Chen et al. *Zool. Res.* 2026, 47(3): 880−894
https://doi.org/10.24272/j.issn.2095-8137.2025.395
https://cstr.cn/32248.14.zr.2025.395

ZR | Zoological Research

Article Open Access

A near telomere-to-telomere genome assembly of the Amur stickleback (*Pungitius sinensis*) reveals structural variation and parallel evolution of sticklebacks

Bao-Hua Chen[1,2], Dong-Ya Wu[1], Dan Yu[1], Quan-Yu Chen[1], Wei Dai[3], Yan-Lin Lu[1], Dan-Yang Sun[1], Guo-Jie Zhang[1,*]

[1] *School of Basic Medical Sciences*, *Zhejiang University School of Medicine*, *Hangzhou* 310058, *China*

[2] *Center for Evolutionary & Organismal Biology*, *Liangzhu Laboratory*, *Zhejiang University Medical Center*, *Hangzhou* 311121, *China*

[3] *BGI Research*, *Wuhan*, 430074, *China*

ABSTRACT

Fish species in the genus *Pungitius* exhibit substantial diversity in morphology and behavior. These species have undergone parallel evolution with three-spined sticklebacks, producing convergent traits. *Pungitius* genomes have a complex evolutionary history with extensive chromosomal rearrangements and structural variations, which is divided into two main clades: the *pungitius* clade, predominantly found in North America and Europe, and the *sinensis* clade endemic to Northeast Asia. While a genome reference has been available for the *pungitius* clade (i.e. ninespine stickleback *Pungitius pungitius*), genomic data for the *sinensis* clade remains lacking, which limits the exploration of genetic changes underlying their morphological diversity, geographic and ecological adaptation, and parallel evolution with the three-spined stickleback (*Gasterosteus aculeatus*). Here, we present a high-quality genome assembly for Amur stickleback (*Pungitius sinensis*) from the *sinensis* clade, which represents the first near telomere-to-telomere stickleback genome assembly using PacBio high-fidelity, Oxford Nanopore Technology ultra-long, and Hi-C sequencing data. Comparative chromosome-scale analyses indicate adaptive introgression from *G. aculeatus* into *P. sinensis*, and reveal shared genomic regions with the three-spined stickleback associated with dorsal spine and pelvic morphology. This genome also provides novel insights into chromosome rearrangements among sticklebacks and reveals unique characteristics in the evolution of sex chromosome systems within Gasterosteidae. Together, these results highlight the parallel evolutionary mechanisms shaping the morphology and behavior of *Pungitius* and *Gasterosteus*, offering a valuable resource for advancing research in ecological adaptation and genome evolution.

Keywords: *Pungitius sinensis*; Near telomere-to-telomere genome assembly; Genome evolution; Stickleback

INTRODUCTION

Sticklebacks refer to species within the family Gasterosteidae and are widely distributed across freshwater and marine environments in North America, Europe, and Northeast Asia. These species have adapted to diverse ecological conditions, resulting in remarkable morphological and behavioral variation (Arai et al., 2020). Such diversity has established sticklebacks as a key model group for studying morphological variation and behavioral adaptation (Feng et al., 2024; Reid et al., 2021; Roberts Kingman et al., 2021). Recent studies have shown that genomic evolutionary events, such as chromosome fusions (Liu et al., 2022) and sex chromosome turnovers (Ross et al., 2009), have occurred frequently in this lineage, further highlighting its importance in genomic research. The stickleback family comprises five genera, with *Gasterosteus* being a well-established model in evolutionary and genetic research (Reid et al., 2021). While *Gasterosteus* has been extensively studied as a genetic model, the closely related *Pungitius* remains less explored despite its remarkable ecological and morphological diversity.

Pungitius is the largest genus within Gasterosteidae, accounting for approximately half of the species in the family. *Pungitius* and *Gasterosteus* may show convergent changes in morphological and behavioral traits in response to similar environmental pressures. For example, heightened predation risk may lead to increased dorsal spine numbers (Keivany, 1996), more well-developed pelvic structures (Herczeg et al., 2010), and shifts in foraging behaviours (Herczeg et al., 2009). Given this potential for intergeneric convergent evolution, it is also important to examine how diversification occurs within *Pungitius* itself. *Pungitius* can be further divided into two

Received: 13 November 2025; Accepted: 11 December 2025; Online: 12 December 2025

Foundation items: This work was supported by China National Postdoctoral Program for Innovative Talents (BX20220269 to D.W.).

*Corresponding author, E-mail: guojiezhang@zju.edu.cn

clades: the *pungitius* clade, represented by *Pungitius pungitius*, which is primarily found in North America and Europe; and the *sinensis* clade, represented by *Pungitius sinensis* and comprising species endemic to Northeast Asia (Wang et al., 2022). Although these sister taxa share considerable morphological overlaps, they can be distinguished by key diagnostic traits: the *sinensis* individuals typically possess fewer dorsal spines (7 to 8 vs. 9 to 11 in *pungitius*), an elongated caudal peduncle, and a complete series of bony lateral plates. These divergent morphological adaptations likely reflect ecological specialization in their respective habitats—the *sinensis* clade to thermally stable lowland rivers and the *pungitius* clade to colder, hydrologically variable systems (Herczeg et al., 2010). The notable diversity in distribution, morphology, and behavior of *Pungitius* make it a fascinating model for genetic research. However, despite their potential, the genomic underpinnings of adaptation and divergence in this genus remain poorly understood. In particular, genomic comparisons between the *pungitius* and *sinensis* clades are essential to elucidate how similar ecological pressures have shaped their distinct evolutionary trajectories.

Building on these insights, understanding whether similar adaptive phenotypes—both within *Pungitius* and between *Pungitius* and *Gasterosteus*—arise through shared or distinct molecular mechanisms requires high-quality genomic resources across these lineages. However, such comparative genomic analyses have been constrained by limited reference data. Long-read sequencing has recently delivered chromosome-level and telomere-to-telomere genomes for a growing range of teleosts (Hu et al., 2025; Wang et al., 2025), yet such high-quality references remain unevenly distributed across lineages.

Although the reference genome of *P. pungitius* has been established (Varadharajan et al., 2019), genomic information for the *sinensis* clade remains largely unavailable. As a result, studies in this clade remain dependent on divergent reference genomes (*P. pungitius* or *Gasterosteus aculeatus*) (Dixon et al., 2019; Mattern, 2004; McLennan & Mattern, 2001), potentially obscuring lineage-specific adaptations. Here, we report the first reference genome assembly of *P. sinensis* generated using PacBio high-fidelity (HiFi), Oxford Nanopore Technologies (ONT) ultra-long, and Hi-C sequencing data. The assembled genome has a total length of 482.99 Mb, comprising 21 chromosomes with only 12 gaps in total, and 34 telomeres identified at the chromosome ends, representing the first near telomere-to-telomere (T2T) assembly achieved in sticklebacks. This assembly offers exceptional completeness and continuity that provide critical insights into genome evolution and phenotypic diversification in this group. Comparative genomic analyses further illuminate the evolutionary history of the *Pungitius* genome, revealing structural and sequence changes potentially associated with increased numbers of dorsal spines and pelvic reduction. These findings underscore the evolutionary importance of this genome assembly and its broad potential for advancing research in evolutionary biology.

MATERIALS AND METHODS

Sample collection and nucleic acid extraction

A male *P.sinensis* individual was collected from a stream in the Yalu River Basin, China (41°47'12"N, 125°35'09"E), on 22 November 2024, for genome sequencing and assembly. The fish was euthanized by immersion in 100 mg/L MS-222 for 15 minutes before dissection. Immediately after euthanasia, muscle, gill, brain, liver, intestine, and skin tissues were collected. All animal procedures were performed in accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health, and were approved by the Committee on the Management and Ethics of Experimental Animals, Zhejiang University (approval No. ZJU20240342). All samples were flash-frozen in liquid nitrogen and stored at –80°C. High-molecular-weight (HMW) genomic DNA (gDNA) was extracted from muscle tissue using a modified phenol-chloroform method for PacBio HiFi and ultra-long Oxford Nanopore Technologies (ONT) sequencing. Total RNA was extracted from gill, liver, intestine, brain, and skin tissues using the TRIzol Reagent (Invitrogen, catalog no. 15596026, USA) following the manufacturer's instructions. Genomic DNA was removed using deoxyribonuclease I (Takara, catalog no. 2270A, Japan).

Library preparation and sequencing

For PacBio HiFi sequencing, HMW gDNA was fragmented into ~15 kb segments using a g-TUBE device (Covaris, USA). The SMRTbell HiFi library was prepared with the SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, USA), following the manufacturer's protocol. Subsequently, primers and Sequel II DNA polymerase were annealed and combined with the SMRTbell templates. The final libraries were sequenced on the PacBio Sequel II platform in Circular Consensus Sequencing (CCS) mode with a runtime of 30 hours (Wenger et al., 2019). Ultra-long ONT libraries were generated from size-selected DNA fragments (>100 kb) using the SageHLS HMW library system (Sage Science, USA) and the Ligation Sequencing Kit (SQK-LSK109; Oxford Nanopore Technologies, UK). Approximately 750 ng of each library was sequenced on a PromethION platform (Oxford Nanopore Technologies, UK). Hi-C libraries were prepared from cross-linked liver tissue. Chromatin was digested with DpnII, filled in with biotin-labeled nucleotides, and ligated to generate chimeric junctions. After reversal of cross-links and purification, the DNA was fragmented to 300–600 bp and sequenced on the MGI-2000 platform (MGI, China).

A total amount of 2 μg RNA was used as input material for RNA sequencingpreparations. Sequencing libraries were generated using the NEBNext® Ultra™ II RNA Library Prep Kit for Illumina (NEB, #E7770L, USA) following the manufacturer's recommendations and index codes were added to attribute sequences. The target products were retrieved and subjected to polymerase chain reaction (PCR) amplification, after which the library construction was completed. The clustering of the index-coded samples was performed on a cBot cluster generation system using the HiSeq PE Cluster Kit v4-cBot-HS (Illumina, USA) according to the manufacturer's instructions.

Chromosome assembly using Hi-C technology

PacBio HiFi reads, ultra-long ONT reads, and Hi-C data were processed with Hifiasm v.0.19.8-r603 (Cheng et al., 2024) to generate primary contig graphs under default settings. Scaffolding was performed using Juicer v.1.6 (Durand et al., 2016b) in combination with 3D-DNA v.201008 (Dudchenko et al., 2017). Initially, restriction enzyme cut sites were identified from the contig-level assembly. After quality trimming of the Hi-C reads with TrimGalore (powered by Cutadapt v.0.6.10) (Martin, 2011), the cleaned reads were

aligned to the contigs using Juicer (Durand et al., 2016b). The 3D-DNA pipeline was then applied to construct a preliminary chromosome-level assembly following standard protocols. Manual refinement, including boundary adjustments and scaffold corrections, was carried out using Juicebox v.2.20.00 (Durand et al., 2016a). To validate the quality and completeness of this assembly and assemblies of related species for comparison, the k-mers frequency of each assembly was computed by Meryl v.1.3. The genome assembly continuity was then measured at single-base resolution with GCI v.1.0 (Chen et al., 2024) using default parameters.

Transposable element annotation

Transposable elements (TEs) were identified using a combination of *de novo* and homology-based approaches. For homology-based annotation, RepeatMasker v.4.1.2-p1 and RepeatProteinMask v.4.09 were used to align the *P. sinensis* genome against the RepBase database (release 21.01) (Bao et al., 2015). EDTA v.2.2.0 (Ou et al., 2019) was then employed to construct a *de novo* TE library, with the *species* parameter set to "others" and the neutral mutation rate set to 2e-9. To eliminate gene-related sequences from the TE library, TE-free coding sequences (CDS) from 11 related species (Supplementary Table S1) were provided to EDTA. Annotation and soft masking of TEs was performed using RepeatMasker with the RMBLAST v.2.14.1+ search engine. Additionally, tandem repeats were detected using Tandem Repeat Finder v.4.09 (Benson, 1999).

Gene prediction and functional annotation

Gene structures were annotated using BRAKER3 v.3.0.8 (Gabriel et al., 2024), which integrates homology-based, transcriptome-based, and *de novo* approaches into a comprehensive workflow. For homology-based annotation, protein sequences from the same 11 related species (Supplementary Table S1) were aligned to the *P. pungitius* genome using DIAMOND v.2.1.9) (Buchfink et al., 2021). The RNA-seq reads were mapped to the newly assembled genome using STAR v.2.7.11a (Dobin et al., 2013) with default parameters. Transcripts were then assembled based on these alignments using StringTie2 v.2.2.1 (Kovaka et al., 2019) to provide additional hints for exon-intron boundaries and alternative splicing. GeneMark-ETP v.1.02 (Brůna et al., 2024) was subsequently trained with these RNA-seq data and protein hints to generate a preliminary gene set. AUGUSTUS v.3.5.0 (Stanke et al., 2006) was then trained on the gene set with high confidence, and generated its own predictions. Finally, TSEBRA v.1.1.2.5 (Gabriel et al., 2021) was employed to combine the outputs of both tools to refine the final gene annotations. Functional annotation was performed using eggNOG-mapper v.2.1.6 (Cantalapiedra et al., 2021) to identify their potential functions based on homology. In addition, genes without a credible annotation were searched against the NCBI non-redundant (NR) protein and Ensembl (release 110) (Harrison et al., 2024) databases using blastx v.2.15.0+ (Camacho et al., 2009) for complement.

Gene family clustering and phylogenetic analysis

We selected four closely related species and *Danio. rerio* (Supplementary Table S1) alongwith *P. sinensis* for gene family analysis. All protein-coding genes with canonical start codon (ATG) were subjected to all-versus-all sequence similarity searches using DIAMOND v.2.1.9 (Buchfink et al., 2021) with an e-value threshold of "1e-5" to identify homologous protein sequences. The resulting similarities were clustered into gene families using OrthoFinder v.2.5.5 with default parameters (Emms & Kelly, 2019). For each single-copy orthologous gene family, the coding sequences (CDSs) of the longest transcripts from each species were retrieved and aligned using MAFFT v.7.526 with default parameters (Nakamura et al., 2018). Within each CDS, four-fold degenerate (4D) sites were identified from alignable codons and concatenated in coding order to produce a 4D-only pseudo-gene sequence. These pseudo-genes within each orthologous group were then used to infer gene trees with IQ-TREE2 v.2.3.4 (Minh et al., 2020) (with parameters "-m MFP -B 1000 -T AUTO"), rooted by Danio rerio. Finally, all gene trees were concatenated and a majority-rule extended (MRE) consensus species tree was inferred using RAxML-NG v.1.2.2 (Kozlov et al., 2019) with parameter "--consense MRE", in which branch support values indicate the occurrence frequency of each clade among all gene trees.

Alignment of short reads and variant calling

Variant calling was performed following the guidance of GATK Best Practices (Koboldt, 2020). Genome-scale short reads were downloaded from NCBI SRA database and then aligned to reference genomes using BWA-MEM2 v.2.2.1 (Vasimuddin et al., 2019) with default parameters. The resulting SAM files were processed using Samtools v1.18-15-g9a59467 (Danecek et al., 2021) and Picard (embedded in GATK) in sequential steps to convert them into BAM format, fill in mate pair information, sort by coordinate, and mark PCR duplicates. The final BAM files generated through this workflow were used as input for GATK v.4.6.1 (https://gatk.broadinstitute.org). Related filtering parameters for GATK were set as "QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < −12.5 || ReadPosRankSum < −8.0". Whole-genome SNPs were further filtered (with parameters "-maf 0.2, -geno 0.1, -hwe 0.0001") using Plink2 v.2.0.0-a.6.4LM (Chang et al., 2015).

Gene expression quantification

Raw RNA-seq reads were trimmed and filtered using TrimGalore (powered by Cutadapt v.0.6.10 (Martin, 2011)). Quasi-mapping and gene expression quantification were then performed using Salmon v.1.10.3 (Patro et al., 2017) with parameters "-l A –validateMappings --seqBias –gcBias". Pearson correlation analysis was performed with the "pearsonr" function provided in the scipy package v.1.12.0 (Virtanen et al., 2020).

Identification of syntenic regions and genome rearrangements

Protein sequences from *G. aculeatus*, *Apeltes quadracus*, *P. pungitius*, and *P. sinensis* (Supplementary Table S1) were blasted against each other using DIAMOND v.2.1.9 (Buchfink et al., 2021) with an e-value threshold of 1e-5. Protein sequences from two homoeologous chromosomes were extracted to queried against each other, followed by *c*-score filtering (*c*-score ≥ 0.99). Synteny blocks containing at least five homologous genes were identified using JCVI v.1.4.22 (Tang et al., 2024). Additionally, intergenomic alignments were generated with MUMmer v.4.0.0beta2 (Marçais et al., 2018) and subsequently processed by SyRI v.1.7.0 (Goel et al., 2019) with default parameters to identify rearrangements, including insertions, deletions, inversions (<1 Mb), and translocations (>5000 bp). Finally, plotsr v.1.1.5 (Goel & Schneeberger, 2022) was used to visualize regions of conserved synteny and rearrangements.

Detection of introgression via *D*-statistics (ABBA–BABA)

Sequencing reads were aligned to the newly assembled *P. sinensis* reference genome using BWA-MEM2 v.2.2.1 (Vasimuddin et al., 2019), followed by coordinate sorting and duplicate marking. *D*-statistics were estimated using the "doAbbababa2" module of ANGSD v.0.940 (Korneliussen et al., 2014) under a four-taxon configuration with P1=*P. pungitius*, P2=*P. sinensis* (China and Japan populations analyzed separately), P3=*G. aculeatus*, and *Gasterosteus nipponicus* as the outgroup. Only reads with mapping quality≥ 30 and bases with Phred quality≥20 were retained, and standard ANGSD read-level filters were applied (with parameters "remove_bads, uniqueOnly, BAQ, indel-proximal mismatch down-weighting with C=50, and underFlowProtect"). Genomic intervals of interest and the remaining callable genome were processed in parallel using distinct region masks. Block-based standard errors and Z-scores were computed using a chromosome size file and summarized with "jackknife.R". Comparisons of blockwise *D* distributions across genomic partitions and between the two *P. sinensis* populations were performed using two-sided Mann-Whitney *U* tests implemented in Python (SciPy v.1.15.3) (Gommers et al., 2025).

Ancestral chromosomal reconstruction and evolutionary history inference

This study elucidates the chromosomal evolutionary history within Gasterosteales through integrative multi-omics approaches. Initially, orthologous gene clusters identified by OrthoFinder (as described in the "Gene family clustering and phylogenetic analysis" section, with species listed in Supplementary Table S1) and their chromosomal coordinates were processed in Agora v.0.9.1 (Muffato et al., 2023) under default parameters to establish cross-species homologous synteny blocks. Subsequently, the syntenic matrix and a preconstructed species phylogeny were jointly analyzed in Anges v.3.2.2 (Chauve et al., 2010), which iteratively performed maximum-likelihood-based ancestral chromosomal reconstruction at all phylogenetic nodes under default settings. Finally, through syntenic block tracing, both extant and reconstructed ancestral chromosomes were color-annotated according to their corresponding homologous blocks in the most recent common ancestor of Gasterosteales.

Identification and genome-wide scanning of tandem repeat motifs

Candidate regions were analyzed using the MEME program from the MEME Suite v.5.5.7 (Bailey et al., 2009) with default parameters, allowing the discovery of statistically significant tandem repeat motifs. The identified tandem repeat motifs were then scanned across the entire genome using FIMO v.5.5.7 (Grant et al., 2011) also from the MEME Suite. The genome-wide scanning was conducted using default parameters to locate significant occurrences of these motifs. Potential transcription factors (TFs) regulating the target genes were predicted by analysing transcription factor binding sites (TFBSs) within genomic regions spanning±2 000 bp around the transcription start site (TSS) of each gene. The hTFtarget webtool (http://bioinfo.life.hust.edu.cn/hTFtarget) (Zhang et al., 2020) were used for *de novo* motif scanning on these TSS-proximal sequences to identify high-confidence TF binding motifs.

Real-time quantitative polymerase chain reaction (qPCR)

Fifteen *P. pungitius* individuals were collected from a stream in the Yalu River Basin, China (41°47′12″N, 125°35′09″E) on 16 January 2025. Total RNA was extracted from pelvic armors, pelvic spines, and opercula using FastPure Cell/Tissue Total RNA Isolation Kit V2 (Vazyme, China) according to the manufacturer's instructions. The quantitative PCR experiments were conducted using ChamQ Universal SYBR MasterMix (Vazyme, China) according to the manufacturer's instructions. Primer sequences used in this study were listed in Supplementary Table S2. All reactions were conducted in triplicate across three independent experiments. Relative expression levels were determined using the $2^{-\Delta\Delta Ct}$ method.

RESULTS

A near T2T genome assembly of *Pungitius sinensis*

A total of 21.61 Gb PacBio HiFi reads, 10.67 Gb ONT ultra-long reads, and 99.15 Gb Hi-C sequencing data was generated to assemble the genome of *Pungitius sinensis*. The average read lengths for PacBio HiFi and ONT ultra-long sequencing data were 17.07 kb and 88.0 kb, respectively (Supplementary Table S3). The initial assembly spanning 482.99 Mb and 73 primary contigs was obtained using hifiasm (Cheng et al., 2024), with contig N50 and N90 of 21.94 Mb and 14.72 Mb, and L50 and L90 values of 10 and 20, respectively.

These contigs were further scaffolded into a chromosome-level assembly using a Hi-C contact map, resulting in 21 chromosomes with a total length of 479.68 Mb, accounting for 99.31% of the total contig length (Figure 1A). Remarkably, only 12 gaps remained within the chromosome sequences, and 11 chromosomes were completely gapless. Additionally, 34 telomeres were identified at chromosome ends, suggesting a near T2T quality level of completeness (Figure 1A). This genome assembly outperforms all previously released assemblies of stickleback genomes with various quality metrics (Figure 1A; Supplementary Table S4). Based on the "eukaryote_odb10" database, benchmarking universal single-copy ortholog (BUSCO) analysis revealed that 99.04% of 3 640 BUSCOs were complete (98.65% single-copy and 0.38% duplicated) in genome model, and 96.43% of the BUSCOs were complete (79.51% single-copy and 16.92% duplicated) in protein model (Supplementary Table S5). Recently, more advanced parameters have been employed to evaluate the quality and completeness of T2T and near-T2T level assemblies. We compared our assembly with reference genome assemblies of closely related species using the Genome Continuity Inspector (GCI) score. The results demonstrated that our assembly surpassed other assemblies in this metric (Figure 1B).

A total of 117.08 Mb repetitive sequences were identified, representing 24.41% of the genome assembly. DNA transposons were the most abundant type, accounting for 10.31% of the genome, followed by unknown type, LTR, and LINE repeat elements (Supplementary Table S6). After masking repetitive sequences, a total of 20 225 protein-coding genes were predicted, with an average transcript length of 10.508 kb. Functional annotation using eggnog-mapper provided functional information for 19 483 genes, while the remaining 742 genes were designated as novel.

Gene families were constructed using protein sequences from *P. sinensis* and six related species: *P. pungitius*, *Apeltes quadracus*, *Gasterosteus aculeatus*, *Epinephelus lanceolatus*, *Takifugu rubripes* and *Danio rerio*. A total of 23 912 gene families were identified, of which 12 147 were shared across

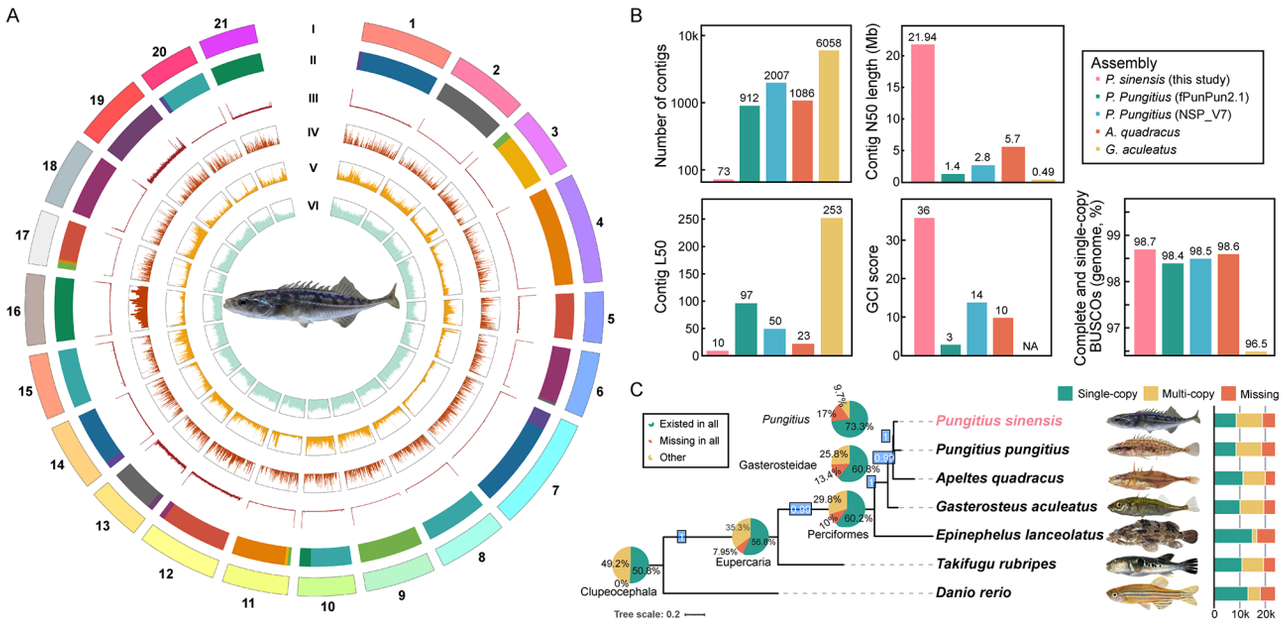

Figure 1 Basic overview of the near telomere-to-telomere genome assembly of *Pungitius sinensis*

A: Circos plot showing the basic overview of the *P. sinensis* genome assembly. Each concentric track from outside to inside represents the following: 21 chromosomes (I), contigs constituting the chromosomes (II), 34 out of 42 telomeres assembled (III), gene abundance (IV), transposable element abundance (V), and GC content (VI). All abundances are calculated using a 50 kb window with a 10 kb step size. B: Comparison of basic quality metrics among the new assembly of *P. sinensis* and other genome assemblies in Gasterosteidae. The quality metrics includes number of contigs, contig N50 length, contig L50, GCI score, and percent of complete and unduplicated benchmarking universal single-copy orthologs (BUSCOs). C: Phylogenetic tree constructed from single-copy homologous gene families. The pie charts on the branches show the presence of orthogroups within that branch, categorized as present in all species within the branch (green), absent (orange), and partially absent (yellow). Bootstrap support values from 1 000 iterations are represented by the numbers in light blue boxes adjacent to nodes. Presence of orthogroups in each species, as shown in the bar graph to the right of the phylogenetic tree, is categorized as single-copy (green), multi-copy (yellow), and missing (orange).

all species, 3 201 were specific to the Gasterosteidae family, and 4 054 were unique to the genus *Pungitius*. Among single-copy orthologous gene families, the shared families among all teleost species, Gasterosteidae, and *Pungitius* were 3 830, 10 740, and 14 157, respectively (Supplementary Figure S1). Phylogenetic analysis based on four-fold degenerate sites of 1 844 single-copy orthologous genes revealed the evolutionary relationship of the Gasterosteidae family, which is consistent with previous studies (Figure 1C) (Mattern, 2004; McLennan & Mattern, 2001).

Lineage-specific chromosome rearrangements between two *Pungitius* species

To investigate chromosomal conservation and structural variation, we conducted pairwise genome comparisons among *P. pungitius*, *P. sinensis*, and *G. aculeatus*. Alignable regions were classified into four categories: syntenic (SYN), inverted (INV), translocated (TRANS), and duplicated (DUP) (Figure 2). We identified approximately 18 Mb of genomic regions in *G. aculeatus* that are syntenically conserved across all three species. However, the total length of the syntenic regions unique to *P. sinensis* is approximately 1.5-fold greater than that of *P. pungitius* (94.40 Mb vs. 60.91 Mb; Figure 2A, B). These *P. sinensis*-unique syntenic regions (hereafter PS-uSYNs) were distributed on six chromosomes (Chr2, Chr4, Chr8, Chr13, Chr16, and Chr20; Figure 2C).

We further examined regions showing structural differences between *P. pungitius* and *P. sinensis* but synteny between one of the *Pungitius* species and *G. aculeatus*. The results revealed that the total length of all types of structural variation unique to *P. pungitius* is greater than that in *P. sinensis* (Figure 2B). Notably, long tandem-repeat-rich regions were frequently observed near the breakpoints flanking PS-uSYNs on Chr2 and Chr16 (Figure 2D, E). These lineage-specific structural differences reflect distinct evolutionary trajectories between the two *Pungitius* species and also have practical implications for comparative genomic analyses. In particular, large-scale rearrangements can influence read-mapping efficiency and variant detection accuracy across species (see Supplementary Materials), underscoring the importance of considering chromosomal context when interpreting interspecific genomic patterns.

Several exceptionally long PS-uSYNs were detected between *P. sinensis* and *G. aculeatus* but were absent in *P. pungitius* (Figure 2C). To examine whether cross-species introgression may have contributed to the allele sharing among these species, we conducted *D*-statistics (ABBA–BABA) analyses using *P. sinensis* (from China and Japan, P2), *P. pungitius* (P1), *G. aculeatus* (P3), and *Gasterosteus nipponicus* as the outgroup (Figure 2F; Supplementary Table S7). PS-uSYNs exhibited significantly higher *D*-statistics than the rest of the genome in both *P. sinensis* populations (Mann–Whitney *U* test, China: 0.2103 vs. 0.0044, *P*-value=3.3×10^{-20}, Japan: 0.2095 vs. 0.1198, *P*-value=4.9×10^{-22}; Figure 2G; Supplementary Table S8). In addition, the magnitude of *D*-statistics within PS-uSYNs did not differ significantly between the Japanese and Chinese *P. sinensis* populations (Mann–Whitney *U* test, *P*-value=0.78), whereas *D*-statistics in the remaining genomic regions were markedly higher in Japan than in China (Mann–Whitney *U* test, *P*-value=1.1×10^{-6}; Figure 2H; Supplementary Table S8).

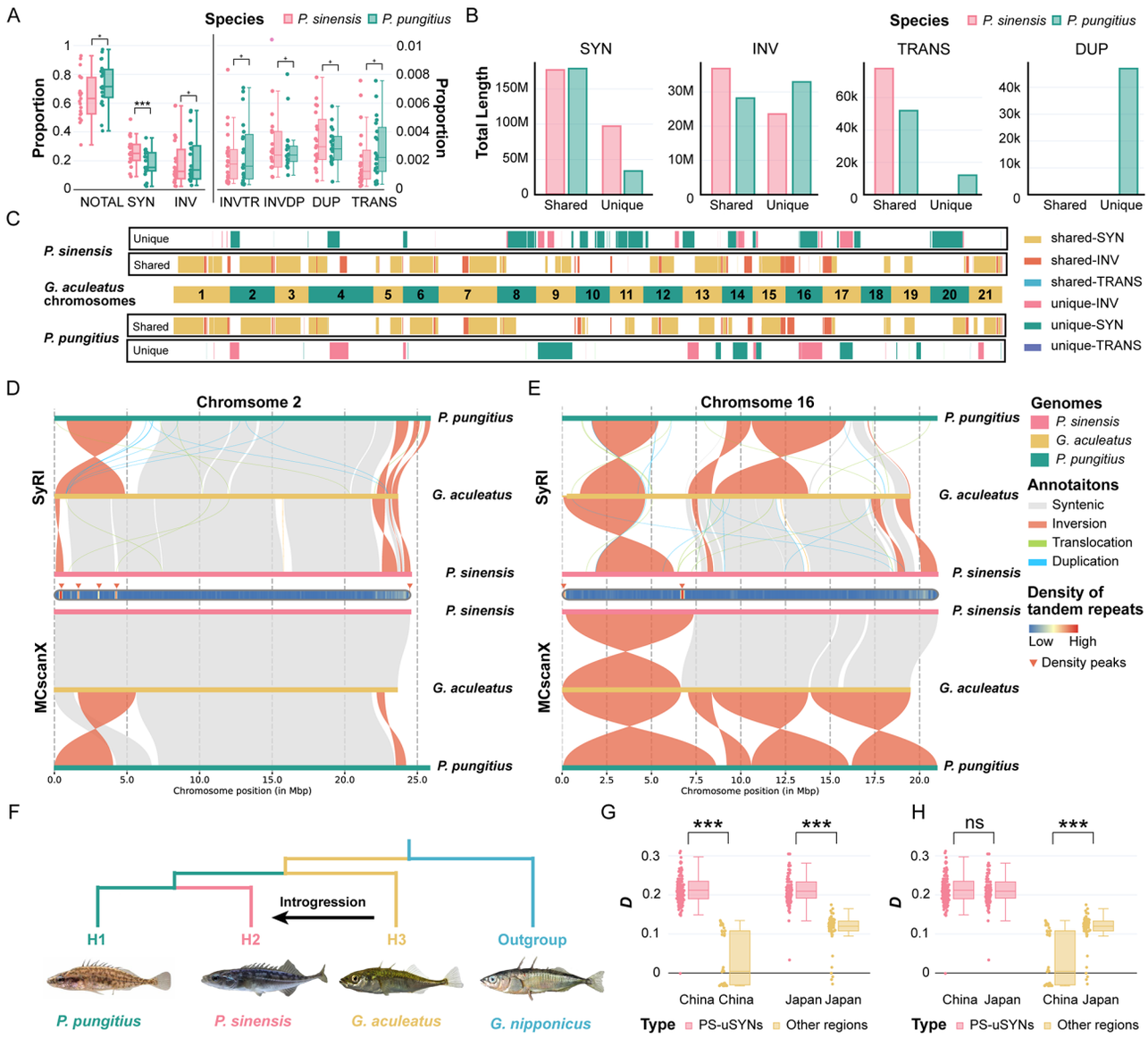

Figure 2 Lineage-specific chromosome rearrangements between two *Pungitius* species

A: Comparative genome synteny analysis of *P. sinensis* (pink) and *P. pungitius* (green) using *Gasterosteus aculeatus* as a reference. The boxplot compares the total length of different types of genome regions. Genome regions are categorized based on whether rearrangements are present between the *P. sinensis* and *P. pungitius* genomes. From left to right: inverted translocations (INVTR), inverted duplications (INVDP), duplications (DUP), translocations (TRANS), syntenic regions (SYN), not aligned regions (NOTAL), and inversions (INV). B: By comparing with the *G. aculeatus* genome, we calculated the summed lengths of syntenic regions, inversions, translocations, and duplications in the *P. sinensis* (pink) and *P. pungitius* (green) genomes, grouped by those present in both species (shared) and those present in only one species (unique). Abbreviations for genome region types follow those in panel (A). C: Presence of uniquely present and shared synteny regions and rearrangements in the *P. pungitius* and *P. sinensis* genomes. Abbreviations for genome region types follow those in panels (A and B). D, E: Detailed synteny analysis results for Chr2 (D) and Chr16 (E) across the three species. For clarity, only the four most abundant types shown here are: syntenic regions (gray), inversions (orange), translocations (light green), and duplications (blue). F: Schematic representation of the *D*-statistics (ABBA–BABA) test design using *P. pungitius* (H1), *P. sinensis* (China or Japan; H2), *G. aculeatus* (H3), and *Gasterosteus nipponicus* as the outgroup. G: Comparison of *D*-statistics between *P. sinensis*-specific syntenic regions (PS-uSYNs) and the rest of the genome. PS-uSYNs exhibited significantly higher *D*-statistics in both *P. sinensis* populations (Mann–Whitney *U* test, China: 0.2103 vs. 0.0044, *P*-value=3.3×10^{-20}, Japan: 0.2095 vs. 0.1198, *P*-value=4.9×10^{-22}). H: Comparison of *D*-statistics between Chinese and Japanese *P. sinensis* populations. Within PS-uSYNs, *D*-statistics did not differ significantly between the two populations, whereas in the remaining genomic regions (Mann–Whitney *U* test, *P*-value=0.78), *D*-statistics were markedly higher in Japan than in China (Mann–Whitney *U* test, *P*-value=1.1×10^{-6}). ns: not significant; *: $P < 0.05$; **: $P < 0.01$; ***: $P < 0.001$; ****: $P < 0.0001$.

Loss of a lncRNA may enhance the allele-specific expression of the extra-spine variant in *hoxd11b*

Compared to three-spined sticklebacks, *Pungitius* species exhibit a notable increase in dorsal spine numbers, ranging from eight to nine spines. Interestingly, some individuals with an extra dorsal spine (referred to as the “extra-spine” mutation), have also been observed in natural populations of three-spined sticklebacks. Wucherpfennig et al. (2022) mapped this trait to several single nucleotide variations (SNVs) within the *hoxdb* genes, including 10, 10, and 8 SNVs in *hoxd11b*, *hoxd9b*, and *hoxd4b* associated with extra-spine in *G. aculeatus*, respectively. These extra-spine-associated alleles exhibit significantly higher expression in the dorsal spines compared to the alternative alleles.

Among them, the most significant variation, located in the 3′ untranslated region 3'-UTR of *hoxd11b* (Figure 3A), showed that nearly all detectable RNA sequence reads from the first three dorsal spines were derived from the extra-spine *Gasterosteus* allele (Wucherpfennig et al., 2022). This SNP (chr6:17 756 571G>C) lies within a highly conserved region among stickleback species and exhibits pronounced allele-specific expression (ASE) in the first three spines of three-spined sticklebacks (mean ASE levels were 3.210, 4.881, 4.200, respectively, Figure 3E) with an extra spine. Whole-genome resequencing of *P. sinensis* and *P. pungitius* revealed that both species carry identical genotypes at this SNP site compared to extra-spine *G. aculeatus* individuals (Figure 3B; Supplementary Figures S2, S3).

We identified a lncRNA (ENSGACG00000024060) located between *hoxd9b* and *hoxd4b*, which is truncated in *P. pungitius* and absent in *P. sinensis* (Figure 3C). Our qRT-PCR experiments confirmed extremely low expression of this lncRNA in the dorsal spines of *P. sinensis* (Supplementary Figure S4). In *P. pungitius*, deletion of the first exon and its upstream regulatory elements likely disrupts its transcriptional activity (Supplementary Figure S5). Reanalysis of three-

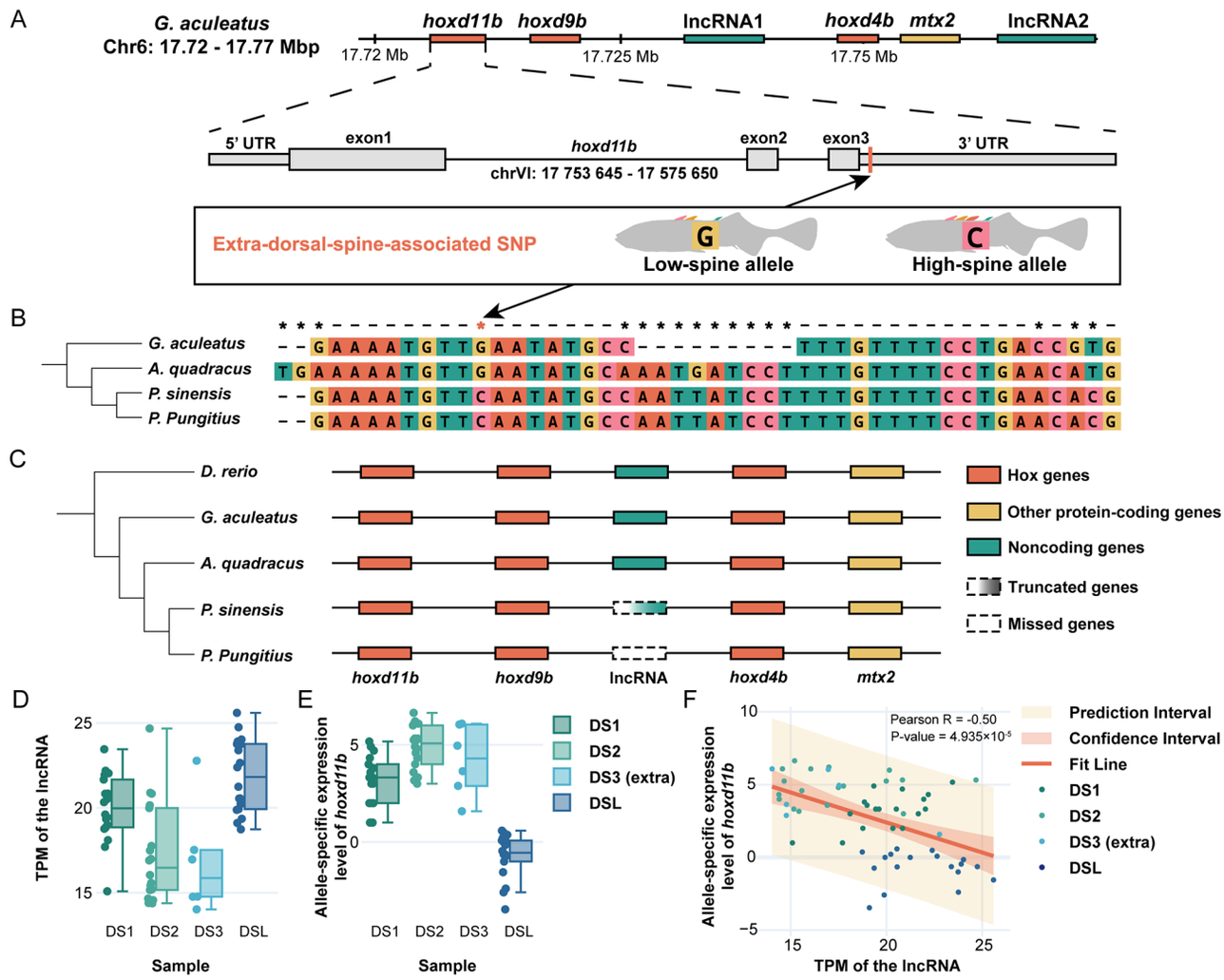

Figure 3 Loss of a lncRNA may enhance allele-specific expression of the extra-spine variant in *hoxd11b*

A: The phenotype of an additional dorsal spine in the three-spined stickleback (extra-spine mutation) is associated with the allele-specific expression of certain *hoxdb* genes. The most significant effect is observed at the 3' UTR of *hoxd11b*, where extra-spine three-spined sticklebacks have genotypes change from G to C. B: The genotypes of *P. pungitius* and *P. sinensis* at this SNP site are identical to those of the extra-spine three-spined stickleback. C: Gene synteny analysis reveals intact synteny of *hoxdb* genes across stickleback species and zebrafish. The colored blocks represent different types of genes: *hox* genes (orange), other protein-coding genes (red), and noncoding genes (green). Missing genes are shown as white blocks, while truncated genes are indicated by a white gradient. The results indicate that a lncRNA located between *hox9b* and *hoxd4b* is either lost or truncated in the genomes of *Pungitius* species. D: Expression levels of the aforementioned lncRNA and *hoxdb* genes across different dorsal spines, along with $\log_2$-fold changes between extra-spine and normal-spine alleles. Extra-spine three-spined sticklebacks possess four dorsal spines, labeled as DS1, DS2, DS3, and DSL, with DS3 corresponding to the additional spine in the low-spine variant. E: Mean allele-specific expression (ASE) levels of *hoxdb1b* in four spines, calculated as the $\log_2$-transformed ratio between counts of the two alleles. The mean ASE values were 3.210, 4.881, 4.200, and −0.696 for DS1, DS2, DS3, and DSL, respectively. F: Linear correlation between lncRNA expression and ASE of *hoxdb* genes across different dorsal spines. The lncRNA exhibits a moderate correlation with the allele-specific expression of *hoxd11b* (Pearson correlation test, R=−0.50, P-value=4.93×10^{-5}) and a weak correlation with *hoxd9b* (Pearson correlation test, R=−0.32, P-value=1.34×10^{-2}). The color of the points indicates the sample, while the orange solid line represents the fitted function. The shaded areas denote confidence intervals (orange) and prediction intervals (yellow). Allele-specific expression data for three-spined sticklebacks were retrieved from the source data provided by Wucherpfennig et al. (https://doi.org/10.6084/m9.Figureshare.20033063).

spined stickleback RNA-seq data shows that lncRNA expression decreases progressively from DS1 to DS3, but is anomalously high in posterior spines (DSL) (Figure 3D). Furthermore, lncRNA expression is negatively correlated with the ASE level of *hoxd11b* (Pearson correlation test, R=–0.50, P-value=4.93×10^{-5}; Figure 3F) and to a lesser extent, with *hoxd9b* (Pearson correlation test, R=–0.32, P-value=1.34×10^{-2}; Supplementary Figure S6). These correlations suggest a potential inhibitory role of this lncRNA on the two adjacent protein coding genes which might be associated with a high number of spines.

Chromosomal fusion events across Gasterosteidae genomes

Whole genome synteny analysis revealed that the two *Pungitius* species show no chromosomal fusions, breaks, or interchromosomal translocations, indicating highly conserved overall synteny (Figure 4A, B). However, several chromosome fusion events have occurred during the diversification of the Gasterosteidae. Based on chromosome-level genome assemblies of the four Gasterosteidae species, *Aulorhynchus flavidus* closely related to Gasterosteidae, and the distant outgroup *Epinephelus lanceolatus*, we reconstructed the ancestral chromosomes at multiple evolutionary nodes. The result indicates that the Gasterosteidae species underwent five fusion events (Fu1–Fu5; Figure 4A).

The ancestral Gasterosteales (Anc1) possessed 24 chromosomes. A fusion event (Fu1) reduced this to 23 chromosomes in the common ancestor of Gasterosteidae (Anc2). Subsequent lineage-specific rearrangements affected five chromosomes of Anc2 (Figure 4A). In *G. aculeatus* and both *Pungitius* species, Chr4 fused with Chr22 (Fu2 and Fu3, respectively), though the fusions occurred at opposite ends of Anc2 Chr 4. Besides, Chr23 fused with Chr7 to form a new Chr7 in *G. aculeatus* (Fu4), whereas in *Pungitius spp.*, it fused with Chr12 to form a new Chr12 (Fu5). These events were confirmed by genome synteny analysis (Figure 4B; Supplementary Figure S7).

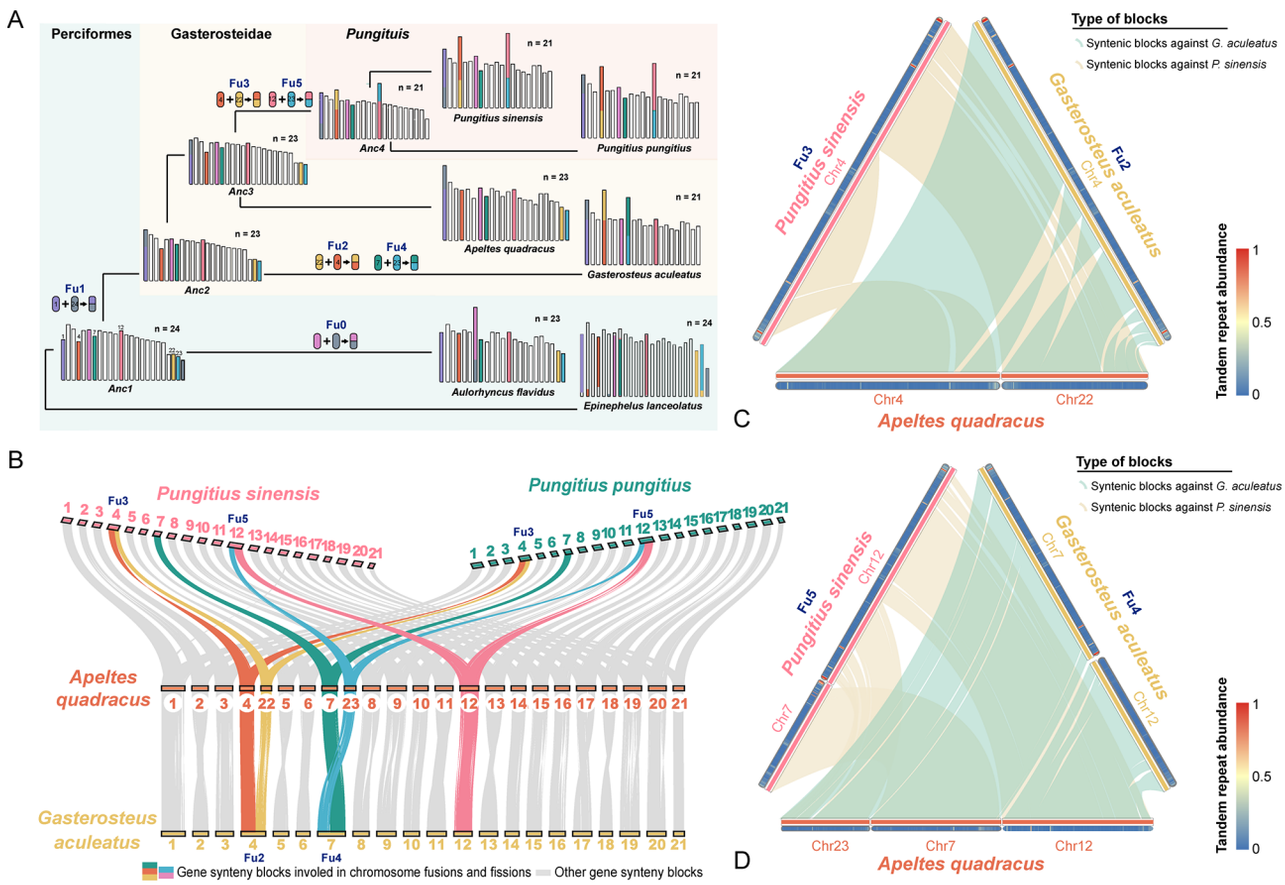

Figure 4 Chromosome fusion among Gasterosteidae species

A: Six existing chromosome-level genome assemblies were used to reconstruct the multi-tiered ancestral chromosomal evolution of the infraorder Gasterosteales, which includes species from the families Gasterosteidae and Aulorhynchidae, with *Epinephelus lanceolatus* as the outgroup. Chromosomes that have undergone fusion or fission events are highlighted in different colors, while those that remained structurally unchanged are shown in gray. The timing of fusion or fission events is annotated along the phylogenetic branches. Ancestral chromosomal levels are defined as Anc1 for ancestral chromosomes of Gasterosteales, Anc2 for Gasterosteidae, Anc3 for the common ancestor of *Pungitius* and *Apeltes*, and Anc4 for *Pungitius*. B: Chromosomal synteny analysis using *Apeltes quadracus* and *G. aculeatus* as references. Colored stripes represent synteny blocks on chromosomes involved in fusion events, with gray stripes indicating synteny on other chromosomes. C: Detailed synteny blocks among the chromosomes involved in Fu2 and Fu3 in the genomes of *Apeltes quadracus*, *P. sinensis*, and *G. aculeatus*. Synteny blocks associated with ancestral chromosomes in *A. quadracus* are shown in yellow, whereas those corresponding to homologous chromosomes in *G. aculeatus* are indicated in green. Heatmap bars adjacent to the chromosomes depict tandem repeat abundance on the fused chromosomes and their respective ancestral and homologous chromosomes. D: Detailed synteny blocks among the chromosomes involved in Fu4 and Fu5 in the genomes of *Apeltes quadracus*, *P. sinensis*, and *G. aculeatus*. Color codes are the same as in C. The abbreviation “Fu” followed by a number (Fu2–Fu5) designates distinct chromosome fusion events, which are labeled beside the corresponding chromosomes in B, C, and D.

Detailed comparisons of chromosomal structures among *A. quadracus*, *P. sinensis*, and *G. aculeatus* showed that the chromosomes involved in Fu2 and Fu3 exhibited higher synteny completeness compared to those involved in Fu4 and Fu5. Chr4 of *P. sinensis* and *G. aculeatus* (involved in Fu2 and Fu3) had only four and seven synteny breakpoints relative to their homologs in *A. quadracus*, whereas *G. aculeatus* Chr7 and Chr12 (involved in Fu4) and *P. sinensis* Chr12 (involved in Fu5) exhibited 12 and 15 breakpoints, respectively (Figure 4C, D).

All four fusion events identified in Gasterosteidae (Fu2–Fu5) are end-to-end fusions (EEF) rather than nested chromosome fusions (NCF; Figure 4B–D). In addition, tandem repeat-enriched regions were consistently observed near the synteny breakpoints in *G. aculeatus*, *P. sinensis*, and *A. quadracus* (Figure 4C, D), and their genomic positions remained conserved regardless of fusion status.

Structural variation at the *Gasterosteus aculeatus* chromosome 7 fusion boundary implicates *pitx* loss in pelvic reduction

Pelvic elements are frequently reduced in freshwater three-spined sticklebacks, and partial pelvic bone degeneration is also documented in the largely marine/brackish *A. quadracus* (Nelson, 1971; Shapiro et al., 2004). Similar reductions also occur in *P. pungitius* (Peichel & Marques, 2017), whereas *P. sinensis* has not been reported to lose the pelvis (Keivany, 1996). In three-spined sticklebacks, such reductions have been linked to loss of *pitx1* expression (Chan et al., 2010).

Motivated by these observations, we constructed a comparative dataset spanning pelvic- reduced and pelvic-complete sticklebacks using four genomes: a pelvic-reduced *G. aculeatus* from Paxton Lake (RefSeq ID: GCF_016920845.1) (Chan et al., 2010), a pelvic-complete *G. aculeatus* from Bamfield Inlet (GenBank ID: GCA_006232265.1) (Fraser & El-Sabaawi, 2022), a pelvic-reduced *A. quadracus* (GenBank ID: GCA_021346845.1) and a pelvic-complete *P. pungitius* (RefSeq ID: GCF_949316345.1). A structurally variable region at the end of Chr23 in *A. quadracus* was identified by comparative genomic analyses (Figure 5A). This region is involved in an end-to-end fusion with Chr7 in *G. aculeatus* and an inversion in *Pungitius* species (Figure 5A).

The *pitx1* locus lies within this structurally variable region (Figure 5B, C). In pelvic-reduced *G. aculeatus*, this region carries a deletion of 13 genes, including *pitx1*. In *A. quadracus*, a smaller deletion removes four genes, again including *pitx1*. By contrast, the pelvic-complete *G. aculeatus* genome retains more conserved synteny around *pitx1*, resembling the organization in pelvic-complete *P. pungitius* and *P. sinensis* (Figure 5C). Notably, *Pungitius* genomes harbor a lineage-specific gene in this interval, *pcdb2*, which is absent from other sticklebacks (Figure 5C). Consistent with the gene synteny at this locus, qRT-PCR revealed significantly higher expression of *pitx1* in the pelvic spines and armors of *P. sinensis* than in the gill covers and anal fins (Mann–Whitney *U* test, *P*-value=6.74×10^{-7}, and 6.40×10^{-15}, respectively; Figure 5D), and a similar pattern for *pcdb2* (Mann–Whitney *U* test, *P*-value=4.47×10^{-3}, and 5.56×10^{-4}, respectively; Figure 5D). Transcription-factor (TF) binding-site analysis indicated substantial regulatory overlap, with 57 TFs shared between *pitx1* and *pcdb2*—exceeding the TFs unique to each gene (27 and 20, respectively)—suggesting partially convergent inputs within this structural hotspot (Figure 5E).

Genomic rearrangements and repeat expansion in sex-linked chromosome of *Pungitius pungitius*

The sex chromosome of *P. pungitius* maps to Chr12, but its internal structure had been resolved only against the divergent *G. aculeatus* reference (Dixon et al., 2019). As sex chromosomes and sex-determining regions in teleosts are increasingly resolved by sex-balanced whole-genome resequencing (Li et al., 2024), we reanalyzed whole-genome sequencing data from 78 individuals of both sexes across three species (30 *P. pungitius*, 22 *P. sinensis*, and 26 *P. tymensis* individuals), with males and females represented in roughly equal proportions (Supplementary Table S7). Reads were independently aligned to both the *P. pungitius* and *P. sinensis* reference genomes.

Across both references, *P. pungitius* exhibited a non-recombining region (NRR) on Chr12 with a length of about 19 Mb (Figure 6A, E). In *P. sinensis* and *P. tymensis*, no sex-based differentiation was detected (Supplementary Figures S8, S9). Within *P. pungitius* NRR, six structural rearrangements were identified (Figure 6D; Supplementary Figure S10), including three inversions (0.23 Mb, 0.26 Mb, and 12.46 Mb) and two translocations (1.10 Mb and 2.45 Mb). In addition, a 2.21 Mb female-specific region (FSR) was detected on *P. pungitius* Chr12, positioned between the two translocations (Figure 6F), and is absent from *P. sinensis* Chr12 (Figure 6B). Repeat analysis showed that 1.24 Mb of the FSR comprises tandem repeats, here termed the *P. pungitius*–specific tandem-repeat-rich region (pTRR; Figure 6G). In the orthologous *P. sinensis* interval, two tandem-repeat-rich regions (sTRR1 ~120 kb and sTRR2 ~60 kb) were identified (Figure 6C). Motif analysis indicated homology between sTRR1 and pTRR (Figure 6H), and sTRR2 has no detectable counterpart in *P. pungitius* (Figure 6I).

Among the analyzed stickleback species, *P. pungitius* exhibits the most divergent gene organization within the FSR, despite its close phylogenetic relationship with *P. sinensis* (Figure 6J). This species shows four unique gene gains (with orthologs found only in distant taxa such as zebrafish), a tandemly duplicated gene, *gr1k1-like*, and loss of *tpx2* from the FSR with a relocated copy present at the distal end of Chr12 (31.93–31.94 Mb).

DISCUSSION

Adaptive introgression and evolutionary implications

The chromosomal evolution of Gasterosteidae species has long provided a model for exploring how genome architecture drives adaptive radiation and ecological diversification (Liu et al., 2022; Urton et al., 2011). Structural rearrangements can bring adaptive loci into tighter linkage and locally reduce recombination, thereby helping maintain co-adapted gene complexes (Liu et al., 2022; Roberts Kingman et al., 2021). In *Pungitius*, our comparisons support this view through the identification of large, conserved syntenic compartments unique to *P. sinensis* (PS-uSYNs, Figure 2A–C). These regions exhibit significantly elevated *D*-statistics when comparing *P. sinensis* with *G. aculeatus* (relative to the genome-wide background; Figure 2G), indicating an excess of shared derived alleles that is inconsistent with incomplete lineage sorting alone. The concentration of this signal with PS-uSYNs, coupled with its absence in other genomic regions, suggested introgression restricted to specific chromosomal regions (Durand et al., 2011; Martin et al., 2015). This pattern aligns with theoretical predictions that structural

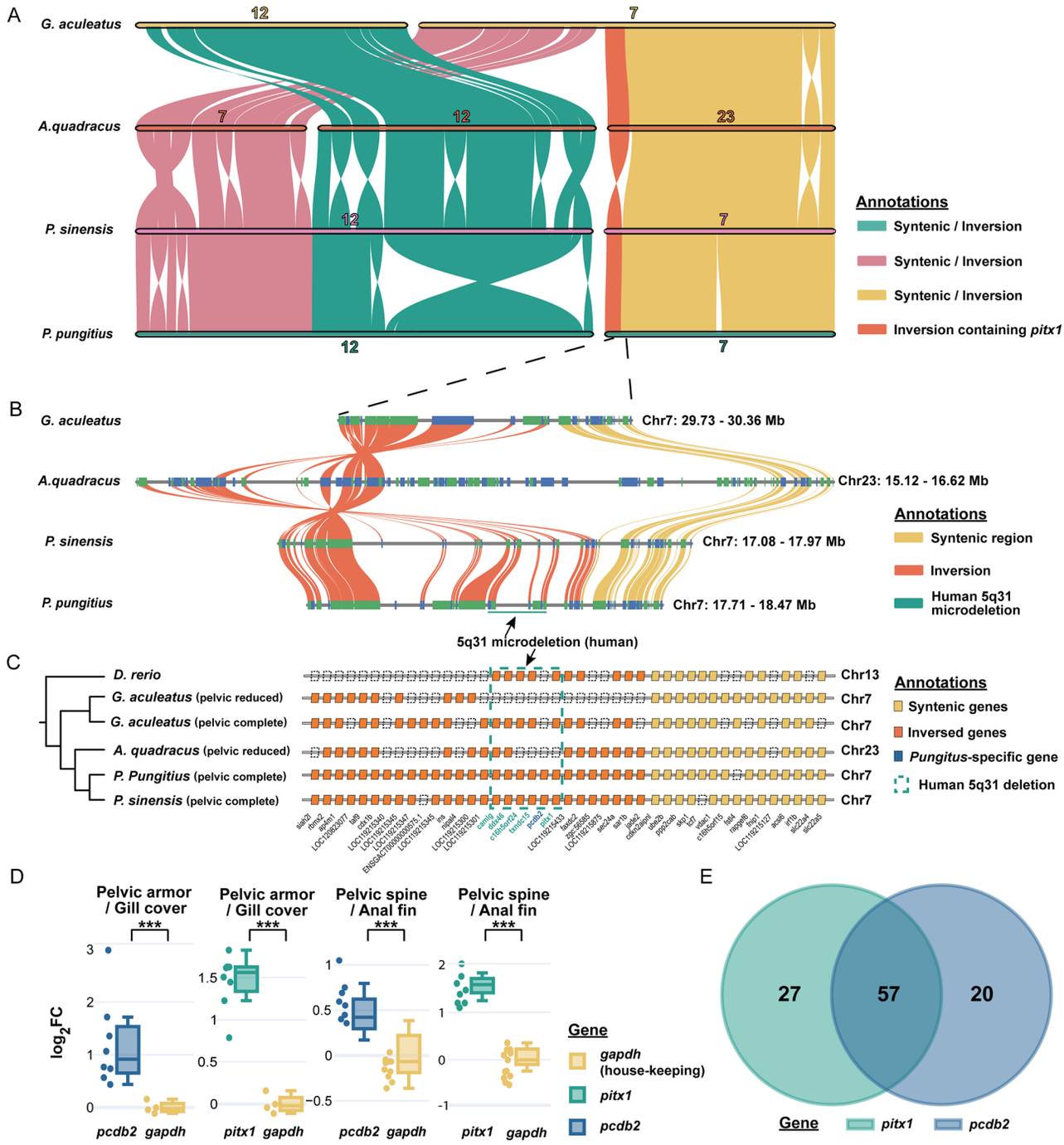

Figure 5 A deletion containing the *pitx1* gene is located at the chromosome fusion boundary on *G. aculeatus* Chromosome 7

A: At the boundary where *A. quadracus* Chr7 and Chr23 join to form *G. aculeatus* Chr7, a highly variable chromosomal region ranking approximately 1.8 Mb in length is observed. Syntenic regions and inversions shown in different chromosomes with different colors, with the inversion containing *pitx1* highlighted in orange. B: Detailed gene synteny illustrates the region deleted from *P. sinensis* and *P. pungitius* within this inversion. Blocks in syntenic regions are colored yellow, while blocks in the inversion are highlighted in orange. The green line at the bottom represents genes in the human and mouse 5q31 microdeletion, which is associated with lower limb deformities. C: Loss and presence of the deletion and nearby genes. Gene colors indicate their genomic location: yellow for syntenic regions, orange for inversions. The green dashed box and green gene names mark genes associated with the human 5q31 deletion. The blue gene names highlight *pcdb2*, a gene specific to *Pungitius* species. D: Expression level of *pcdb2* (blue), *pitx1* (green), and house-keeping gene *gapdh* (yellow) in the pelvic armor and pelvic spine of *P. sinensis* measured by qRT-PCR. The expression levels in pelvic armors and pelvic spines are normalized to thosein the gill cover and anal fins, respectively. E: Venn diagram highlights the shared and unique transcription factors predicted to regulate *pitx1* (green) and *pcdb2* (blue).

rearrangements, particularly those flanked by repeat-rich breakpoints, can reduce effective recombination and thereby slow haplotype erosion, promoting the co-inheritance of introgressed variants.

Importantly, PS-uSYNs showed similar *D*-statistics between Japanese and Chinese *P. sinensis* populations, whereas other genomic regions exhibited higher values in Japan (Figure 2H; Supplementary Table S8), suggesting that introgressed variation within these conserved compartments has been stably maintained across populations. This stability contrasts

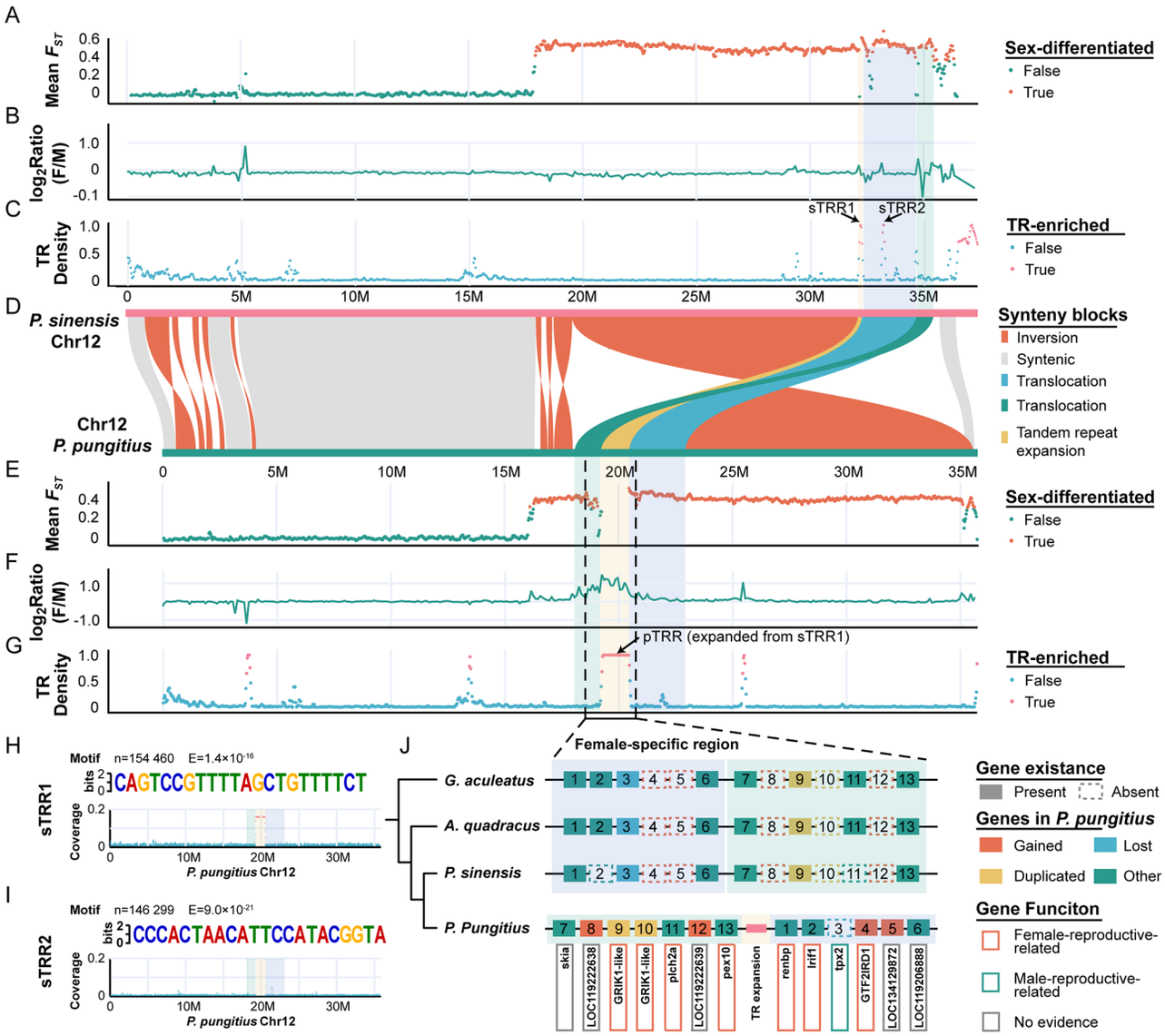

Figure 6 Genomic evidence of X chromosome evolution in *Pungitius* species

Whole-genome sequencing data from female and male *P. pungitius* individuals were mapped to the genomes of *P. sinensis* and *P. pungitius*, respectively. A,E: The fixation index (F_{ST}) scan identifies a non-recombining region (NRR) of approximately 19 Mb on Chr12 in both *P. sinensis* (A) and *P. pungitius* (E). Differentiated regions (F_{ST}>0.5) are highlighted in orange, while other regions are shown in green. B, F: Log_2-transformed fold change of mapping ratios (female/male) reveals a female-specific region (FSR, outlined with dashed lines) in *P. pungitius* (F), which is absent in *P. sinensis* (B). C,G: Tandem repeat density scans identify a large tandem-repeat-rich region (pTRR) within the *P. pungitius* FSR (G), whereas two smaller tandem-repeat-rich regions (sTRR1 and sTRR2) are observed in the corresponding region of *P. sinensis* (C). Tandem-repeat-rich regions are highlighted in pink, while non-rich regions are shown in blue. D: Sequence synteny analysis reveals that the NRR contains a large inversion (orange), two major translocations (blue and green), a tandem-repeat-enriched region (yellow), and a smaller syntenic region (gray). H, I: These TRRs exhibit significantly distinct motifs. The motif of sTRR1 is significantly enriched in pTRR (H), whereas the motif of sTRR2 is absent on *P. pungitius* Chr12 (I). J: Synteny of genes involved in the *P. pungitius* FSR in the genomes of stickleback species. Solid squares indicate present genes, while dashed squares indicate missing genes. The pink solid box indicates pTRR. Color of genes indicates that they are involved in different types of *P. pungitius*-specific events, including gaining (orange), loss (blue), duplications (yellow), and not involved (green). Function associations are indicated alongside gene names: female-reproductive-related genes (orange), male-reproductive-related genes (green), and genes with unknown functions (gray). In panels A, B, C, E, F, G, H, I, and J, green and blue shading corresponds to translocations in D, while yellow shading highlights tandem-repeat-rich regions.

with the more variable signals elsewhere in the genome, which likely reflect recent or geographically restricted gene flow (Nelson & Cresko, 2018; Roesti et al., 2013). Similar couplings between introgressive gene flow and chromosome-level modifications have been reported in other stickleback systems (Nedoluzhko et al., 2022, 2025), reinforcing the notion that chromosome architecture both reflects the history of gene flow and constrains which introgressed segments persist to contribute to local adaptation.

A potential *cis*-regulatory mechanism linking lncRNA loss to dorsal spine number variation

Building on the identification of a segregating SNP in the 3′ UTR of *hoxd11b* associated with dorsal spine number in sticklebacks (Wucherpfennig et al., 2022), we observe that both *P. sinensis* and *P. pungitius* carry the same high-spine

allele found in *G. aculeatus*. Because 3′ UTR variants can modulate allele-specific expression (ASE) via effects on mRNA stability or miRNA binding (Griesemer et al., 2021; Mayr, 2017), this parallelism strengthens the view that this variant contributes to spine diversification.

Our comparative analysis of the *hoxdb* locus across Gasterosteidae reveals a potential *cis*-regulatory mechanism linking dorsal spine number variation to the expression dynamics of *hoxd11b*. Previous work identified an axial enhancer (AxE) within this interval that modulates *hoxdb* gene expression and influences spine number in three-spined sticklebacks (Wucherpfennig et al., 2022), though quantitative trait locus mapping indicates that AxE accounts for only ~23% of phenotypic variance in *A. quadracus* (Herbert et al., 2024). This incomplete explanation suggests that additional regulatory elements contribute to *hoxdb* regulation. Despite the conserved synteny of the *hoxdb* locus (Figure 3C) (Hoegg et al., 2007; Wucherpfennig et al., 2022), we identified a lncRNA within the *hoxdb* locus that has lost its promoter and first exon in *P. pungitius* and is entirely absent in *P. sinensis*, which both have higher dorsal spine numbers than *G. aculeatus* (Figure 3C; Supplementary Figure S5). Re-analysis of *G. aculeatus* RNA-seq data showed that this lncRNA decreases from the first to third dorsal spines but remains unusually high in the last spine, and its expression is negatively correlated with *hoxd11b* allele-specific expression (ASE) as defined by a segregating 3′-UTR SNP (Figure 3D–F). Similar *cis*-regulatory effects of lncRNAs on the ASE of nearby genes have been described previously. For instance, the lncRNA *ASAR6* acts in *cis* to maintain proper replication timing and monoallelic expression of adjacent loci (Heskett et al., 2020), whereas deletion of another lncRNA, *Bendr*, disrupts the allele-specific transcriptional balance of a neighboring gene in mouse embryonic stem cells (Engreitz et al., 2016). While our findings are correlative and require experimental validation, they point to a multifactorial *cis*-regulatory architecture at the *hoxdb* locus. We propose that enhancer activity, lncRNA transcription, and 3'-UTR variation may collectively contribute contribute to the regulation of dorsal spine number in a lineage-specific manner..

Mechanisms and evolutionary implications of chromosomal fusions

The reconstruction of Gasterosteidae chromosomal evolution trajectory indicates that four end-to-end fusion (EEF) events (Fu2–Fu5) have shaped the lineage, leading to distinct karyotypes among modern sticklebacks (Figure 4). The conserved synteny and low number of breakpoints in fusions Fu2 and Fu3 (Figure 4C) suggest that these events represent ancient, structurally stable rearrangements. In contrast, the more complex rearrangements in Fu4 and Fu5 (Figure 4D) may indicate more recent or less stabilized fusions, possibly reflecting ongoing genomic flexibility in certain lineages. Notably, while EEF-type fusions generally retain ancestral telomeric sequences at the breakpoints as observed in human chromosome 2 (Hillier et al., 2005), no such traces were detected in the *P. sinensis* genome despite its near T2T assembly quality (Figure 1A).This absence suggests that either the species has evolved a highly efficient telomere-clearance mechanism, or that telomere loss itself may act as a driver for chromosomal end-to-end fusion, as previously observed in mammalian systems when telomere protection fails (Capper et al., 2007). Clarifying this causal relationship will require high-contiguity telomere-to-telomere (T2T) assemblies and broader comparative analyses across closely related taxa.

The presence of conserved tandem repeat-enriched regions near synteny breakpoints across species, regardless of fusion status (Figure 4C, D), implies a potential role of centromere-like elements in facilitating chromosomal stability post-fusion. Such structural features may act as organizing centers that maintain proper segregation and integrity of fused chromosomes. Interestingly, similar patterns of frequent end-to-end fusions have been reported in invertebrates with holocentric chromosomes, such as butterflies and *Rhynchospora* species (Cicconardi et al., 2021; Hill et al., 2019; Hofstatter et al., 2022; Mandrioli & Manicardi, 2020). Although vertebrates are typically monocentric, the observed stability of fusion events in Gasterosteidae may represent a parallel trend toward increased chromosomal flexibility, potentially contributing to the group's remarkable adaptive radiation.

Possible regulatory interactions between *pcdb2* and *pitx1* in pelvic structure maintenance

Across sticklebacks, pelvic morphology covaries with the integrity of the Chr7 boundary harboring *pitx1*, where lineages with deletions in this region tend to lack pelvic elements, whereas those retaining conserved synteny maintain them. The *pitx1* locus resides within a structurally dynamic genomic interval that has undergone lineage-specific rearrangements among sticklebacks. In pelvic-reduced species such as *G. aculeatus* and *A. quadracus*, deletions encompassing *pitx1* are observed, whereas pelvic-complete lineages retain conserved synteny. In *Pungitius*, this region additionally harbors a lineage-specific gene, pterin-4-alpha-carbinolamine dehydratase 2-like (*pcdb2*), absent from other sticklebacks. The proximity of *pcdb2* to *pitx1*, together with their coordinated expression and extensive overlap in predicted transcription-factor binding profiles, points to potential regulatory coupling between the two loci (Figure 5C–E). Although our data do not establish a mechanism, the co-expression of *pcdb2* and *pitx1* and their shared TF inputs, nominate a candidate axis for sustaining pelvic cell proliferation and maintenance in *P. sinensis*. While *pitx1* is often characterized as a developmental regulator, several studies demonstrate its expression and functional importance in adult skeletal tissues. For example, Picard et al. (2013) reported high *PITX1* expression in adult human knee cartilage, with reduced expression being linked to osteoarthritis. Moreover, mouse models with bone-specific *Pitx1* overexpression and knockout indicate that *Pitx1* modulates bone and cartilage integrity in adult joints (Karam et al., 2019; Khalid et al., 2022). The co-expression pattern and shared regulatory inputs nominate a candidate axis for sustaining pelvic cell proliferation and maintenance in *P. sinensis*, though determining whether *pcdb2* contributes to pelvic development and how it interacts with *pitx1* will require targeted analyses across developmental stages.

Hybrid origin and progressive remodeling of ChrX under sex-biased regulation

Our species-wide, sex-balanced resequencing analysis supports an interspecific origin for the *P. pungitius* sex chromosome system and refines the underlying structure beyond earlier inferences based on *G. aculeatus* alignments (Dixon et al., 2019). We identified a large X-linked NRR on *P. pungitius* Chr12 (ChrX) (Figure 6A, E) that resolves not as a single inversion but as a mosaic of multiple inversions and

translocations, consistent with stepwise remodeling of the ancestral architecture (Figure 6D). Within this remodeled interval, we also observe a female-specific region (FSR) accompanied by a marked expansion of tandem repeats (Figure 6F–I), indicating local sequence accretion concurrent with structural change. Large inversions on homologous chromosomes may promote repeat expansion during meiosis when inversion loops or anti-parallel pairing trigger DNA repair through non-homologous end joining, microhomology-mediated end joining, or break-induced replication (Carbone et al., 2014; Hastings et al., 2009; Termolino et al., 2019).

Within the remodeled interval, most genes affected by gene-content changes are LOC-annotated and uncharacterized, while two cases are particularly notable for sex-biased biology (Figure 6J). Notably, the transcription factor, *gtf2ird1*, has been recruited into this interval. It is enriched in unfertilized oocytes (Morel et al., 2007; Yi et al., 2024), implicated in estrogen-responsive transcription and cell proliferation (Ogura et al., 2006; Tammimies et al., 2012) and its loss of expression has been linked to defects in maternal behaviors (Nygaard, 2022). Conversely, *tpx2*, a key regulator of sperm spindle assembly (Casanova et al., 2008) has been deleted from the FSR toward the distal end of ChrX. The recruitment of an oocyte-biased regulator and the displacement of a male-germline-associated factor from the X-linked NRR are consistent with more efficient sex-specific regulation. These patterns align with theoretical expectations for early sex-chromosome differentiation, where structural rearrangements facilitate the accumulation of sex-beneficial alleles and the resolution of sexually antagonistic selection (Meisel, 2022; Vicoso, 2019).

Taken together, these findings indicate that the *P. pungitius* ChrX have undergone progressive and dynamic remodeling since their hybrid origin, coupling structural innovation with the refinement of sex-biased regulatory networks that characterize the early stages of sex-chromosome differentiation.

CONCLUSION

In this study, we present a high-quality, near T2T genome assembly for *Pungitius sinensis*, addressing a significant gap in genomic resources for species within the *sinensis* clade of *Pungitius* genus. Employing advanced sequencing technologies, including PacBio HiFi, ultra-long ONT, and Hi-C, we have developed a comprehensive reference genome that offers critical insights into the evolutionary biology of sticklebacks. These insights include adaptive introgression from *Gasterosteus aculeatus* into restricted genomic regions, the genetic basis of spine patterning, independent chromosomal fusion events in *Pungitius* and *Gasterosteus*, and the structural variation underlying phenotypic adaptations such as pelvic reduction. Additionally, we identified a more complex structure of the non-recombining region NRR on chromosome 12, involving large inversions and translocations, thereby enhancing our understanding of sex chromosome evolution. These findings establish *P. sinensis* as a valuable model for investigating parallel evolution, ecological adaptation, and genomic structural variation. This contribution advances the field of fish genomics by providing a valuable resource for unraveling the complexities of adaptation and speciation in *Pungitius*. It also underscores the potential of modern sequencing technologies to overcome limitations in the study of non-model organisms, paving the way for genomic exploration of other understudied species.

DATA AVAILABILITY

The genome assembly and annotation for *P. sinensis* have been deposited in the Genome Warehouse (GWH) at the National Genomics Data Center (NGDC) under accession number GWHFSNL00000000.1, and GenBank at the National Center for Biotechnology Information (NCBI) with accession number JBNCOF000000000. Raw sequencing data have been uploaded to the Genome Sequence Archive (GSA) at NGDC, with accession number CRA024527. All datasets are additionally available at Science Data Bank at https://doi.org/10.57760/sciencedb.39068.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

G.Z. designed the study and supervised all analyses. D.Y., D.S., and Y.L. collected all biological samples. D.Y., D.S., Y.L., and W.D. performed nucleic acid extraction. W.D. conducted the RNA-seq experiment. D.W. assembled the primary contigs. B.C. annotated the assembly and performed further comparative and evolutionary genomic analyses with important suggestions from D.W. D.Y. performed qRT-PCR experiments. D.Y. and B.C. performed the statistical analysis of experiments. B.C., G.Z., and D.W. wrote the manuscript. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We thank Guangji Chen (Zhejiang University) for valuable discussions on the comparative genomic analyses. We also gratefully acknowledge the support of the K. C. Wong Education Foundation. We also thank Hangjun Wu at the Center of Cryo-electron Microscopy (CCEM), Zhejiang University, for his technical assistance on computational resources.

REFERENCES

Arai T, Ueno D, Kitamura T, et al. 2020. Habitat preference and diverse migration in threespine sticklebacks, *Gasterosteus aculeatus* and *G. nipponicus*. *Scientific Reports*, **10**(1): 14311.

Bailey TL, Boden M, Buske FA, et al. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*, **37**(S2): W202–W208.

Bao WD, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, **6**(1): 11.

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

Brůna T, Lomsadze A, Borodovsky M. 2024. GeneMark-ETP significantly improves the accuracy of automatic annotation of large eukaryotic genomes. *Genome Research*, **34**(5): 757–768.

Buchfink B, Reuter K, Drost HG. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods*, **18**(4): 366–368.

Camacho C, Coulouris G, Avagyan V, et al. 2009. BLAST+: architecture and applications. *BMC Bioinformatics*, **10**(1): 421.

Cantalapiedra CP, Hernández-Plaza A, Letunic I, et al. 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Molecular Biology and Evolution*, **38**(12): 5825–5829.

Capper R, Britt-Compton B, Tankimanova M, et al. 2007. The nature of telomere fusion and a definition of the critical telomere length in human cells. *Genes & Development*, **21**(19): 2495–2508.

Carbone L, Alan Harris R, Gnerre S, et al. 2014. Gibbon genome and the fast karyotype evolution of small apes. *Nature*, **513**(7517): 195–201.

Casanova CM, Rybina S, Yokoyama H, et al. 2008. Hepatoma up-regulated protein is required for chromatin-induced microtubule assembly independently of TPX2. *Molecular Biology of the Cell*, **19**(11): 4900–4908.

Chan YF, Marks ME, Jones FC, et al. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science*,

327(5963): 302–305.
Chang CC, Chow CC, Tellier LC, et al. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, **4**(1): 7.
Chauve C, Gavranovic H, Ouangraoua A, et al. 2010. Yeast ancestral genome reconstructions: the possibilities of computational methods II. *Journal of Computational Biology*, **17**(9): 1097–1112.
Chen QY, Yang CT, Zhang GJ, et al. 2024. GCI: a continuity inspector for complete genome assembly. *Bioinformatics*, **40**(11): btae633.
Cheng HY, Asri M, Lucas J, et al. 2024. Scalable telomere-to-telomere assembly for diploid and polyploid genomes with double graph. *Nature Methods*, **21**(6): 967–970.
Cicconardi F, Lewis JJ, Martin SH, et al. 2021. Chromosome fusion affects genetic diversity and evolutionary turnover of functional loci but consistently depends on chromosome size. *Molecular Biology and Evolution*, **38**(10): 4449–4462.
Danecek P, Bonfield JK, Liddle J, et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience*, **10**(2): giab008.
Dixon G, Kitano J, Kirkpatrick M. 2019. The origin of a new sex chromosome by introgression between two stickleback fishes. *Molecular Biology and Evolution*, **36**(1): 28–38.
Dobin A, Davis CA, Schlesinger F, et al. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**(1): 15–21.
Dudchenko O, Batra SS, Omer AD, et al. 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*, **356**(6333): 92–95.
Durand EY, Patterson N, Reich D, et al. 2011. Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, **28**(8): 2239–2252.
Durand NC, Robinson JT, Shamim MS, et al. 2016a. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Systems*, **3**(1): 99–101.
Durand NC, Shamim MS, Machol I, et al. 2016b. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Systems*, **3**(1): 95–98.
Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, **20**(1): 238.
Engreitz JM, Haines JE, Perez EM, et al. 2016. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature*, **539**(7629): 452–455.
Feng XY, Merilä J, Löytynoja A. 2024. Secondary contact, introgressive hybridization, and genome stabilization in sticklebacks. *Molecular Biology and Evolution*, **41**(2): msae031.
Fraser AL, El-Sabaawi R. 2022. Characterizing phenotypic diversity in marine populations of the threespine stickleback. *Scientific Reports*, **12**(1): 17923.
Gabriel L, Brůna T, Hoff KJ, et al. 2024. BRAKER3: fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Research*, **34**(5): 769–777.
Gabriel L, Hoff KJ, Brůna T, et al. 2021. TSEBRA: transcript selector for BRAKER. *BMC Bioinformatics*, **22**(1): 566.
Goel M, Schneeberger K. 2022. plotsr: visualizing structural similarities and rearrangements between multiple genomes. *Bioinformatics*, **38**(10): 2922–2926.
Goel M, Sun HQ, Jiao WB, et al. 2019. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biology*, **20**(1): 277.
Gommers R, Virtanen P, Haberland M, et al. 2025. scipy/scipy: SciPy 1.15.0. *Zenodo*. doi: 10.5281/zenodo.14593523.
Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics*, **27**(7): 1017–1018.
Griesemer D, Xue JR, Reilly SK, et al. 2021. Genome-wide functional screen of 3′UTR variants uncovers causal variants for human disease and evolution. *Cell*, **184**(20): 5247–5260. e19.
Harrison PW, Amode MR, Austine-Orimoloye O, et al. 2024. Ensembl 2024. *Nucleic Acids Research*, **52**(D1): D891–D899.
Hastings PJ, Ira G, Lupski JR. 2009. A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genetics*, **5**(1): e1000327.
Herbert AL, Lee D, McCoy MJ, et al. 2024. Genetic mechanisms of axial patterning in *Apeltes quadracus*. *Evolution Letters*, **8**(6): 893–901.
Herczeg G, Gonda A, Merilä J. 2009. Evolution of gigantism in nine-spined sticklebacks. *Evolution*, **63**(12): 3190–3200.
Herczeg G, Turtiainen M, Merilä J. 2010. Morphological divergence of North-European nine-spined sticklebacks (*Pungitius pungitius*): signatures of parallel evolution. *Biological Journal of the Linnean Society*, **101**(2): 403–416.
Heskett MB, Smith LG, Spellman P, et al. 2020. Reciprocal monoallelic expression of ASAR lncRNA genes controls replication timing of human chromosome 6. *RNA*, **26**(6): 724–738.
Hill J, Rastas P, Hornett EA, et al. 2019. Unprecedented reorganization of holocentric chromosomes provides insights into the enigma of lepidopteran chromosome evolution. *Science Advances*, **5**(6): eaau3648.
Hillier LW, Graves TA, Fulton RS, et al. 2005. Generation and annotation of the DNA sequences of human chromosomes 2 and 4. *Nature*, **434**(7034): 724–731.
Hoegg S, Boore JL, Kuehl JV, et al. 2007. Comparative phylogenomic analyses of teleost fish Hox gene clusters: lessons from the cichlid fish *Astatotilapia burtoni*. *BMC Genomics*, **8**(1): 317.
Hofstatter PG, Thangavel G, Lux T, et al. 2022. Repeat-based holocentromeres influence genome architecture and karyotype evolution. *Cell*, **185**(17): 3153–3168. e18.
Hu SM, Zhou ZX, Yang JY, et al. 2025. Evolutionary balance between genomic conservation and coral reef adaptation in the yellow boxfish (*Ostracion cubicus*). *Zoological Research*, **46**(3): 661–674.
Karam N, Lavoie JF, St-Jacques B, et al. 2019. Bone-specific overexpression of PITX1 induces senile osteoporosis in mice through deficient self-renewal of mesenchymal progenitors and Wnt pathway inhibition. *Scientific Reports*, **9**(1): 3544.
Keivany Y. 1996. Taxonomic Revision of the Genus *Pungitius* with Emphasis on *P. hellenicus*. Master thesis, University of Alberta, Edmonton.
Khalid S, Ekram S, Salim A, et al. 2022. Transcription regulators differentiate mesenchymal stem cells into chondroprogenitors, and their *in vivo* implantation regenerated the intervertebral disc degeneration. *World Journal of Stem Cells*, **14**(2): 163–182.
Koboldt DC. 2020. Best practices for variant calling in clinical sequencing. *Genome Medicine*, **12**(1): 91.
Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics*, **15**(1): 356.
Kovaka S, Zimin AV, Pertea GM, et al. 2019. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biology*, **20**(1): 278.
Kozlov AM, Darriba D, Flouri T, et al. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, **35**(21): 4453–4455.
Li WJ, Zhang YM, Li S, et al. 2024. Investigation of sex determination in starry flounder (*Platichthys stellatus*) reveals sex chromosome evolution in Pleuronectiformes and identifies a sex-specific marker. *Zoological Research*, **45**(6): 1347–1356.
Liu ZY, Roesti M, Marques D, et al. 2022. Chromosomal fusions facilitate adaptation to divergent environments in threespine stickleback. *Molecular Biology and Evolution*, **39**(2): msab358. Mack KL, Square TA, Zhao B, et al. 2023. Evolution of spatial and temporal *cis*-regulatory divergence in sticklebacks. *Molecular Biology and Evolution*, **40**(3): msad034.
Mandrioli M, Manicardi GC. 2020. Holocentric chromosomes. *PLoS Genetics*, **16**(7): e1008918.
Marçais G, Delcher AL, Phillippy AM, et al. 2018. MUMmer4: a fast and versatile genome alignment system. *PLoS Computational Biology*, **14**(1): e1005944.
Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, **17**(1): 10–12.
Martin SH, Davey JW, Jiggins CD. 2015. Evaluating the use of ABBA–BABA statistics to locate introgressed loci. *Molecular Biology and*

Evolution, **32**(1): 244–257.
Mattern MY. 2004. Molecular phylogeny of the Gasterosteidae: the importance of using multiple genes. *Molecular Phylogenetics and Evolution*, **30**(2): 366–377.
Mayr C. 2017. Regulation by 3′-untranslated regions. *Annual Review of Genetics*, **51**: 171–194.
McLennan DA, Mattern MY. 2001. The phylogeny of the Gasterosteidae: combining behavioral and morphological data sets. *Cladistics*, **17**(1): 11–27.
Meisel RP. 2022. Ecology and the evolution of sex chromosomes. *Journal of Evolutionary Biology*, **35**(12): 1601–1618.
Minh BQ, Schmidt HA, Chernomor O, et al. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, **37**(5): 1530–1534.
Morel F, Laudier B, Guérif F, et al. 2007. Meiotic segregation analysis in spermatozoa of pericentric inversion carriers using fluorescence in-situ hybridization. *Human Reproduction*, **22**(1): 136–141.
Muffato M, Louis A, Nguyen NTT, et al. 2023. Reconstruction of hundreds of reference ancestral genomes across the eukaryotic kingdom. *Nature Ecology & Evolution*, **7**(3): 355–366.
Nakamura T, Yamada KD, Tomii K, et al. 2018. Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics*, **34**(14): 2490–2492.
Nedoluzhko A, Sharko F, Tsygankova S, et al. 2022. Intergeneric hybridization of two stickleback species leads to introgression of membrane-associated genes and invasive TE expansion. *Frontiers in Genetics*, **13**: 863547.
Nedoluzhko AV, Sharko FS, Rastorguev SM. 2025. Intergeneric introgression enhances the adaptive potential of nine-spined stickleback (*Pungitius pungitius*). *Acta Naturae*, **17**(1): 110–113.
Nelson JS. 1971. Comparison of the pectoral and pelvic skeletons and of some other bones and their phylogenetic implications in the Aulorhynchidae and Gasterosteidae (Pisces). *Journal of the Fisheries Research Board of Canada*, **28**(3): 427–442.
Nelson TC, Cresko WA. 2018. Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations. *Evolution Letters*, **2**(1): 9–21.
Nygaard KR. 2022. Extensive Behavioral Phenotyping of Williams Syndrome Locus Relevant Mouse Models to Assess Contributions of Oxytocin and Gtf2ird1. Ph. D. dissertation, Washington University in St. Louis, St. Louis.
Ogura Y, Azuma M, Tsuboi Y, et al. 2006. TFII-I down-regulates a subset of estrogen-responsive genes through its interaction with an initiator element and estrogen receptor α. *Genes to Cells*, **11**(4): 373–381.
Ou SJ, Su W, Liao Y, et al. 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology*, **20**(1): 275.
Patro R, Duggal G, Love MI, et al. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, **14**(4): 417–419.
Peichel CL, Marques DA. 2017. The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **372**(1713): 20150486.
Picard C, Pellicelli M, Taheri M, et al. 2013. Nuclear accumulation of Prohibitin 1 in osteoarthritic chondrocytes down-regulates *PITX1* expression. *Arthritis & Rheumatism*, **65**(4): 993–1003.
Reid K, Bell MA, Veeramah KR. 2021. Threespine stickleback: a model system for evolutionary genomics. *Annual Review of Genomics and Human Genetics*, **22**: 357–383.
Roberts Kingman GA, Vyas DN, Jones FC, et al. 2021. Predicting future from past: the genomic basis of recurrent and rapid stickleback evolution. *Science Advances*, **7**(25): eabg5285.
Roesti M, Moser D, Berner D. 2013. Recombination in the threespine stickleback genome—patterns and consequences. *Molecular Ecology*, **22**(11): 3014–3027.
Ross JA, Urton JR, Boland J, et al. 2009. Turnover of sex chromosomes in the stickleback fishes (Gasterosteidae). *PLoS Genetics*, **5**(2): e1000391.
Shapiro MD, Marks ME, Peichel CL, et al. 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature*, **428**(6984): 717–723.
Stanke M, Keller O, Gunduz I, et al. 2006. AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Research*, **34**(S2): W435–W439.
Tammimies K, Tapia-Páez I, Rüegg J, et al. 2012. The rs3743205 SNP is important for the regulation of the dyslexia candidate gene *DYX1C1* by estrogen receptor β and DNA methylation. *Molecular Endocrinology*, **26**(4): 619–629.
Tang HB, Krishnakumar V, Zeng XF, et al. 2024. JCVI: a versatile toolkit for comparative genomics analysis. *iMeta*, **3**(4): e211.
Termolino P, Falque M, Aiese Cigliano R, et al. 2019. Recombination suppression in heterozygotes for a pericentric inversion induces the interchromosomal effect on crossovers in Arabidopsis. *The Plant Journal*, **100**(6): 1163–1175.
Urton JR, McCann SR, Peichel CL. 2011. Karyotype differentiation between two stickleback species (Gasterosteidae). *Cytogenetic and Genome Research*, **135**(2): 150–159.
Varadharajan S, Rastas P, Löytynoja A, et al. 2019. A high-quality assembly of the nine-spined stickleback (*Pungitius pungitius*) genome. *Genome Biology and Evolution*, **11**(11): 3291–3308.
Vasimuddin M, Misra S, Li H, et al. 2019. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. *In*: 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS). Rio de Janeiro: IEEE, 314–324.
Vicoso B. 2019. Molecular and evolutionary dynamics of animal sex-chromosome turnover. *Nature Ecology & Evolution*, **3**(12): 1632–1641.
Virtanen P, Gommers R, Oliphant TE, et al. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, **17**(3): 261–272.
Wang C, Gu XD, Lin GL, et al. 2026. Haplotype-resolved genome of *Schizothorax nukiangensis* sheds light on polyploidy and altitudinal adaptation in freshwater fishes. *Zoological Research*. doi: 10.24272/j.issn.2095-8137.2025.354.
Wang YN, Wang Y, Zhao YH, et al. 2022. Phylogenomics of Northeast Asian *Pungitius* sticklebacks. *Diversity and Distributions*, **28**(12): 2610–2621.
Wenger AM, Peluso P, Rowell WJ, et al. 2019. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature Biotechnology*, **37**(10): 1155–1162.
Wucherpfennig JI, Howes TR, Au JN, et al. 2022. Evolution of stickleback spines through independent *cis*-regulatory changes at *HOXDB*. *Nature Ecology & Evolution*, **6**(10): 1537–1552.
Yi XL, Wang DD, Reid K, et al. 2024. Sex chromosome turnover in hybridizing stickleback lineages. *Evolution Letters*, **8**(5): 658–668.
Zhang Q, Liu W, Zhang HM, et al. 2020. hTFtarget: a comprehensive database for regulations of human transcription factors and their targets. *Genomics, Proteomics & Bioinformatics*, **18**(2): 120–128.