Shao et al. *Zool. Res.* 2023, 44(2): 349−352
https://doi.org/10.24272/j.issn.2095-8137.2022.387

ZR | Zoological Research

Letter to the editor Open Access

Genome of striped eel catfish (*Plotosus lineatus*) provides insights into the adaptive evolution of amphidromous fish

DEAR EDITOR,

Dramatic changes in habitat can have a range of effects on organisms (Bi & Zhang, 2021). *Plotosus lineatus* is a representative fish within Ostariophysi showing adaptation from freshwater habitats to the ocean. Here, we sequenced the genome of *P. lineatus* using long-read and Hi-C sequencing technology. Based on the high-quality chromosome-level genome, we explored the molecular basis of various *P. lineatus* features, including large genome with fewer chromosomes, novel dendritic organ (DO), and amphidromous and venomous nature. Our results showed that changes in the *P. lineatus* habitat may have resulted in genomic enlargement, owing to the wide range of transposon activity caused by genomic shock. Increased copy number of transposons may have promoted chromosome fusion and translocation. Positive selection and gene family expansion also played a role in the adaptive evolution of *P. lineatus*, particularly in immunity and gametogenesis. Tissue-specific expression analysis indicated that some developmental stages of the DO may be similar to those of the kidney, and existing genes may play an important role in the formation of novel organs. We identified candidate genes associated with toxins in *P. lineatus* and found that these genes may have undergone rapid evolution. Tandem duplication of the *stonustoxin* gene may be one reason for *P. lineatus* toxicity. Our findings provide novel insights into the evolution of the striped eel catfish genome, chromosomes, adaptations, novel organs, and virulence.

Using Nanopore and HiFi long-read technologies, we assembled the *P. lineatus* genome, which was approximately 1.33 Gb in size and consisted of 111 contigs with a contig N50 of 36.8 Mb (Supplementary Table S1). These contigs were further assembled into 24 pseudochromosomes at the chromosome scale with an assembly mount ratio of 99.74% (Supplementary Figure S1 and Table S2). Benchmarking Universal Single-Copy Orthologs (BUSCO) assessment showed that 97.69% of the genes were complete (metazoa_odb10 database), indicating high completeness of the assembled genome (Supplementary Table S3). We identified 21 969 protein-coding genes with a BUSCO assessment of 92.80% (actinopterygii_odb9 database), of which 20 378 (92.76%) were functionally annotated (Supplementary Tables S4, S5). Transposable elements (TEs) accounted for 65.44% of the genome (Supplementary Table S6). DNA and long interspersed nuclear element (LINE) transposons were significantly increased in *P. lineatus*. The increase in LINE transposons was primarily due to the increase in the *RTE* superfamily (Supplementary Figure S2). The *P. lineatus* genome is the largest among published catfish genomes (https://www.ncbi.nlm.nih.gov/assembly/?term=Siluriformes), which is likely due to the proliferation of transposons. According to the genomic shock hypothesis proposed by McClintock (1984), stress and regulatory interference due to environmental changes in the habitat may lead to the mobilization of TEs. Within Ostariophysi, *P. lineatus* is a representative fish showing adaptation from freshwater habitats to the ocean, thus experiencing dramatic changes in habitat during its evolutionary history. This may have caused genomic shock and transposon activation in the genome of its ancestors, thereby affecting genome size.

Chromosome number varies greatly among catfish species (Arai, 2011). Our results showed that *P. lineatus* has fewer chromosomes compared to other catfish with published genomes. To explain this, we used chromosome data from different species (*Cranoglanis bouderius*, *Hemibagrus wyckioides*, *Ictalurus punctatus*, *Leiocassis longirostris*, *Silurus meridionalis*, *Tachysurus fulvidraco*, and *Electrophorus electricus*) and reconstructed 34 proto-chromosomes (Figure 1A; Supplementary Table S7) for the last common ancestor (LCA) of Siluriformes and identified many more chromosome fusion events in *P. lineatus* (Supplementary Figure S3). We selected 15 fusion and translocation sites (FTSs, Figure 1A; Supplementary Table S8), and found that TE content was significantly higher in seven FTSs (c, d, h, i, k, n, and o) than in other regions (Figure 1B; Supplementary Table S9). Only *Tc1-Mariner* and *RTE* were at higher concentrations (Figure 1C; Supplementary Table S10), and the Mann-Whitney test was used to compare differences in *Tc1-Mariner* and *RTE* content among the seven sites and other FTSs by examining the main factors causing differences in transposon content at sites. The *RTE* superfamily (P=0.00373) played a dominant role in the enrichment of transposons at FTSs. We speculate that transposons may participate in chromosome fusion and translocation in *P. lineatus*, with *RTE* playing an important role in these processes. Ectopic recombination is an important cause of chromosomal rearrangements; for example, non-allelic homologous recombination between different ancestral muntjac chromosomes is thought to have led to tandem

Received: 03 December 2022; Accepted: 20 February 2023; Online: 20 February 2023
Foundation items: This work was supported by the National Natural Science Foundation of China (31872204) and Fundamental Research Funds for the Central Universities (SWU120049)

fusions (Yin et al., 2021). Furthermore, the copy number of transposons is positively correlated with the frequency of heterotopic recombination (Petrov et al., 2003). This process may be exhibited in *P. lineatus*, whereby transition from freshwater to seawater habitat potentially caused genomic shock, resulting in large-scale transposon activation, increased transposon copy number, and increased heterotopic recombination, and finally the fusion and translocation of *P. lineatus* chromosomes. Synteny analysis (Figure 1D) indicated the occurrence of many chromosomal rearrangements during catfish evolution, possibly explaining the diversity of chromosome numbers in these species.

A phylogenetic tree was constructed based on 3 774 single-copy orthologous genes from *P. lineatus* and nine other fish genomes (Supplementary Figure S4). To detect the potential genetic basis of *P. lineatus* traits compared to other species, we performed positive selection and gene family analyses. We identified various positively selected genes (PSGs) that may be involved in osmotic pressure regulation. Notably, *efhc2* is crucial for the distal segmentation of the pronephros (primitive urinary organ) in zebrafish. In addition, *pdzk1* is involved in the regulation of proximal tubular Na (+)-dependent inorganic phosphate co-transport. The protein encoded by *atp1b1* is a non-catalytic component of the active enzyme that catalyzes adenosine triphosphate (ATP) hydrolysis, coupled with the exchange of Na^+ and K^+ ions across the plasma membrane. Interestingly, we identified several PSGs that may be involved in gametogenesis, specifically genes associated with spermatogenesis, such as *crtap*, *ccdc136*, *cfap43*, and *iho1*. Furthermore, the *marf1* gene, which is involved in oogenesis, was shown to be under positive selection. We also identified several possible genes related to nocturnal adaptation in *P. lineatus*, including *ciart*, which is associated with circadian rhythm, and *crba1*, which is associated with lens development.

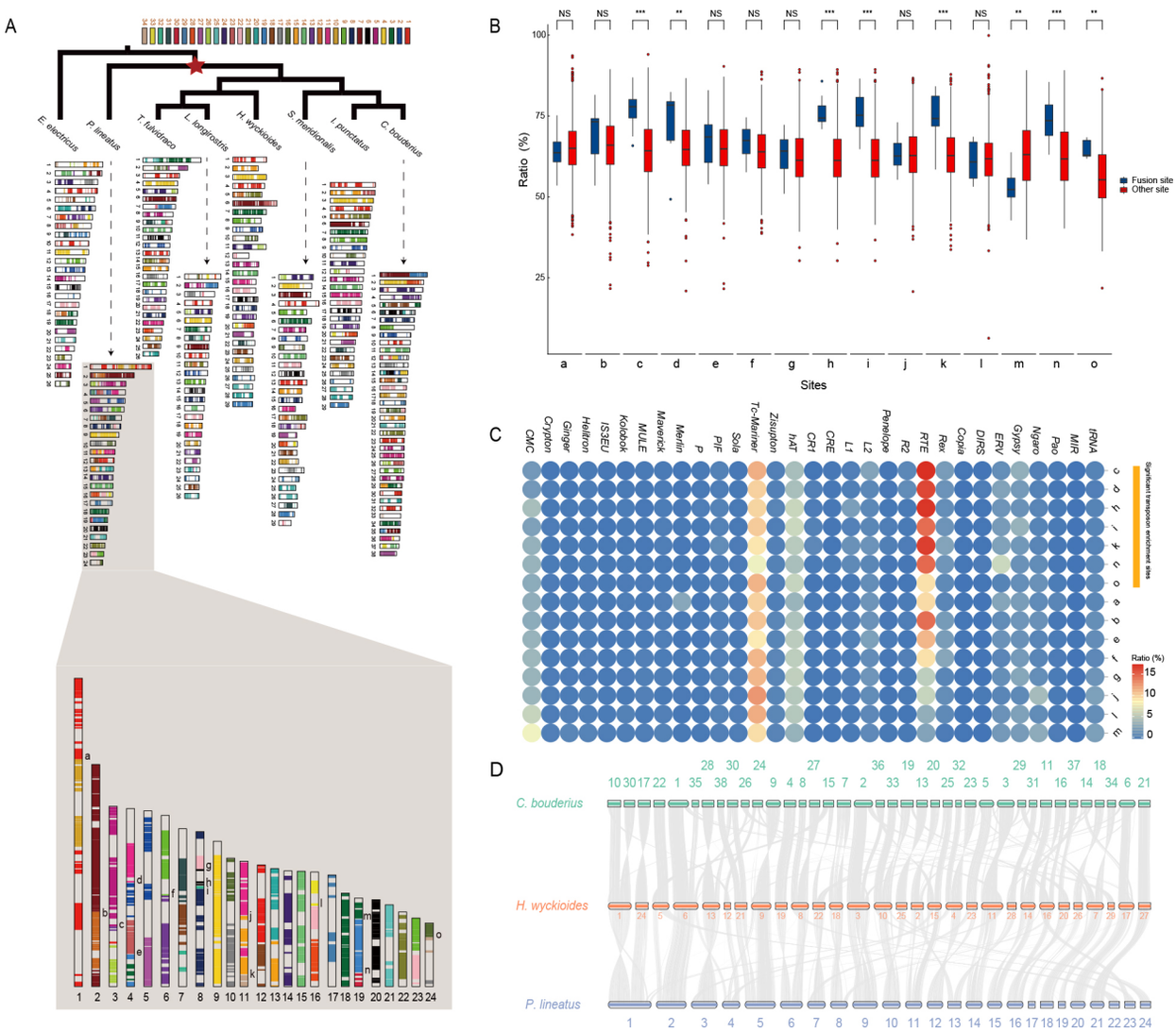


Figure 1 Chromosomal evolution of *P. lineatus*

A: Ancestral chromosome construction of Siluriformes. Conserved blocks are color-coded according to their chromosomal source in Siluriformes ancestor. Designated FTSs (distance between adjacent different colored blocks of less than 1 Mb) are identified by letters in enlarged figure. B: Comparison of TE content between selected fusion site and other sites on corresponding chromosome (100 kb window, significant differences: NS: Not significant; **: P<0.01; ***: P<0.001). Letters on horizontal axis indicate different fusion sites, and different colors indicate selected fusion site and other sites on corresponding chromosome. C: Heatmap of TE superfamily content at selected fusion site. D: Pairwise whole-genome alignments across *C. bouderius*, *H. wyckioides*, and *P. lineatus* genomes. Each horizontal bar represents one chromosome, and chromosome IDs of the three species are labeled.

The above PSGs and related references are presented in Supplementary Table S11. In addition, 345 and 2 942 gene families were expanded and contracted in *P. lineatus*, respectively, including 63 significantly expanded gene families (Supplementary Table S12). Immune-related genes were considerably enriched based on Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) analysis (Supplementary Tables S13, S14). Many GO terms were related to virus defense processes (Supplementary Figure S5A), suggesting that *P. lineatus* may have significantly enhanced viral defense capability, consistent with the viruses encountered in seawater and freshwater habitats. Furthermore, circadian rhythm-related pathways and terms were also enriched, which may be related to their nocturnal behavior. The *sox* gene family was also expanded (43 members) in *P. lineatus* compared with that in *C. bouderius* and *H. wyckioides* (26 and 25 members, respectively), which was mainly due to *sox30* (Supplementary Figure S6 and Table S15). The *sox30* gene is involved in spermatogenesis in fish (Wei et al., 2022), suggesting that *P. lineatus* may also have adapted to the marine environment during germ cell processes.

Plotosidae catfish have an extra-branchial salt-secreting DO organ, which is unique among catfish and all marine bony fish (Kolbadinezhad et al., 2018). We sequenced the transcriptomes of 13 tissues from *P. lineatus* to study DO formation. Consequently, we obtained 305 genes specifically expressed in DO tissues (Supplementary Table S16), which were thus subjected to GO and KEGG enrichment analyses (Supplementary Tables S17, S18). Results showed that the genes were enriched in two KEGG pathways, i.e., “Proximal tubule bicarbonate reclamation” and “Mineral absorption”, and in GO terms related to embryonic and kidney development (Supplementary Figure S5B). These findings are consistent with previous studies showing that the DO is osmoregulatory (Kolbadinezhad et al., 2018) and suggest similarities between DO and kidney development, as specifically expressed DO genes were enriched in kidney development. Tissue-specific expression analysis also indicated that existing genes (DO specific-expressed genes related to renal tubule development) may play an important role in novel organ development. This is essentially similar to our previous findings in lungfish, where the *sftpb* gene, important in respiratory system evolution, already existed in the common ancestors of bony fish and thus is not novel (Wang et al., 2021). We also identified transposon activation in 91 of the 305 purposely expressed genes, including 33 genes specifically in the DO (Supplementary Table S19). Of these 33 genes, *six1b*, *six4*, and *chs2* were enriched in “renal tubule morphogenesis/development”, and the corresponding activated transposons were long terminal repeat type (LTR, unknown superfamily), *Ngaro*, and *RTE*, respectively. These results suggest that transposons may be involved in DO tissue-specific expression of genes as well as DO formation.

Plotosus lineatus has received increasing attention due to its possible high toxicity. However, among catfish with published genomes, the virulence of most species is ambiguous, with only *Ameiurus melas* found to be non-toxic (Wright, 2009). We comprehensively annotated venom-like genes in *A. melas* and *P. lineatus* using a multistep approach and identified 86 and 98 venom-like genes in their genomes, respectively (Supplementary Table S20). We evaluated the molecular evolutionary characteristics of venom-like genes in *A. melas* and *P. lineatus* by calculating the number of synonymous (*Ks*) and nonsynonymous (*Ka*) nucleotide substitutions per site for each pair of venom-like and other genes in the two species. Results showed that the *Ka*/*Ks* substitution ratios of venom-like gene pairs were significantly higher than those of other gene pairs (Supplementary Figure S7A–C and Table S21), suggesting that rapid evolution may be one way to form venom-like genes. Comparing venom-like genes in the two species, *stonustoxin* genes were found to be more abundant in *P. lineatus*. Stonustoxin is an extremely potent cytolytic toxin, mainly found in stonefish (Ellisdon et al., 2015). Here, we identified 12 *stonustoxin* genes in the *P. lineatus* genome and nine in *A. melas*. Importantly, we found tandem duplications in nine *stonustoxin* genes in the *P. lineatus* genome (Supplementary Figure S7D). In contrast, *A. melas* had only five *stonustoxin* genes at the corresponding locations (two very close clusters). We also observed the distribution of *stonustoxin* genes in corresponding regions of *T. fulvidraco*, a species with toxicity characteristics, and identified 10 *stonustoxin* genes, although no continuous tandem repeats. Tandem repeat genes can affect expression and evolution; for example, duplicated genes often accumulate more transcripts to enhance expression (Picart-Picolo et al., 2020). However, among the toxin candidate genes in *P. lineatus*, some were immune-related genes. For example, there were 14 lectin-related genes in *P. lineatus* and 10 lectin-related genes in *A. melas*, as well as six *aep1* genes in *P. lineatus* and three *aep1* genes in *A. melas*. These results suggest that gene number expansion may be an important pathway for toxin formation in fish. Notably, there appears to be a link between immune-related genes and toxin genes (Ellisdon et al., 2015).

DATA AVAILABILITY

The striped eel catfish chromosome-level genome and genome sequencing data used in the present study, including Nanopore, PacBio, MGISEQ2000, Hi-C, and transcriptome data, are available from the Sequence Read Archive (SRA) with the BioProjectID PRJNA720277. The above data were also archived in the Science Data Bank database (DOI: 10.57760/sciencedb.06933) and Genome Sequence Archive (GSA) database (CRA009258).

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS’ CONTRIBUTIONS

F.S. and Z.G.P. conceived the study and wrote the manuscript. F.S. and P.L. conducted field sample collection and prepared samples for sequencing. F.S. conducted bioinformatics analysis. Z.C.W. manually checked the genes under study. Y.X. participated in the construction of the ancestral genome of catfish. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We thank Lu-Yun Ni, Min-Zhi Zeng, Yan-Qiu Zhang, Huai-Yi Fang, and Xue-Lan Chen for their help with sample collection. We are grateful to Drs. Cheng-Long Zhu, Xu-Peng Bi, and Long Zhou for help with the reconstruction of ancestral karyotype. We also thank Drs. Lei Chen, Ru Zhang, Ze-Shan Lin, Bao Wang, Zhong-Kai Wang, Yan-Dong Ren, Yuan Yin, and Luo-Hao Xu for helpful discussions.

Feng Shao[1], Ze-Chen Wu[1], Yuan Xu[1], Ping Li[2], Zuo-Gang Peng[1,3,*]

[1] *Key Laboratory of Freshwater Fish Reproduction and Development (Ministry of Education)*, *Southwest University School of Life Sciences*, *Chongqing* 400715, *China*

[2] *College of Fisheries*, *Southwest University*, *Rongchang*, *Chongqing* 402460, *China*

[3] *Academy of Plateau Science and Sustainability*, *Qinghai Normal University*, *Xining*, *Qinghai* 810008, *China*

*Corresponding author, E-mail: pzg@swu.edu.cn

REFERENCES

Arai R. 2011. Fish Karyotypes: A Check List. Tokyo: Springer.

Bi XP, Zhang GJ. 2021. Ancestral developmental potentials in early bony fish contributed to vertebrate water-to-land transition. *Zoological Research*, **42**(2): 135.

Ellisdon AM, Reboul CF, Panjikar S, et al. 2015. Stonefish toxin defines an ancient branch of the perforin-like superfamily. *Proceedings of the National Academy of Sciences of the United States of America*, **112**(50): 15360−15365.

Kolbadinezhad SM, Coimbra J, Wilson JM. 2018. Osmoregulation in the Plotosidae catfish: role of the salt secreting dendritic organ. *Frontiers in Physiology*, **9**: 761.

Mcclintock B. 1984. The significance of responses of the genome to challenge. *Science*, **226**(4676): 792−801.

Petrov DA, Aminetzach YT, Davis JC, et al. 2003. Size matters: non-LTR retrotransposable elements and ectopic recombination in drosophila. *Molecular Biology and Evolution*, **20**(6): 880−892.

Picart-Picolo A, Grob S, Picault N, et al. 2020. Large tandem duplications affect gene expression, 3D organization, and plant-pathogen response. *Genome Research*, **30**(11): 1583−1592.

Wang K, Wang J, Zhu CL, et al. 2021. African lungfish genome sheds light on the vertebrate water-to-land transition. *Cell*, **184**(5): 1362−1376.e18.

Wei L, Tang YH, Zeng XH, et al. 2022. The transcription factor Sox30 is involved in Nile tilapia spermatogenesis. *Journal of Genetics and Genomics*, **49**(7): 666−676.

Wright JJ. 2009. Diversity, phylogenetic distribution, and origins of venomous catfishes. *BMC Evolutionary Biology*, **9**: 282.

Yin Y, Fan HZ, Zhou BT, et al. 2021. Molecular mechanisms and topological consequences of drastic chromosomal rearrangements of muntjac deer. *Nature Communications*, **12**(1): 6858.

ZR Zoological Research

Supplementary Materials

Genome of striped eel catfish (*Plotosus lineatus*) provides insights into the adaptive evolution of amphidromous fish

Feng Shao [1], Ze-Chen Wu [1], Yuan Xu [1], Ping Li [2], Zuo-Gang Peng [1,3,*]

[1] Key Laboratory of Freshwater Fish Reproduction and Development (Ministry of Education), Southwest University School of Life Sciences, Chongqing 400715, China
[2] College of Fisheries, Southwest University, Rongchang, Chongqing 402460, China
[3] Academy of Plateau Science and Sustainability, Qinghai Normal University, Xining, Qinghai 810008, China

*Corresponding author, E-mail: pzg@swu.edu.cn

SUPPLEMENTARY MATERIALS AND METHODS

Sample collection and sequencing

Samples for genome sequencing were collected from female *Plotosus lineatus* from Beibu Gulf, Qin Zhou, Guangxi Province, China (21°44′19.79″ N, 108°34′41.97″ E). The study of *P. lineatus* has been approved by Southwest University IACUC (Institutional Animal Care and Use Committee), IACUC NO. Approved: 20200715-01. Blood, brain, dendritic organ (DO), fins, gills, gonads, gut, heart, head kidney, liver, muscle, skin, and spleen were collected and immediately frozen in liquid nitrogen.

Blood samples were collected for genome sequencing, and DNA was prepared using a QIAGEN® Genomic Kit (Cat No. ID:13343, QIAGEN, Hilden, Germany). *De novo* genome sequencing was performing using high-throughput sequencing, Nanopore PromethION, PacBio Sequel II (circular consensus sequence, CCS), and Hi-C technologies. For MGISEQ 2000 sequencing, we constructed one paired-end library with insert sizes of 200–400 bp, followed by filtering using fastp (Chen et al., 2018) to generate 61.75 Gb of clean data for genome survey and correction. For Nanopore PromethION sequencing, we constructed two long-read libraries. Raw reads in fastq format with a mean_qscore_template<7 were then filtered to generate 115.93 Gb of clean data. For PacBio Sequel II sequencing, we constructed one long-read library (10–20 kb). Raw data were analyzed using Smrtlink software (https://www.pacb.com/support/software-downloads/) to generated 10.28 Gb of clean data. We constructed three Hi-C libraries for Hi-C sequencing and generated clean data at a depth of 122.14 Gb (filtered using fastp).

RNA analysis involved 13 tissues (brain, DO, fin, gill, gonad, gut, heart, head kidney, kidney, liver, muscle, skin, and spleen) extracted using an RNeasy Plus Mini Kit (QIAGEN, Hilden, Germany). Illumina paired-end sequencing (validated RNA samples, Illumina HiSeq4000, 150 bp) was used to prepare a cDNA library with a TruSeq Sample Preparation Kit (Illumina, San Diego, USA). Qualified RNA from 10 tissues (blood, brain, fin, gill, gonad, heart, head kidney, liver, muscle, and spleen) was mixed in equal amounts and reverse transcribed using SQK-PCS109 for ONT (Oxford Nanopore Technologies) library preparation and sequencing (Nanopore PromethION). Raw data were filtered in the same manner as for DNA.

Genome assembly and evaluation

Quality-filtered reads were subjected to 35 mer frequency distribution analysis using Jellyfish (Marçais & Kingsford, 2011). We analyzed the 35 mer depth distribution of clean sequencing reads using FindGSE (Sun et al., 2018) and KMC (Kokot et al., 2017) to estimate *P. lineatus* genome size and heterozygosity (Supplementary Figure S8). Long-read (ONT) assembly was performed using NextDenovo (reads_cutoff: 1 k, seed_cutoff: 32 k, https://github.com/Nextomics/NextDenovo). To improve assembly accuracy, the contigs were refined using default parameters in NextPolish (https://github.com/Nextomics/NextPolish) for long and short reads. To anchor the hybrid scaffolds to the chromosome, they were further clustered, ordered, and oriented on chromosomes using LACHESIS (Korbel & Lee, 2013) with the parameters CLUSTER_MIN_RE_SITES=100, CLUSTER_MAX_LINK_DENSITY=2.5, NONINFORMATIVE RATIO=1.4, ORDER MIN N RES IN TRUNK=60, and ORDER MIN N RES IN SHREDS=60. Finally, the interactive heatmap of the initial assembly results of LACHESIS was drawn according to the interactions between different scaffolds. The position and direction of scaffolds that obviously did not meet the chromosome interaction characteristics in the figure were adjusted. Of note, if a scaffold itself did not meet chromosome interaction characteristics, the scaffold was interrupted. The scaffolds were then adjusted separately until the overall heatmap conformed to the characteristics of chromosome interaction.

To evaluate *P. lineatus* genome assembly quality, we identified single-copy orthologs and conserved eukaryotic core genes in metazoa_odb10 based on the BUSCO database and filtered RNA sequencing (RNA-seq) short reads aligned to the reference genome using hisat2 (Kim et al., 2015) with default parameters (Supplementary Table S22).

Genome annotation

We annotated tandem repeats using GMATA (Wang & Wang, 2016) and TRF (Benson, 1999). For transposable element (TE) annotation, an *ab initio* repeat library for *P. lineatus* was initially

predicted using MITE-hunter (Han & Wessler, 2010) and RepeatModeler2 (Flynn et al., 2020) with default parameters, followed by LTR_FINDER (Xu & Wang, 2007), LTRharverst (Ellinghaus et al., 2008), and LTR_retriver (Ou & Jiang, 2018). RepeatMasker (http://www.repeatmasker.org/) was used to search for known and novel TEs by mapping sequences against the *de novo* repeat and RepBase TE libraries.

RNA-seq-based gene prediction was based on the RNA-seq short-read mapping results (described above). Transcripts were assembled using StringTie (Pertea et al., 2015). Full-length reads were identified and oriented from the sequencing reads using the Pychopper tool (https://github.com/nanoporetech/pychopper) with default parameters. Full-length reads were aligned to the *P. lineatus* reference genome using minimap2 (Li, 2018) with "-ax splice-uf". The aligned full-length reads were clustered using pinfish software (https://github.com/nanoporetech/pinfish) following the official recommendations of Nanopore. Redundancy was removed using cDNA_Cupcake software (https://github.com/Magdoll/cDNA_Cupcake) and polished using a reference genome sequence. The assembled transcripts based on full-length and short reads were merged with open reading frames (ORFs) and predicted using PASA (Haas et al., 2008). For *de novo* prediction, RNA-seq reads were assembled using StringTie (Pertea et al., 2015) and analyzed with PASA (Haas et al., 2008) to produce a training set. AUGUSTUS (Stanke et al., 2008) with default parameters was used for *ab initio* gene prediction using the training set. Homologous proteins of other species (*Bagarius yarrelli*, *Danio rerio*, *Glyptosternon maculatum*, *Hemibagrus* wyckioides, *Homo sapiens*, *Ictalurus punctatus*, *Oryzias latipes*, *Tachysurus fulvidraco*, and *Takifugu rubripes*) were also used to predict protein-coding genes using GeMoMa (Keilwagen et al., 2016). A protein-coding gene dataset was then generated by integrating the three results using EvidenceModeler (Haas et al., 2008). Furthermore, TransposonPSI (http://transposonpsi.sourceforge.net) was used to remove genes containing TEs. To evaluate *P. lineatus* genome annotation quality, we identified single-copy orthologs and conserved eukaryotic core genes in actinopterygii_odb9 based on the BUSCO database. To annotate gene function, the predicted gene sequences were searched against five databases: i.e., SwissProt (Uniprot Consortium, 2021), NR (https://www.ncbi.nlm.nih.gov/refseq/about/nonredundantproteins/), KEGG (Kanehisa & Goto, 2000), KOG (Tatusov et al., 2003), and GO (Gene Ontology Consortium, 2017).

Reconstruction of ancestral karyotype and fusion site analysis

We performed whole-genome alignments across all chromosome-level catfish genomes (*Cranoglanis bouderius*, *H. wyckioides*, *I. punctatus*, *Leiocassis longirostris*, *P. lineatus*, *Silurus meridionalis*, and *T. fulvidraco*) and an outgroup species (*Electrophorus electricus*). First, pairwise whole-genome alignments were obtained between *P. lineatus* and the other species using LASTZ (Harris, 2007) with parameters 'H=2 000, Y=3 400, L=6 000, K=2 200'. Chain/Net was performed using lavToPsl, axtChain, chainPreNet, and chainNet (Kuhn et al., 2013). We then applied the DESCHRAMBLER algorithm (https://github.com/jkimlab/DESCHRAMBLER), chains, and nets generated by LASTZ to generate conserved syntenic fragments across the eight species. Syntenic fragments ≥500 kb in length were used to construct conserved syntenic fragments. Finally, the ancestral genome structures were constructed using ANGES (Jones et al., 2012), and the results less than 1M were filtered out. The final results were drawn in SVG. Rearrangement events in each lineage were inferred using GRIMM (Tesler, 2002). Ancestral chromosome blocks on chromosomes of different species are shown in different colors. Adjacent blocks of different ancestors on the same chromosome are mainly caused by fusion and translocation. In this study, the distance between adjacent color blocks of less than 1 Mb was defined as a fusion and translocation site (FTS). For these FTSs in *P. lineatus*, 500 kb was taken from the center of the two-color blocks upstream and downstream for a total of 1 Mb for subsequent analysis. Combined with the transposon annotation results, we calculated the transposon content in FTSs and other regions of the chromosome (100 kb window), performed rank sum Mann-Whitney tests on FTSs and non-fusion sites, and performed statistical analysis of transposon types in FTSs.

Comparative genomic analysis

We used the MUMmer (Marçais et al., 2018) tool nucmer to perform pairwise whole-genome alignment with the parameter "-b 400." The alignments were filtered to retain the one-to-one best

hits using delta-filt from the MUMmer package. Unanchored scaffolds were excluded from the alignments. The alignments were formatted using the MCscan pipeline (https://github.com/tanghaibao/jcvi) for synteny visualization.

We identified homologous relationships between *P. lineatus* and other species (*Astyanax mexicanus*, *C. bouderius*, *D. rerio*, *E. electricus*, *G. maculatum*, *H. wyckioides*, *L. oculatus*, *O. latipes*, and *S. meridionalis*) by downloading their protein sequences and aligning them using OrthoMCL (Li et al., 2003). Molecular phylogenetic analysis was performed using shared single-copy genes based on orthologous gene sets identified with OrthoMCL. Each ortholog group was aligned multiple times using MAFFT (Kasahara et al., 2007). Poorly aligned sequences were then eliminated using Gblocks (http://molevol.cmima.csic.es/castresana/Gblocks.html), and the GTRGAMMA substitution model in RAxML (Stamatakis, 2014) was used for phylogenetic tree construction with 1 000 bootstrap replicates. Three fossil calibration times were obtained from TimeTree (Kumar et al., 2022). The control time with divergence time parameters was "((*O. latipes*, (*D. rerio*, ((*E. electricus*, (*P. lineatus*, (((*H. wyckioides*, *G. maculatum*), *C. bouderius*), *S. meridionalis*))), *A. mexicanus*)) 'B (126 179)') 'B (205 255)', *L. oculatus*) 'B (291 338)'". To detect the potential genetic basis of the unique characteristics of *P. lineatus*, we performed positive selection analysis based on phylogenetic trees using the branch-site model implemented in PAML (Yang, 1997). A likelihood ratio test (LRT) was conducted to compare models that allowed sites to be under positive selection for *P. lineatus*. *P*-values were computed based on chi-square statistics, and genes with $P<0.05$ were treated as candidate positively selected genes (PSGs). According to the OrthoMCL results, gene family expansions and contractions were detected using CAFE (De Bie et al., 2006) and enrichment tests were performed using ClusterProfiler (Yu et al., 2012), with the same tools used for other enrichment analyses.

For phylogenetic tree reconstruction of the *sox* gene family, we aligned all protein sequences (MUSCLE) and constructed a maximum-likelihood tree using MEGA 11 (Tamura et al., 2021).

Analysis of DO-specific expressed gene

Based on the gene structure results and 13 tissue RNA-seq short-read mapping results, StringTie was used to calculate the expression value of each gene. The expression value in each tissue was then used to calculate the tissue specificity index (TAU) (Yanai et al., 2005) for each gene. Tissue-specific expression was defined as a gene that showed at least two-fold higher expression than in any other tissue, with the highest RPKM (Reads Per Kilobase per Million mapped reads) >1 and TAU>0.8.

To determine whether transposons participate in DO development, RNA-seq data and transposon annotation results of the 13 tissues were used to predict activated transposons using TE candidates (Ohtani et al., 2013) with very demanding parameters (-c=0.9 -l=300). We screened for transposons inserted into genes specifically expressed and activated by the DO.

Identification of venom-related genes and molecular evolution

To study venom-related genes in *P. lineatus*, we used research methods applied for snake venom (Li et al., 2021). We first combined two venom-related protein databases: i.e., manually reviewed sequences from the animal toxin annotation project115 (http://www.uniprot.org/program/Toxins), which systematically annotated venom proteins and toxins from all venomous species, and venom-related proteins reported in a previous article (Xie et al., 2016), which were annotated using Swiss-Prot (Boeckmann et al., 2003). All predicted protein sequences of *P. lineatus* were queried against the combined dataset using BLASTP (Camacho et al., 2009), with a similarity cut-off E-value of 1e-5. Redundant alignments were removed. For each candidate venom-related protein identified, a venom-related gene was manually curated according to the annotations of the closest homologous protein hit in Swiss-Prot. The Swiss-Prot annotation results of the query and target sequences were similar. For comparison, the same analysis was performed for *Ameiurus melas*.

To assess the diversification rates of major venom-related genes in *P. lineatus* and *A. melas*, we first ran all-vs-all BLASTN (E-value≤1e-5) in two species pairs using coding sequences (CDSs) to obtain gene pairs with the highest similarity. Kalign (Lassmann, 2020) was then used to perform pairwise alignment of the gene pairs and output AXT files, which were imported into KaKs_Calculator (Wang et al., 2010). We computed nonsynonymous and synonymous substitution

rates (Ka and Ks) for protein-coding regions of each gene pair using previously reported methods (Yang & Nielsen, 2000). Modified YN (Zhang et al., 2006) and Nei-Gojobori (Nei & Gojobori, 1986) models were also applied simultaneously when invalid Ka/Ks values were generated. Significant differences between venom-like and non-venom gene families were assessed using the Mann-Whitney test.

The genes we focused on (*sox* gene family and *stonustoxin* gene) were based on the manual inspection results according to the collinearity between genes. Briefly, we used Fgenesh+ to re-predict collinear position (region between two gene flanks) of these genes.

References

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

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

Camacho C, Coulouris G, Avagyan V, et al. 2009. BLAST+: architecture and applications. *BMC Bioinformatics*, **10**: 421.

Chen SF, Zhou YQ, Chen YR, et al. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, **34**(17): i884–i890.

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

Ellinghaus D, Kurtz S, Willhoeft U. 2008. *LTRharvest*, an efficient and flexible software for *de novo* detection of LTR retrotransposons. *BMC Bioinformatics*, **9**: 18.

Flynn JM, Hubley R, Goubert C, et al. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America*, **117**(17): 9451–9457.

Haas BJ, Salzberg SL, Zhu W, et al. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology*, **9**(1): R7.

Han YJ, Wessler SR. 2010. MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Research*, **38**(22): e199.

Harris RS. 2007. Improved Pairwise Alignment of Genomic DNA. Ph.D. dissertation, The Pennsylvania State University, State College.

Jones BR, Rajaraman A, Tannier E, et al. 2012. ANGES: reconstructing ANcestral GEnomeS maps. *Bioinformatics*, **28**(18): 2388–2390.

Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, **28**(1): 27–30.

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

Keilwagen J, Wenk M, Erickson JL, et al. 2016. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Research*, **44**(9): e89.

Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, **12**(4): 357–360.

Kokot M, Długosz M, Deorowicz S. 2017. KMC 3: counting and manipulating *k*-mer statistics. *Bioinformatics*, **33**(17): 2759–2761.

Korbel JO, Lee C. 2013. Genome assembly and haplotyping with Hi-C. *Nature Biotechnology*, **31**(12): 1099–1101.

Kuhn RM, Haussler D, Kent WJ. 2013. The UCSC genome browser and associated tools. *Briefings in Bioinformatics*, **14**(2): 144–161.

Kumar S, Suleski M, Craig JM, et al. 2022. TimeTree 5: an expanded resource for species divergence times. *Molecular Biology and Evolution*, **39**(8): msac174.

Lassmann T. 2020. Kalign 3: multiple sequence alignment of large datasets. *Bioinformatics*, **36**(6): 1928–1929.

Li A, Wang JJ, Sun K, et al. 2021. Two reference-quality sea snake genomes reveal their divergent evolution of adaptive traits and venom systems. *Molecular Biology and Evolution*, **38**(11): 4867–4883.

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

Li L, Stoeckert CJ Jr, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research*, **13**(9): 2178–2189.

Marçais G, Delcher AL, Phillippy AM, et al. 2018. MUMmer4: a fast and versatile genome alignment system. *PLoS Computational Biology*, **14**(1): e1005944.

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

Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*, **3**(5): 418–426.

Ohtani H, Iwasaki YW, Shibuya A, et al. 2013. DmGTSF1 is necessary for Piwi-piRISC-mediated transcriptional transposon silencing in the *Drosophila* ovary. *Genes & Development*, **27**(15): 1656–1661.

Ou SJ, Jiang N. 2018. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiology*, **176**(2): 1410–1422.

Pertea M, Pertea GM, Antonescu CM, et al. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, **33**(3): 290–295.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**(9): 1312–1313.

Stanke M, Diekhans M, Baertsch R, et al. 2008. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics*, **24**(5): 637–644.

Sun HQ, Ding J, Piednoël M, et al. 2018. *findGSE*: estimating genome size variation within human and *Arabidopsis* using *k*-mer frequencies. *Bioinformatics*, **34**(4): 550–557.

Tamura K, Stecher G, Kumar S. 2021. MEGA11: molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution*, **38**(7): 3022–3027.

Tatusov RL, Fedorova ND, Jackson JD, et al. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*, **4**: 41.

Tesler G. 2002. GRIMM: genome rearrangements web server. *Bioinformatics*, **18**(3): 492–493.

The Gene Ontology Consortium. 2017. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Research*, **45**(D1): D331–D338.

The Uniprot Consortium. 2021. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, **49**(D1): D480–D489.

Wang DP, Zhang YB, Zhang Z, et al. 2010. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics*, **8**(1): 77–80.

Wang XW, Wang L. 2016. GMATA: an integrated software package for genome-scale SSR mining, marker development and viewing. *Frontiers in Plant Science*, **7**: 1350.

Xie B, Li XF, Lin ZL, et al. 2016. Prediction of toxin genes from Chinese yellow catfish based on transcriptomic and proteomic sequencing. *International Journal of Molecular Sciences*, **17**(4): 556.

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

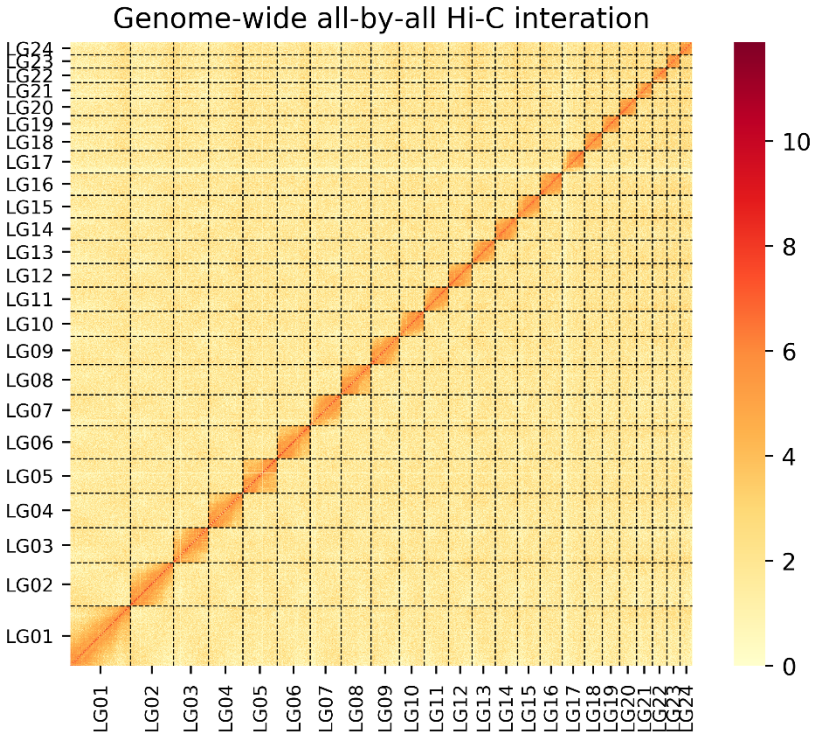
Yanai I, Benjamin H, Shmoish M, et al. 2005. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics*, **21**(5): 650–659.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences*, **13**(5): 555–556.

Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution*, **17**(1): 32–43.

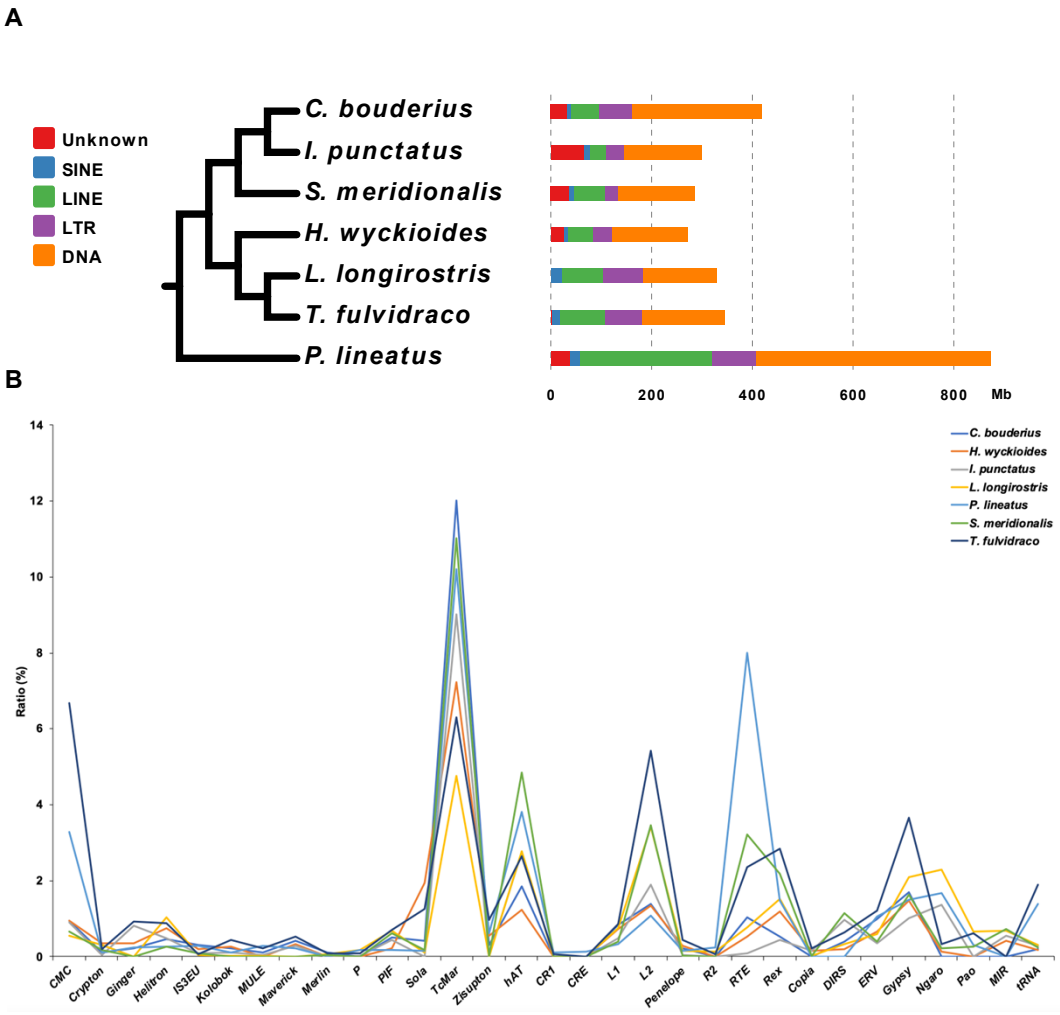
Yu GC, Wang LG, Han YY, et al. 2012. ClusterProfiler: an R package for comparing biological themes among gene clusters. *Omics*, **16**(5): 284–287.

Zhang Z, Li J, Yu J. 2006. Computing Ka and Ks with a consideration of unequal transitional substitutions. *BMC Evolutionary Biology*, **6**: 44.



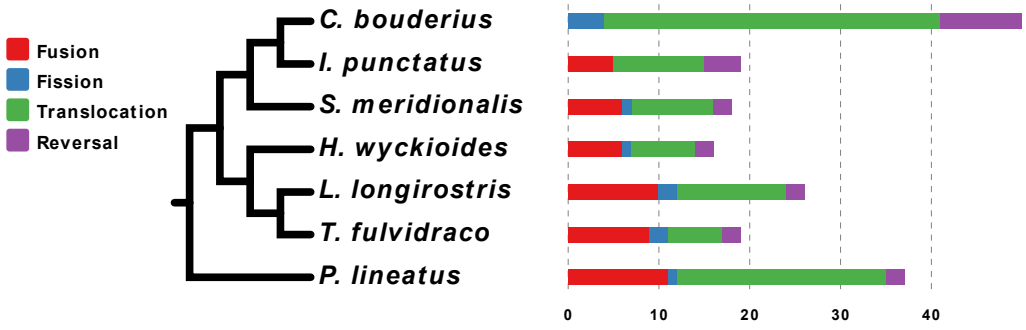
Supplementary Figure S1 Heatmap of Hi-C interaction signals

Intensity of intrachromosomal interactions is higher than that between chromosomes, indicating good assembly quality.



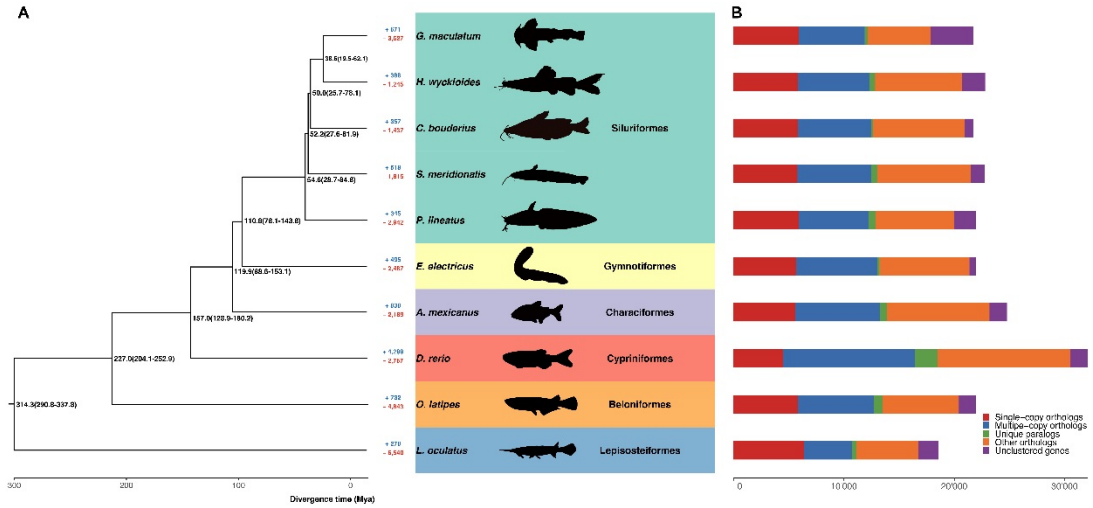
Supplementary Figure S2 TEs in *C. bouderius*, *H. wyckioides*, *I. punctatus*, *L. longirostris*, *P. lineatus*, *S. meridionalis*, and *T. fulvidraco*

A: Length of TEs of different types in three genomes. Different types of transposons are shown in different colors. B: Ratio of TEs of different types in three genomes. Different colored lines represent different species, TE superfamilies with concentrations greater than 0.1% in at least one selected species are shown.



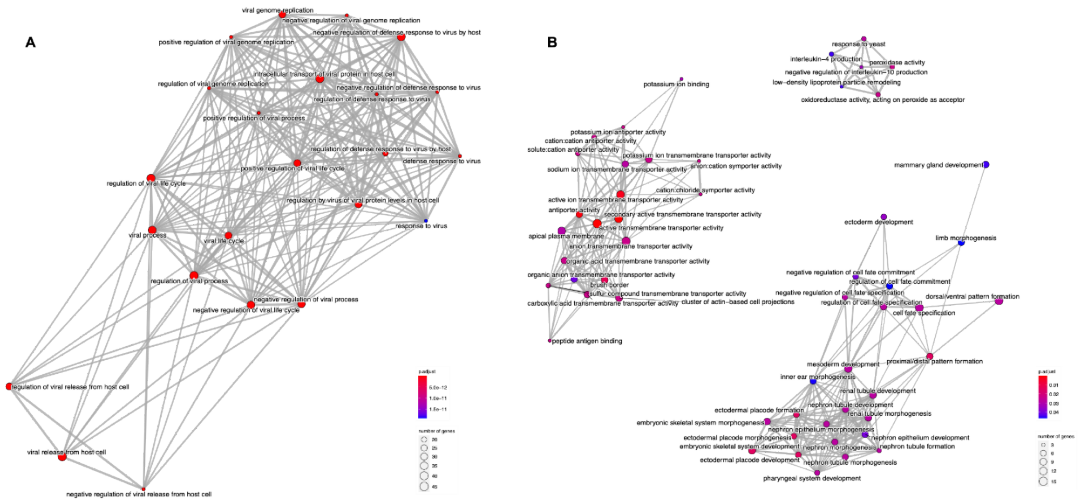
Supplementary Figure S3 Estimated number of chromosomal rearrangements from last common ancestor (LCA) of Siluriformes to each species

Different colors represent different types of rearrangement.



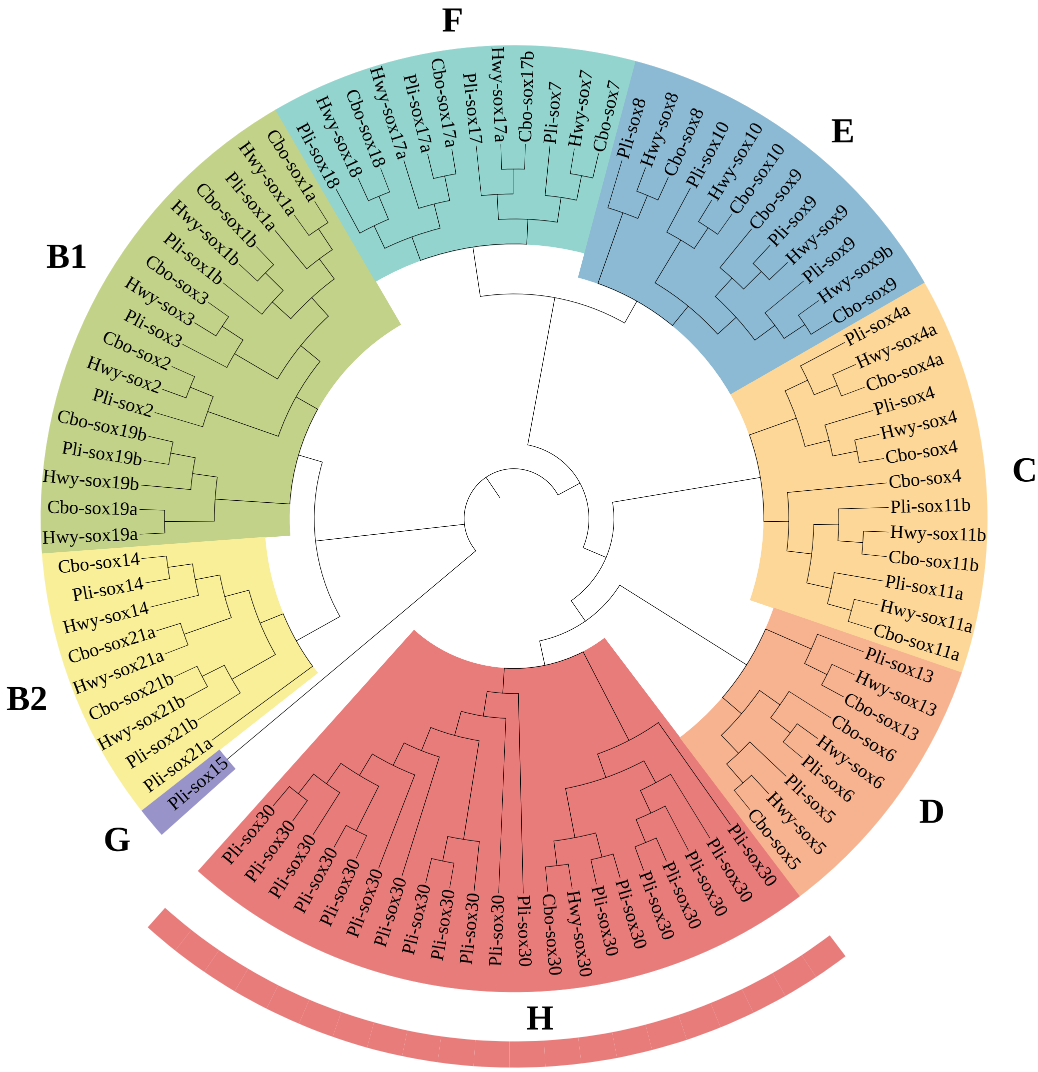
Supplementary Figure S4 Evolutionary genomics of *P. lineatus*

A: Phylogenetic relationships among selected Siluriformes. Numbers on nodes represent time of species differentiation. Numbers at end of branches represent number of expanded gene families and contracted gene families of corresponding species. B: Statistics of orthologous genes between selected Siluriformes.

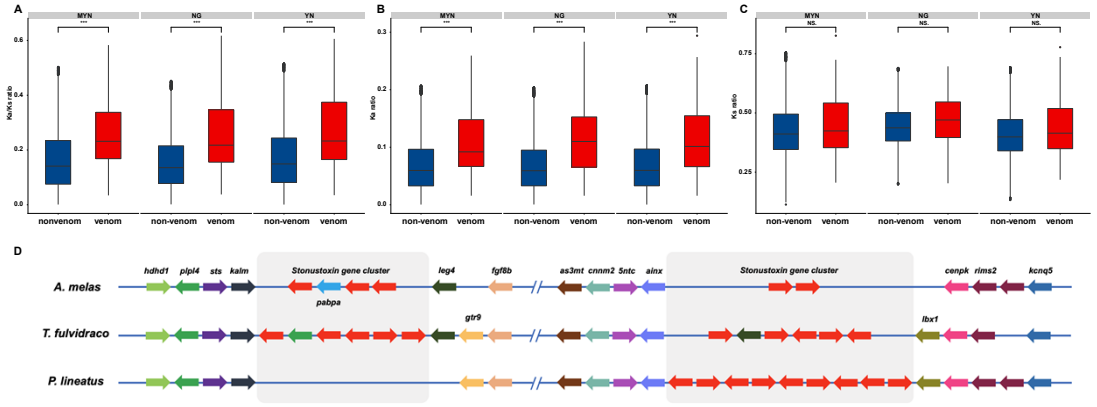


Supplementary Figure S5 GO enrichment results

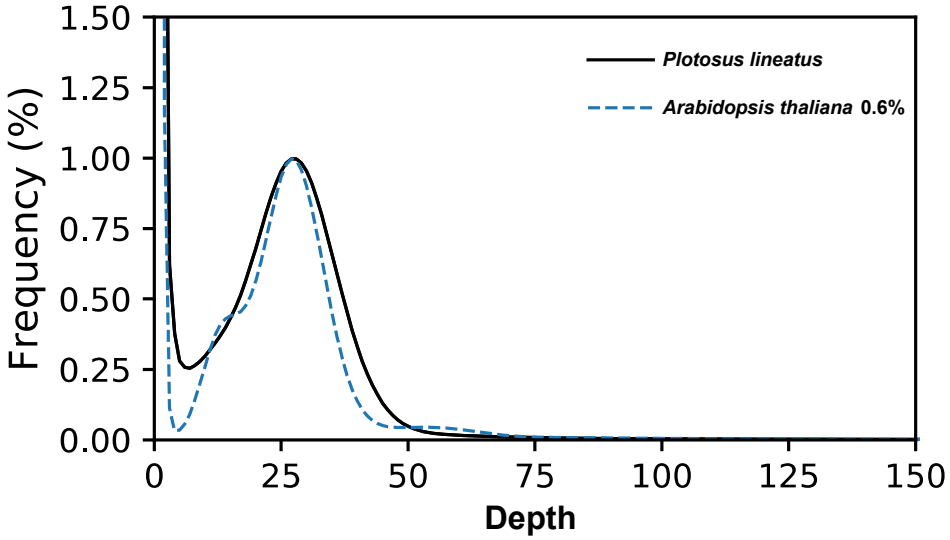
A: GO enrichment results of *P. lineatus* expanded gene families associated with viral response. B: GO enrichment results of DO tissue-specific expressed genes.



Supplementary Figure S6 Phylogenetic tree of Sox proteins from *P. lineatus* (Pli), *C. bouderius* (Cbo), and *H. wyckioides* (Hwy)



Supplementary Figure S7 Box plots comparing distributions of Ka/Ks (A), Ka (B), and Ks (V) values of venom-like and non-venom genes in *P. lineatus* and *A. melas*. Venom represents venom-like gene pairs of *P. lineatus* and *A. melas* obtained by homologous alignment, while non-venom represents gene pairs of *P. lineatus* and *A. melas*, excluding venom-like gene pairs. (D) Tandem duplication of *stonustoxin* genes in *P. lineatus*; *A. melas*, *T. fulvidraco*, and *P. lineatus* represent *stonustoxin* and adjacent genes. Synteny of genes around *stonustoxin* genes is illustrated by different colors. Gray areas connect *stonustoxin* gene clusters across species.



Supplementary Figure S8 Frequency distribution of 35 mer graph analysis used to estimate size of *Plotosus lineatus* (black line) and *K-mer* curve fitting was used to estimate heterozygosity of *P. lineatus* (blue dotted line).

Supplementary Tables

Supplementary Table S1 Summary of final *P. lineatus* genome assembly

Assembly index	Value
Contig number	111
N50	36,883,020
N90	8,315,259
Longest contig length	91,334,651
Total	1,333,853,200

Supplementary Table S2 Summary of 24 pseudochromosomes in *P. lineatus* genome

Linkage groups	Length	Contig number
LG01	128,135,270	6
LG02	92,328,563	4
LG03	74,941,363	5
LG04	73,950,024	3
LG05	73,028,679	6
LG06	71,125,833	1
LG07	65,578,422	6
LG08	64,393,414	3
LG09	60,384,906	4
LG10	53,384,567	9
LG11	52,100,826	2
LG12	50,546,027	1
LG13	49,018,779	8
LG14	48,077,731	2
LG15	47,936,828	1
LG16	47,594,762	6
LG17	46,357,040	23
LG18	38,807,430	2
LG19	36,773,966	2
LG20	36,115,174	2
LG21	33,994,089	2
LG22	31,023,330	4
LG23	28,471,796	2
LG24	26,424,324	1
Total	1,330,493,143	105

Supplementary Table S3 BUSCO assessment of *P. lineatus* genome assembly

Type	Number	Percent(%)
Complete BUSCOs (C)	932	97.69
Complete and single-copy BUSCOs (S)	909	95.28
Complete and duplicated BUSCOs (D)	23	2.41
Fragmented BUSCOs (F)	11	1.15
Missing BUSCOs (M)	11	1.15
Total	954	100

Supplementary Table S4 BUSCO assessment of *P. lineatus* gene structure annotation

Type	Number	Percent(%)
Complete BUSCOs (C)	4,254	92.80
Complete and single-copy BUSCOs (S)	4,026	87.83
Complete and duplicated BUSCOs (D)	228	4.97
Fragmented BUSCOs (F)	134	2.92
Missing BUSCOs (M)	196	4.28
Total	4,584	100.00

Supplementary Table S5 Statistics of gene functional annotation

Type	Number	Percent(%)
Complete BUSCOs (C)	4,254	92.80
Complete and single-copy BUSCOs (S)	4,026	87.83
Complete and duplicated BUSCOs (D)	228	4.97
Fragmented BUSCOs (F)	134	2.92
Missing BUSCOs (M)	196	4.28
Total	4,584	100.00

Supplementary Table S6 Statistics of repeat element annotation

Class	Order	Super family	Number of elements	Length of sequence (bp)	Percentage of sequence (%)
Class I			1,590,662	369,377,499	27.69
	LINE		722,123	260,706,981	19.55
		L2	113,009	14,543,261	1.09
		Unknown	237,150	101,196,416	7.59
		L1	28,766	2,235,818	0.17
		RTE-BovB	183,239	106,665,935	8.00
		Rex-Babar	50,034	19,618,757	1.47
		Penelope	33,866	2,082,438	0.16
		L1-Tx1	14,218	2,303,091	0.17
		CRE	8,690	1,841,167	0.14
		CR1	11,175	1,517,045	0.11
		R2-Hero	5,603	3,190,948	0.24
		Other	36,373	5,512,105	0.41
	LTR		681,180	87,083,630	6.53
		ERV1	111,638	11,959,426	0.90
		Gypsy	128,823	19,870,449	1.49
		Unknown	151,073	25,408,540	1.90
		Ngaro	217,173	22,246,332	1.67
		Pao	23,614	4,145,511	0.31
		ERVK	34,057	1,979,040	0.15
		Other	14,802	1,474,332	0.11
	SINE		187,359	21,586,888	1.62
		tRNA-Core-RTE	39,169	6,576,533	0.49

		tRNA-V-RTE	26,022	3,512,466	0.26
		Unknown	23,302	1,920,394	0.14
		tRNA-V	78,449	8,438,436	0.63
		Other	20,417	1,139,059	0.09
Class II			3,398,873	465,801,625	34.92
	DNA		3,000,878	380,196,452	28.50
		Zisupton	96,382	8,173,467	0.61
		hAT-Tip100	77,208	6,093,231	0.46
		CMC-EnSpm	387,936	43,703,814	3.28
		TcMar-Tigger	480,738	80,098,091	6.00
		Ginger	54,158	3,314,830	0.25
		Unknown	669,125	98,565,048	7.39
		hAT-Ac	168,534	21,372,343	1.60
		PIF-Harbinger	48,904	2,336,425	0.18
		Kolobok-T2	34,856	1,645,200	0.12
		Maverick	68,086	2,843,623	0.21
		Crypton-V	35,204	1,391,936	0.10
		hAT-Charlie	124,832	23,440,649	1.76
		TcMar-Tc1	240,666	56,024,697	4.20
		P	42,462	2,395,966	0.18
		MULE-MuDR	51,040	3,904,970	0.29
		Sola-1	35,084	2,076,380	0.16
		IS3EU	39,570	3,796,079	0.28
		Other	346,093	19,019,703	1.43
	RC		73,648	3,553,055	0.27
		Helitron	73,648	3,553,055	0.27
	MITE		324,347	82,052,118	6.15

		Unknown	324,347	82,052,118	6.15
Total TEs			4,989,535	835,179,124	62.61
Tandem Repeats			205,231	9,285,759	0.70
	SSR		105,478	1,406,934	0.11
	Tandem repeat		99,753	7,878,825	0.59
Unknown			228,854	31,488,600	2.36
Simple repeats			22,217	3,402,620	0.26
Other			45,145	6,140,295	0.46
Low complexity			2,477	371,538	0.03
Total Repeats			5,493,459	885,867,936	66.41

Supplementary Table S7 Conserved regions and ancestral chromosomes of Siluriformes

Supplementary Table S8 Selected fusion and translocation sites (FTSs)

Supplementary Table S9 FTSs and corresponding chromosome transposon content (100 kb window)

Supplementary Table S10 Content of different transposon superfamilies in FTSs

Supplementary Table S11 Positively selected genes

Supplementary Table S12 Genes within expanded gene families

Supplementary Table S13 GO enrichment results of expanded gene families

Supplementary Table S14 KEGG enrichment results of expanded gene families

Supplementary Table S15 *Plotosus lineatus*, *Cranoglanis bouderius*, and *Hemibagrus wyckioides sox* gene family CDSs

Supplementary Table S16 Tissue-specific genes expressed in DO

Supplementary Table S17 GO enrichment analysis of DO tissue-specific expressed genes

Supplementary Table S18 KEGG enrichment analysis of DO tissue-specific expressed genes

Supplementary Table S19 Genes of transposons specifically expressed in DO tissue

Supplementary Table S20 *Plotosus lineatus* and *Ameiurus melas* annotated venom-related CDSs

Supplementary Table S21 KaKs calculator results of *P. lineatus* vs *A. melas*

Supplementary Table S22 Clean RNA-seq data of different tissues and mapping ratios

Supplementary Tables S7–S22 are listed as a separate excel file due to their large size.