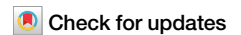


<https://doi.org/10.1038/s42003-025-08029-4>

Chromosome-level genome assembly reveals adaptive evolution of the invasive Amazon sailfin catfish (*Pterygoplichthys pardalis*)



Yunyun Lv^{1,4,5}✉, Yanping Li^{1,4}, Miao Fang², Yi Liu², Yuanyuan Wang², Yexin Yang², Yuanchao Zou¹, Qiong Shi^{1,3,5}✉ & Xidong Mu^{1,2,5}✉

Catfish represents a diverse lineage with variable number of chromosomes and complex relationships with humans. Although certain species pose significant invasive threats to native fish populations, comprehensive genomic investigations into the evolutionary adaptations that contribute to their invasion success are lacking. To address this gap, our study presents a high-quality genome assembly of the Amazon sailfin catfish (*Pterygoplichthys pardalis*), a member of the armored catfish family, along with a comprehensive comparative genomic analysis. By utilizing conserved genomic regions across different catfish species, we reconstructed the 29 ancestral chromosomes of catfish, including two microchromosomes (28 and 29) that show different fusion and breakage patterns across species. Our analysis shows that the Amazon sailfin catfish genome is notably larger (1.58 Gb) with more than 40,000 coding genes. The genome expansion was linked to early repetitive sequence expansions and recent gene duplications. Several expanded genes are associated with immune functions, including antigen recognition domains like the Ig-v-set domain and the tandem expansion of the *CD300* gene family. We also identified specific insertions in CNEs (conserved non-coding elements) near genes involved in cellular processes and neural development. Additionally, rapidly evolving and positively selected genes in the Amazon sailfin catfish genome were found to be associated with collagen formation. Moreover, we identified multiple positively selected codons in *hoxb9*, which may lead to functional alterations. These findings provide insights into molecular adaptations in an invasive catfish that may underlie its widespread invasion success.

Catfish represents a diverse and globally distributed group of ray-finned fish, characterized by significant species diversity and morphological variability¹. Certain species, such as yellow catfish (*Tachysurus fulvidraco*)² and channel catfish (*Ictalurus punctatus*)³, have formed close bonds with humans, becoming vital sources of protein produced through aquaculture. However, some catfish species have gained notoriety as invasive species, posing

significant threats to local ecosystems⁴ and complicating conservation and management efforts⁵.

Invasive catfish can swiftly colonize new environments, disrupting the structural integrity of local ecosystems and thus threatening the viability of native species⁶. Although these invasive species are often characterized by high reproductive rates and resilience⁷, the molecular underpinnings of

¹Key Laboratory of Sichuan Province for Fishes Conservation and Utilization in the Upper Reaches of the Yangtze River, College of Life Science, Neijiang Normal University, Neijiang, 641100, China. ²Key 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, China. ³Laboratory of Aquatic Genomics, College of Life Sciences and Oceanography, Shenzhen University, Shenzhen, 518067, China. ⁴These authors contributed equally: Yunyun Lv, Yanping Li. ⁵These authors jointly supervised this work: Yunyun Lv, Qiong Shi, Xidong Mu.

✉ e-mail: lyyunyun_sci@foxmail.com; shiqiong@szu.edu.cn; muxd@prfi.ac.cn

these traits, including chromosomal structural variations, transposon composition, gene evolution, and variations in conserved non-coding elements (CNEs), remain insufficiently studied.

Catfish species have a wide range of chromosome numbers, ranging from $2n = 48$ to $2n = 60$ ⁸. Recent genomic advances have enabled the sequencing of several catfish chromosomes^{2,3,8}, paving the way for more comprehensive studies on their chromosomal evolution. However, comparative analyses of chromosomal structures across species remain limited⁹, emphasizing the need for a comprehensive understanding of these structural variations, especially given their critical role in speciation processes^{10–12}.

Preservation of ecological integrity requires protecting local ecosystems from invasive species⁶. Certain catfish species, infamous for their disruptive effects, have complicated environmental conservation efforts⁵. For instance, species of catfish in the genus *Pterygoplichthys*, often released by aquarium hobbyists, have established feral populations in warm waters worldwide¹³. Among them, *Pterygoplichthys pardalis*, commonly known as the Amazon sailfin catfish, is native to the Amazon River basin in Brazil and Peru¹⁴. This tropical freshwater fish thrives in a pH range of 7.0–7.5¹⁵ and is typically benthic, though it can breathe air during dry periods or when dissolved oxygen levels are low¹⁶.

Having been introduced to various countries beyond its native habitat, the Amazon sailfin catfish has become a serious concern due to its adverse ecological effects¹⁷. Reports indicate its presence in the Marikina River in the Philippines¹⁸, where it is colloquially known as the “cleaner fish” and has become a local pest. In the United States, its populations disrupt aquatic food chains, reduce native biodiversity, and degrade wetlands and shorelines by burrowing activities¹⁹. Similarly, in China, the introduction of the Amazon sailfin catfish poses significant risks to local species and ecosystems, particularly in Guangdong Province, a major invasion hotspot²⁰.

The management of Amazon sailfin catfish invasions has become a global challenge. As a representative invasive species, Amazon sailfin catfish has become a key target for ecological management across various countries⁴. While some morphological and physiological studies have identified distinctive anatomical features and behaviors—such as armored skin²¹, air-breathing ability¹⁵, pollution resistance²², and low natural mortality²², that facilitate its ecological invasiveness, molecular-level studies have predominantly focused on mitochondrial genome analysis^{21,23}. Although a preliminary draft genome of *P. pardalis* was released in comparative genomics studies with the channel catfish³, its gene annotations are currently unavailable.

The primary aim of this study was to investigate the genome of the Amazon sailfin catfish and compare its chromosomal sequences with those of other catfish species, to gain insights into the chromosomal evolution within the catfish lineage. Comparative genomics is an effective approach for uncovering species-specific adaptations. For instance, comparative genomic analyses have been used to explain the evolutionary adaptations of the hadal snailfish to deep-sea environments²⁴, revealing pseudogenization of the rhythm-related *ANNAT* gene²⁵. Similarly, in the *Sinocyclocheilus* cavefish, comparative genomics has shed light on genetic variations in visual receptor genes that facilitate adaptation to cave environments²⁶. For mudskippers—amphibious fish—comparative genomics has identified genetic variations in visual receptors that underpin their adaptive evolution to both aquatic and terrestrial environments²⁷. In seahorses, the loss of *tbx4* (T-box transcription factor 4) has been identified through comparative genomics, and gene knockout experiments have confirmed the relationship between the presence of this gene and the development of pelvic fins. This highlights how such approaches can uncover the genetic basis of phenotypic traits²⁸.

In this study, we conducted comprehensive comparative genomic analyses, examining transposons, gene families, gene sequences, and CNEs to identify molecular-specific variations within the Amazon sailfin catfish genome. Ultimately, the findings from this research will shed light on the molecular adaptations that confer invasion success in invasive species and contribute to enhanced conservation efforts.

Results and Discussion

High-Quality Genome Assembly and Annotation of the Amazon Sailfin Catfish

We generated approximately 90 Gb of sequencing data, including around 33 Gb of high-accuracy long-read data and 57 Gb of short reads (Supplementary Table 1). This dataset is significantly deeper than the previous Amazon sailfin catfish (*Pterygoplichthys pardalis*) sequencing data³.

After meticulously correcting sequencing errors and assembling the sequences, we obtained a highly contiguous genome assembly of the Amazon sailfin catfish, with a total size of 1.58 Gb (Supplementary Table 2). The scaffold N50 length is 52 Mb, and the contig N50 length is 7.4 Mb (Supplementary Table 2), representing a more than 1000-fold improvement over the previous version (contig N50 ~ 4 Kb)³. Using a K-mer-based method, we assessed the genome completeness at 96.2%, with a QV index of 42.8, where QV represents the transformation of base errors according to the formula $QV = -10 \log(E)$, where E is the error rate at each base position.

We successfully identified 26 distinct clusters representing the Amazon sailfin catfish haploid 26 chromosomes ($2n = 52$) using Hi-C data for chromosomal-level assembly (Fig. 1B, Supplementary Table 3), consistent with previously reported data on the karyotype of this fish²⁹. An assessment of genome assembly completeness using the BUSCO software revealed that the current genome assembly reached approximately 96% completeness (Supplementary Table 4), which is consistent with the genome completeness estimate obtained using the K-mer method. Our analysis of repetitive sequences showed that the Amazon sailfin catfish has a remarkably high repetitive sequence content of 71.08% (Supplementary Table 5), encompassing a total length of 1.13 Gb, which is significantly greater than that found in other catfish species.

A repetitive sequence analysis, using a 5 Mb sliding window across each chromosome, revealed that the distribution of repetitive elements varied markedly at different chromosomal locations. Further examination of short tandem repeats showed significant biases in their distribution, potentially correlating with centromeric and telomeric regions on the chromosomes (Fig. 1C).

Gene annotation identified a total of 40,485 protein-coding genes. BUSCO assessments were used to evaluate the completeness of the annotated protein-coding genes, and the results indicated a high level of completeness, reaching 97.50% (Supplementary Table 6). Also, through synteny analysis of the annotated genes we identified duplicated gene pairs and their distribution across chromosomes. We found that many of these gene pairs had low Ks (synonymous substitutions rate) values of around 0.375 in average (Fig. 1C), suggesting that a portion of the genes in the Amazon sailfin catfish genome have undergone recent segmental duplications.

Comparative analysis of the Amazon sailfin catfish and six other representative catfish species, including hipogean catfish (*Trichomycterus rosablanca*), North African catfish (*Clarias gariepinus*), yellow catfish (*Tachysurus fulvidraco*), striped catfish (*Pangasianodon hypophthalmus*), channel catfish (*Ictalurus punctatus*), and blue catfish (*Ictalurus furcatus*), as well as zebrafish (*Danio rerio*) (Supplementary Table 7), indicated varying degrees of chromosomal conservation among catfish species. Taking zebrafish chromosome numbering as a reference, significant interspecies chromosomal rearrangements were observed in chromosomes 1–7, 11, and 19–22 (Fig. 1D), while other chromosomes remained relatively conserved across different species. These findings indicate that the chromosomal evolution of catfish varies among species, reflecting unique lineage characteristics.

Evolutionary History of Ancestral Chromosomes in Catfish

By leveraging conserved genomic regions across multiple catfish species, we reconstructed 29 ancestral chromosomes for the common ancestor of catfish (Fig. 2A). Notably, we identified two chromosomes (28 and 29) that harbor a relatively small number of genes compared to other chromosomes (Fig. 2A). Ancestral chromosome 28 contains 28 genes (24 with annotated functions and 4 uncharacterized), while ancestral chromosome 29 contains 21 genes (16 with annotated functions and 5 uncharacterized), suggesting that these

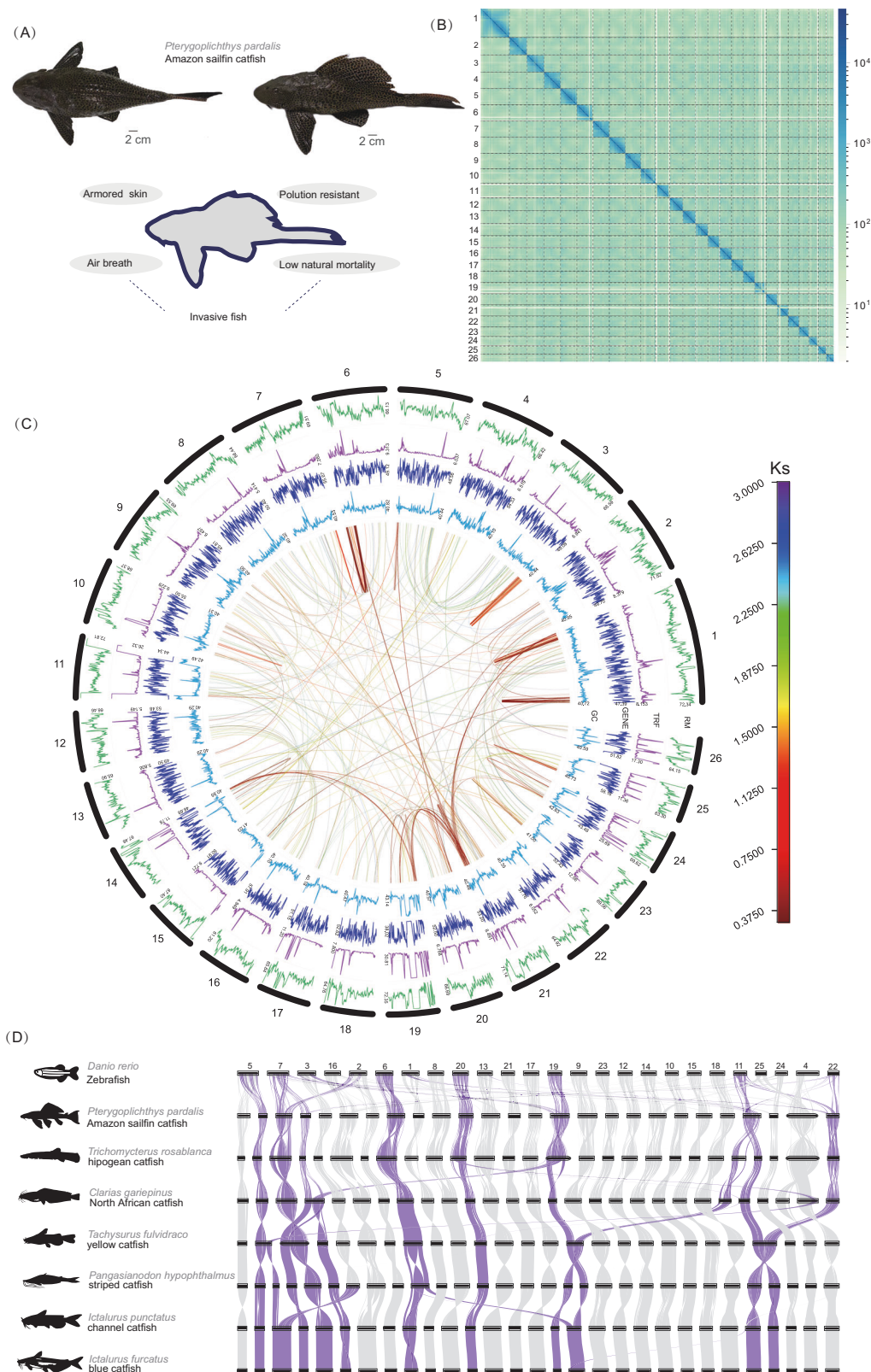


Fig. 1 | Genomic Features of the Amazon Sailfin Catfish. **A** Sample images of the Amazon sailfin catfish, showing a dorsal view at the top left side and a lateral view at the top right side. The morphological and physiological traits reported in the Amazon sailfin catfish as an invasive species are shown at the bottom. **B** Heatmap of the intensity of inter- and intra-chromosomal interactions in the Amazon sailfin catfish. **C** Circos plot illustrating genomic features of the Amazon sailfin catfish. From outer to inner layers: repeat sequence content, tandem repeat sequence

content, gene density, and GC content in 500 kb windows with a 100 kb step size. The innermost layer shows identified segmental duplication blocks, where color intensity represents Ks values (darker indicating higher Ks values). **D** Whole-genome synteny analysis between the Amazon sailfin catfish and other catfish species. Chromosomes shaded in purple indicate those that have undergone chromosomal rearrangements within catfish species.

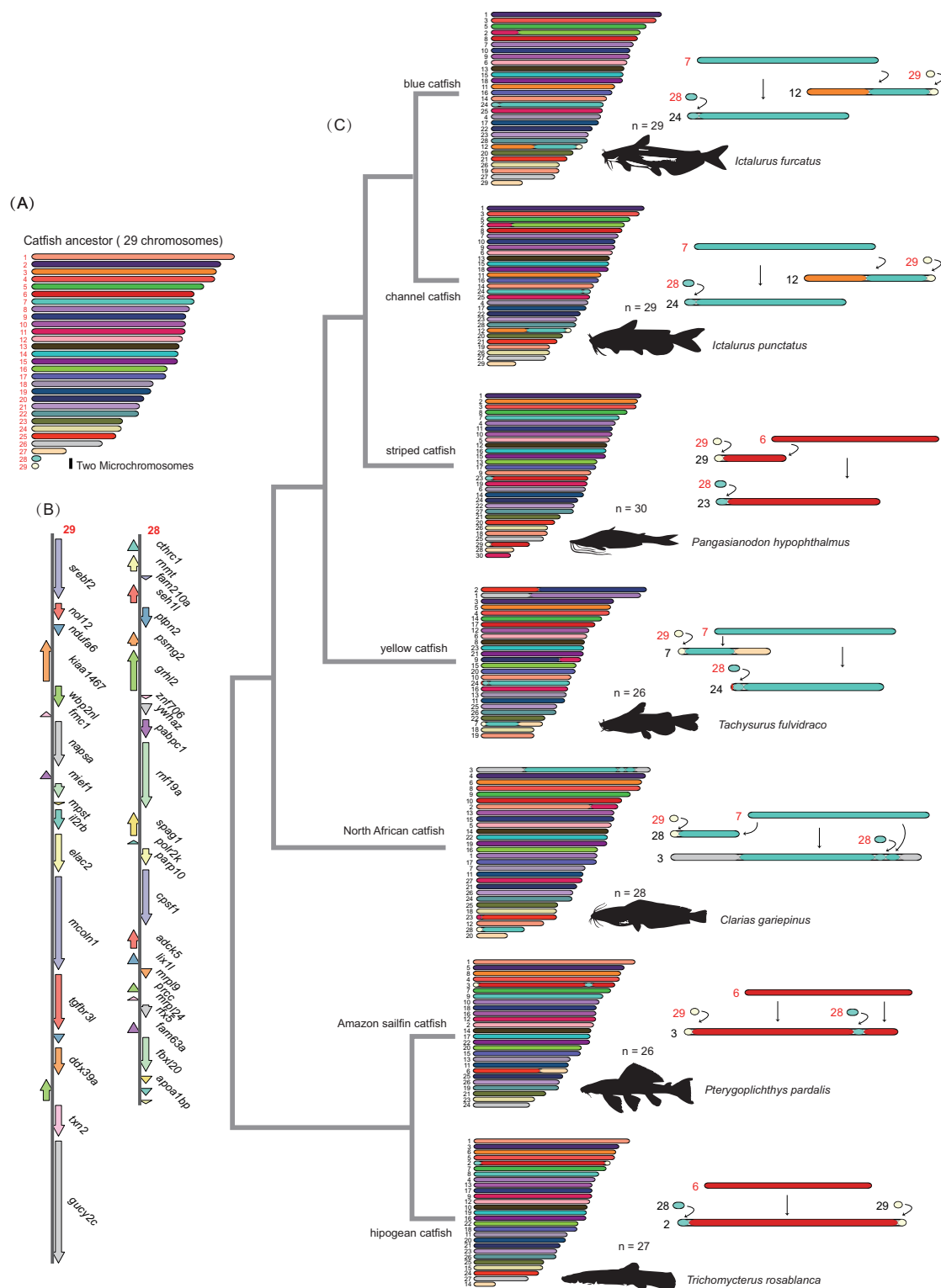


Fig. 2 | Ancestral Catfish Chromosomes and Their Rearrangement Patterns Across Catfish Species. **A** Reconstruction of the 29 ancestral catfish chromosomes, with chromosomes 28 and 29 containing relatively few genes and identified as ancestral microchromosomes. **B** Gene distribution on the two (28 and 29) microchromosomes. **C** Chromosomal rearrangement patterns across seven representative catfish species, highlighting the distinct rearrangements of the two microchromosomes in different catfish lineages. Numbers in red indicate ancestral chromosomes, while numbers in black represent the chromosomes of modern catfish species.

may correspond to two microchromosomes (hereafter called ancestral microchromosome 28 and 29) in the common ancestor of catfish (Fig. 2B).

The 24 genes with annotated functions on ancestral microchromosome 28 include *cthr1*, *mrpl*, *fam210a*, *seh1l*, *ptpn2*, *psmg2*, *grhl2*, *znf706*, *ywhaz*, *pabpc1*, *rnf19a*, *spag1*, *polr2k*, *parp10*, *cpsf1*, *adck5*, *lix1l*, *mrpl9*, *prcc*,

mrpl24, *rxf5*, *fam63a*, *fbxl20*, and *apoa1bp* (Fig. 2B). Meanwhile, the 16 genes with annotated functions on ancestral microchromosome 29 include *sreb2*, *nol12*, *ndufa6*, *kiaa1467*, *wbp2nl*, *fmc1*, *napsa*, *mief1*, *mpst*, *il2rb*, *elac2*, *mcoln1*, *tgfr3l*, *ddx39a*, *txn2*, and *gucy2c* (Fig. 2B). The genes located on ancestral microchromosome 28 are primarily involved in biological

processes such as metabolism, signal transduction, transcriptional regulation, and DNA repair. For instance, *cthrcl* is involved in tissue remodeling, including increasing bone mass and preventing myelination³⁰, *nmmt* participates in RNA processing³¹, *ptpn2* regulates cellular signaling³², and *grhl2* influences epidermal development³³. Other proteins encoding by these genes, like *polr2k*, are involved in transcription³⁴, *parp10* plays a role in DNA repair³⁵, indicating that this microchromosome may be critical for regulating basic cellular metabolism and genomic stability.

On the other hand, the genes with annotated functions on ancestral microchromosome 29 are more closely linked to mitochondrial function, signaling, and immune response processes. For instance, *sreb2* is involved in cholesterol metabolism³⁶, *ndufa6* and *mief1* are part of the mitochondrial electron transport chain^{37,38}, *txn2* plays a role in mitochondrial redox balance³⁹, and *il2rb* is involved in immune signaling and tolerance⁴⁰.

Remarkably, the rearrangement patterns of these two ancestral microchromosomes vary across different catfish species. In the blue catfish and channel catfish, the evolution of chromosomes 12 and 24 involved breaks in ancestral chromosome 7, which subsequently fused with ancestral microchromosomes 28 and 29 (Fig. 2C). In striped catfish, chromosomes 29 and 23 resulted from the integration of microchromosomes 28 and 29 with a broken ancestral chromosome 6 (Fig. 2C). In yellow catfish, chromosomes 7 and 24 also derived from breaks in ancestral chromosome 7, which then combined with microchromosomes 28 and 29 (Fig. 2C). In the North African catfish, ancestral chromosome 7 underwent a break, with one part linking to ancestral microchromosome 29 to form North African catfish chromosome 28, while the other part of ancestral chromosome 7 and ancestral microchromosome 28 combined to form North African catfish chromosome 3 (Fig. 2C). In the Amazon sailfin catfish, the distribution of ancestral microchromosomes 28 and 29 occurred on chromosome 3, where microchromosome 28 inserted into ancestral chromosome 6 (Fig. 2C). Also, in hipogean catfish, chromosome 2 forms analogously to that of the Amazon sailfin catfish, resulting from the fusion of ancestral chromosome 6 with microchromosomes 28 and 29, but with microchromosome 28 positioned centrally within the hipogean catfish (Fig. 2C).

Chromosomal rearrangements can reduce gene flow within populations, thus decreasing the exchange of genetic information^{41,42}. By reconstructing the ancestral chromosomes in catfish (Fig. 2), we illustrate the transformations undergone by these chromosomes during speciation among different catfish species, with a particular focus on the varying evolutionary patterns of two ancestral microchromosomes. Microchromosomes have been observed across different animal species, including turtles, birds, lizards, and some ancestral fishes^{43–46}. These microchromosomes tend to cluster together in the center of the cell, suggesting they may have functional coherence⁴³. Although microchromosomes have completely disappeared in mammals, some chromosomes of basal species like the platypus are aligned with several microchromosomes, indicating that they are components of atypically variable chromosomes in mammals⁴³.

Our findings confirm that in catfish, ancestral microchromosomes 28 and 29 also act as building blocks of atypically variable chromosomes, following a similar evolutionary pattern. We also observed that the structure of these two ancestral microchromosomes undergoes different rearrangements as species diverge, suggesting that the formation of microchromosomes could be closely linked to speciation. Therefore, the identification of these two ancestral microchromosomes in catfish provides important insights into the evolutionary biology of microchromosomes. The functional genes located on these chromosomes serve as candidate loci for further investigation of the origin and function of microchromosomes.

Recent Genome Expansion in the Amazon Sailfin Catfish

The number of annotated genes in the Amazon sailfin catfish is significantly higher than that in other catfish species, leading us to hypothesize that this species has undergone gene family expansion. We confirmed this hypothesis through molecular clock analysis and assessment of gene family dynamics (Fig. 3A), which indicated that the Amazon sailfin catfish had the

highest gene family expansion rate, estimated at 0.1, while other catfish species showed rates ranging from -0.03 to 0.01. Additionally, the Amazon sailfin catfish has the largest genome among catfish species, with a length of 1.58 Gb (Fig. 3B).

Repetitive sequence annotation revealed a substantial amplification of transposable elements in the Amazon sailfin catfish compared to other catfish species (Fig. 3C). We hypothesize that this amplification is due to a recent expansion of transposable sequences in this species. Further analysis of the distribution of the divergence of repetitive sequences across all catfish revealed a specific burst of transposons in the Amazon sailfin catfish (Fig. 3D), which may contribute to the expansion of its genome. The divergence distribution suggests that these transposable elements are relatively recent in origin, reinforcing the notion that this burst may be a key factor in gene family expansion.

We also compared the distribution of synonymous substitution rates (Ks) among conserved syntenic gene pairs in the Amazon sailfin catfish, six other catfish species, and zebrafish (Supplementary Fig. 1). Our findings revealed that the Ks values for the top four species, including Amazon sailfin catfish, hipogean catfish, zebrafish, and yellow catfish, fall within the range of 0 to 1, with the Amazon sailfin catfish having the highest Ks value (Fig. 3E). Notably, the number of gene pairs in the Amazon sailfin catfish is twice that of the second highest, the hipogean catfish, suggesting that recent gene duplication events have led to the higher gene counts in this species compared to other catfish and zebrafish.

We not only compared the evolutionary changes in gene families between the Amazon sailfin catfish and other catfish species, as well as zebrafish, but also estimated the evolutionary rates of protein-coding loci by identifying single-copy genes (Fig. 3F, G). Our estimates revealed that the rate of protein-coding sequence changes in the Amazon sailfin catfish is not higher than that in other catfish species, with the blue catfish having the highest rate. It is likely that protein evolutionary rates increase as catfish species diversify within the catfish lineage.

Transposon bursts driving genome size expansion have been reported in many species and are considered one of the forces behind evolution^{47,48}. In many species, transposon activity is associated with periods of rapid genomic change, often coinciding with speciation events or environmental shifts⁴⁹. These transposon bursts can lead to genomic instability, but they also provide raw material for evolutionary innovation by altering gene expression, creating new regulatory elements, and contributing to the emergence of new genes⁵⁰.

The impact of transposon-driven genome expansion varies across taxa. For example, plant genomes, such as maize and wheat, often experience significant expansion due to transposon activity⁵¹. In some animals, including certain fish and amphibians, transposons are also associated with genome size variation^{52,53}. However, in species with robust genome defense mechanisms, such as some mammals, transposon activity is more controlled, resulting in more stable genome size⁵⁴.

The recent transposon burst identified in the Amazon sailfin catfish in this study likely explains why its genome is significantly larger than that of other catfish species. This recent burst of transposable elements has created opportunities for gene movement and duplication, dramatically reshaping the genome. It provides the Amazon sailfin catfish with greater potential for genetic innovation and may offer opportunities for molecular adaptations that contribute to the invasion success of the Amazon sailfin catfish in diverse environments.

Specific Amplification of Immune System-related Genes in the Amazon Sailfin Catfish

The Amazon sailfin catfish has a higher number of gene family members compared to other catfish species (Fig. 3F). Therefore, we focused on regions with significant tandem gene duplications, some of which are markedly enriched in immune system-related genes (Fig. 4A and Supplementary Table 8). These genes encode the variable region (V) of immunoglobulin (Ig) proteins (Fig. 4B), crucial for recognizing antigens and activating the specific immune response of the body⁵⁵.

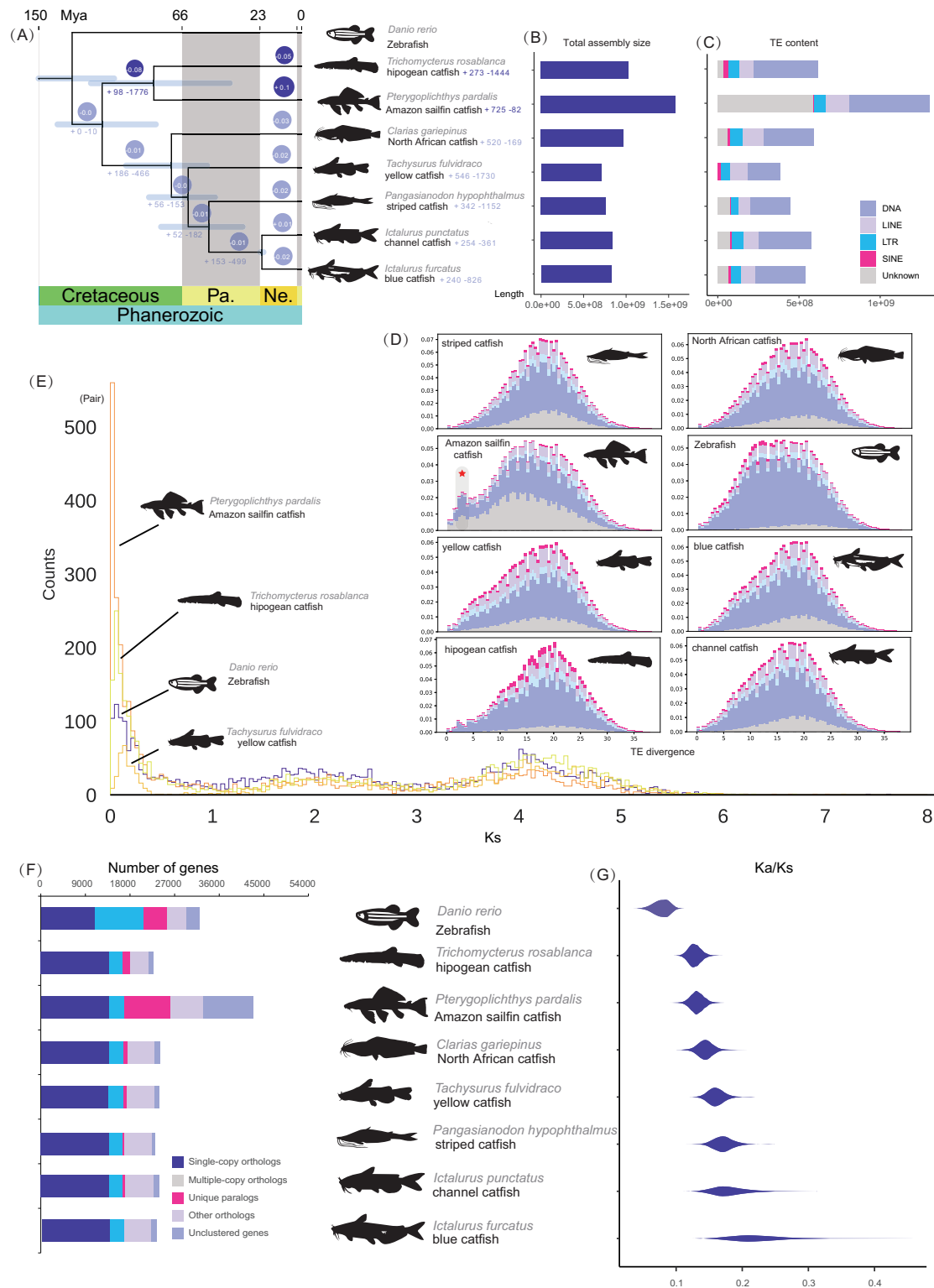


Fig. 3 | Comparative Genomic Evolution of the Amazon Sailfin Catfish. A Gene family expansion and contraction analysis. The circles above each node represent the evolutionary rate of gene families, while the two numbers (left and right) below or to the right of each node denote the number of expanded and contracted gene families, respectively. The horizontal bars within the nodes represent 95% confidence intervals for speciation events. B Bar chart comparing genome sizes across different catfish species, showing that the genome of the Amazon sailfin catfish is significantly larger than that of other catfish species. C Comparison of repeat sequence annotation and accumulation across different catfish genomes, where the Amazon sailfin catfish has a much higher repeat sequence content than other catfish species.

D Transposon divergence comparison across catfish species, revealing a clear recent transposon burst in the Amazon sailfin catfish on the far left side of the divergence spectrum. This burst is absent in other catfish and zebrafish genomes. E A selection of four species with the highest number of collinear genes within the Ks range of 0-1. The Amazon sailfin catfish has the highest number, more than double that of the second-highest species, the hipogean catfish. F Cumulative bar chart of gene family clustering across different catfish species, where the Amazon sailfin catfish has far more gene families than other species. G Violin plot comparing protein evolutionary rates across different catfish species.

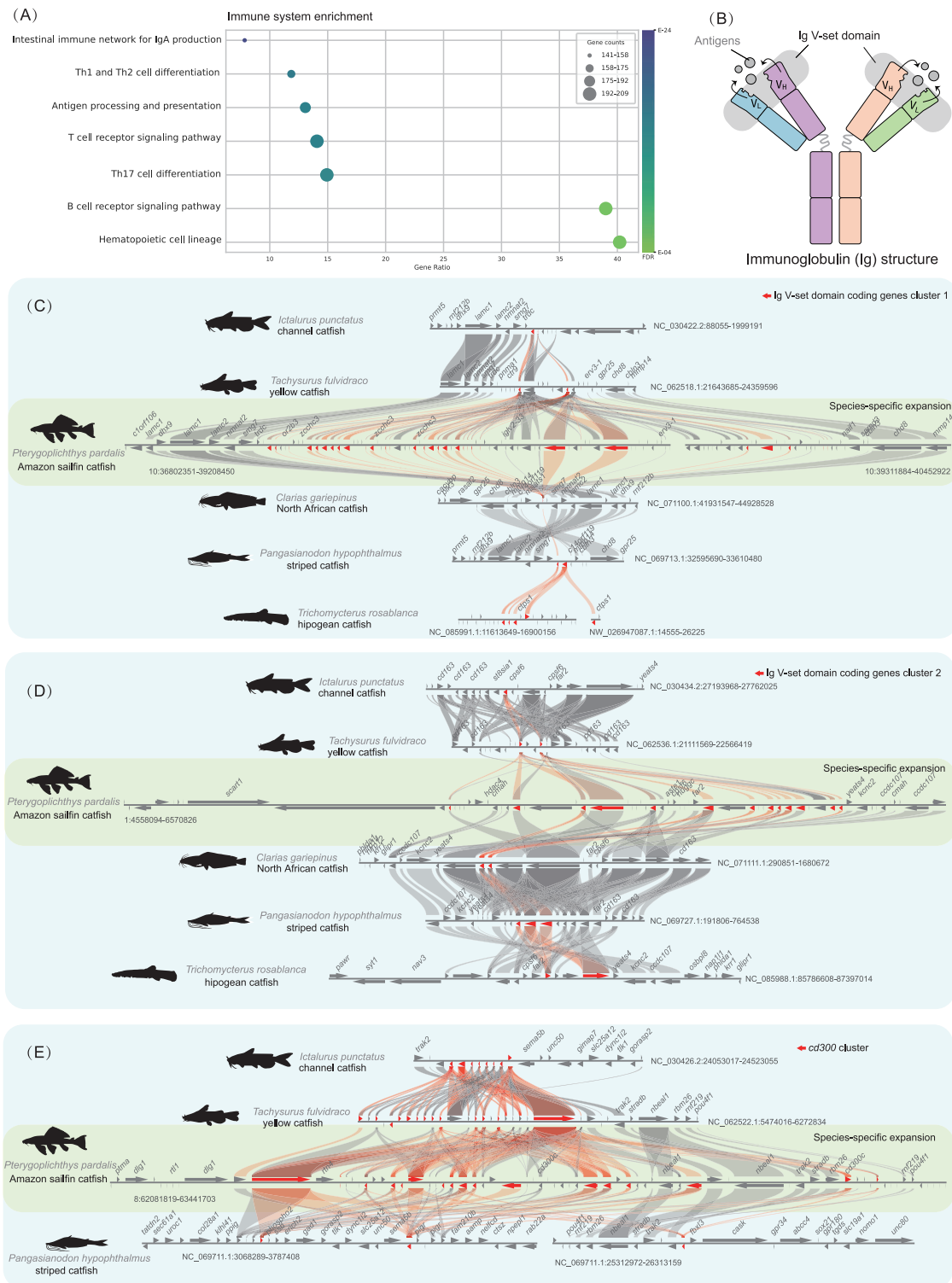


Fig. 4 | Expansion of Immune System-related Genes in the Amazon Sailfin Catfish. **A** Gene expansion in the Amazon sailfin catfish enriched in immune system-related functional categories. **B** Structure of immunoglobulins and the mechanism by which the Ig V-set domain recognizes different antigens. **C** Specific tandem expansion of Ig V-set domains on chromosome 10 of the Amazon sailfin

catfish, with a comparison of nearby genes in other catfish species. **D** Specific tandem expansion of Ig V-set domains on chromosome 1 of the Amazon sailfin catfish, and comparison with nearby genes in other catfish species. **E** Specific tandem expansion of the CD300 gene cluster on chromosome 8 of the Amazon sailfin catfish, with a comparison of neighboring genes in other catfish species.

In the genome of the Amazon sailfin catfish, we identified clusters of genes encoding the immunoglobulin variable (Ig V) region, which have undergone extensive duplication in two conserved genomic regions (Fig. 4C, D). This phenomenon has resulted in a significantly higher number of these genes compared to other species, which likely enhances the ability of

catfish to generate a more diverse immune response and recognize a broader range of pathogens. Additionally, we also observed an increased copy number of the CD300 gene cluster (Fig. 4E). The CD300 family encodes immune receptors that play a crucial role in regulating immune responses, cellular signaling, and pathogen interactions⁵⁶. These receptors are

important for regulating the activity of T cells, B cells, and other immune cells, influencing immune tolerance and the clearance of pathogens⁵⁶.

Gene duplications related to immunity are a common theme across various taxa, contributing to the evolution of more complex immune systems⁵⁷. For instance, during early vertebrate evolution, a greater number of immune genes were identified in jawed vertebrates compared to jawless vertebrates, suggesting that the evolution of immune genes may correlate with the increasing complexity of pathogens in their environments⁵⁸.

Similarly, in zebrafish, expansions in the toll-like receptor (TLR) gene family enhance the ability to detect a broader range of pathogens, a critical adaptation in water-borne environments that are rich in diverse microbial life⁵⁹. In mammals, duplication events in immune-related genes, such as killer cell immunoglobulin-like receptors (KIRs), are well-documented in primates⁶⁰. These duplications provide greater flexibility to the immune system, enabling more effective responses to evolving pathogens.

In the Amazon sailfin catfish, the duplication of the Ig V region and *CD300* genes likely represents a similar adaptive mechanism. These expansions of these genes create more opportunities for genetic innovation, strengthening the immune defenses of the species and enabling it to better adapt to diverse environmental conditions. This may explain why the Amazon sailfin catfish is able to survive in polluted environments²², where the ability to combat a wide range of pathogens is crucial. Such traits may be key to the success of invasive species in adapting to and thriving in new environments.

Specific Changes in Molecular Function

Through the identification of CNEs (conserved non-coding elements) and specific insertions within these intergenic regions, we found that the Amazon sailfin catfish has over 1000 specific insertions. Genes within 2000 bp of these insertions are associated with several functional categories annotated in humans and other organisms. These categories include nervous system development, regulation of cell differentiation, regulation of developmental processes, multicellular organism development, cell differentiation, cocaine addiction, hepatitis B, the FoxO signaling pathway, gap junctions, and the estrogen signaling pathway (Fig. 5A and Supplementary DATA 1). These specific insertions may introduce new transcription factor binding sites (TFBS), leading to changes in gene expression patterns. Such changes could regulate the expression of genes involved in nervous system development, cell differentiation, and the overall development of multicellular organisms, thereby enhancing the adaptability of the species.

The functional categories related to nervous system development suggest that these insertions may influence neurological functions and behaviors of the Amazon sailfin catfish. For example, alterations in neuronal development and synapse formation could enhance the responsiveness of the fish to environmental stimuli. Additionally, the genes involved in cell differentiation and developmental regulation indicate that these insertions might affect the formation and function of different cell types during individual development, allowing the fish to better adapt to varying environmental conditions, such as by optimizing physiological functions in response to habitat changes.

By improving the regulatory capacity of pathways related to FoxO signaling, estrogen signaling, and gap junctions, these insertions may allow species to be more flexible during reproduction, growth, and stress responses, further supporting its survival in diverse ecological niches. Remarkably, KEGG pathway analysis revealed that some of these insertions are associated with functions related to cocaine addiction and hepatitis B in humans and other mammals. This finding potentially highlights both the biomedical research relevance of the Amazon sailfin catfish and its adaptability to specific environmental pressures, such as disease.

The specific enrichment of genes related to the FoxO signaling pathway, estrogen signaling, and gap junctions suggests a role in enhancing growth, reproduction, and stress responses. The FoxO pathway is crucial for regulating oxidative stress, metabolism, and longevity, all of which are critical for the resilience of an organism in challenging environments⁶¹. The impact of these insertions on the FoxO pathway may help the Amazon

sailfin catfish manage oxidative stress more effectively, allowing it to survive in polluted or low-oxygen waters.

Estrogen signaling, often associated with reproductive and developmental processes, is another important pathway. Modifications in non-coding regions that regulate estrogen-responsive genes could enhance reproductive success by adjusting the timing of reproductive cycles or increasing fertility under environmental stress⁶². In addition, gap junctions are vital for intercellular communication and coordination of cellular responses, especially during development and tissue repair⁶³. Enhanced regulation of gap junction proteins may provide the fish with greater plasticity in tissue function and regeneration, further contributing to its ecological adaptability.

Besides exploring potential regulatory changes in functionally related genes, we directly assessed functional enrichment among rapidly evolving genes (REGs) and positively selected genes (PSGs) in the Amazon sailfin catfish (Fig. 5B and Supplementary Table 9). We found that these genes were enriched in categories such as collagen trimer complexes, collagen networks, network-forming collagen trimers, basement membrane collagen trimers, extracellular matrix (ECM)-receptor interactions, protein digestion and absorption, focal adhesion, and small cell lung cancer pathways.

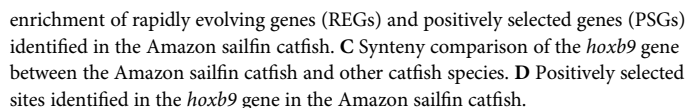
Catfishes typically have exposed, scaleless skin¹. However, the Amazon sailfin catfish has developed armored skin consisting of a series of dermal scutes, forming a dense, highly-mineralized protective tissue⁶⁴. Previous studies have indicated that the armored skin of the Amazon sailfin catfish is rich in collagen⁶⁵. The enrichment of REGs and PSGs identified in the Amazon sailfin catfish within pathways related to collagen production and interaction suggests the formation of a unique collagen structure. Studies on armored skin in other species have similarly pointed to collagen as a major component in the formation of durable and flexible skin tissue⁶⁶. In the Amazon sailfin catfish, the enrichment of genes involved in collagen pathways indicates the formation of a unique collagenous structure that may contribute to its specialized armor.

Furthermore, we identified multiple positively selected sites in the *Hoxb9* gene in the Amazon sailfin catfish (Fig. 5C, D). *Hoxb9* is part of the well-known *Hox* gene family, which plays a crucial role in controlling body plan development in vertebrates⁶⁷. Changes in positively selected sites within *Hox* genes, such as *Hoxb9*, are often associated with the evolution of new structural adaptations⁶⁸. Recent studies have shown that *Hoxb9* can regulate the expression of the *SPPI* gene⁶⁹, whose encoded protein, secreted phosphoprotein 1 (also known as osteopontin), is critical for mineralization⁷⁰. In zebrafish, knocking out *SPPI* leads to a skeletal structure characterized by cartilage-like features⁷¹. In the Amazon sailfin catfish, the modifications in *Hoxb9* may be linked to the development of its distinctive, highly mineralized armored skin through the regulation of *SPPI* or other genes related to specialized morphological traits.

These findings suggest that both collagen modification and changes in developmental genes, as well as *Hoxb9*, are key evolutionary mechanisms driving adaptation of the Amazon sailfin catfish as an invasive fish.

Conclusion

This study provides a comprehensive analysis of the evolutionary history of catfish chromosomes, with a particular focus on the genomic structural changes that may underlie the invasive success of the Amazon sailfin catfish. By reconstructing ancestral chromosomes, we reveal striking differences in chromosomal rearrangement patterns and conserved regions across various catfish species. These differences illustrate the dynamic evolution of genetic information during speciation, suggesting that chromosomal rearrangements contribute to reduced gene flow and promote differentiation and adaptation among species. Specifically, we document substantial genome expansion and the insertion of specific genomic elements that may enhance the adaptability of Amazon sailfin catfish to diverse environmental pressures. Notably, the identified genetic changes strengthen the immune system, improving responses to environmental fluctuations and helping Amazon sailfin catfish thrive in new habitats. Furthermore, the enrichment of genes near CNEs with specific insertions and collagen-related genes



mechanisms underlying the adaptive evolution that contributes to the invasive success of Amazon sailfin catfish in new environments. While further analyses may be needed to directly link these genetic adaptations to invasion success, the findings offer valuable insights into fish chromosome

evolution and the role of genomic structural changes in ecological adaptation.

Methods and materials

Sample Collection

We collected and selected a female sample of the invasive Amazon sailfin catfish from the National Freshwater Genetic Resource Center (Latitude: 23.07336, Longitude: 113.22624), located in Guangzhou, Guangdong Province, China. Anatomical examination confirmed that the specimen was an adult female (Fig. 1A and Supplementary Fig. 2). Muscle tissue from this specimen was then used as the source for HiFi sequencing, next-generation sequencing (NGS), and Hi-C sequencing.

Inclusion and Ethics

All the procedures applied for this study were approved by the Institutional Ethics Committee of the Animal Ethical and Welfare Committee of Neijiang Normal University (permit: NJNU20240923), and all the methods were carried out in accordance with the Code of Practice for the Care and Handling of animal guidelines. This study is reported in compliance with the ARRIVE guidelines. We have complied with all relevant ethical regulations for animal use.

Genome Library Construction and Sequencing

Genomic DNA was isolated from muscle tissue via phenol/chloroform extraction followed by ethanol precipitation. For long-length reads, genomic DNA was randomly fragmented into approximately ~15 kb fragments using a Covaris g-Tubes (Covaris LLC., Woburn, MA, USA). We then constructed SMRTbell HiFi sequencing libraries using the SMRTbell Express Template Prep kit 2.0 (Pacific Biosciences, Menlo Park, CA, USA), as follows: A total of 15 µg of DNA was used for the library construction. After enzymatically removing protruding single-stranded ends, DNA damage was repaired using the SMRTbell Damage Repair Kit - SPv3 (Pacific Biosciences). Next, after performing end repairs and adding poly(A)-tails to the double-stranded ends, the T-overhangs of the SMRTbell adapters were ligated to the ends at 20 °C for 15 h, and then the library was purified using 1× AMPure PB magnetic beads (Pacific Biosciences). Subsequently, the fragment size and concentration of the library samples were determined using the Agilent FEMTO Pulse automated pulsed-field capillary electrophoresis system (Agilent Technologies Inc., Santa Clara, CA, USA) and a Qubit 3.0 Fluorometer (Thermo Fisher Scientific Inc., Waltham, MA, USA). Then, after eliminating any SMRTbell fragments below 15 kb using the BluePippin system (Sage Science Inc., Beverly, MA, USA) for DNA size selection, the library was subjected to a second purification step using 1× AMPure PB magnetic beads. Afterwards, the size and quality of the library were determined using the Agilent FEMTO Pulse system and the Qubit dsDNA HS assay kit (Thermo Fisher Scientific Inc.). Then, sequencing primers and Sequel II DNA polymerase were annealed and bound, respectively, to the completed SMRTbell library. Finally, the prepared SMRTbell library was sequenced on the PacBio Sequel II platform (Pacific Biosciences) at a concentration of 120 pM, with a running time of 30 h. Afterwards, high-accuracy sequencing of the Amazon sailfin catfish genome was performed using the PacBio Revio sequencing platform (Pacific Biosciences).

For short-length reads, we constructed short DNA fragment libraries using 1 µg of DNA and the MGIEasy DNA Universal Library Prep Set (MGI Tech Co. Ltd., Shenzhen, China), strictly following the manufacturer's protocol. To ensure accurate sample identification during sequencing, a unique barcode was added to the sample. After library construction, we determined the concentration and fragment size distribution of the samples using a Qubit 3.0 Fluorometer (Thermo Fisher Scientific Inc.) and a Bioanalyzer 2100 system (Agilent Technologies Inc.). These measurements informed our pooling strategy for sequencing. Once all libraries passed quality control checks, sequencing was performed on the MGI-SEQ 2000 platform (MGI Tech Co. Ltd.) at the Nanjing Sequencing Center, Novogene BioTechnology Co., Ltd. (Nanjing, China).

Genome Assembly

The initial genome assembly with HiFi reads from PacBio Revio sequencing platform was performed using the Hifiasm (v0.19.9-r616) assembly tool⁷². Afterwards, the haplotigs present in the completed assembly were removed using the purge_haplotigs v1.1.3⁷³ tool. Subsequently, we polished the genome using the Pilon (v1.24) software⁷⁴ for three rounds, using next-generation sequencing reads to correct specific errors at individual loci in the Amazon sailfin catfish genome, thereby enhancing its accuracy. Next, we aligned Hi-C reads to the assembled genome using Chromap (v0.2.7) software⁷⁵, which enabled us to obtain interaction information between chromatin regions across contigs. We then performed chromosomal-level assembly of the Amazon sailfin catfish genome using the YaHS (v1.1) tool⁷⁶. Finally, we manually corrected the errors in the chromosome-level assembly in the Juicebox (v2.20.00) tool⁷⁷ and performed a back-alignment of the Hi-C reads to the chromosomal-level assembly to assess the clarity of domain structures across different chromosomes using the HiCExplorer (v3.7.2) tool⁷⁸.

The final chromosome-level genome assembly was evaluated using BUSCO (v3.1.0)⁷⁹, based on the actinopterygii_odb10 gene set. Additionally, genome quality was assessed through K-mer analysis with a K-length of 21 bp using of Merquery (v1.4.1)⁸⁰.

Annotation of Repetitive Sequences and Protein-coding Genes

We constructed a de novo repeat library using the RepeatModeler (v2.0.1) software⁸¹ and used to start the genome annotation process. Afterwards, we referenced the Repbase database⁸² to identify lineage-specific repeats in catfish. These libraries were merged and subsequently searched against the assembled genomes using the RepeatMasker (v4.1.1) software⁸³ with parameters set to “-nolow -gff -poly -a -inv -e rmbast.” This approach allowed us to compile a comprehensive list of repetitive elements. In addition to annotating the repetitive sequences of the Amazon sailfin catfish genome, we used the same methodology to annotate the published genomes of other species. Ultimately, we compared the content and divergence levels of repetitive sequences across different species to determine the specific changes in transposable element sequences within the Amazon sailfin catfish.

To predict protein-coding gene regions within the assembled genomes, we used three different approaches: (1) de novo prediction using the Augustus (v3.3.3) software⁸⁴ and Snap (v2006-07-28) software⁸⁵; (2) homolog-based prediction with the Tblastn (v2.11.0) tool⁸⁶ and Exonerate (v2.2.0) software⁸⁷ using protein sequences from seven representative species; (3) transcript-based annotation. Initially, a de novo transcriptome assembly was generated from RNA-seq data of mixed tissues (brain, kidney, liver, heart, spleen and muscle, see Supplementary Table 2) using Trinity (v2.11.0) software⁸⁸. The raw transcripts were then aligned to the assembled genome using the BLAST-Like Alignment Tool (BLAT, v36) software⁸⁹ and Gmap (v2017-11-15) software⁹⁰, followed by validation and integration using the Program to Assemble Spliced Alignments (PASA, v2.3.3) pipeline⁹¹. Finally, we used the EvidenceModeler (v1.1.1) software⁹² to combine results from all three prediction strategies, carefully filtering out erroneous predictions that included stop codons within coding regions. Functional annotation and characterization of protein biological functions were performed using the EggNOG-mapper (v2.67) tool⁹³. The completeness of the protein-coding genes was also assessed using the actinopterygii_odb10 gene set with BUSCO software⁷⁹.

Gene Family Analysis

To investigate the dynamic changes in gene family evolution within the Amazon sailfin catfish, we obtained genome and annotation files from other representative catfish species and the model organism zebrafish and performed comparative gene analysis. We used the BLASTP (v2.11.0) software to perform an all-vs.-all protein sequence comparison across all species. In order to establish different gene families, we used the OrthoMCLight (v1.4) software⁹⁴ to cluster the sequences based on similarity. We also analyzed variations in gene copy numbers to quantify both single-copy and multi-copy gene families.

All identified single-copy genes were extracted and aligned using the MAFFT (v7.4.75) program⁹⁵, to generate a super-length alignment matrix for further analysis. The topology of the species tree was initially constructed using Bayesian inference (BI) with the MarBayes (v3.2.2) program⁹⁶, and was further validated by Maximum Likelihood (ML) analysis using the Randomized Axelerated Maximum Likelihood (RAXML, v8.0.17) program⁹⁷. Both methods used the GTR + I + G model for accuracy. The Bayesian analysis included two parallel runs of 2 million generations, using three chains per run, with samples taken every 100 generations and discarding the initial 25% of samples. For the ML analysis, we performed fast bootstrapping with 1000 replicates to compute branch support values.

We calibrated the species tree topology using the MCMCtree (v4.9) program⁹⁸, incorporating two time-calibrated points: node 1 (the common ancestor of blue catfish and channel catfish) estimated at 22.1 to 22.9 million years ago (Mya); node 2 (the common ancestor of channel catfish and yellow catfish) estimated at 19 to 44.2 Mya; node 3 (corresponding to the red-tailed pipe snake and sunbeam snake) estimated at 50.9 to 97 Mya, with data sourced from the TimeTree database⁹⁹. A log-normal independent molecular clock model was used to generate 1,000,000 samples, discarding the initial 25% of samples. Visualization of credible intervals for species divergence times at each node was achieved by generating the species tree using the MCMCtree R package¹⁰⁰.

To refine our gene family clusters, we removed potential false results where some species had over 200 paralogs while others had none. The remaining gene clusters were analyzed for member changes using the CAFÉ (v2.0) software¹⁰¹, with the significance threshold set at $p < 0.05$. We particularly focused on gene families that showed significant expansions, specifically those associated with tandem repeats. To identify these tandem repeat genes in the Amazon sailfin catfish genome, we used the MCScanX software with default parameters¹⁰². Finally, we overlapped the significantly expanded gene families with the identified tandem repeat genes and performed functional enrichment analysis on the obtained gene set.

Ks Analysis

We initially performed a synteny analysis of each fish genome using the MCScanX software¹⁰² to identify pairs of duplicated genes within each genome. We then performed protein sequence alignments of the syntenic gene pairs using MAFFT program⁹⁵. The amino acid alignments of the proteins were converted into codon-based nucleotide alignments using the PAL2NAL program¹⁰³. Subsequently, the Ks values (synonymous substitutions per site) for each syntenic gene pair were calculated using the KaKs_Calculator (v2.0) software¹⁰⁴, specifically using the Yang-Nielsen (YN) method¹⁰⁵. We constructed a histogram of gene pairs across Ks values ranging from 0 to 10, with 0.1 units, and compared the differences in gene pair distributions among the various species.

Ancestral Chromosome Restructuring

We initially identified candidate orthologous genes across all genomes using reciprocal best hits (RBH) with MMseq2 v15-6f452¹⁰⁶, which served as potential markers for reconstructing ancestral chromosomes. Subsequently, we employed MCScanX¹⁰² for synteny analysis, retaining genes that exhibited syntenic relationships in both catfish and zebrafish (with a minimum of five syntenic genes per module) as effective unique markers for reconstructing the ancestral chromosomes of catfish. We then marked the ancestral node for catfish within the phylogenetic tree of catfish and zebrafish, organizing the effective unique markers based on their chromosomal locations across species. Finally, we reconstructed the ancestral chromosomes of catfish using AnGes v1.01^{107,108}, applying the parameters `acs_sa=1`, `acs_ra=1`, `acs_sci=1`, `acs_mci=1`, and `c1p_linear=1`.

Analysis of Conserved Non-Coding Elements (CNEs)

Using the soft-masked genomes of all species subjected to repeat masking, we performed the whole-genome alignment with the Cactus (v2.8.2) program^{109,110}. The `hal2maf` command in the Cactus toolkit was used to generate a raw genome alignment of eight fish species in the multiple

alignment format (MAF). From this whole-genome alignment, we extracted four-fold degenerate sites and estimated a neutral phylogenetic model (non-conserved model) using the phyloFit program¹¹¹.

To estimate the expected substitution rate of conserved elements relative to neutrality (ρ), we used the phastCons package¹¹¹ with the options ‘-target-coverage 0.25, -expected-length 20, -estimate_rho.’ The input alignment was processed in non-overlapping 10⁷ bp windows. The conserved models for each window were combined using the phyloBoot program¹¹¹, which allowed us to predict the initial conserved elements. We generated CNEs by excluding exon regions from the highly conserved elements using the ‘subtract’ command in BedTools (v2.31.1) toolkit¹¹².

Our primary focus was on CNEs located in intergenic regions, within introns, and within the 20 kb upstream or downstream flanking regions of the genes. We identified specific insertion sequences longer than 15 bp present in the CNEs of the Amazon sailfin catfish, as these could become potential new transcription factor binding sites that may alter proteins regulatory functions. We performed functional enrichment analysis using the R package clusterProfiler v4.0¹¹³ on genes with specific insertions located upstream and downstream, to evaluate the functional impact of these CNEs on the Amazon sailfin catfish genome. We ultimately selected the top five enriched gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) terms to present the associated genes and their functional descriptions.

Identification of Rapidly Evolving Genes (REGs) and Positively Selected Genes (PSGs)

Further analysis involved extracting a subset of orthologous genes between the Amazon sailfin catfish and other fishes, including various catfish species and zebrafish, to identify PSGs and rapidly evolving genes REGs in the Amazon sailfin catfish genome. To detect REGs, we used the CODEML branch model in the PAML (v4.9) package⁹⁸. The null model assumed that the non-synonymous to synonymous substitution ratio (Ka/Ks) was uniform across branches, whereas the alternative model allowed for branch-specific variation in Ka/Ks within the ancestral branches of Amazon sailfin catfish. We calculated *p-values* using the likelihood ratio test (LRT) to compare the two models, determining significant REGs based on a corrected *p-value* < 0.05 using the Bonferroni method.

In addition, we used the CODEML branch-site model in the PAML (v4.9) package to identify potential PSGs. Under the null hypothesis, the Ka/Ks value for each site on each branch was constrained to 1, while the alternative hypothesis allowed variable dN/dS values at specific sites within Amazon sailfin catfish. The *p-values* from the LRT were obtained following the CODEML analysis. PSGs were defined based on a nominal 5% traditional *p-value* (< 0.05) and the presence of at least one positively selected site with a posterior probability > 0.90 , as determined using an empirical Bayesian approach. We selected the top five enriched GO and KEGG terms to present the associated genes and their functional descriptions for PSGs and REGs.

Data availability

Raw reads, genome assembly and annotation were deposited into National Genomics Data Center with project number PRJCA030928 for *Pterygoplichthys pardalis*. The genome assembly, mRNA sequences, protein sequences, and gene annotations are also available at Figshare¹¹⁴ at <https://doi.org/10.6084/m9.figshare.27080665.v2>.

Received: 26 October 2024; Accepted: 1 April 2025;
Published online: 16 April 2025

References

1. Chandra Segaran, T. et al. Catfishes: a global review of the literature. *Heliyon* **9**, e20081 (2023).
2. Gong, G. et al. Chromosomal-level assembly of yellow catfish genome using third-generation DNA sequencing and Hi-C analysis. *Gigascience* **7**, <https://doi.org/10.1093/gigascience/giy120> (2018).

3. Liu, Z. et al. The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts. *Nat. Commun.* **7**, 11757 (2016).
4. Orfinger, A. B. & Douglas Goodding, D. The Global Invasion of the Suckermouth Armored Catfish Genus *Pterygoplichthys* (Siluriformes: Loricariidae): Annotated List of Species, Distributional Summary, and Assessment of Impacts. *Zool. Stud.* **57**, e7 (2018).
5. Ranjan, R. Protecting endemic species from African Catfish invasion when community behavioral responses get in the way. *PLoS One* **13**, e0209009 (2018).
6. Britton, J. R. Contemporary perspectives on the ecological impacts of invasive freshwater fishes. *J. Fish. Biol.* **103**, 752–764 (2023).
7. Havel, J. E., Kovalenko, K. E., Thomaz, S. M., Amalfitano, S. & Kats, L. B. Aquatic invasive species: challenges for the future. *Hydrobiologia* **750**, 147–170 (2015).
8. Ye, H. et al. Chromosome-level genome assembly of the largefin longbarbel catfish (*Hemibagrus macropterus*). *Front Genet.* **14**, 1297119 (2023).
9. Zhu, C. et al. Insights into chromosomal evolution and sex determination of *Pseudobagrus ussuriensis* (Bagridae, Siluriformes) based on a chromosome-level genome. *DNA Res* **29**, <https://doi.org/10.1093/dnares/dsac028> (2022).
10. Berdan, E. L. et al. Structural variants and speciation: multiple processes at play. *Cold Spring Harb. Perspect. Biol.* **16**, <https://doi.org/10.1101/cshperspect.a041446> (2024).
11. Zhang, L., Reifova, R., Halenkova, Z. & Gompert, Z. How important are structural variants for speciation? *Genes (Basel)* **12**, <https://doi.org/10.3390/genes12071084> (2021).
12. Fuller, Z. L., Koury, S. A., Phadnis, N. & Schaeffer, S. W. How chromosomal rearrangements shape adaptation and speciation: Case studies in *Drosophila pseudoobscura* and its sibling species *Drosophila persimilis*. *Mol. Ecol.* **28**, 1283–1301 (2019).
13. Paul, P., Gupta, N. K. & Mondal, A. A. Gautam %J Aquatic Ecology. The Sailfin Catfish (*Pterygoplichthys* spp.) as a non-native aquarium pet. *Assess. potential Distrib. public Percept.* **58**, 89–105 (2024).
14. Hossain, M. Y., Vadas, R. L., Ruiz-Carus, R. & Galib, S. M. Amazon Sailfin Catfish *Pterygoplichthys pardalis* (Loricariidae) in Bangladesh. *A Crit. Rev. Its Invasive Threat Nativ. Endem. Aquat. Species* **3**, 14 (2018).
15. Shiels, H. A., Santiago, D. A. & Galli, G. L. Hypercapnic acidosis reduces contractile function in the ventricle of the armored catfish, *Pterygoplichthys pardalis*. *Physiol. Biochem Zool.* **83**, 366–375 (2010).
16. Lardizabal, C. C., Benitez, E. M. & Matamoros, W. A. Record of the Non-native Suckermouth armored catfish hybrid *Pterygoplichthys pardalis* (Castelnau, 1985) x *Pterygoplichthys disjunctivus* (Weber, 1991) (Siluriformes: Loricariidae) in Honduras. *Zootaxa* **4778**, zootaxa 4778 4773 4710, <https://doi.org/10.11646/zootaxa.4778.3.10> (2020).
17. Raj, S., Kumar, A. B., Raghavan, R. & Dahanukar, N. Amazonian invaders in an Asian biodiversity hotspot: Understanding demographics for the management of the armoured sailfin catfish, *Pterygoplichthys pardalis* in Kerala, India. *J. Fish. Biol.* **96**, 549–553 (2020).
18. Chavez, J. M. et al. New Philippine record of south american sailfin catfishes (Pisces: Loricariidae). **1109** (2006).
19. Nico, L. G. & Martin, R. T. J. T. S. N. The South American suckermouth armored catfish. *Pterygoplichthys anisitsi* (Pisces: Loricariidae), *Tex., comments foreign fish. introductions Am. Southwest* **46**, 98–104 (2001).
20. Wei, H. et al. The distribution, establishment and life-history traits of non-native sailfin catfishes *Pterygoplichthys* spp. *Guangdong Prov. China* **12**, 241–249 (2017).
21. Wu, L.-W., Liu, C.-C. & Lin, S.-M. J. Z. s. Identification of exotic sailfin catfish species (*Pterygoplichthys*, Loricariidae) in Taiwan based on morphology and mtDNA sequences. **50**, 235–246 (2011).
22. Elfidasari, D. et al. Intestinal bacteria diversity of suckermouth catfish (*Pterygoplichthys pardalis*) in the Cd, Hg, and Pb contaminated Ciliwung River, Indonesia. *Heliyon* **9**, e14842 (2023).
23. Xia, W. et al. Next-generation sequencing yields the complete mitochondrial genome of *Pterygoplichthys pardalis* (Loricariidae; Siluriformes). *Mitochondrial DNA B Resour.* **6**, 3209–3211 (2021).
24. Xu, W. et al. Chromosome-level genome assembly of hadal snailfish reveals mechanisms of deep-sea adaptation in vertebrates. *Elife* **12**, <https://doi.org/10.7554/eLife.87198> (2023).
25. Lv, Y. et al. A comparative genomics study on the molecular evolution of serotonin/melatonin biosynthesizing enzymes in vertebrates. *Front Mol. Biosci.* **7**, 11 (2020).
26. Yang, J. et al. The *Sinocyclocheilus* cavefish genome provides insights into cave adaptation. *BMC Biol.* **14**, 1 (2016).
27. You, X. et al. Mudskipper genomes provide insights into the terrestrial adaptation of amphibious fishes. *Nat. Commun.* **5**, 5594 (2014).
28. Lin, Q. et al. The seahorse genome and the evolution of its specialized morphology. *Nature* **540**, 395–399 (2016).
29. Dos Santos Guimaraes, A., Maciel, L. A. M., de Souza, M. F. B. & Rodrigues, L. R. R. Karyotypic and Molecular Analysis of *Pterygoplichthys pardalis* (Castelnau 1855) from the Lower Amazon River. *Animals (Basel)* **13**, <https://doi.org/10.3390/ani13091533> (2023).
30. Mei, D., Zhu, Y., Zhang, L. & Wei, W. The Role of CTHRC1 in regulation of multiple signaling and tumor progression and metastasis. *Mediators Inflamm.* **2020**, 9578701 (2020).
31. Liang, S., Almohammed, R. & Cowling, V. H. The RNA cap methyltransferases RNMT and CMTR1 co-ordinate gene expression during neural differentiation. *Biochem Soc. Trans.* **51**, 1131–1141 (2023).
32. Veenstra, C. et al. The effects of PTPN2 loss on cell signalling and clinical outcome in relation to breast cancer subtype. *J. Cancer Res Clin. Oncol.* **145**, 1845–1856 (2019).
33. Gasperoni, J. G., Fuller, J. N., Darido, C., Wilanowski, T. & Dworkin, S. Grainyhead-like (Grhl) target genes in development and cancer. *Int. J. Mol. Sci.* **23**, <https://doi.org/10.3390/ijms23052735> (2022).
34. Howe, K. J. RNA polymerase II conducts a symphony of pre-mRNA processing activities. *Biochim Biophys. Acta* **1577**, 308–324 (2002).
35. Khatib, J. B., Dhoonmoon, A., Moldovan, G. L. & Nicolae, C. M. PARP10 promotes the repair of nascent strand DNA gaps through RAD18 mediated translesion synthesis. *Nat. Commun.* **15**, 6197 (2024).
36. Madison, B. B. Srebp2: A master regulator of sterol and fatty acid synthesis. *J. Lipid Res.* **57**, 333–335 (2016).
37. Rathore, A. et al. MIEF1 microprotein regulates mitochondrial translation. *Biochemistry* **57**, 5564–5575 (2018).
38. Alston, C. L. et al. Bi-allelic mutations in NDUFA6 establish its role in early-onset isolated mitochondrial complex I deficiency. *Am. J. Hum. Genet.* **103**, 592–601 (2018).
39. Holzerova, E. et al. Human thioredoxin 2 deficiency impairs mitochondrial redox homeostasis and causes early-onset neurodegeneration. *Brain* **139**, 346–354 (2016).
40. Malek, T. R. & Castro, I. Interleukin-2 receptor signaling: at the interface between tolerance and immunity. *Immunity* **33**, 153–165 (2010).
41. Rogers, R. L. Chromosomal rearrangements as barriers to genetic homogenization between archaic and modern humans. *Mol. Biol. Evol.* **32**, 3064–3078 (2015).
42. Mackintosh, A. et al. Chromosome fissions and fusions act as barriers to gene flow between brenthis fritillary butterflies. *Mol. Biol. Evol.* **40**, <https://doi.org/10.1093/molbev/msad043> (2023).

43. Waters, P. D. et al. Microchromosomes are building blocks of bird, reptile, and mammal chromosomes. *Proc. Natl. Acad. Sci. USA* **118**, <https://doi.org/10.1073/pnas.2112494118> (2021).
44. Huang, Z. et al. Evolutionary analysis of a complete chicken genome. *Proc. Natl. Acad. Sci. USA* **120**, e2216641120 (2023).
45. Thompson, A. W. et al. The bowfin genome illuminates the developmental evolution of ray-finned fishes. *Nat. Genet.* **53**, 1373–1384 (2021).
46. Braasch, I. et al. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat. Genet.* **48**, 427–437 (2016).
47. Belyayev, A. Bursts of transposable elements as an evolutionary driving force. *J. Evol. Biol.* **27**, 2573–2584 (2014).
48. Hara, Y. et al. Shark genomes provide insights into elasmobranch evolution and the origin of vertebrates. *Nat. Ecol. Evol.* **2**, 1761–1771 (2018).
49. Serrato-Capuchina, A. & Matute, D. R. The role of transposable elements in speciation. *Genes (Basel)* **9**, <https://doi.org/10.3390/genes9050254> (2018).
50. Pourrajab, F. & Hekmatimoghaddam, S. Transposable elements, contributors in the evolution of organisms (from an arms race to a source of raw materials). *Heliyon* **7**, e06029 (2021).
51. Hassan, A. H., Mokhtar, M. M. & El Allali, A. Transposable elements: multifunctional players in the plant genome. *Front Plant Sci.* **14**, 1330127 (2023).
52. Canapa, A., Barucca, M., Biscotti, M. A., Forconi, M. & Olmo, E. Transposons, genome size, and evolutionary insights in animals. *Cytogenet. Genome Res.* **147**, 217–239 (2015).
53. Sotero-Caio, C. G., Platt, R. N. 2nd, Suh, A. & Ray, D. A. Evolution and diversity of transposable elements in vertebrate genomes. *Genome Biol. Evol.* **9**, 161–177 (2017).
54. Platt, R. N. 2nd, Vandeweghe, M. W. & Ray, D. A. Mammalian transposable elements and their impacts on genome evolution. *Chromosome Res.* **26**, 25–43 (2018).
55. Pennell, M., Rodriguez, O. L., Watson, C. T. & Greiff, V. The evolutionary and functional significance of germline immunoglobulin gene variation. *Trends Immunol.* **44**, 7–21 (2023).
56. Borrego, F. The CD300 molecules: an emerging family of regulators of the immune system. *Blood* **121**, 1951–1960 (2013).
57. Muller, V., de Boer, R. J., Bonhoeffer, S. & Szathmari, E. An evolutionary perspective on the systems of adaptive immunity. *Biol. Rev. Camb. Philos. Soc.* **93**, 505–528 (2018).
58. Vinkler, M. et al. Understanding the evolution of immune genes in jawed vertebrates. *J. Evol. Biol.* **36**, 847–873 (2023).
59. Meijer, A. H. et al. Expression analysis of the Toll-like receptor and TIR domain adaptor families of zebrafish. *Mol. Immunol.* **40**, 773–783 (2004).
60. Parham, P., Norman, P. J., Abi-Rached, L. & Guethlein, L. A. Human-specific evolution of killer cell immunoglobulin-like receptor recognition of major histocompatibility complex class I molecules. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 800–811 (2012).
61. Martins, R., Lithgow, G. J. & Link, W. Long live FOXO: unraveling the role of FOXO proteins in aging and longevity. *Aging Cell* **15**, 196–207 (2016).
62. Li, M., Sun, L. & Wang, D. Roles of estrogens in fish sexual plasticity and sex differentiation. *Gen. Comp. Endocrinol.* **277**, 9–16 (2019).
63. Rossello, R. A. & Kohn, D. H. Gap junction intercellular communication: a review of a potential platform to modulate craniofacial tissue engineering. *J. Biomed. Mater. Res B Appl Biomater.* **88**, 509–518 (2009).
64. Lowe, A., Summers, A. P., Walter, R. P., Walker, S. & Misty Paig-Tran, E. W. Scale performance and composition in a small Amazonian armored catfish, *Corydoras trilineatus*. *Acta Biomater.* **121**, 359–370 (2021).
65. Nurubhasha, R., Sampath Kumar, N. S., Thirumalasetti, S. K., Simhachalam, G. & Dirisala, V. R. Extraction and characterization of collagen from the skin of *Pterygoplichthys pardalis* and its potential application in food industries. *Food Sci. Biotechnol.* **28**, 1811–1817 (2019).
66. Sire, J. Y. & Huysseune, A. Formation of dermal skeletal and dental tissues in fish: a comparative and evolutionary approach. *Biol. Rev. Camb. Philos. Soc.* **78**, 219–249 (2003).
67. Ozernyuk, N. & Schepetov, D. HOX-gene cluster organization and genome duplications in fishes and mammals: transcript variant distribution along the anterior-posterior axis. *Int. J. Mol. Sci.* **23**, <https://doi.org/10.3390/ijms23179990> (2022).
68. Pearson, J. C., Lemons, D. & McGinnis, W. Modulating Hox gene functions during animal body patterning. *Nat. Rev. Genet.* **6**, 893–904 (2005).
69. Han, J. et al. HOXB9 promotes osteosarcoma cell survival and malignancy under glucose starvation via upregulating SPP1 expression. *Biochem. Pharm.* **224**, 116208 (2024).
70. Kawasaki, K., Suzuki, T. & Weiss, K. M. Genetic basis for the evolution of vertebrate mineralized tissue. *Proc. Natl. Acad. Sci. USA* **101**, 11356–11361 (2004).
71. Venkatesh, B. et al. Elephant shark genome provides unique insights into gnathostome evolution. *Nature* **505**, 174–179 (2014).
72. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175 (2021).
73. Roach, M. J., Schmidt, S. A. & Borneman, A. R. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinforma.* **19**, 460 (2018).
74. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963 (2014).
75. Zhang, H. et al. Fast alignment and preprocessing of chromatin profiles with Chromap. *Nat. Commun.* **12**, 6566 (2021).
76. Zhou, C., McCarthy, S. A. & Durbin, R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics* **39**, <https://doi.org/10.1093/bioinformatics/btac808> (2023).
77. Durand, N. C. et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).
78. Wolff, J. et al. Galaxy HiCEXplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. *Nucleic Acids Res.* **46**, W11–W16 (2018).
79. Manni, M., Berkeley, M. R., Seppey, M., Simao, F. A. & Zdobnov, E. M. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**, 4647–4654 (2021).
80. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merquy: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 245 (2020).
81. Flynn, J. M. et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. USA* **117**, 9451–9457 (2020).
82. Jurka, J. et al. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467 (2005).
83. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* **Chapter 4**, 4.10.11–4.10.14, <https://doi.org/10.1002/0471250953.bi0410s25> (2009).
84. Stanke, M. et al. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439 (2006).
85. Korf, I. Gene finding in novel genomes. *BMC Bioinforma.* **5**, 59 (2004).
86. Gertz, E. M., Yu, Y. K., Agarwala, R., Schaffer, A. A. & Altschul, S. F. Composition-based statistics and translated nucleotide

- searches: improving the TBLASTN module of BLAST. *BMC Biol.* **4**, 41 (2006).
87. Slater, G. S. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinforma.* **6**, 31 (2005).
88. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
89. Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
90. Wu, T. D. & Watanabe, C. K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859–1875 (2005).
91. Haas, B. J. et al. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003).
92. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EvidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
93. Cantalapiedra, C. P., Hernandez-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**, 5825–5829 (2021).
94. Li, L., Stoeckert, C. J. Jr. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).
95. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
96. Ronquist, F. et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
97. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
98. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
99. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).
100. Puttick, M. N. MCMCtreeR: functions to prepare MCMCtree analyses and visualize posterior ages on trees. *Bioinformatics* **35**, 5321–5322 (2019).
101. Hahn, M. W., Demuth, J. P. & Han, S. G. Accelerated rate of gene gain and loss in primates. *Genetics* **177**, 1941–1949 (2007).
102. Wang, Y. et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).
103. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–W612 (2006).
104. Wang, D., Zhang, Y., Zhang, Z., Zhu, J. & Yu, J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteom. Bioinforma.* **8**, 77–80 (2010).
105. Yang, Z. & Nielsen, R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* **17**, 32–43 (2000).
106. Hauser, M., Steinegger, M. & Soding, J. MMseqs software suite for fast and deep clustering and searching of large protein sequence sets. *Bioinformatics* **32**, 1323–1330 (2016).
107. Chauve, C. & Tannier, E. A methodological framework for the reconstruction of contiguous regions of ancestral genomes and its application to mammalian genomes. *PLoS Comput. Biol.* **4**, e1000234 (2008).
108. Chauve, C., Gavranovic, H., Ouangraoua, A. & Tannier, E. Yeast ancestral genome reconstructions: the possibilities of computational methods II. *J. Comput Biol.* **17**, 1097–1112 (2010).
109. Armstrong, J. et al. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587**, 246–251 (2020).
110. Paten, B. et al. Cactus: Algorithms for genome multiple sequence alignment. *Genome Res.* **21**, 1512–1528 (2011).
111. Hubisz, M. J., Pollard, K. S. & Siepel, A. PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief. Bioinform.* **12**, 41–51 (2011).
112. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
113. Wu, T. et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innov. (Camb.)* **2**, 100141 (2021).
114. Yunyun, L. *Pterygoplichthys pardalis* (Amazon sailfin catfish) genome project. figshare. Dataset. <https://doi.org/10.6084/m9.figshare.27080665.v2> (2024).

Acknowledgements

We are grateful for the funding provided to this project by the following project funds. This work was supported by the Open Fund of Key Laboratory of Prevention and Control for Aquatic Invasive Alien Species, Ministry of Agriculture and Rural Affairs, Pearl River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Guangzhou 510380, PR China (No. 20230303), National Natural Science Foundation of China (No. 32200347), Sichuan Science and Technology Program (No. 2023NSFSC1238 and 2025ZNSFSC1077), the Major Project in Neijiang Normal University (no. 2024ZDZ05), the Neijiang Normal University Genomic Innovation Research Team (No. X23B0002), the Rural Revitalization Strategy Special Provincial Organization and Implementation Project Funds (No. 2023SBH00001), National Freshwater Genetic Resource Center (No. FGRC18537), Sichuan Provincial Funding for Freshwater Fish Innovation (No. SCCXTD-2024-15), National Key R&D Program of China (No. CJW2024236), and Finance Special Fund of Chinese Ministry of Agriculture and Rural Affairs of the People's Republic of China (Fisheries resources and environment survey in the key water areas of Southwest China).

Author contributions

Y.Y.L., Q.S., and X.D.M. supervised the project. Y.Y.L., Y.P.L., M.F., Y.L., Y.Y.W., Y.X.Y., and Y.C.Z. performed main experiments and analyses. Y.Y.L., Y.P.L., Q.S., and X.D.M. wrote and revised the manuscript. All authors read and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42003-025-08029-4>.

Correspondence and requests for materials should be addressed to Yunyun Lv, Qiong Shi or Xidong Mu.

Peer review information *Communications Biology* thanks Gabriel M. Yazbeck and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Michele Repetto.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025, corrected publication 2025