

Article

Open Access

Haplotype-resolved genome of *Schizothorax nukiangensis* sheds light on polyploidy and altitudinal adaptation in freshwater fishes

Chong Wang^{1,2,#}, Xue-Die Gu^{1,#}, Guo-Liang Lin^{1,#}, Shu-Wei Liu³, Guo-Zheng Liao¹, Yu-Hui Du¹, Bang-Yang Sun¹, Ting-Chu Wei^{1,4}, Cheng-Zhi Ding⁵, Ya-Ping Zhang^{1,6}, Yan Li^{1,*}, Xiao-Yong Chen^{3,*}, Jing Luo^{1,4,*}

¹ Bio-X Center for Interdisciplinary Innovation, School of Ecology and Environment, School of Life Sciences, Yunnan University, Kunming, Yunnan 650500, China

² MOE Key Laboratory of Freshwater Fish Reproduction and Development, School of Life Sciences, Southwest University, Chongqing 400715, China

³ Yunnan Key Laboratory of Biodiversity and Ecological Conservation of Gaoligong Mountain, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650223, China

⁴ Southwest United Graduate School, Kunming, Yunnan 650092, China

⁵ Institute of International Rivers and Eco-Security, Yunnan University, Kunming, Yunnan 650500, China

⁶ State Key Laboratory of Genetic Evolution and Animal Models, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650223, China

ABSTRACT

Polyploidy results in genome redundancy and drives ecological diversification in freshwater fishes, yet the genomic architecture and adaptive significance of recent polyploidization in high-altitude ecosystems remain poorly understood. Herein, after determining genome size and karyotype, we tackle the long-standing challenge of resolving highly similar homologous chromosomes in autopolyploid vertebrates. By integrating long-read sequencing, Hi-C scaffolding, and haplotype-specific phasing, we assemble a haplotype-resolved genome of the autotetraploid *Schizothorax nukiangensis*, comprising 100 chromosomes grouped into 25 homologous sets, with a genome size of 3.59 Gb per cell. Evidence consistently supports its autotetraploid origin, with a lineage-specific polyploidization event dated to ~0.68 million years ago. Comparative genomics reveals substantial expansions of gene families in innate immunity, redox regulation, and transposon activity, suggesting genome remodeling might have contributed to adaptation in hypoxic, pathogen-rich, and fast-flowing mountain rivers. Population genomics across three elevational zones reveal subtle yet significant structure and identify candidate genes under selection related to skeletal development, immune defense, and ciliary function, consistent with adaptation to the torrents or

rapids in fast-moving rivers characteristic of steep mountain ranges. Transcriptomic profiling further shows altitude-associated upregulation of genes related to immune responses and oxidative stress mitigation, which is consistent with other reported schizothoracines, highlighting convergent molecular strategies for altitude adaptation. Collectively, these demonstrate that recent autotetraploidization, coupled with gene family innovation and homeolog-specific expression divergence, has facilitated the ecological success of *S. nukiangensis* in extreme environments of the Salween River. This work provides important insights into the genomic imprints underlying adaptation in high-altitude freshwater vertebrates.

Keywords: *Schizothorax nukiangensis*; Whole-genome

Received: 17 November 2025; Accepted: 19 April 2026; Online: 20 April 2026

Foundation items: This work was supported by the National Key R&D Program of China (2022YFC260250302), National Natural Science Foundation of China (32293253, U1902201), Yunnan Fundamental Research Projects (202301BF070001-019, 2022KF004) from State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan University, Scientific Research Fund Project of Yunnan Education Department (2025Y0075), Scientific Research and Innovation Project of Postgraduate Students in the Academic Degree of Yunnan University (KC-242410939), National Key R&D Program of China (2022YFC2602500), and the Second Tibetan Plateau Scientific Expedition and Research Program (2024QZKK0200)

*Authors contributed equally to this work

*Corresponding authors, E-mail: liyan0910@ynu.edu.cn; chenxy@mail.kiz.ac.cn; jingluo@ynu.edu.cn

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

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

INTRODUCTION

Whole-genome duplication (WGD) is a key driver of genomic variation, rapidly increasing genome size and resulting in gene redundancy and greater genetic diversity (Otto, 2007; Sémon & Wolfe, 2007; Vlček et al., 2025). WGD can also trigger genomic shock, causing gene loss, chromosomal rearrangements, and alterations in gene expression regulation, thereby contributing to the emergence of novel phenotypic traits (Almeida-Silva & Van De Peer, 2023; Hämälä et al., 2024; Yu et al., 2024). Many studies have shown that WGD events often coincide with major climatic and geological transitions, indicating that WGD plays a crucial role in adaptation and species diversification (Cai et al., 2019; Carruthers et al., 2024; David, 2022; Van De Peer et al., 2021).

Although WGD is more prevalent in plants, hundreds of allopolyploid and autopolyploid species have also been documented in animals, particularly in freshwater fish (Leggatt & Iwama, 2003; Lien et al., 2016; Mezzasalma et al., 2023; Zhou & Gui, 2017). Snow carp (Cyprinidae: Schizothoracinae), also called naked carps, one of the youngest known polyploid lineages, are distributed across the Qinghai-Xizang Plateau and adjacent regions and represent one of the few freshwater fish taxa adapted to extreme plateau environments (Du et al., 2024). Their diversification is primarily associated with recent and recurrent polyploidization, supporting the view that polyploidy may facilitate adaptation, particularly under global climate change (Li et al., 2024c). Moreover, fossil evidence indicates that diversification in snow carp corresponds closely to the uplift history of the Qinghai-Xizang Plateau, with fossil distribution patterns aligning with current spatial distributions, reflecting long-term biological adaptation to plateau uplift (Deng et al., 2020; Wang & Chang, 2010). All known snow carp are polyploid, with chromosome numbers ranging from $2n=66$ to $2n=470$, further emphasizing the important role of polyploidization in their evolutionary history (Dai & Han, 2018; Li et al., 2024c; Yu & Yu, 1990).

The Hengduan Mountains, located at the southeastern margin of the Qinghai-Xizang Plateau, arose from complex regional tectonic processes and form a critical part of the southeastern extension of the Himalayas (Wang et al., 2022). Their steep altitudinal gradients and rugged topography have profoundly influenced the Salween River Basin, creating a series of microhabitats along its course. Acting as a climatic barrier, the Hengduan Mountains intercept warm, moist air from the Indian Ocean, resulting in dramatic variations in temperature, precipitation, and humidity over short distances. These microclimatic heterogeneities have facilitated the coexistence of tropical, subtropical, temperate, and alpine ecosystems within a narrow geographic range, fostering exceptional biodiversity and high levels of endemism among both flora and fauna (Chaplin, 2005; Li et al., 2024b; Liu et al., 2020; Zhou & Li, 2024).

Equally important, the dynamic aquatic environment of the Salween River imposes strong selective pressures on aquatic species. The river exhibits a pronounced longitudinal gradient in hydrological conditions, including fast-flowing rapids, deep pools, and braided channels (Yang et al., 2021). The ecological environment of the Salween River varies greatly along its elevation gradient. At higher altitudes, the water flows rapidly, temperatures are lower, and food resources are

limited, creating harsher survival conditions for aquatic organisms. In contrast, lower-altitude sections are characterized by slower water, warmer temperatures, and abundant food resources. Seasonal fluctuations in discharge, driven by monsoonal rains and glacial meltwater, introduce substantial variability in flow velocity, sediment load, and turbidity (Yao et al., 2012). Furthermore, dissolved oxygen concentrations decrease with increasing altitude and fluctuating temperatures, posing significant physiological challenges for aquatic organisms in high-elevation reaches. Collectively, these complex hydrological and physicochemical factors generate a mosaic of aquatic habitats that, in conjunction with the region's topography and climate, shape the adaptive landscape of native freshwater species (Kawecki & Ebert, 2004; Zhou & Li, 2024).

From a biogeographical perspective, the combined effects of climatic barriers, rugged terrain, and dynamic hydrology have promoted population fragmentation, genetic divergence, and local adaptation to heterogeneous environmental conditions (Wang & Bradburd, 2014). Endemic freshwater fishes of the Salween River, including *Schizothorax nukiangensis*, a representative snow carp in this river, has shown specialized morphological, physiological, and behavioral traits to cope with strong currents, low oxygen availability, and fluctuating hydrological regimes. Among these, *S. nukiangensis* exhibits a distinctive shovel-like lower jaw that enables firm scratch algae on rocks in turbulent streams, facilitating efficient feeding (Chen, 2013). Distributed across elevations ranging from 612 to nearly 3890 m above sea level (asl), *S. nukiangensis* exhibits remarkable ecological plasticity, thriving in diverse habitats along the Salween River (Ding et al., 2019; Xu et al., 2023). This extensive altitudinal habitats within a single river system raises important questions about intraspecific variation: have geographically structured populations emerged along different elevations, and how differentiated are these populations in terms of genetic composition and adaptive traits? However, the genomic basis of this broad altitudinal and ecological adaptability, as well as the evolutionary implications of its recent autotetraploidization, remains largely unexplored.

Resolving homologous chromosome sets in recent autopolyploid vertebrates poses significant technical challenges due to high sequence similarity between homeologs and pervasive genome redundancy. Previous studies on polyploid fish genomes have often resulted in collapsed assemblies or unresolved haplotypes, obscuring evolutionary and functional insights. In this work, we present the first haplotype-resolved, chromosome-level assembly of *S. nukiangensis*. By leveraging high-fidelity long-read sequencing, Hi-C scaffolding, and homeolog-specific phasing, we successfully partitioned the genome into 100 chromosomes corresponding to 25 homologous chromosome sets. By integrating genome resequencing of populations inhabiting distinct altitudinal gradients, we examined patterns of genetic diversity, population structure, and historical demography. Additionally, we identified candidate genes and genomic regions potentially involved in adaptation to the extreme and heterogeneous environments of the Hengduan Mountains, and the dynamic aquatic habitats of the Salween River. Our findings offer new insights into the genetic mechanisms underlying high-altitude and hydrological resilience in freshwater fishes and advance our understanding

of how geological, climatic, and aquatic processes jointly shape biodiversity in mountainous river systems.

MATERIALS AND METHODS

Tissue sampling and sequencing for genome assembly of *Schizothorax nukiangensis*

Samples of *S. nukiangensis* (sample ID is YPQ230120 for de novo genome assembly, more details in Supplementary Table S1) were collected from the Salween River Basin in Xizang and Yunnan Province, China (Supplementary Table S1). Muscle tissues were flash-frozen in liquid nitrogen for PacBio HiFi (CCS) long-read sequencing, DNBSEQ-T7 short-read sequencing, and Hi-C chromatin conformation capture sequencing. A total of 128.35 Gb of PacBio HiFi reads and 112.88 Gb (~71.31×) of paired-end short reads (150 bp) were generated for the *de novo* genome assembly of *S. nukiangensis*. Hi-C library preparation, including DpnII restriction enzyme digestion, was followed by sequencing on an DNBSEQ-T7 platform, yielding 192.26 Gb (~106.81×) of clean paired-end reads (Supplementary Table S2). To support genome annotation, both full-length transcriptome and short-read RNA sequencing were performed using pooled RNA samples from multiple tissues, including heart, liver, gill, kidney, swim bladder, muscle, skin, fin, brain, eye, spleen, and intestine. For full-length transcriptome sequencing, a total of 1.04 Gb of Iso-Seq circular consensus sequencing (CCS) data were generated on the PacBio platform. Additionally, short-read RNA-seq libraries were constructed and sequenced on the Illumina platform, generating 15.63 Gb of paired-end reads (150 bp) (Supplementary Table S2). These transcriptome data were integrated into genome annotation to improve gene model prediction.

Population sampling and sequencing

For population studies, a total of 37 wild *S. nukiangensis* individuals were collected from various locations along the Salween River, spanning altitudes from 683 m asl to 1987 m asl (Supplementary Table S1) and sequenced. Genomic DNA was extracted from muscle tissue, and whole-genome sequencing (WGS) was conducted on the DNBSEQ-T7 platform, producing 150-bp paired-end reads. The average sequencing depth of the samples was approximately 30×. Detailed information on sample collection, including sampling sites, altitude, and sequencing depth, is listed in Supplementary Table S1.

Genome size estimation and assembly of *Schizothorax nukiangensis*

The genome size of *S. nukiangensis* was estimated by flow cytometry using fresh peripheral blood, with chicken (*Gallus gallus domesticus*, 1C=1.25 pg) as an internal reference. Nuclei suspensions from fish and chicken blood were mixed, stained with propidium iodide, and analyzed on a BD FACSymphony A1 flow cytometer (BD, USA). Genome size was calculated based on the fluorescence intensity ratio between the sample and the reference, with measurements performed in triplicate.

To further estimate genome size, heterozygosity, and repeat content with sequences, Jellyfish v.2.3.1 (parameters: -m 31 -s 1G -t 4 -C) (Marçais & Kingsford, 2011) and GenomeScope v.2.0 (Ranallo-Benavidez et al., 2020) were used to generate the k-mer frequency distribution. The genome was primarily assembled from 128.35 Gb of CCS reads and 192.26 Gb of

clean Hi-C reads using hifiasm v.0.24.0-r703 with the parameters -l 0 --n-hap 4 (Cheng et al., 2021, 2022). Chromosome assignment was performed using Hi-C technology. High-quality paired-end reads were aligned to contigs using BWA v.0.7.17-r1188 (Li & Durbin, 2009) to calculate contact frequency. Following the YaHS v.1.2.2 pipeline with default parameters (Zhou et al., 2022), the process began with an initial round of error correction, followed by the generation of an interaction matrix using directional brackets in conjunction with Juicer, which enabled subsequent inspection and manual correction via the Juicebox assembly tool (Durand et al., 2016).

Multiple strategies were applied to evaluate genome assembly quality. The phasing accuracy of haplotypes was assessed using Merqury v.1.4.1 (Rhie et al., 2020) with default parameters. Genome completeness was evaluated using BUSCO v.5.8.0 with the actinopterygii_odb10 (Tegenfeldt et al., 2025). In addition, both HiFi and Hi-C reads were realigned to the assembly to calculate the mapping ratio and assess the reliability of the genome assembly.

Annotation of repeat elements and protein-coding regions

A combination of homology-based and *de novo* approaches was used to identify repeat elements. Known transposable elements (TEs) were identified using RepeatMasker v.4.1.8 (parameters: -xsmall, -e rmbblast, -nolow, -no_is, -norna) (<https://www.repeatmasker.org/>) (Tarailo-Graovac & Chen, 2009) with the Repbase TE library (Jurka et al., 2005). Additionally, a *de novo* repeat library was constructed using RepeatModeler v.2.0.5 (Flynn et al., 2020). The two libraries were then merged and used to screen the genome for repetitive sequences with RepeatMasker.

Protein-coding genes were predicted using the BRAKER3 v.3.0.6 pipeline with default parameters (Gabriel et al., 2024), which integrates *ab initio* gene prediction via GeneMark-ETP v.4.71 (Brúna et al., 2024), TSEBRA v.1.1.2.3 (Gabriel et al., 2021), and AUGUSTUS v.3.5.0 (Stanke et al., 2008) with RNA-seq evidence to improve annotation accuracy. RNA-seq alignments provided extrinsic hints to guide gene structure prediction, while GeneMark-ETP v.4.71 enabled model training based on transcript evidence. For homology-based predictions, GeMoMa v.1.9 (Keilwagen et al., 2018) was used to compare the genome of *S. nukiangensis* with 10 related species, including *Culter alburnus* (GCA_040182925.1), *Danio rerio* (GRCz11 release-112, https://ftp.ensembl.org/pub/release-112/fasta/danio_rerio/dna), *Gasterosteus aculeatus* (GCA_016920845.1), *Megalobrama amblycephala* (GCA_018812025.1), *Oryzias latipes* (GCA_002234675.1), *Cyprinus carpio* (GCA_018340385.1), *Carassius auratus* (Luo et al., 2020), *Ctenopharyngodon idella* (GCA_019924925.1), *Oreochromis niloticus* (GCA_001858045.3), and *Poropuntius huangchuchieni* (Chen et al., 2021).

For RNA-seq data, both *de novo* and reference-guided transcriptome assemblies were performed. RNA-seq reads were aligned to the genome using HISAT2 v.2.2.1 with default parameters. (Kim et al., 2019), and genome-guided transcript assembly was carried out with StringTie v.1.3.3b (Shumate et al., 2022). TransDecoder v.5.7.1 was employed to identify candidate coding regions by predicting open reading frames (ORFs), providing protein-coding evidence to support gene annotation. In parallel, Trinity v.2.8.5 (Grabherr et al., 2011) was used for *de novo* transcriptome assembly.

To improve gene structure annotation, a comprehensive

transcriptome database was constructed using both RNA-seq and Iso-Seq assemblies following the Program to Assemble Spliced Alignments (PASA v.2.4.1) (Haas et al., 2008) pipeline with default parameters. Transcripts were aligned to the reference genome using GMAP v.2021.08.25 and BLAT v.36, and PASA v.2.4.1 subsequently constructed high-quality gene models, annotated alternative splice isoforms, and refined UTR and exon-intron structures.

EVidenceModeler v.2.1.0 (Haas et al., 2008) was used to integrate all predicted gene sets into a nonredundant, comprehensive, and high-confidence gene annotation set. Finally, the PASA pipeline was used to update the EVidenceModeler output by adding UTR annotations and alternative splicing models. The resulting protein-coding genes were functionally annotated using the eggNOG-mapper (Cantalapiedra et al., 2021) comprehensive database.

Phylogenetic reconstruction of haplotype chromosomes based on single-copy orthologs

Phylogenetic relationships among the 25 chromosome sets of each haplotype were inferred using single-copy orthologous genes. First, protein-coding sequences from *S. nukiangensis* and comparison species (*D. rerio*, *Ca. auratus*, and *Cy. carpio*) were clustered using OrthoFinder v.2.5.4 (Emms & Kelly, 2019) to identify single-copy orthologous gene families. The resulting orthologues were then partitioned into 25 groups based on their chromosome of origin. For each group, coding sequences were aligned using MUSCLE v.3.8.31 (Edgar, 2022). Maximum-likelihood (ML) phylogenies were reconstructed with RAxML v.8.2.12 (Stamatakis, 2014), under the JTT+I+G+F substitution model and with 1000 bootstrap replicates to assess node support. The resulting ML trees were exported in Newick format and interactively visualized using FigTree v.1.4.4.

Genome comparison and synteny analyses

To investigate relationships of *S. nukiangensis* to other cyprinids, we conducted a comparative genomics analysis using the published genomes of 10 species including *Cirrhinus molitorella* (Tu et al., 2025), *D. rerio*, *Ct. idella*, *Cu. alburnus*, *M. amblycephala*, *P. huangchuchieni*, *Sinocyclocheilus grahami* (Yin et al., 2021), *Oxygymnocypris stewartii* (Liu et al., 2019), *Ca. auratus*, and *Cy. carpio*. Orthologous gene families were identified using OrthoFinder based on protein sequences. Single-copy orthogroups shared across all species were retained for phylogenetic inference. Protein sequences from these orthogroups were aligned using MUSCLE v.3.8.31, and the resulting alignments were stringently trimmed using Gblocks v.0.91b (Talavera & Castresana, 2007) to remove ambiguously aligned and poorly conserved regions.

Phylogenetic tree reconstruction was performed using the ML method implemented in RAxML, with automatic selection of the best-fit substitution model and 1000 ultrafast bootstrap replicates to assess node support. Divergence times were estimated using a Bayesian relaxed molecular clock implemented in the MCMCtree module of PAML v.4.9j (Yang, 2007), calibrated with fossil records and divergence time priors obtained from the TimeTree database (Kumar et al., 2017).

Gene family expansion and contraction analyses were carried out using café (Mendes et al., 2021), based on a birth-death process model mapped onto the phylogeny. Significantly expanded or contracted gene families (false

discovery rate (FDR)<0.05) were subjected to Gene Ontology (GO) and KEGG pathway enrichment analysis using Fisher's exact test with multiple testing correction. Finally, genome synteny between *S. nukiangensis* and other species was analyzed using the jcv v.1.5.4 (Tang et al., 2024) toolkit to identify conserved collinear blocks across genomes.

Whole-genome duplication (WGD) evidence

We investigated and compared WGD events between the *S. nukiangensis* genome and 8 Cyprinidae species (*D. rerio*, *Ct. idella*, *Cu. alburnus*, *M. amblycephala*, *P. huangchuchieni*, *O. stewartii*, *Ca. auratus*, and *Cy. carpio*). Paralogous gene pairs were identified using BLAST-based methods, and syntenic paralogs were detected using WGD1 (Sun et al., 2022). The number of synonymous substitutions per synonymous site (Ks) was calculated for gene pairs using the WGD1 pipeline. To estimate the timing of WGD events, we applied a synonymous substitution rate of 3.51×10^{-9} mutations per site per year, as previously established for common carp (David et al., 2003).

Single nucleotide polymorphism (SNP) calling

For variant calling, paired-end reads from each accession were mapped to the four haplotype genomes using BWA v.0.7.17-r1188. After filtering out low-quality alignments, SNP calling was conducted using the HaplotypeCaller module of the Genome Analysis Toolkit (GATK) (McKenna et al., 2010). High-confidence single nucleotide polymorphisms (SNPs) were obtained using VCFtools v.0.1.13 (Danecek et al., 2011) with the following parameters: --max-missing 0.8, --maf 0.05, --minDP 10, --maxDP 500, --minQ 30, --minGQ 30, --min-alleles 2, --max-alleles 2 (Lu et al., 2024). SNP annotation was performed based on genomic location and predicted coding effects, using the *S. nukiangensis* genome annotation as a reference, with SnpEff v.4.1g (Cingolani et al., 2012).

Population genetics analyses

Phylogenetic analysis was conducted using a SNP dataset derived from the *S. nukiangensis* genome with IQ-TREE v.2.2.2.6 (Minh et al., 2020). A ML phylogenetic tree was reconstructed using 1 000 ultrafast bootstrap replicates. Model selection was carried out with ModelFinder, implemented within IQ-TREE, which evaluated 242 candidate nucleotide substitution models. The best-fitting model, TVM+F+ASC+R2, was selected based on the bayesian information criterion (BIC) and subsequently applied for tree inference. Population structure analysis (K=2–5) were performed using PLINK v.1.90 (Chang et al., 2015) and TBtools v.2.482 (Chen et al., 2023) respectively, based on the same SNP dataset.

Demographic history

To investigate the demographic history of *S. nukiangensis*, the effective population size (N_e) over the past 10 million years was estimated using the pairwise sequentially markovian coalescent (PSMC) model (Li & Durbin, 2011). The parameters were set as follows: -N25 -t15 -r5 -p "4+25*2+4+6", with a generation time (g) of 6 years (Ma, 2022; Mu et al., 2025). And Mutation rate was calculated through a linear regression of the relationship between log-genome size and log-mutation rate. The best fitting model for mutation rate μ in bp was $\mu = 2.59 \times 10^{-10} \times g^{0.584}$ for genome size g in Mb (genome size=3585.88 Mb), giving an estimate of 3.08×10^{-8} per bp per generation (Crellen et al., 2016; Lynch, 2010).

Selection analyses

To detect signals of selective sweeps between the southern and northern groups, F_{ST} values were calculated using VCFtools with a sliding window size of 50 kb and a step size of 10 kb. XP-CLR (Chen et al., 2010) scores were also computed using the same window and step sizes. Genes overlapping with the identified sweep regions were extracted and considered candidate selective sweep genes.

Quantitative real-time PCR (qRT-PCR) Analysis

Total RNA was extracted from tissues using TRIzol reagent (Invitrogen, USA) and reverse-transcribed into cDNA using the PrimeScript RT reagent kit (Takara, Japan) following the manufacturer's instructions. Quantitative real-time PCR was performed on a CFX96 Real-Time PCR System (Bio-Rad, USA) using SYBR Green Master Mix (Takara, Japan). The relative expression levels of target genes were calculated using the $2^{-\Delta\Delta C_t}$ method, with *ef1a* as the internal reference.

RNA sequencing and differential expression analyses

Total RNA was extracted from intestinal tissues of *S. nukiangensis* individuals collected from three distinct altitudinal zones, with each elevation group represented by three biological replicates ($n=9$ in total) (Supplementary Table S3). We grouped the samples based on the individuals' sampling altitudes (Supplementary Table S3), with the following specific groupings: Group 1: Samples from an altitude of 1068 m asl; Group 2: Samples from an altitude of 718 m asl; Group 3: Samples from an altitude of 1255 m asl. All transcriptome data were used for haplotype-specific expression bias analysis. For the haplotype-specific expression bias analysis, we identified 1:1:1:1 homologous gene pairs across hapA, hapB, hapC, and hapD using OrthoFinder v.2.5.4. Group 2 and Group 3 samples were used for differential gene expression analysis. RNA libraries were constructed and sequenced on the Illumina platform. Clean reads were aligned to the reference genome using HISAT2 with default parameters. Gene expression levels were quantified for each sample using featureCounts v.2.0.3 (Liao et al., 2014), and differential expression analysis was performed using DESeq2 v.1.50.2 (Love et al., 2014). Genes with a fold change ≥ 2 and a FDR < 0.05 were considered significantly differentially expressed.

RESULTS

High-Quality haplotype-resolved genome assembly and annotation of autotetraploid *Schizothorax nukiangensis*

We determined the genome size and karyotype of *S. nukiangensis* for the first time. Flow cytometry estimated its nuclear DNA content at 2.15 pg (Genome size ≈ 1.80 Gb), using Chicken as a reference (Supplementary Figure S1A; Supplementary Table S4). Karyotype analysis showed 100 chromosomes (Supplementary Figure S1B), a count often associated with tetraploidy in cyprinids. K-mer analysis conducted with Jellyfish confirmed the autotetraploid nature of *S. nukiangensis* (Figure 1A; Supplementary Figure S1C). *De novo* assembly using hifiasm produced 853 scaffolds totaling 3585.89 Mb, with an N50 of 35.11 Mb. Integration of Hi-C data anchored 3463.55 Mb (96.59% of the assembly) onto 100 chromosomes, resulting in a haplotype-resolved genome assembly (Figure 1B; Supplementary Figure S2; Supplementary Table S5). These 100 chromosomes were organized into 25 homologous groups, each comprising four

homologous chromosomes corresponding to haplotypes A, B, C, and D (Figure 1C; Supplementary Table S5).

Multiple validation approaches demonstrated the high accuracy and reliability of the *S. nukiangensis* genome assembly. K-mer spectra analysis with Merqury clearly confirmed the tetraploid nature and accurate phasing of the assembled haplotypes (Figure 1D; Supplementary Figure S3). Base-level quality assessment revealed an exceptionally low estimated error rate of 3.13×10^{-7} and a consensus quality value (QV) of 65.05, indicating excellent agreement between the HiFi reads and the assembly (Supplementary Table S6; Supplementary Figure S4). Hi-C contact maps exhibited strong diagonal signals, supporting correct phasing, ordering, and orientation of chromosomes (Figure 1B). Mapping of clean reads yielded high alignment rates, exceeding 99% for both PacBio HiFi and Hi-C reads, and over 95% for RNA-seq reads (Supplementary Table S7). BUSCO analysis demonstrated 99.76% completeness of the vertebrata_odb10 gene set across haplotypes (Supplementary Table S8; Supplementary Figure S5).

Repetitive elements constitute more than half of the *S. nukiangensis* genome, with comparable proportions across all four haplotypes (Supplementary Table S5). Comprehensive repeat annotation revealed that repetitive sequences account for 52.84%, 51.33%, 50.75%, and 50.94% of haplotypes A, B, C, and D, respectively, resulting in an overall repeat content of approximately 51.96% (Supplementary Table S5). Transposable elements (TEs) predominated in the genome, accounting for 34.89% of the assembly, with retrotransposons contributing 13.32% of this total (Supplementary Table S9). This high repeat proportion is consistent with the large genome size and autotetraploid nature of *S. nukiangensis*. Divergence analyses of TEs showed that DNA transposons, LINEs, long terminal repeats (LTRs), and short interspersed nuclear elements (SINEs) have all substantially contributed to genome expansion (Supplementary Figure S6). Among these, DNA transposons and LINEs exhibited prominent divergence peaks at $\sim 2\%$ – 5% , indicating recent bursts of TE amplification. Gene annotation revealed a high number of protein-coding genes per haplotype and near-complete coverage of the vertebrata_odb10 gene set. Specifically, 25956, 25096, 25878, and 26193 protein-coding genes were identified in haplotypes A, B, C, and D, respectively, collectively covering 98.48% of vertebrate core genes (Supplementary Table S8; Figure S5). Functional annotation of these genes was performed using the comprehensive eggNOG-mapper database. In total, 39869, 38137, 39883, and 40139 transcripts were predicted for in haplotypes A, B, C, and D, respectively, corresponding to an average of 1.52–1.54 splice isoforms per gene (Figure 1C; Supplementary Table S5).

Evidence for autotetraploid origin of *Schizothorax nukiangensis*

Among Cyprinidae species, polyploidy is common (Li et al., 2024c), with origins spanning both autopolyploid and allopolyploid events. Notably, well-studied allotetraploid species such as *Ca. auratus* (Luo et al., 2020) and *Cy. carpio* (Xu et al., 2019a) exhibit pronounced subgenomic divergence, enabling clear partitioning into distinct subgenomes. To elucidate the polyploid origin of *S. nukiangensis*, we conducted k-mer frequency analysis, collinearity assessment, and phylogenetic reconstruction. The 31-mer frequency plot

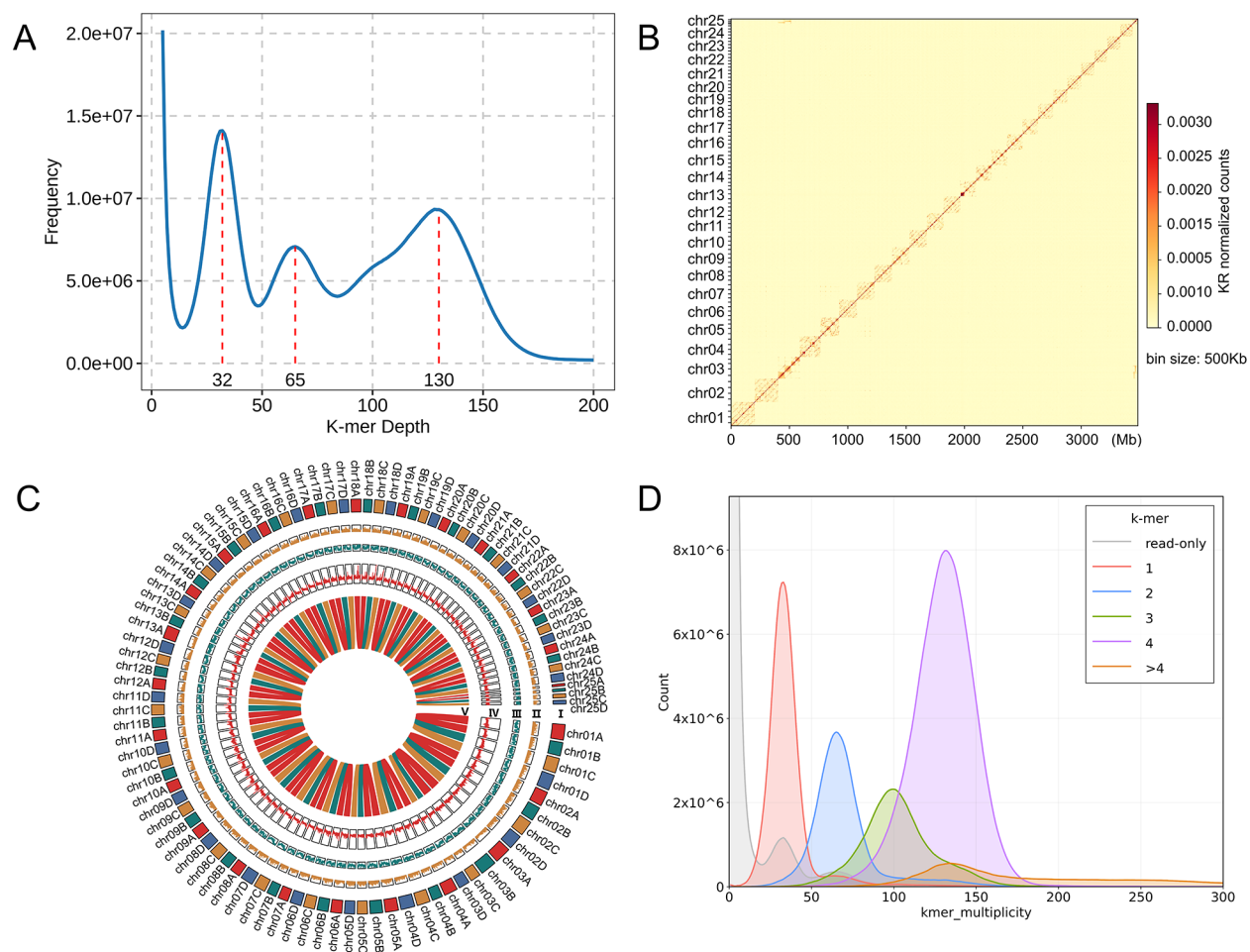


Figure 1 Genome assembly and annotation of *Schizothorax nukiangensis*

A: K-mer frequency distribution ($k=31$) showing peaks at depths of 32, 65, and 130, consistent with autotetraploid genome characteristics. B: Hi-C contact heatmap of the assembled chromosomes, illustrating chromosomal interactions and validating assembly continuity and phasing accuracy. C: Circos plot depicting the genomic landscape of *Schizothorax nukiangensis*: I. 100 chromosomes organized into 25 homologous groups; II. GC content; III. repeat density; IV. gene density; V. syntenic relationships between chromosomes visualized by curved links. D: K-mer multiplicity spectrum comparing distinct k-mers absent from the assembly (black) and present with different copy numbers: 1 (red), 2 (blue), 3 (green), 4 (purple), and ≥ 4 times (orange) in the assembly relative to raw HiFi reads.

displayed three distinct peaks at depths of approximately $32\times$, $65\times$, and $128\times$ (Figure 1A), a pattern consistent with an autotetraploid genome. Analysis of nucleotide heterozygosity revealed a high proportion of homozygous (aaaa, 97.95%) and aaab-type k-mers, with 1.19% aaab and 0.57% aabb heterozygous k-mers (Supplementary Table S10). This distribution matches the expected pattern for autotetraploids, in which aaab-type heterozygosity exceeds aabb, further supporting the autotetraploid nature of *S. nukiangensis*. Synteny analysis revealed extensive conservation of collinear blocks among the four haplotypes, as demonstrated by both the synteny dot plot and the collinearity alignment (Figures 2A, B). This high degree of genomic homology strongly supports the hypothesis of autopolyploidy.

To further validate this inference, we constructed a ML phylogenetic tree using genome-wide orthologs from eight closely related cyprinid genomes along with the four haplotypes of *S. nukiangensis*. The resulting tree showed that the four haplotypes clustered tightly together with high bootstrap support (Figure 2C), consistent with an autotetraploid origin. Additional individual phylogenetic trees constructed for each of the 25 chromosomes exhibited

congruent topologies, reinforcing this conclusion (Supplementary Figure S7). In contrast, allotetraploid species such as *Cy. carpio* and *Ca. auratus* display distinct subgenomic topologies, with duplicated haplotypes partitioned into separate clades corresponding to different progenitor lineages, which is not observed in *S. nukiangensis* (Figure 2C; Supplementary Figure S7). Furthermore, orthogroup analysis among the four haplotypes revealed extensive genetic similarity and a high degree of shared gene content. A total of 24466 orthogroups were identified, of which 17460 were shared by all four haplotypes, 4403 by three haplotypes, 2386 by two haplotypes. Only 188 orthogroups were unique to a single haplotype (hapA: 65; hapB: 45; hapC: 43; hapD: 35) (Figure 2D). This high level of gene content uniformity provides additional evidence for an autotetraploid origin via WGD within a single lineage.

Phylogeny and adaptive genome evolution of autotetraploid *Schizothorax nukiangensis*

An ML phylogenetic tree was reconstructed using 698 single-copy orthologous genes from *S. nukiangensis* and ten other cyprinids: *Cu. alburnus*, *D. rerio*, *Ct. idella*, *Ci. molitorella*, *M. amblycephala*, *Cy. carpio*, *Ca. auratus*, *O. stewartii*, *P.*

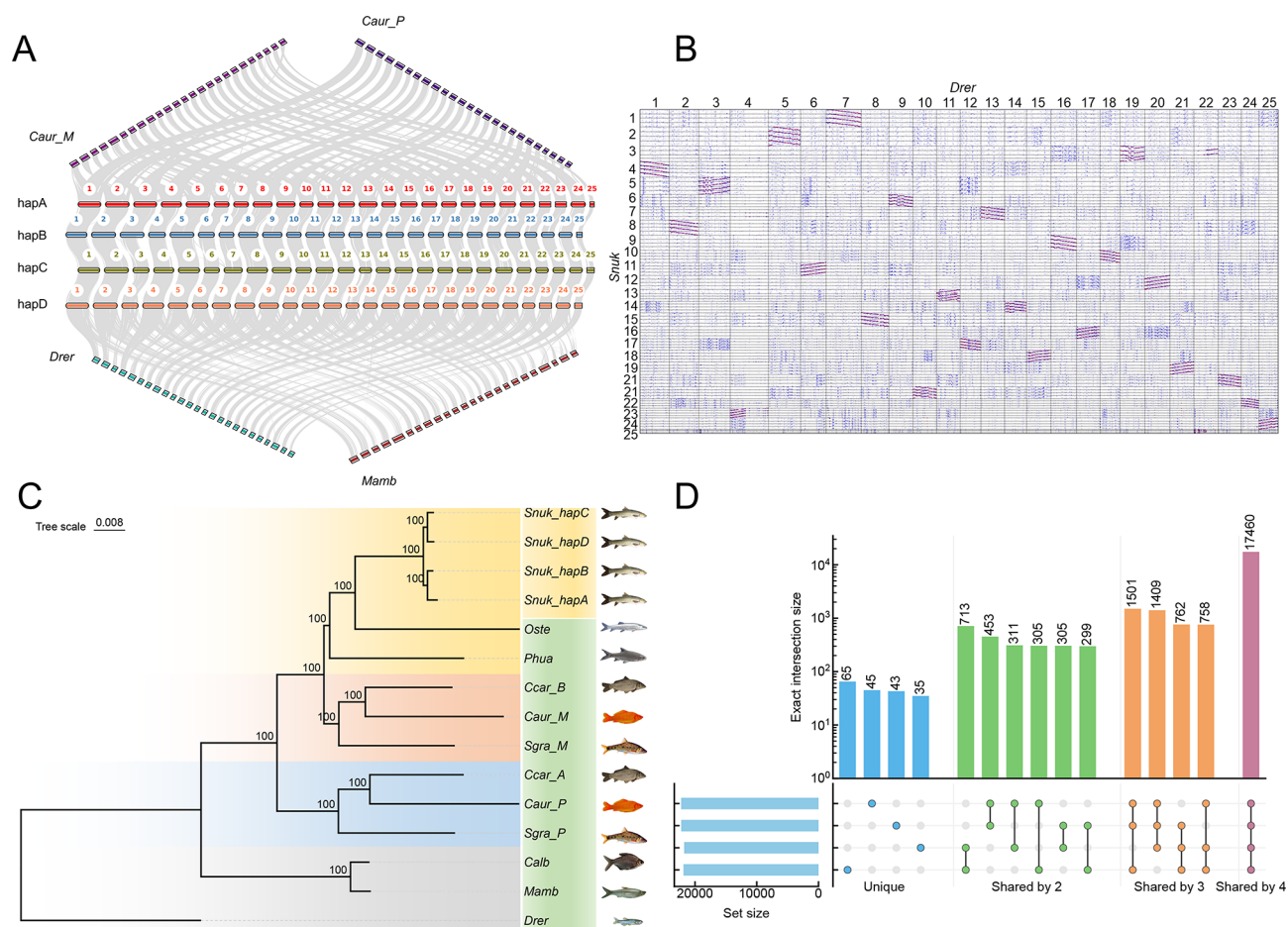


Figure 2 Evidence supporting the autotetraploid origin of *Schizothorax nukiangensis*

A: Syntentic relationships among the four haplotypes and representative cyprinid genomes. Comparison of syntentic blocks among *Ca. auratus* (Caru_M and Caru_P), *D. rerio* (Drer), *M. amblycephala* (Mamb), and the four haplotypes (hapA-hapD) of *Schizothorax nukiangensis*. B: Dot plot illustrating the syntentic blocks among the four haplotypes of *Schizothorax nukiangensis* and the reference genome of *D. rerio*. C: Phylogenetic tree of four *Schizothorax nukiangensis* haplotypes (hapA-hapD) and eight related cyprinid species, constructed from genome-wide orthologs; yellow blocks show *Schizothorax nukiangensis* and the sister clade of *Schizothorax nukiangensis*, orange blocks show *Cy. carpio*, *Ca. auratus*, *Si. grahami* M subgenome clade and blue blocks show *Cy. carpio*, *Ca. auratus*, *Si. grahami* P subgenome clade, grey blocks show *Culter alburnus*, *D. rerio*, *M. amblycephala*, respectively; red block represents the data generated in this work (hapA, hapB, hapC, and hapD represent the 4 haplotypes, respectively), and the green block represents downloaded species data. D: Upset plot illustrating the distribution of orthogroups shared across haplotypes, including unique and shared orthologs.

huangchuchieni, and *Si. grahami* (Figure 3A). All nodes in the resulting phylogeny were supported by strong bootstrap values (100%). Divergence time estimation using the Bayesian MCMCtree method indicated that *S. nukiangensis* diverged from the Barbinae lineage approximately 7.58 million years ago (Ma), with a subsequent split from *O. stewartii* around 5.33 Ma. These divergence events coincide with major uplift phases of the Qinghai-Xizang Plateau (Royden et al., 2008), suggesting a potential role for geological processes in shaping the evolutionary trajectory of *S. nukiangensis* (Li et al., 2024c).

In total, 1127 gene families were expanded, collectively gaining 4515 genes, while 1470 families were contracted, resulting in the loss of 1641 genes. This net expansion is reflected by a positive average expansion index (0.1257), suggesting substantial gene family innovation following WGD events (Figure 3A). GO and KEGG enrichment analyses of the 702 significantly expanded gene families highlighted biological processes central to genome plasticity and innate immunity, including DNA recombination ($p < 0.01$), transposon activity ($p < 0.01$), and signaling pathways such as NF-kappa B

($p < 0.01$) and NOD-like receptor signaling ($p < 0.01$) (Supplementary Figure S8). These findings suggest that TE activation and immune-related gene family expansions may have played a role in facilitating rapid adaptation to the pathogen-rich, high-altitude river environments inhabited by *S. nukiangensis* (Chuong et al., 2017; Oliver & Greene, 2009) (Supplementary Table S11; Supplementary Table S12).

Orthologous gene family clustering across seven representative cyprinid species revealed a shared core of 9 119 gene families conserved across all species, reflecting essential biological functions maintained throughout Cyprinidae evolution (Figure 3B). Notably, *S. nukiangensis* possessed 441 unique gene families absent from the other genomes analyzed, representing lineage-specific genetic innovations potentially underlying ecological specialization. These gene families were enriched in functions related to transposon activity, nucleic acid metabolism, and developmental regulation, which may influence lineage-specific traits. While speculative, such features could be associated with physiological or morphological plasticity relevant to dynamic riverine habitats (Supplementary Table

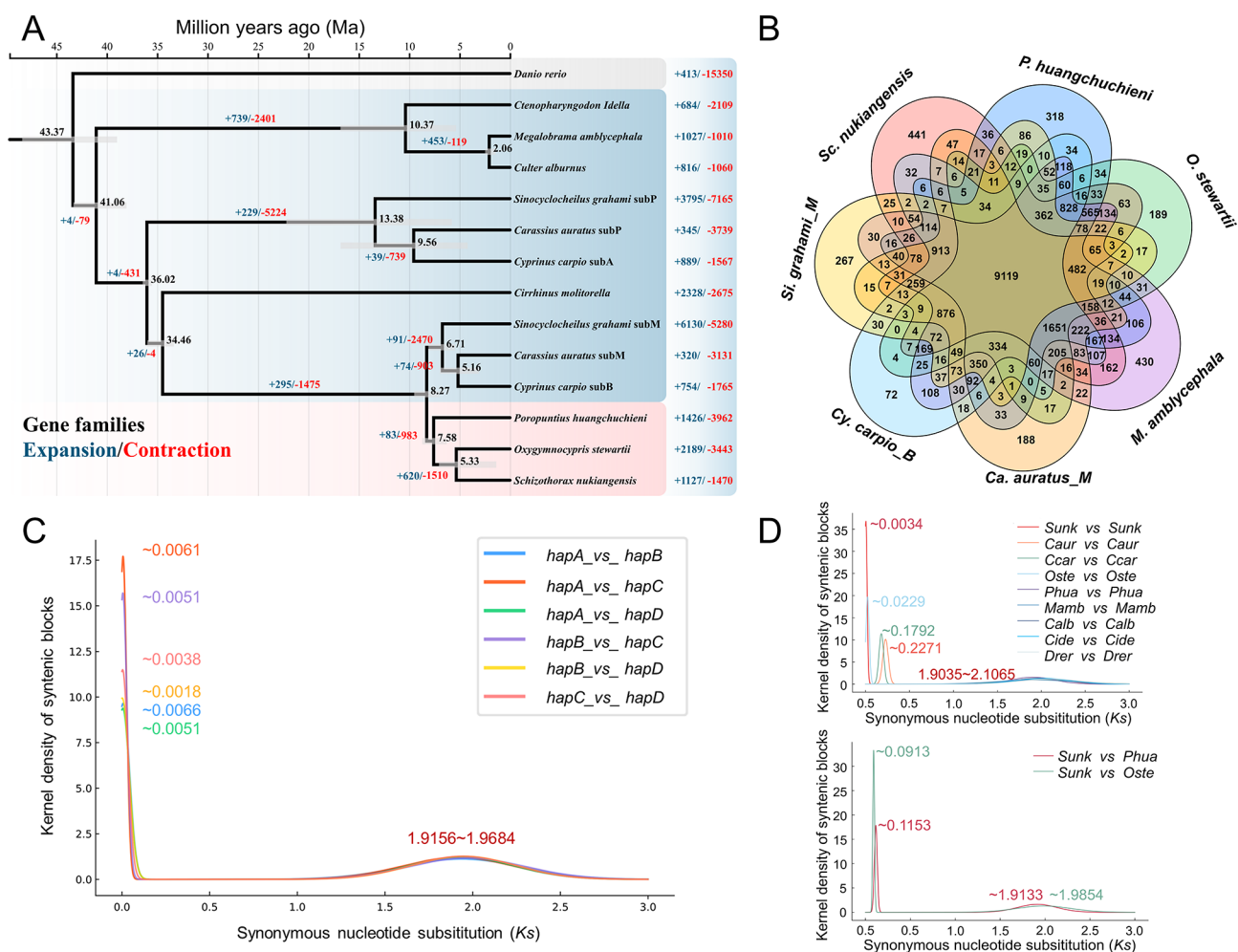


Figure 3 Phylogenetic relationships, comparative genomics, and evolutionary analyses of *Schizothorax nukiangensis*

A: Phylogenetic tree with divergence times and gene family expansions (blue) and contractions (red) for 11 cyprinid species, including *Cu. alburnus*, *D. rerio*, *Ct. idella*, *Ci. molitorella*, *M. amblycephala*, *Cy. carpio*, *Ca. auratus*, *O. stewartii*, *P. huangchuchieni*, *Si. grahami*. **B:** Venn diagram of orthologous gene families of *Cy. carpio*, *Ca. auratus*, *Si. grahami*, *Schizothorax nukiangensis*, *M. amblycephala*, *O. stewartii* and *P. huangchuchieni*. **C:** Kernel density plots of Ks distributions for paralogous gene pairs within *Schizothorax nukiangensis* haplotypes. **D:** Ks distributions of orthologous gene pairs among *Schizothorax nukiangensis* and related species; the upper panel shows distributions consistent with Cyprinidae-wide WGD, while the lower panel highlights recent speciation between *Schizothorax nukiangensis*, *P. huangchuchieni*, and *O. stewartii*.

S13). Additionally, we identified a set of positively selected genes associated with high-altitude adaptation. These genes are primarily involved in cell cycle regulation, oxygen sensing, signal transduction, and stress response. Nitric oxide synthase (NOS) and natriuretic peptide receptor genes likely regulate nitric oxide production and vascular dilation under hypoxic conditions. Phospholipase D2 and TRP channels play key roles in calcium homeostasis and cellular stress responses. Cell division cycle genes and CDK2 contribute to cell proliferation and tissue repair, whereas EGLN1 regulates the HIF-1 pathway under low-oxygen stress. Collectively, these genes facilitate adaptation to high-altitude environments by enhancing oxygen utilization, promoting cellular repair, and improving stress tolerance (Supplementary Table S14).

Meanwhile, WGD analysis was conducted using WGDI, which identified paralogous gene pairs among haplotype assemblies. Kernel density estimation of synonymous substitution rates (Ks) for these paralogous pairs revealed a prominent peak at $Ks \approx 0.0018$ – 0.0066 , indicative of a recent large-scale duplication event likely corresponding to the tetraploidization of *S. nukiangensis*. Based on the Ks distribution, this WGD event is estimated to have occurred

approximately 0.68 Ma, suggesting a relatively recent timeframe for the origin of autotetraploidy in *S. nukiangensis* (Figure 3C; Supplementary Figure S9). Analysis of gene pairs from nine representative cyprinid genomes revealed distinct patterns of WGD history across species. In several species, a single major Ks peak corresponded to the well-characterized third round of WGD (3R-WGD), a duplication event conserved across all teleost. In contrast, other cyprinid species displayed bimodal Ks distributions, reflecting both this ancestral event and an additional, lineage-specific fourth round of WGD (4R-WGD) (Figure 3D). In *S. nukiangensis*, a clear bimodal Ks distribution was observed, with a prominent peak at low Ks values (≈ 0.0034), consistent with a recent duplication event, and a broader secondary peak at higher Ks values (≈ 1.9035 – 2.1065), reflecting the ancient 3R-WGD shared by all teleost (Figure 3D, upper panel). Additionally, Ks distributions for collinear orthologs between *S. nukiangensis* and *P. huangchuchieni* or *O. stewartii* also revealed bimodal patterns. Primary peaks occurred at $Ks \approx 0.1153$ (*Snuk_vs_Phua*) and $Ks \approx 0.0913$ (*Snuk_vs_Oste*), corresponding to more recent speciation events, while broader secondary peaks at $Ks \approx 1.9133$ and ≈ 1.9854 , respectively

(Figure 3D, lower panel), reflect deeper evolutionary relationships. The similarity of these secondary peaks suggests a shared ancient WGD event in the common ancestor of these species. Collectively, these findings indicate that *S. nukiangensis* has undergone both recent lineage-specific genome duplication and retains genomic signatures of an ancient WGD shared with *P. huangchuchieni* and *O. stewartii*. These results illuminate the complex evolutionary history of schizothoracine fishes and underscore the pivotal role of polyploidization and subsequent diversification in shaping cyprinid genomic architecture.

Population structure and demographic history of autotetraploid *Schizothorax nukiangensis* along an altitudinal gradient

To investigate the population genetic structure of *S. nukiangensis* along the Salween River, we sampled 37 individuals across a pronounced elevational gradient (632–1987 m asl) in The Hengduan Mountains (Figure 4A; Supplementary Table S1). The sampling sites represented habitats with distinct vertical stratification, reflecting the species' adaptation to heterogeneous, high-altitude riverine environments. Whole-genome resequencing of each individual was performed at an average depth of ~33.54×, generating

approximately 60.37 Gb of clean data per sample (Supplementary Table S1). Following read alignment to the *S. nukiangensis* reference genome, we identified 6553219 high-confidence SNPs. Based on these genome-wide SNPs, we constructed a ML unrooted phylogenetic tree, which revealed three major genetic clusters, designated as populations H, M, and L (popH, popM, and popL) (Figure 4B; Supplementary Table S1). PopH was predominantly distributed at high elevations (1793–1987 m asl), popM at mid-elevations (1190–1286 m asl), and popL at low elevations (632–683 m asl). Nonetheless, several individuals from each group were found outside their representative elevational ranges, indicating overlapping distributions and potential gene flow among populations. Population structure analysis based on genome-wide admixture modeling further revealed complex genetic patterns across elevations (Figure 4C). Although the lowest cross-validation (CV) error was observed at K=1, suggesting overall weak genome-wide differentiation, higher K-values (K=2–5) uncovered pronounced clustering patterns that aligned with the phylogenetic results. These patterns indicate the presence of cryptic population structure, likely shaped by ecological or historical factors rather than strict altitudinal isolation.

PSMC analysis of individual genomes from eight sampling

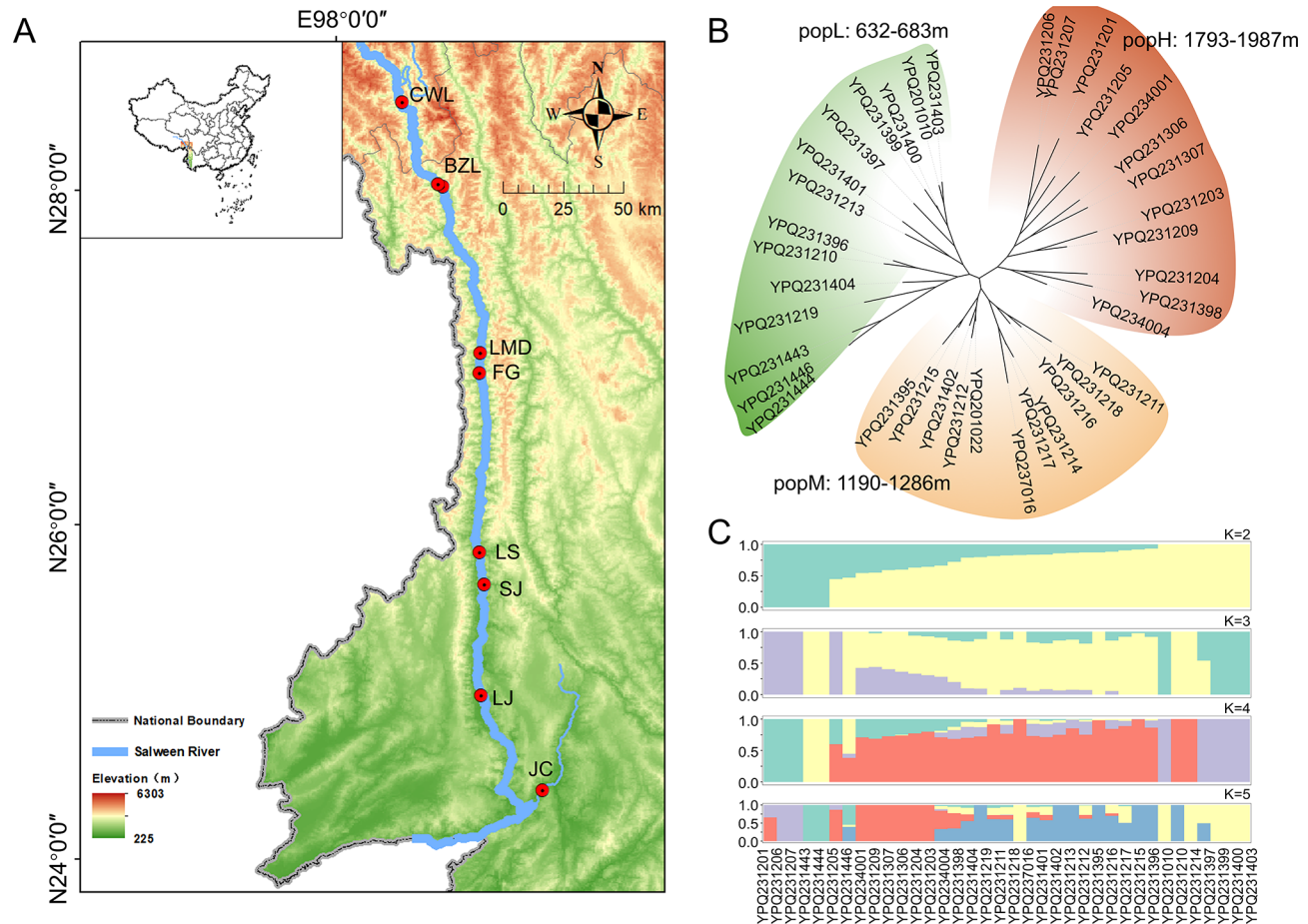


Figure 4 Sampling locations, phylogenetic relationships, and population structure of *Schizothorax nukiangensis* along an altitudinal gradient of the Salween River.

A: Sampling sites of 37 *Schizothorax nukiangensis* individuals along the Salween River in the region of The Hengduan Mountains, southwestern China. Elevations of sampling sites range from 632 to 1987 m asl, spanning habitats with pronounced vertical stratification. B: ML unrooted phylogenetic tree constructed from genome-wide SNPs, revealing three major genetic population (H, M, L) with individuals from different elevations intermixed within clusters. C: Population structure analysis inferred from genome-wide SNPs using admixture modeling at K=2 to K=5, showing varying degrees of clustering among individuals.

localities revealed highly consistent historical demographic trajectories across populations (Supplementary Figure S10). Estimates of effective population size (N_e) indicated a substantial decline beginning approximately 1 Ma, culminating in pronounced bottlenecks during the late Pleistocene ($\sim 10^4$ – 10^5 years ago), coinciding with major glaciation events on the Qinghai-Xizang Plateau. Following these bottlenecks, all populations exhibited clear signals of demographic recovery, with N_e gradually increasing toward the present. These results suggest that *S. nukiangensis* experienced synchronous population contractions and subsequent expansions in response to Pleistocene climate oscillations, which have shaped the species' present-day genetic diversity and distribution along the Salween River.

Population genomic analyses reveal candidate regions of altitudinal adaptation of autotetraploid *Schizothorax nukiangensis*

F_{ST} comparison analysis revealed consistent and strong differentiation signals on chromosome chr03A between popH (high altitude) and the mid- and low-altitude populations (popM and popL). Genomic regions exceeding the top 1% F_{ST} threshold (red dashed line) were designated as candidate selective sweeps (Figure 5A; Supplementary Figures S11A, B). By intersecting the top 1% F_{ST} peaks from both comparisons on chromosome chr03A, we identified a robust set of candidate genes predominantly enriched in transcriptional regulation (Supplementary Table S15). GO enrichment analysis of these candidate genes revealed significant molecular functions related to DNA binding and transcription factor activity. In particular, several key functions related to transcription regulation were identified, including nucleic acid binding transcription factor activity (GO:0001071), sequence-specific DNA binding (GO:0003700), and RNA polymerase II core promoter proximal region sequence-specific DNA binding (GO:0000978). These findings indicate that the genes within this region are involved in essential regulatory processes, such as the control of transcription and the binding of specific DNA sequences, which may contribute to adaptive responses to high-altitude environments. Additional functions like scavenger receptor binding (GO:0005124) and N-formyl peptide receptor activity (GO:0004982) further suggest roles in immune response and environmental stress adaptation.

Parallel Cross Population Composite Likelihood Ratio (XP-CLR) analyses comparing popH with both popM and popL consistently revealed strong signals of selection on chromosome chr04A. Genomic regions exceeding the top 1% threshold (red dashed line) were designated as candidate selective sweeps (Figure 5B; Figures S11C, D). Overlapping XP-CLR peaks revealed several candidate genes on chr04A, including genes associated with critical cellular functions such as protein interaction, receptor signaling, and gene regulation (Supplementary Table S16). Notably, genes such as transmembrane protein 131-like and endothelin receptor B (ETB) are implicated in cellular signaling pathways and may play roles in stress adaptation, as transmembrane proteins and receptors are often involved in transmitting signals in response to environmental changes. Tripartite motif-containing protein 2-like and mitogen-activated protein kinase kinase kinase 4-like are involved in protein regulation and immune response pathways, where tripartite motif proteins are known to regulate various cellular processes, and MAPK

signaling plays a critical role in stress responses. Genes like zinc finger protein basophilin-2-like and zinc finger protein 501-like indicate regulatory roles in gene expression, where zinc finger proteins are key in regulating transcription and protein synthesis. These findings suggest that the genes on chr04A are involved in a broad range of adaptive processes, including immune modulation, cellular signaling, and stress response mechanisms, all potentially contributing to adaptation in high-altitude environments.

To further refine the signals of local adaptation and identify robust candidate genes, we integrated three genomic metrics, F_{ST} , XP-CLR, and the nucleotide diversity ratio (π -ratio, $\log(\pi_L/\pi_H)$), by comparing popH and popL. Scatterplots of F_{ST} versus π -ratio (highlighted in red) delineated genomic regions exceeding the top 1% thresholds for both metrics (Figure 5C), and intersecting these with XP-CLR signals further strengthened evidence for selection. This integrative approach identified 11 robust candidate genes likely involved in altitudinal adaptation (Figure 5B; Supplementary Table S17), including *sdk1a* (cell adhesion), *fgf23* and *fgf6b* (skeletal and muscle development), and *mx* (immune defense). Notably, *fgf23* and *fgf6b* are associated with mineral metabolism and muscle development, respectively. These functions may underlie morphological or physiological traits that facilitate local adaptation to high-altitude river environments, such as strong currents and fluctuating environmental pressures. For example, enhanced muscle growth mediated by *fgf6b* could support locomotor performance in turbulent waters, while *fgf23*-related mineral regulation may influence skeletal maintenance under variable resource availability. These adaptations likely contribute to the ecological success of *S. nukiangensis* in its extreme habitat.

Detailed regional analyses of two representative candidate regions on Chr05B and Chr23B revealed coincident peaks in XP-CLR, F_{ST} , and $\log(\pi$ -ratio) profiles (Figures 5D, E), reinforcing the likelihood of strong local selection at these loci. Furthermore, a genotype heatmap for the candidate gene *ift70* (Figure 5F) showed distinct allele frequency patterns among popH, popM, and popL. Notably, popH exhibited a high frequency of homozygous mutations, suggestive of altitude-associated selective pressure.

Interestingly, F_{ST} and XP-CLR analyses, each performed as pairwise comparisons between popH and both popM and popL, highlighted divergent regions on different chromosomes: intersections of top F_{ST} peaks were enriched on Chr03A, while XP-CLR intersections were concentrated on Chr04A. This non-overlapping distribution suggests that each metric captures distinct signatures of selection, underscoring the complementary value of combining multiple approaches to detect selective sweeps. Final candidate genes were identified by integrating F_{ST} , XP-CLR, and $\log(\pi_L/\pi_H)$ metrics, specifically focusing on the comparison between popH and popL. This integrative analysis yielded a focused and robust set of loci potentially involved in altitudinal adaptation in *S. nukiangensis*, including several significantly differentiated genes such as *fgf6b*, *fgf23*, *mx*, and *ift70*.

Homeolog expression bias and altitude-associated transcriptional responses in the autotetraploid *Schizothorax nukiangensis*

To investigate expression bias among the four haplotypes of *S. nukiangensis*, we quantified transcript abundance for 15 063 homeologous gene groups present in a complete 1:1:1:1

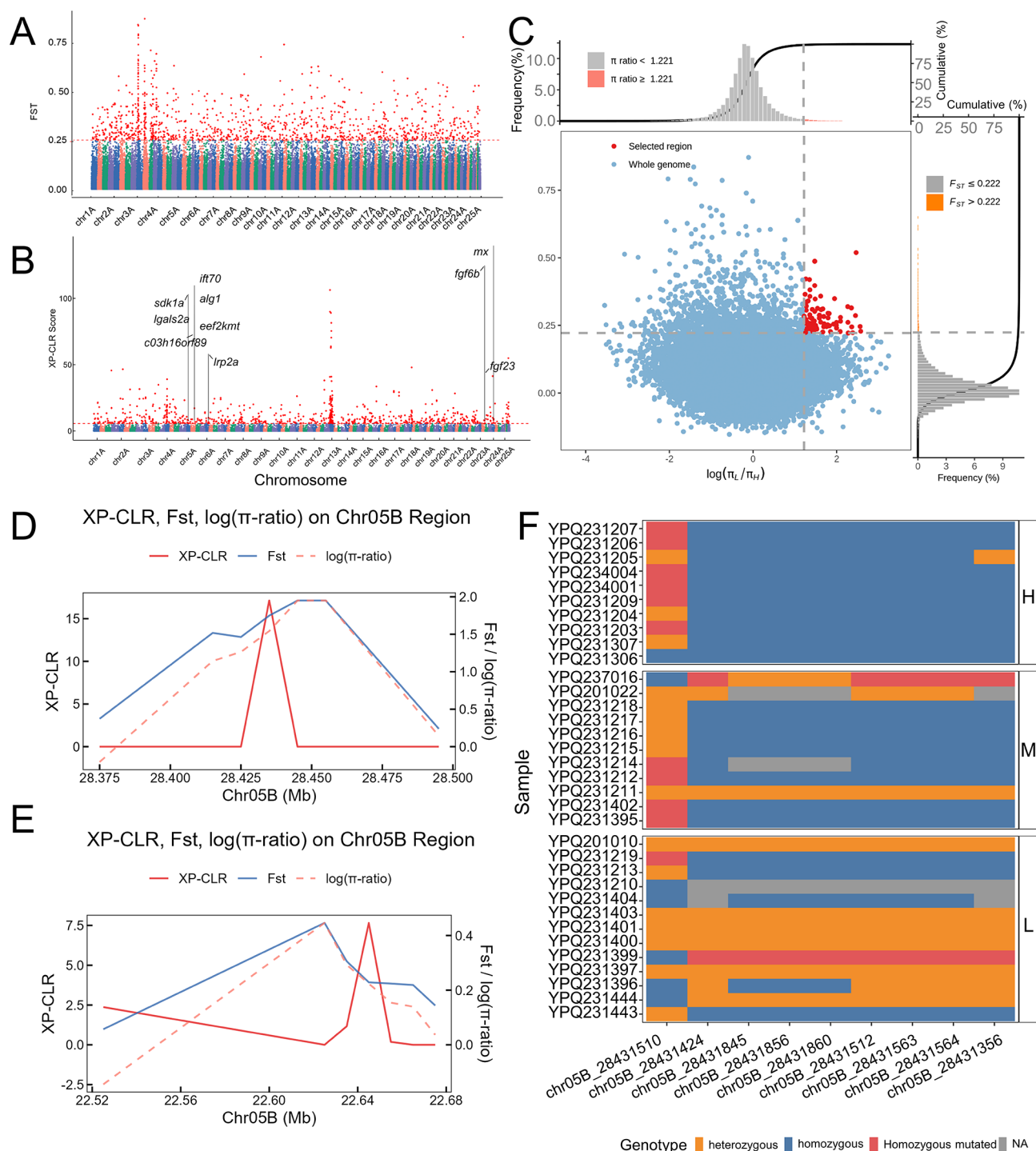


Figure 5 Genomic signatures of selection and candidate genes in *Schizothorax nukiangensis* populations

A: Manhattan plot of genome-wide F_{ST} values between high-altitude (popH) and low-altitude (popL) populations. Red dashed line indicates top 1% F_{ST} threshold. B: Manhattan plot of XP-CLR scores for the same population comparison; top 1% threshold marked with a red dashed line. C: Scatterplot of F_{ST} versus $\log(\pi_L/\pi_H)$, highlighting candidate regions (red) exceeding top 1% thresholds in both metrics. Marginal histograms show the distribution of each variable across the genome. D: Line plots of XP-CLR, F_{ST} , and $\log(\pi\text{-ratio})$ in candidate regions on Chr5B, respectively, showing co-localized peaks of selection signals. E: Line plots of XP-CLR, F_{ST} , and $\log(\pi\text{-ratio})$ in candidate regions on Chr23B, respectively, showing co-localized peaks of selection signals. F: Genotype heatmap of candidate gene *ift70*, illustrating distribution of heterozygous, homozygous, and homozygous mutant genotypes across individuals from popH, popM, and popL.

correspondence across hapA, hapB, hapC, and hapD (Figure 6A; Supplementary Table S18). Expression bias was defined as a ≥ 2 -fold increase ($\log_2(\text{fold change}) \geq 1$) in expression of a single haplotype compared to the other three. Analysis revealed that expression dominance was relatively rare: only 315 homeolog groups (2.1%) displayed significant

bias toward one haplotype, whereas the vast majority (97.9%) showed no detectable bias, indicating largely balanced expression among homeologs (Figure 6B). GO enrichment analysis of these biased genes revealed significant overrepresentation of functions related to mitochondrial activity, oxidative phosphorylation, and ribosomal structure.

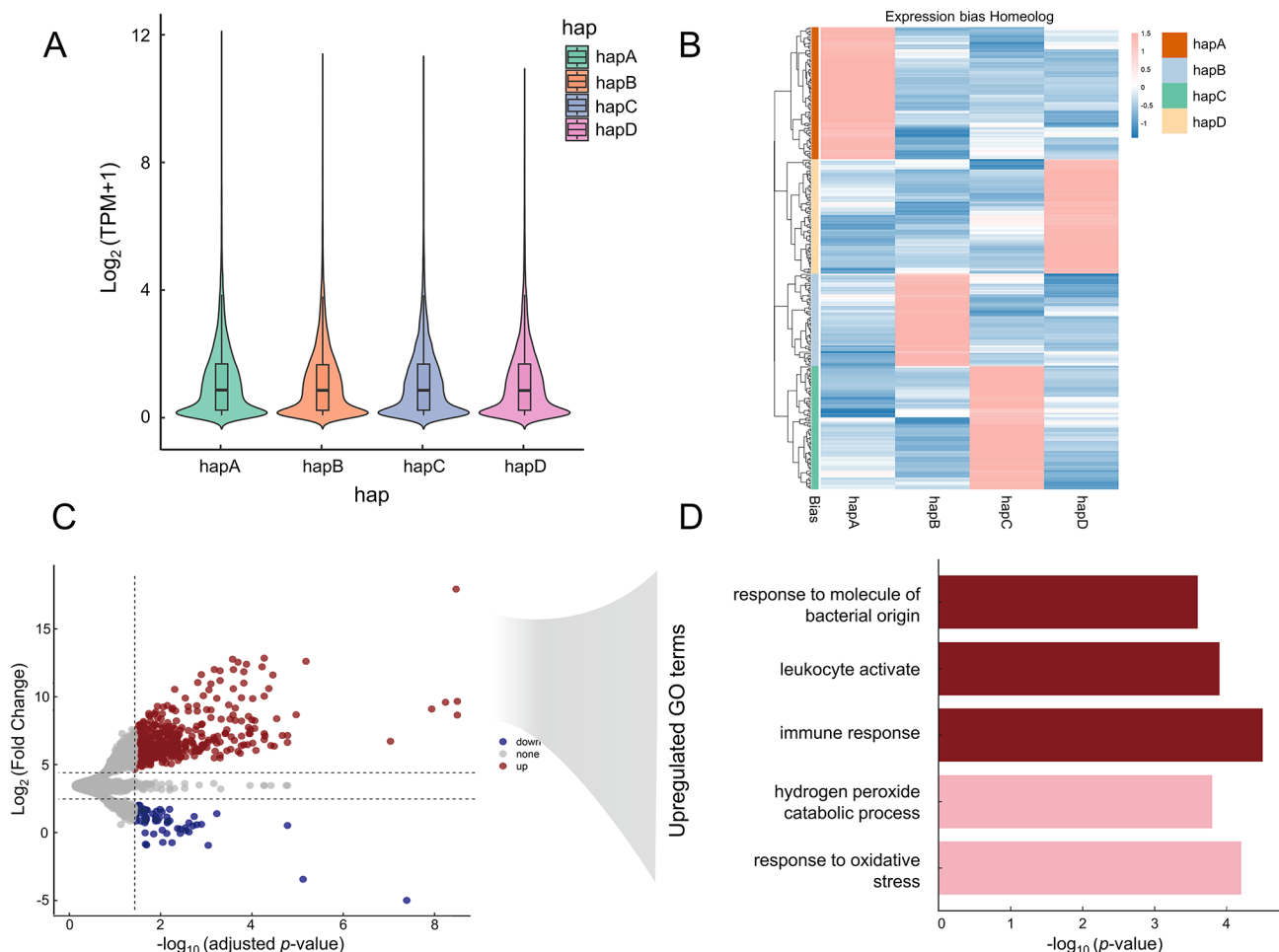


Figure 6 Homeolog expression bias and altitude-associated transcriptional responses in *Schizothorax nukiangensis*

A: Violin plots showing overall gene expression levels ($\log_2(\text{TPM}+1)$) across four haplotypes (hapA, hapB, hapC, hapD) in 15063 homeolog groups. B: Heatmap of 315 homeolog groups with significant expression bias ($\log_2(\text{fold-change}) \geq 1$) towards a single haplotype; color intensity indicates relative expression levels. C: Volcano plot of differential expression analysis in intestinal transcriptomes comparing high- (1255 m asl) and low-altitude (718 m asl) groups; red dots denote upregulated, blue dots downregulated genes ($P\text{-adjusted} < 0.05$, $|\log_2(\text{fold change})| \geq 1$). D: GO enrichment analysis of upregulated genes from high-altitude samples, highlighting terms related to immune response, leukocyte activation, and oxidative stress adaptation.

Specifically, enriched terms included respiratory chain (GO:0070469), mitochondrial inner membrane (GO:0005743), and ATP metabolic process (GO:0046034), indicating a functional preference toward energy production and protein synthesis. Notably, several components of the electron transport chain, including NADH dehydrogenase complex and respiratory chain complex I, were also significantly enriched. These biased genes may reflect subgenome-specific optimization of energy metabolism and translational control, potentially contributing to the physiological demands of surviving in cold, oxygen-variable, high-altitude, and high-flow river environments such as the Salween River (Supplementary Table S19).

To explore transcriptional mechanisms underlying local adaptation along the altitudinal gradient, we performed differential expression analysis of intestinal transcriptomes from two elevation groups: group 1 (1068 m asl) and group 2 (718 m asl) (Figure 6C). We identified 481 significantly upregulated and 60 downregulated genes ($|\log_2(\text{fold change})| \geq 1$, $P\text{-adjusted} < 0.05$) in the high-altitude group relative to the low-altitude group. GO enrichment analysis of upregulated genes revealed significant overrepresentation of biological processes related to innate and adaptive immunity, including

response to bacterial molecules, leukocyte activation, and immune signaling pathways, as well as oxidative stress responses such as hydrogen peroxide catabolism (Figure 6D; Supplementary Table S20). These enrichments suggest that *S. nukiangensis* individuals inhabiting higher altitudes enhance immune function and oxidative stress defenses to cope with increased environmental stressors, such as hypoxia, temperature fluctuations, and pathogen pressures, representative of fast-flowing mountain streams.

DISCUSSION

Polyploidy might be a major driving force in teleost diversification; yet understanding recent polyploidization events in high-altitude species remains challenging due to genome complexity and high redundancy. Within the family Cyprinidae, both autopolyploidy and allopolyploidy have occurred repeatedly, contributing significantly to species richness and ecological diversity (Glasauer & Neuhauss, 2014; Leggett & Iwama, 2003). However, despite the importance of autopolyploidy, the genomes have remained poorly characterized due to technical difficulties in resolving highly similar homologous sequences. Genome assembly of

polyploid species has been particularly challenging; early efforts often resulted in collapsed assemblies or artificial merging of homologous regions, especially in autopolyploids where sequence similarity among homeologs is high. Here, we present the first high-quality, haplotype-resolved genome assembly of the autotetraploid *S. nukiangensis* (Figures 1B, C), providing a foundational resource for investigating polyploid genome evolution in schizothoracine fishes. The exceptional continuity, completeness, and phasing accuracy of our assembly demonstrate the feasibility of generating high-quality, haplotype-resolved genomes for complex polyploid species (Figure 1D). This assembly not only enriches genomic resources for Cyprinid but also establishes a robust foundation for future studies on polyploidy in teleost fishes.

Polyploidy has emerged as a key driver of diversification in fishes inhabiting extreme environments, with recent studies highlighting its repeated occurrence among schizothoracine fishes of the Qinghai-Xizang Plateau (Li et al., 2024c; Liu et al., 2019; Xiao et al., 2020). Our finding of an autotetraploid origin for *S. nukiangensis* is supported by multiple lines of evidence (Figure 2; Supplementary Figure S1C). The genome size of *S. nukiangensis* is 1.80 Gb (Supplementary Figure S1A; Supplementary Table S4), approximately twice that of diploid cyprinids, consistent with other known tetraploid species such as *Cy. carpio* (Xu et al., 2019a) and *Ca. auratus* (Luo et al., 2020). Its chromosome number ($2n=100$) is also comparable to that of other autotetraploid cyprinids (Figure 1B; Supplementary Table S21). K-mer analysis shows an autotetraploid pattern, with four copies of most loci (Figure 1D; Supplementary Figure S1C, S3). The high proportion of homozygous k-mers and the low heterozygosity further support an autotetraploid origin (Supplementary Table S10). Haplotype analysis demonstrates that the four assembled haplotypes of *S. nukiangensis* have similar genome sizes and gene numbers and cluster closely in phylogenetic trees (Figure 2C; Supplementary Figure S7; Supplementary Table S5), which is consistent with an autotetraploid origin. In contrast, allotetraploid species such as *Cy. carpio* and *Ca. auratus* display distinct subgenomic topologies, with subgenomes grouping into different phylogenetic clades. These results, supported by chromosome-level phylogenetic trees, confirm that *S. nukiangensis* is an autotetraploid species, highlighting the importance of polyploidization in its evolutionary history. The autotetraploidy of *S. nukiangensis* may have arisen through multiple mechanisms. One possibility is the formation of unreduced gametes caused by meiotic errors, producing gametes with double the normal chromosome number, which subsequently fused to form a tetraploid organism (Leggatt & Iwama, 2003). Alternatively, spontaneous genome duplication may have occurred due to mitotic errors. Environmental factors may also have facilitated polyploidization (Otto, 2007). Cooling events on the Qinghai-Xizang Plateau could have forced populations into geothermal refugia, such as hot springs. Temperature shocks during reproduction in these environments may have induced genome duplication in fertilized eggs, promoting polyploid formation (Gu et al., 2024; Yan et al., 2022).

Our *Ks*-based analyses of *S. nukiangensis* revealed clear signatures of both ancient and recent WGD events (Figure 3C; Supplementary Figure S9). A prominent *Ks* peak at very low divergence (~ 0.0018 – 0.0066) indicates a recent autotetraploidization event estimated at ~ 0.68 Ma (Figure 3C), while broader secondary peaks align with the ancient 3R-

WGD shared by all teleosts (Leggatt & Iwama, 2003). Similar patterns were reported for *Schizothorax o'connori*, in which Xiao et al. identified a *Ks* peak at ~ 0.0086 corresponding to a recent WGD ~ 1.23 Ma (Xiao et al., 2020). These independent but temporally proximate polyploidization events occurred during Pleistocene uplift phases of the Qinghai-Xizang Plateau, a period marked by dramatic climatic and geological changes that imposed strong selective pressures (Wang et al., 2016; Zhang et al., 2018). Recent studies on snow carps across the plateau further revealed multiple independent and recurrent polyploidization events within schizothoracine fishes, each associated with population divergence and ecological adaptation (Li et al., 2024c; Wang et al., 2016). Climate change on the Qinghai-Xizang Plateau, particularly cooling and glacial expansion during the Pleistocene, profoundly influenced species distributions, population dynamics, and adaptive evolution. As temperatures declined and glaciers advanced, formerly warm and humid environments became colder and drier, causing habitat contraction and population isolation, which promoted speciation and increased biodiversity (Hewitt, 2000). The tectonic uplift of the plateau, which began in the Miocene and intensified during the Pliocene (Pan et al., 2012), further amplified these climatic shifts by altering hydrological systems and fragmenting habitats. These conditions may have facilitated autotetraploidization in species such as *S. nukiangensis* (Jia et al., 2020). PSMC analysis reveals a population bottleneck in *S. nukiangensis* during the Late Pleistocene (Supplementary Figure S10), consistent with climate-driven fluctuations in population size (Maron et al., 2015; Zhou et al., 2023). This bottleneck likely resulted from repeated cycles of glacial advance and uplift-induced habitat instability. Over the past 1000 years, the continued decline in effective population size corresponds with both climatic variability and intensified human disturbance. Human activities, including hydropower development, habitat destruction, and urban expansion, have further reduced suitable habitats and increased adaptive pressure on the species (Li et al., 2024a).

Expanded gene families and positively selected genes reveal convergent molecular strategies for high-altitude adaptation. In *S. nukiangensis*, gene family expansions were significantly enriched in immunity, nucleic acid metabolism, transposition, cell repair, and metabolic regulation—paralleling patterns observed in other high-altitude schizothoracines like *Schizopygopsis* (Supplementary Figure S8; Supplementary Table S11; Supplementary Table S14) (Zhou et al., 2020, 2023, 2025). Meanwhile, we found that the unique gene family is involved in several key biological processes, particularly transposition, phospholipid metabolism, cell differentiation, and metabolic regulation. These functions may influence cell activity, genome stability, and immune regulation (Supplementary Table S13). The recent expansion of immune-related genes and bursts of TEs suggest a potential link between genome instability and rapid adaptation, consistent with the hypothesis that WGD provides genomic raw material under environmental stress (Leggatt & Iwama, 2003). However, compared with other schizothoracine species, the genome of *S. nukiangensis* shows more pronounced changes in transposon activity and genomic remodeling. These changes are likely linked to its more recent autotetraploidization event and the resulting higher genome instability. Such instability may have generated greater adaptive variation, facilitating the rapid evolution of *S.*

nukiangensis in the highly dynamic high-altitude environment. Positive selection analyses further identified genes involved in neural development, neurotransmission, and immune regulation, echoing signatures from *S. o'connori* (Gao et al., 2024) and *Schizopygopsis* (Zhou et al., 2023). These convergent genomic patterns point to a shared adaptive strategy involving nervous system modulation, energy homeostasis, and immune enhancement to cope with the hypoxia, cold, and pathogen challenges of high-altitude river environments.

Although *S. nukiangensis* is continuously distributed along the Salween River, population genomic analyses reveal clear genetic structuring into high-, mid-, and low-altitude population (Figure 4). The F_{ST} values reveal varying degrees of genetic differentiation along the altitudinal gradient (Supplementary Table S22). While modest in magnitude, the F_{ST} values still reflect some genetic differentiation between the high-altitude population and the medium- and low-altitude groups, likely reflecting adaptive responses of *S. nukiangensis* to harsher high-altitude conditions such as reduced oxygen availability and greater temperature fluctuations. In contrast, the relatively low F_{ST} value between the medium- and low-altitude populations suggests weaker genetic differentiation, potentially due to higher gene flow or similar environmental pressures between these populations. Together, these patterns demonstrate that altitude and associated environmental factors have played an important role in shaping the genetic structure of *S. nukiangensis* and driving local adaptation. Although the weak population structure observed in the admixture analyses (with the lowest CV error at $K=1$) may reduce the resolution of selection scans, the results are still informative for identifying loci under environmental selection, particularly in high-altitude populations. Additionally, the phylogenetic tree constructed from SNP data separates the populations into three major clusters, further supporting the presence of population structure across high-, middle-, and low-altitude groups. This structure is likely shaped by ecological barriers and life-history traits. Its demersal, adhesive-spawning strategy likely limits juvenile dispersal, promoting local differentiation despite geographic continuity (Chen et al., 2015). Occasional mismatches between sampling location and genetic clusters may result from downstream displacement during seasonal floods. This fine-scale structure aligns with previous findings in schizothoracines such as *S. o'connori* (Gao et al., 2024) and *Schizopygopsis younghusbandi* (Zhou et al., 2025), where elevational gradients and river fragmentation shaped divergence (Gao et al., 2024; Zhou et al., 2023).

Genome-wide selective sweep analyses identified candidate genes likely involved in high-altitude adaptation in *S. nukiangensis*, including immune responses (e.g., *sdk1a*, *lgals2a*, *mx*), cellular transport (e.g., *ift70*), and metabolic regulation (e.g., *eef2kmt*, *alg1*). Additionally, *fgf6b* and *fgf23* are associated with muscle growth regulation and phosphate metabolism, respectively (Figures 5A–C; Supplementary Table S17). These genes may support survival in hypoxic, fast-flowing mountain streams. Direct recurrence of the exact same loci in other Schizothoracinae is rare (likely due to lineage-specific demography and heterogeneous selection targets). However, several of our candidates fall into pathways repeatedly implicated in Qinghai-Xizang Plateau fish adaptation. Similar signatures in energy metabolism and DNA repair pathways have been reported in *S. o'connori* and

Schizopygopsis younghusbandi, suggesting convergent molecular adaptations among schizothoracines (Xu et al., 2023; Yu et al., 2017; Zhou et al., 2023). Notably, *mx* gene has been characterized in *Schizothorax richardsonii*, supporting recurrent selection on immune modules in snow carp (Saxena et al., 2018). In teleost, *fgf6b* is strongly expressed in muscle tissues and promotes muscle fiber growth and differentiation (Hu et al., 2023; Xu et al., 2019b; Yang et al., 2024). Given the high mechanical and energetic demands of swimming in turbulent, oxygen-poor plateau rivers, positive selection on *fgf6b* might have contributed to improve muscle performance and endurance, optimizing movement and propulsion in turbulent. Differentiation at loci such as *ift70* aligns with altitude-related expression patterns previously reported in *S. nukiangensis* (Yu et al., 2017). Furthermore, metabolomic differences between high- and low-altitude populations (Xu et al., 2023), particularly in lipid metabolism and oxidative stress, correspond to the functional roles of these candidate genes. We further validated the expression of several candidate genes using qRT-PCR (Supplementary Table S23). With the exception of *lgals2a*, all other genes showed higher expression levels in the high-altitude population compared with the low-altitude population. Notably, *lrp2a* and *fgf6b* exhibited significantly elevated expression in the high-altitude group (Supplementary Figure S12). Together, these findings highlight consistent local adaptation at key functional loci, despite overall weak population structure—underscoring the molecular basis of ecological specialization in high-altitude riverine environments.

Our analysis of expression bias among the four haplotypes of *S. nukiangensis* revealed largely balanced expression, with only 2.1% of homeologous gene groups showing significant bias toward a single haplotype. The WGD event in *S. nukiangensis* occurred relatively recently (0.68 Ma), much later than in other schizothoracine species such as *Gymnocypris przewalskii* (2.55 Ma) (Tian et al., 2022) and *S. o'connori* (1.23 Ma) (Xiao et al., 2020). In those species, subgenome differentiation and functional specialization have occurred following WGD. For example, in *G. przewalskii*, gene expression analyses revealed subfunctionalization and neofunctionalization of duplicated genes, whereas in *S. o'connori*, bursts of transposable elements contributed to rediploidization. In contrast, *S. nukiangensis* may not yet have undergone extensive subgenome divergence, which could explain the lower homeolog expression bias observed in this species. This weak bias may also indicate a higher degree of functional redundancy among homeologs, allowing one copy to compensate for another and maintain balanced gene expression. Additionally, regulatory mechanisms may stabilize gene dosage and prevent harmful over- or under-expression, particularly for genes involved in essential processes such as development and stress response. Collectively, these patterns suggest that *S. nukiangensis* is in an early post-WGD stage and is maintaining genomic stability through balanced homeolog expression.

Differential expression analysis between high- and low-altitude populations identified multiple immune- and oxidative stress-related genes that were upregulated at higher elevations. This suggests that high-altitude individuals experience greater environmental stress, including hypoxia and temperature fluctuations, and have adapted by enhancing immune responses and oxidative stress defenses. These

results reveal the transcriptional mechanisms contributing to local adaptation across the altitudinal gradient and emphasize the role of immune and metabolic processes in surviving the extreme conditions of fast-flowing, high-altitude rivers such as the Salween.

Recent autotetraploidy in *S. nukiangensis*—coincident with late-Pleistocene climatic and geomorphic change—provides the genomic foundation for adaptation to high-altitude rivers. Within this background, TE dynamics and genomic innovation, together with expansions in immunity, metabolism, cellular transport and repair gene families, point to increased genomic flexibility under environmental stress. PSMC indicates bottleneck–expansion cycles consistent with glacial oscillations. Candidate genes associated with high-altitude adaptation are consistent with the physiological demands of cold, fast-flowing, and food-limited habitats. Notably, minimal homeolog expression bias suggests dosage balance and regulatory stability after a recent WGD, potentially buffering essential pathways while permitting targeted divergence at ecologically relevant loci. While these multi-layer patterns are mutually reinforcing, they remain largely correlative. Genome-wide structure is weak, and selection inferences can be sensitive to filtering and linkage, functional validation through gene editing, expression assays, and physiological experiments will be essential to confirm their adaptive roles. Future integrative studies combining genomics with experimental approaches will further clarify the molecular mechanisms underlying these adaptations.

CONCLUSION

At the core of this work is the high-quality, haplotype-resolved genome assembly of *S. nukiangensis*, the first of its kind for an autotetraploid schizothoracine species. This genomic resource provides a foundation for future investigations into polyploidy and high-altitude adaptation. Our results confirm that *S. nukiangensis* is an autotetraploid, with tetraploidization occurring approximately 0.68 Ma. This recent polyploidy event, coupled with active transposable elements, has driven structural variation and genomic innovation, likely facilitating adaptation to the extreme conditions of the Qinghai-Xizang Plateau. Gene family expansions in immune and energy-related pathways support the hypothesis that polyploidy enhances ecological plasticity. Moreover, population genomic analyses along elevational gradients reveal subtle genetic structure shaped by both historical climate fluctuations and local environmental pressures.

In conclusion, *S. nukiangensis* illustrates how recent polyploidy can fuel evolutionary innovation and local adaptation in high-altitude environments, offering key insights into genome evolution in polyploid vertebrates.

SCIENTIFIC FIELD SURVEY PERMISSION INFORMATION

All animal collections complied with the Wildlife Protection Act of China. Collection permits were issued by the Animal Care and Ethics Committee of Yunnan University (YNU20220348).

DATA AVAILABILITY

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive (GSA; <https://ngdc.cncb.ac.cn/gsa>) at the National Genomics Data Center (NGDC), China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (CNCB-NGDC Members & Partners, 2025; Ma et al., 2025), under accession numbers CRA029232 (RNA-seq, Iso-Seq, Hi-C, and PacBio HiFi

data) and CRA020723 (whole-genome resequencing data). The data have also been deposited in the Science Data Bank (<https://doi.org/10.57760/sciencedb.39040>; <https://doi.org/10.57760/sciencedb.j00139.00319>) and the NCBI Sequence Read Archive under BioProject ID PRJNA1475900. The genome assembly has been deposited in the Genome Warehouse (GWH) at NGDC under accession number GWHHAFH000000000.1 and is publicly accessible at <https://ngdc.cncb.ac.cn/gwh>. The genome annotation results have been made available in the Figshare database (<https://doi.org/10.6084/m9.figshare.30507266>).

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

J.L., G.L.L., X.Y.C., Y.L. and Y.P.Z. conceived and designed the work; C.W., S.W.L., G.Z.L., B.Y.S. and C.Z.D. collected the samples and performed key experiments. C.W. conducted the main data analyses and drafted the initial manuscript; X.D.G., Y.H.D. and T.C.W. contributed to part of the data analysis; J.L., X.D.G., G.L.L., Y.L., X.Y.C. revised and finalized the manuscript. All authors read and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

We gratefully acknowledge the Bio-X Center for Interdisciplinary Innovation at Yunnan University for providing research facilities and technical support. We also sincerely thank the Young Elite Scientists Sponsorship Program for PhD Candidates, supported by the China Association for Science and Technology. We thank Kunming Cell Bank, Chinese Academy of Sciences (Kunming, China) and the Animal Bank at the Germplasm Bank of Wild Species (<https://cstr.cn/31121.02.GBOWS.AB>) for providing technical support.

REFERENCES

- Almeida-Silva F, Van De Peer Y. 2023. Whole-genome duplications and the long-term evolution of gene regulatory networks in angiosperms. *Molecular Biology and Evolution*, **40**(7): msad141.
- Brúna 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.
- Cai LM, Xi ZX, Amorim AM, et al. 2019. Widespread ancient whole-genome duplications in malpighiales coincide with eocene global climatic upheaval. *New Phytologist*, **221**(1): 565–576.
- 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.
- Carruthers T, Gonçalves DJP, Li P, et al. 2024. Repeated shifts out of tropical climates preceded by whole genome duplication. *New Phytologist*, **244**(6): 2561–2575.
- Chang CC, Chow CC, Tellier LCAM, et al. 2015. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, **4**: 7.
- Chaplin G. 2005. Physical geography of the gaoligong shan area of southwest china in relation to biodiversity. *Proceedings of the California Academy of Sciences*, **56**(28): 527–556.
- Chen CJ, Wu Y, Li JW, et al. 2023. TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining. *Molecular Plant*, **16**(11): 1733–1742.
- Chen H, Patterson N, Reich D. 2010. Population differentiation as a test for

- selective sweeps. *Genome Research*, **20**(3): 393–402.
- Chen L, Li BJ, Chen BH, et al. 2021. Chromosome-level genome of *Poropuntius huangchuchieni* provides a diploid progenitor-like reference genome for the allotetraploid *Cyprinus carpio*. *Molecular Ecology Resources*, **21**(5): 1658–1669.
- Chen WT, Du K, He SP. 2015. Genetic structure and historical demography of *schizothorax nukiangensis* (cyprinidae) in continuous habitat. *Ecology and Evolution*, **5**(4): 984–995.
- Chen XY. 2013. Checklist of fishes of Yunnan. *Zoological Research*, **34**(4): 281–343. (in Chinese)
- Cheng HY, Concepcion GT, Feng XW, et al. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods*, **18**(2): 170–175.
- Cheng HY, Jarvis ED, Fedrigo O, et al. 2022. Haplotype-resolved assembly of diploid genomes without parental data. *Nature Biotechnology*, **40**(9): 1332–1335.
- Chuong EB, Elde NC, Feschotte C. 2017. Regulatory activities of transposable elements: From conflicts to benefits. *Nature Reviews Genetics*, **18**(2): 71–86.
- Cingolani P, Platts A, Wang LL, et al. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, **6**(2): 80–92.
- CNCB-NGDC Members and Partners. 2025. Database resources of the national genomics data center, china national center for bioinformation in 2025. *Nucleic Acids Research*, **53**(D1): D30–D44.
- Crellen T, Allan F, David S, et al. 2016. Whole genome resequencing of the human parasite *Schistosoma mansoni* reveals population history and effects of selection. *Scientific Reports*, **6**(1): 20954.
- Dai YG, Han HF. 2018. Karyological analysis of two species in the subfamily schizothoracinae (cypriniformes: Cyprinidae) from China, with notes on karyotype evolution in schizothoracinae. *Turkish Journal of Fisheries and Aquatic Sciences*, **18**(1): 175–186.
- Danecek P, Auton A, Abecasis G, et al. 2011. The variant call format and VCFtools. *Bioinformatics*, **27**(15): 2156–2158.
- David KT. 2022. Global gradients in the distribution of animal polyploids. *Proceedings of the National Academy of Sciences of the United States of America*, **119**(48): e2214070119.
- David L, Blum S, Feldman MW, et al. 2003. Recent duplication of the common carp (*cyprinus carpio* L.) genome as revealed by analyses of microsatellite loci. *Molecular Biology and Evolution*, **20**(9): 1425–1434.
- Deng T, Wu FX, Zhou ZK, et al. 2020. Tibetan Plateau: an evolutionary junction for the history of modern biodiversity. *Science China Earth Sciences*, **63**(2): 172–187.
- Ding LY, Tao J, Ding CZ, et al. 2019. Hydrogeomorphic factors drive differences in otolith morphology in fish from the Nu-Salween river. *Ecology of Freshwater Fish*, **28**(1): 132–140.
- Du TQ, Ding CZ, Yang K, et al. 2024. A global dataset on species occurrences and functional traits of schizothoracinae fish. *Scientific Data*, **11**(1): 272.
- Durand NC, Robinson JT, Shamim MS, et al. 2016. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Systems*, **3**(1): 99–101.
- Edgar RC. 2022. Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nature Communications*, **13**(1): 6968.
- Emms DM, Kelly S. 2019. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, **20**(1): 238.
- 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.
- Gabriel L, Brūna 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, Brūna T, et al. 2021. TSEBRA: Transcript selector for BRAKER. *BMC Bioinformatics*, **22**(1): 566.
- Gao K, He Z, Xiong JX, et al. 2024. Population structure and adaptability analysis of *schizothorax o'connori* based on whole-genome resequencing. *BMC Genomics*, **25**(1): 145.
- Glasauer SMK, Neuhauss SCF. 2014. Whole-genome duplication in teleost fishes and its evolutionary consequences. *Molecular Genetics and Genomics*, **289**(6): 1045–1060.
- Grabherr MG, Haas BJ, Yassour M, et al. 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nature Biotechnology*, **29**(7): 644–652.
- Gu HR, Wang S, Yang CH, et al. 2024. Global cooling and hot springs may have induced autotetraploidy and autohexaploidy in schizothorax ancestors, and its implications for polyploid breeding. *Aquaculture*, **584**: 740659.
- 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.
- Hämälä T, Moore C, Cowan L, et al. 2024. Impact of whole-genome duplications on structural variant evolution in *Cochlearia*. *Nature Communications*, **15**(1): 5377.
- Hewitt G. 2000. The genetic legacy of the Quaternary ice ages. *Nature*, **405**(6789): 907–913.
- Hu ZC, Li HD, Liu S, et al. 2023. Assessment of black soldier fly (*Hermetia illucens*) larvae meal as a potential substitute for soybean meal on growth performance and flesh quality of grass carp ctenopharyngodon idellus. *Animal Nutrition*, **14**: 425–449.
- Jia YT, Sui X, Chen YF, et al. 2020. Climate change and spatial distribution shaped the life-history traits of schizothoracine fishes on the Tibetan Plateau and its adjacent areas. *Global Ecology and Conservation*, **22**: e01041.
- Jurka J, Kapitonov VV, Pavlicek A, et al. 2005. Repbase update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research*, **110**(1–4): 462–467.
- Kawecki TJ, Ebert D. 2004. Conceptual issues in local adaptation. *Ecology Letters*, **7**(12): 1225–1241.
- Keilwagen J, Hartung F, Paulini M, et al. 2018. Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinformatics*, **19**(1): 189.
- Kim D, Paggi JM, Park C, et al. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, **37**(8): 907–915.
- Kumar S, Stecher G, Suleski M, et al. 2017. TimeTree: A resource for timelines, timetrees, and divergence times. *Molecular Biology and Evolution*, **34**(7): 1812–1819.
- Leggatt RA, Iwama GK. 2003. Occurrence of polyploidy in the fishes. *Reviews in Fish Biology and Fisheries*, **13**(3): 237–246.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics*, **25**(14): 1754–1760.
- Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature*, **475**(7357): 493–496.
- Li HR, Tan L, Li XF, et al. 2024a. Aquatic protected area system in the Qinghai–Tibet plateau: Establishment, challenges and prospects. *Frontiers*

in *Ecology and Evolution*, **12**: 1204494.

- Li Q, Li XY, Hu WQ, et al. 2024b. Mammals of Gaoligong mountain in China: Diversity, distribution, and conservation. *Zoological Research: Diversity and Conservation*, **1**(1): 3–19.
- Li XX, Wang M, Zou M, et al. 2024c. Recent and recurrent autopolyploidization fueled diversification of snow carp on the Tibetan Plateau. *Molecular Biology and Evolution*, **41**(11): msae221.
- Liao Y, Smyth GK, Shi W. 2014. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**(7): 923–930.
- Lien S, Koop BF, Sandve SR, et al. 2016. The Atlantic salmon genome provides insights into rediploidization. *Nature*, **533**(7602): 200–205.
- Liu HP, Xiao SJ, Wu N, et al. 2019. The sequence and *de novo* assembly of *Oxygymnocypris stewartii* genome. *Scientific Data*, **6**(1): 190009.
- Liu SR, Fornacca D, Yang DD, et al. 2020. Species richness and biodiversity significance of alpine micro-waterbody systems in Gaoligong mountain, northwest Yunnan, China. *Journal of Mountain Science*, **17**(4): 907–918.
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**(12): 550.
- Lu XM, Yu XF, Li GQ, et al. 2024. Genome assembly of autotetraploid *Actinidia arguta* highlights adaptive evolution and enables dissection of important economic traits. *Plant Communications*, **5**(6): 100856.
- Luo J, Chai J, Wen YL, et al. 2020. From asymmetrical to balanced genomic diversification during rediploidization: Subgenomic evolution in allotetraploid fish. *Science Advances*, **6**(22): eaaz7677.
- Lynch M. 2010. Evolution of the mutation rate. *Trends in Genetics*, **26**(8): 345–352.
- Ma R. 2022. The Study on the Biology and Stock Dynamics of *schizothorax nukiangensis*. Master thesis, Southwest University. (in Chinese)
- Ma YK, Zhao XT, Jia YK, et al. 2025. The updated genome warehouse: Enhancing data value, security, and usability to address data expansion. *Genomics, Proteomics & Bioinformatics*, **23**(1): qzaf010.
- Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics*, **27**(6): 764–770.
- Maron M, Mcalpine CA, Watson JEM, et al. 2015. Climate-induced resource bottlenecks exacerbate species vulnerability: a review. *Diversity and Distributions*, **21**(7): 731–743.
- McKenna A, Hanna M, Banks E, et al. 2010. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**(9): 1297–1303.
- Mendes FK, Vanderpool D, Fulton B, et al. 2021. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics*, **36**(22-23): 5516–5518.
- Mezzasalma M, Brunelli E, Odierna G, et al. 2023. Evolutionary and genomic diversity of true polyploidy in tetrapods. *Animals*, **13**(6): 1033.
- 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.
- Mu X, Liu QS, Wu HH, et al. 2025. Domestication and artificial breeding experiment of wild broodstock of *schizothorax nukiangensis*. *Scientific Fish Farming*, (8): 15–17. (in Chinese)
- Oliver KR, Greene WK. 2009. Transposable elements: powerful facilitators of evolution. *Bioessays*, **31**(7): 703–714.
- Otto SP. 2007. The evolutionary consequences of polyploidy. *Cell*, **131**(3): 452–462.
- Pan GT, Wang LQ, Li RS, et al. 2012. Tectonic evolution of the Qinghai-Tibet plateau. *Journal of Asian Earth Sciences*, **53**: 3–14.
- Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. Genomescope 2.0 and smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications*, **11**: 1432.
- Rhie A, Walenz BP, Koren S, et al. 2020. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology*, **21**(1): 245.
- Royden LH, Burchfiel BC, Van Der Hilst RD. 2008. The geological evolution of the Tibetan Plateau. *Science*, **321**(5892): 1054–1058.
- Saxena A, Belwal K, Chauhan A, et al. 2018. Interferon induced Mx protein from Indian snow trout *Schizothorax richardsonii* (gray) lacks critical functional features unlike its mammalian homologues. *Computational Biology and Chemistry*, **73**: 31–40.
- Sémon M, Wolfe KH. 2007. Consequences of genome duplication. *Current Opinion in Genetics & Development*, **17**(6): 505–512.
- Shumate A, Wong B, Pertea G, et al. 2022. Improved transcriptome assembly using a hybrid of long and short reads with stringtie. *PLoS Computational Biology*, **18**(6): e1009730.
- 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 PC, Jiao BB, Yang YZ, et al. 2022. WGDl: a user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. *Molecular Plant*, **15**(12): 1841–1851.
- Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, **56**(4): 564–577.
- Tang HB, Krishnakumar V, Zeng XF, et al. 2024. JCVI: a versatile toolkit for comparative genomics analysis. *iMeta*, **3**(4): e211.
- Tarailo-Graovac M, Chen NS. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*, **25**(1): 4.10. 11–14.10. 14.
- Tegenfeldt F, Kuznetsov D, Manni M, et al. 2025. OrthoDB and BUSCO update: annotation of orthologs with wider sampling of genomes. *Nucleic Acids Research*, **53**(D1): D516–D522.
- Tian F, Liu SJ, Zhou BZ, et al. 2022. Chromosome-level genome of Tibetan naked carp (*Gymnocypris przewalskii*) provides insights into Tibetan highland adaptation. *DNA Research*, **29**(4): dsac025.
- Tu GX, Yan ZY, Zhang L, et al. 2025. A chromosome-level genome assembly of the mud carp (*Cirrhinus molitorella*). *Scientific Data*, **12**(1): 285.
- Van De Peer Y, Ashman TL, Soltis PS, et al. 2021. Polyploidy: an evolutionary and ecological force in stressful times. *The Plant Cell*, **33**(1): 11–26.
- Viček J, Hämälä T, Vives Cobo C, et al. 2025. Whole-genome duplication increases genetic diversity and load in outcrossing *Arabidopsis arenosa*. *Proceedings of the National Academy of Sciences of the United States of America*, **122**(31): e2501739122.
- Wang IJ, Bradburd GS. 2014. Isolation by environment. *Molecular Ecology*, **23**(23): 5649–5662.
- Wang JL, Li N, Stuart FM, et al. 2022. Rapid exhumation processes of the gaoligong mountain range in the southeastern margin of the Qinghai-Tibet plateau since the late cenozoic. *Frontiers in Earth Science*, **10**: 875237.
- Wang N, Chang MM. 2010. Pliocene cyprinids (cypriniformes, teleostei) from kunlun pass basin, northeastern Tibetan Plateau and their bearings on development of water system and uplift of the area. *Science China Earth Sciences*, **53**(4): 485–500.
- Wang XZ, Gan XN, Li JB, et al. 2016. Cyprininae phylogeny revealed independent origins of the Tibetan Plateau endemic polyploid cyprinids and their diversifications related to the neogene uplift of the plateau. *Science China Life Sciences*, **59**(11): 1149–1165.

- Xiao SJ, Mou ZB, Fan DD, et al. 2020. Genome of tetraploid fish *Schizothorax o'connori* provides insights into early re-diploidization and high-altitude adaptation. *iScience*, **23**(9): 101497.
- Xu P, Xu J, Liu GJ, et al. 2019a. The allotetraploid origin and asymmetrical genome evolution of the common carp *Cyprinus carpio*. *Nature Communications*, **10**(1): 4625.
- Xu W, Zhu F, Wang D, et al. 2023. Comparative analysis of metabolites between different altitude *Schizothorax nukiangensis* (cyprinidae, schizothoracine) on the Qinghai-Tibet plateau in Nujiang river. *Water*, **15**(2): 284.
- Xu YY, Tan QS, Hu PC, et al. 2019b. Characterization and expression analysis of FGF6 (fibroblast growth factor 6) genes of grass carp (*Ctenopharyngodon idellus*) reveal their regulation on muscle growth. *Fish Physiology and Biochemistry*, **45**(5): 1649–1662.
- Yan CC, Wu W, Dong WQ, et al. 2022. Temperature acclimation in hot-spring snakes and the convergence of cold response. *The Innovation*, **3**(5): 100295.
- Yang F, Lu H, Yang K, et al. 2021. Hydrological characteristics and changes in the Nu-Salween river basin revealed with model-based reconstructed data. *Journal of Mountain Science*, **18**(11): 2982–3002.
- Yang J, Dong XM, Wen HS, et al. 2024. FGFs function in regulating myoblasts differentiation in spotted sea bass (*lateolabrax maculatus*). *General and Comparative Endocrinology*, **347**: 114426.
- Yang ZH. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**(8): 1586–1591.
- Yao TD, Thompson L, Yang W, et al. 2012. Different glacier status with atmospheric circulations in Tibetan Plateau and surroundings. *Nature Climate Change*, **2**(9): 663–667.
- Yin YH, Zhang XH, Wang XA, et al. 2021. Construction of a chromosome-level genome assembly for genome-wide identification of growth-related quantitative trait loci in *Sinocyclocheilus grahami* (cypriniformes, cyprinidae). *Zoological Research*, **42**(3): 262–266.
- Yu DQ, Ren YD, Uesaka M, et al. 2024. Hagfish genome elucidates vertebrate whole-genome duplication events and their evolutionary consequences. *Nature Ecology & Evolution*, **8**(3): 519–535.
- Yu MC, Zhang DS, Hu P, et al. 2017. Divergent adaptation to Qinghai-Tibetan plateau implicated from transcriptome study of gymnocypris dobula and *Schizothorax nukiangensis*. *Biochemical Systematics and Ecology*, **71**: 97–105.
- Yu X Y, Yu X J. 1990. A schizothoracine fish species, diptychus dipogon, with a very high number of chromosomes. *Chromosome Information Service*, **48**: 17–18.
- Zhang CF, Tong C, Ludwig A, et al. 2018. Adaptive evolution of the Eda gene and scales loss in schizothoracine fishes in response to uplift of the Tibetan Plateau. *International Journal of Molecular Sciences*, **19**(10): 2953.
- Zhou C, Wang XD, Hu ZR, et al. 2023. Comparative analyses reveal potential genetic mechanisms for high-altitude adaptation of *Schizopygopsis* fishes based on chromosome-level genomes. *Journal of Heredity*, **114**(6): 654–668.
- Zhou CW, Xiao SJ, Liu YC, et al. 2020. Comprehensive transcriptome data for endemic schizothoracine fish in the Tibetan Plateau. *Scientific Data*, **7**(1): 28.
- Zhou CW, Zhou Y, Xu LH, et al. 2025. Chromosome-level genome assembly and population genomic analysis provide insights into the genetic diversity and adaption of *Schizopygopsis younghusbandi* on the Tibetan Plateau. *Integrative Zoology*, **20**(6): 1246–1264.
- Zhou CX, McCarthy SA, Durbin R. 2022. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics*, **39**(1): btac808.
- Zhou L, Gui JF. 2017. Natural and artificial polyploids in aquaculture. *Aquaculture and Fisheries*, **2**(3): 103–111.
- Zhou W, Li X. 2024. Fish diversity and selection of taxa for conservation in the Salween and Irrawaddy rivers, southeast Asia. *Scientific Reports*, **14**(1): 2393.