:10.3374/014.065.0101
17. Wang, O., Chin, R., Cheng, X. et al. (2019). Efficient and unique cobarcoding of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and *de novo* assembly. *Genome Res.* **29**:798–808. DOI:10.1101/gr.245126.118
18. Rhoads, A. and Au, K.F. (2015). PacBio Sequencing and Its Applications. *Genom. Proteom. Bioinform.* **13**:278–289. DOI:10.1016/j.gpb.2015.08.002
19. Parey, E., Louis, A., Montfort, J. et al. (2023). Genome structures resolve the early diversification of teleost fishes. *Science* **379**:572–575. DOI:10.1126/science.abq4257
20. Kratochwil, C.F., Kautt, A.F., Nater, A. et al. (2022). An intronic transposon insertion associates with a trans-species color polymorphism in Midas cichlid fishes. *Nat. Commun.* **13**:296. DOI:10.1038/s41467-021-27685-8
21. Bista, I., Wood, J.M.D., Desvignes, T. et al. (2023). Genomics of cold adaptations in the Antarctic notothenioid fish radiation. *Nat. Commun.* **14**:3412. DOI:10.1038/s41467-023-38567-6
22. Jian, X., Jiong-Tang, L., Yanliang, J. et al. (2017). Genomic Basis of Adaptive Evolution: The Survival of Amur Ide (Leuciscus waleckii) in an Extremely Alkaline Environment. *Mol. Biol. Evol.* **34**:145–159. DOI:10.1093/molbev/msw230
23. Wang, K., Wang, J., Zhu, C. et al. (2021). African lungfish genome sheds light on the vertebrate water-to-land transition. *Cell* **184**:1362–1376.e18. DOI:10.1016/j.cell.2021.01.047
24. Keane, M., Semeiks, J., Webb, A.E. et al. (2015). Insights into the evolution of longevity from the bowhead whale genome. *Cell Rep.* **10**:112–122. DOI:10.1016/j.celrep.2014.12.008
25. Lucinda, P.M.F., Elaine, M. and Richard, W. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**:695–716. DOI:10.1038/nature03154
26. Zoonomia Consortium (2020). A comparative genomics multitool for scientific discovery and conservation. *Nature* **587**:240–245. DOI:10.1038/s41586-020-2876-6
27. McLean, C.Y., Reno, P.L., Pollen, A.A. et al. (2011). Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature* **471**:216–219. DOI:10.1038/nature09774
28. Kuderna, L.F.K., Ulirsch, J.C., Rashid, S. et al. (2024). Identification of constrained sequence elements across 239 primate genomes. *Nature* **625**:735–742. DOI:10.1038/s41586-023-06798-8
29. Seki, R., Li, C., Fang, Q. et al. (2017). Functional roles of Aves class-specific cis-regulatory elements on macroevolution of bird-specific features. *Nat. Commun.* **8**:14229. DOI:10.1038/ncomms14229
30. Lindblad-Toh, K., Garber, M., Zuk, O. et al. (2011). A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**:476–482. DOI:10.1038/nature10530
31. Bi, X., Wang, K., Yang, L. et al. (2021). Tracing the genetic footprints of vertebrate landing in non-teleost ray-finned fishes. *Cell* **184**:1377–1391.e14. DOI:10.1016/j.cell.2021.01.046
32. Erwin, D.H. (2015). Novelty and Innovation in the History of Life. *Curr. Biol.* **25**:R930–R940. DOI:10.1016/j.cub.2015.08.019
33. Giammona, F.F. (2021). Form and Function of the Caudal Fin Throughout the Phylogeny of Fishes. *Integr. Comp. Biol.* **61**:550–572. DOI:10.1093/icb/icab127
34. Jønsson, K.A., Fabre, P.-H., Fritz, S.A. et al. (2012). Ecological and evolutionary determinants for the adaptive radiation of the Madagascan vangas. *Proc. Natl. Acad. Sci. USA* **109**:6620–6625. DOI:10.1073/pnas.1115835109
35. Lauder, G.V. (2000). Function of the caudal fin during locomotion in fishes: kinematics, flow visualization, and evolutionary patterns. *Am. Zool.* **40**:101–122. DOI:10.1093/icb/40.1.101
36. Glasauer, S.M.K. and Neuhauss, S.C.F. (2014). Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol. Genet. Genomics.* **289**:1045–1060. DOI:10.1007/s00438-014-0889-2
37. Dornburg, A., Wcisel, D.J., Zapfe, K. et al. (2021). Holosteans contextualize the role of the teleost genome duplication in promoting the rise of evolutionary novelties in the ray-finned fish innate immune system. *Immunogenetics* **73**:479–497. DOI:10.1007/s00251-021-01225-6
38. Meyer, A. and Van de Peer, Y. (2005). From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *Bioessays* **27**:937–945. DOI:10.1002/bies.20293
39. Opazo, J.C., Butts, G.T., Nery, M.F. et al. (2013). Whole-genome duplication and the functional diversification of teleost fish hemoglobins. *Mol. Biol. Evol.* **30**:140–153. DOI:10.1093/molbev/mss212
40. Volff, J.N. (2005). Genome evolution and biodiversity in teleost fish. *Heredity* **94**:280–294. DOI:10.1038/sj.hdy.6800635
41. Pasquier, J., Cabau, C., Nguyen, T. et al. (2016). Gene evolution and gene expression after whole genome duplication in fish: the PhyloFish database. *BMC Genom.* **17**:368. DOI:10.1186/s12864-016-2709-z
42. Baranasic, D., Hörtenhuber, M., Balwierz, P.J. et al. (2022). Multiomic atlas with functional stratification and developmental dynamics of zebrafish cis-regulatory elements. *Nat. Genet.* **54**:1037–1050. DOI:10.1038/s41588-022-01089-w
43. Sedykh, I., Keller, A.N., Yoon, B. et al. (2018). Zebrafish Rfx4 controls dorsal and ventral midline formation in the neural tube. *Dev. Dyn.* **247**:650–659. DOI:10.1002/dvdy.24613
44. Grillner, S. (2021). Evolution of the vertebrate motor system - from forebrain to spinal cord. *Curr. Opin. Neurobiol.* **71**:11–18. DOI:10.1016/j.conb.2021.07.016
45. Albert, J.S. (2001). Species Diversity and Phylogenetic Systematics of American Knifefishes (Gymnotiformes, Teleostei) (Ann Arbor: University of Michigan Museum of Zoology), p. 127 pp, (Miscellaneous Publications 190.)
46. Fink, S.V. and Fink, W.L. (1981). Interrelationships of the ostariophysan fishes (Teleostei). *Zool. J. Linn. Soc.* **72**:297–353. DOI:10.1111/j.1096-3642.1981.tb01575.x
47. Liu, Z., Liu, S., Yao, J. et al. (2016). The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts. *Nat. Commun.* **7**:11757. DOI:10.1038/ncomms11757
48. Orti, G. and Meyer, A. (1996). Molecular evolution of ependymin and the phylogenetic resolution of early divergences among euteleost fishes. *Mol. Biol. Evol.* **13**:556–573. DOI:10.1093/oxfordjournals.molbev.a025616
49. Arcila, D., Ortí, G., Vari, R. et al. (2017). Genome-wide interrogation advances resolution of recalcitrant groups in the tree of life. *Nat. Ecol. Evol.* **1**:1–10. DOI:10.1038/s41559-016-0020
50. Betancur-R, R., Wiley, E.O., Arratia, G. et al. (2017). Phylogenetic classification of bony fishes. *BMC Evol. Biol.* **17**:162. DOI:10.1186/s12862-017-0958-3
51. Chakrabarty, P., Faircloth, B.C., Alda, F. et al. (2017). Phylogenomic Systematics of Ostariophysan Fishes: Ultraconserved Elements Support the Surprising Non-Monophyly of Characiformes. *Syst. Biol.* **66**:881–895. DOI:10.1093/sysbio/syx038
52. Dai, W., Zou, M., Yang, L. et al. (2018). Phylogenomic Perspective on the Relationships and Evolutionary History of the Major Otocephalan Lineages. *Sci. Rep.* **8**:205. DOI:10.1038/s41598-017-18432-5
53. Nakatani, M., Miya, M., Mabuchi, K. et al. (2011). Evolutionary history of Otophysi (Teleostei), a major clade of the modern freshwater fishes: Pangaean origin and Mesozoic radiation. *BMC Evol. Biol.* **11**:177. DOI:10.1186/1471-2148-11-177
54. Hughes, L.C., Ortí, G., Huang, Y. et al. (2018). Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. *Proc. Natl. Acad. Sci. USA* **115**:6249–6254. DOI:10.1073/pnas.1719358115
55. Near, T.J., Eytan, R.I., Dornburg, A. et al. (2012). Resolution of ray-finned fish phylogeny and timing of diversification. *Proc. Natl. Acad. Sci. USA* **109**:13698–13703. DOI:10.1073/pnas.1206625109
56. Melo, B.F., Sidlauskas, B.L., Near, T.J. et al. (2021). Accelerated diversification explains the exceptional species richness of tropical characoid fishes. *Syst. Biol.* **71**:78–92. DOI:10.1093/sysbio/syab040
57. Dornburg, A. and Near, T.J. (2021). The emerging phylogenetic perspective on the evolution of actinopterygian fishes. *Annu. Rev. Ecol. Evol. Syst.* **52**:427–452. DOI:10.1146/annurev-ecolsys-122120-122554
58. Alfaro, M.E., Faircloth, B.C., Harrington, R.C. et al. (2018). Explosive diversification of marine fishes at the Cretaceous–Palaeogene boundary. *Nat. Ecol. Evol.* **2**:688–696. DOI:10.1038/s41559-018-0494-6
59. Friedman, M. (2009). Ecomorphological selectivity among marine teleost fishes during the end-Cretaceous extinction. *Proc. Natl. Acad. Sci. USA* **106**:5218–5223. DOI:10.1073/pnas.0808468106
60. Friedman, M. (2010). Explosive morphological diversification of spiny-finned teleost fishes in the aftermath of the end-Cretaceous extinction. *Proc. Biol. Sci.* **277**:1675–1683. DOI:10.1098/rspb.2009.2177
61. Friedman, M. (2022). The macroevolutionary history of bony fishes: a paleontological view. *Annu. Rev. Ecol. Evol. Syst.* **53**:353–377. DOI:10.1146/annurev-ecolsys-111720-010447
62. Ghezelayagh, A., Harrington, R.C., Burress, E.D. et al. (2022). Prolonged morphological expansion of spiny-rayed fishes following the end-Cretaceous. *Nat. Ecol. Evol.* **6**:1211–1220. DOI:10.1038/s41559-022-01801-3