

Research Article

Chromosome-level genome of *Coptodon zillii* reveals adaptive evolution and invasion mechanisms

Shoujie Tang^{a,b,c}, Zhiyi Yan^{a,b,c}, Ying Xing^{a,b,c}, Keyan Dong^{a,b,c}, Yuxuan Liang^{a,b,c},
Chaoyi Wang^d, Jiaying Zheng^{a,b,c}, Temesgen Tola Geletu^{a,b,c,e}, Chenhong Li^{a,b,c}, Yan Zhao^{a,b,c},
Xiaowu Chen^{a,b,c}, Yaolei Zhang^{f,g,h,*}, Jinliang Zhao^{a,b,c,**}

^a Key Laboratory of Freshwater Aquatic Genetic Resources, Ministry of Agriculture and Rural Affairs, Shanghai Ocean University, Shanghai, 201306, China

^b Shanghai Collaborative Innovation Center for Cultivating Elite Breeds and Green-culture of Aquaculture Animals, Shanghai Ocean University, Shanghai, 201306, China

^c National Demonstration Center for Experimental Fisheries Science Education, Shanghai Ocean University, Shanghai, 201306, China

^d School of Life Sciences, South China Normal University, Guangzhou, 510631, China

^e School of Biological Sciences and Biotechnology, Haramaya University, P.O. Box 138, Dire Dawa, Ethiopia

^f BGI Research, Qingdao, 266555, China

^g BGI Research, Chongqing, 401329, China

^h College of Laboratory Medicine, Chongqing Medical University, Chongqing, 400016, China

ARTICLE INFO

Keywords:

Coptodon zillii
Invasive fish
Chromosome-level genome
Population genomics
Environmental adaptation
Evolution

ABSTRACT

Understanding the adaptive mechanisms of invasive fish is crucial for studying their evolution and designing effective control measures. The redbelly tilapia (*Coptodon zillii*) is a widespread and invasive tilapia species with a strong ability to adapt to new environments where it alters ecosystem structure and function. To investigate the mechanism of its rapid invasion, we generated a chromosome-level genome assembly using Illumina, PacBio, and Hi-C technologies, followed by population resequencing (79 individuals across three native Ethiopian and five invasive Chinese populations). The final genome assembly spanned 980.04 Mb, featuring a 34.72 Mb contig N50. Contigs were mapped to 22 pseudochromosomes (scaffold N50: 40.09 Mb), predicting 29,033 protein-coding genes with 96.87% functional annotation. We identified three significantly expanded gene families linked to adaptation and invasiveness, including olfactory receptors, C-type lectin receptors, and B7 co-stimulatory molecules. Two Immune response-related genes under positive selection may have facilitated rapid environmental adaptation in redbelly tilapia. Population genomic analyses revealed three ancestral sources for China's invasive populations. Selective sweep analysis showed that candidate genes associated with local adaptation were functionally correlated with olfactory signal transduction, immune response, antioxidant defence, and stress response. In conclusion, this high-quality reference genome serves as a valuable resource for studying invasive tilapia genomics, aiding investigations into redbelly tilapia's successful invasion mechanisms. We have also produced genetic resources for developing new management strategies for invasive fishes.

1. Introduction

Biological invasions constitute a major component of global change, causing severe ecological and economic impacts worldwide (Gallardo et al., 2016). Approximately 3500 invasive species have now infiltrated nearly all ecosystem types globally (Li and Yu, 2023). Freshwater ecosystems, vital for human survival (Carpenter et al., 2011), face particular

threats from 273 invasive species, including fish (161 species), molluscs (80), zooplankton (13), and aquatic plants (6), leading to significant ecosystem degradation (Corrales et al., 2020; Ricciardi and Macisaac, 2011). Freshwater fishes represent one of the most introduced taxa globally (Gozlan, 2008), with examples occurring in all biogeographic regions (Leprieux et al., 2008). Among these, tilapias (Cichlidae: *Oreochromis*, *Coptodon*, *Serranochromis*, *Pelmatolapia*, *Sarotherodon*) rank

* Corresponding author. BGI Research, Qingdao, 266555, China.

** Corresponding author. Key Laboratory of Freshwater Aquatic Genetic Resources, Ministry of Agriculture and Rural Affairs, Shanghai Ocean University, Shanghai, 201306, China.

E-mail addresses: jlzhao@shou.edu.cn (Y. Zhang), zhangyaolei@genomics.cn (J. Zhao).

Peer review under the responsibility of Editorial Office of Water Biology and Security.

<https://doi.org/10.1016/j.watbs.2026.100695>

Received 1 December 2025; Received in revised form 6 May 2026; Accepted 31 May 2026

Available online xxxx

2772-7351/© 2026 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

among the most widespread invasive fish groups (Bernery et al., 2022), with 551 non-native freshwater fish species recorded. Native to Africa and the Middle East, they have been introduced globally for aquaculture and other purposes (Canonico et al., 2005; Cishahayo et al., 2022; Yongo et al., 2022). Their invasion success stems from traits like omnivory, parental care (Lowe-McConnell, 2006), and broad physiological tolerances (Courtenay, 1997; Peterson et al., 2005; Martin et al., 2010), often causing severe ecosystem impacts (Deines et al., 2016). *Coptodon zillii* exemplifies these problematic invaders (Canonico et al., 2005).

Coptodon zillii (redbelly tilapia, formerly *Tilapia zillii*) is native to northern Africa and parts of the Middle East (Philippart and Ruwet, 1982). Introduced globally from the 1950s to 1980s for aquatic weed/mosquito control, aquaculture, and recreational fisheries (Geletu et al., 2024), it is now among the most widespread tilapia species (Canonico et al., 2005). Global Biodiversity Information Facility (GBIF, <http://www.gbif.org/>) records document its presence in 36 non-native countries (1613 occurrences), with established populations in 17 nations: China (Gu et al., 2016), Iran (Roobzhfar et al., 2014), Turkey (Innal and Giannetto, 2017), Iraq (AL-Zaidy, 2013), Thailand (Panprommin et al., 2018), Philippines (Abdulmalik-Labe and Quilang, 2019), Japan (Fatsi et al., 2020), Singapore (Ng et al., 1993), USA (Robins et al., 1991), Australia (Hoese et al.,), Madagascar (Vences et al., 2022), Fiji, Mexico, Sri Lanka, Eritrea, Libya, and Tanzania (Seebens et al., 2017 for multiple entries).

C. zillii, a cold-tolerant tilapia, competes with native fish for food and spawning sites (Tarkan, 2022; Costa-Pierce, 2003). It harms ecosystems by degrading water quality and reducing native fish populations in multiple countries: USA (Crutchfield et al., 1992; Crutchfield, 1995; Varela-Romero et al., 2002; Costa-Pierce, 2003; Canonico et al., 2005; Andreu-Soler and Ruiz-Campos, 2013; Cassemiro et al., 2018), Iran (Bavali et al., 2022), Iraq (Mohamed and Al-wan, 2020), Japan (Ishikawa and Tachihara, 2008), and China (Gu et al., 2016; Gu et al., 2018a; Xiong et al., 2023). Ranked among the world's most invasive freshwater fish, its ecological impact is regarded as being only slightly less than that of the devastating common carp (*Cyprinus carpio*) and Mozambique tilapia (*Oreochromis mossambicus*) (Casal, 2006).

China faces severe ecological impacts from invasive *C. zillii* (Gu et al., 2016; Gu et al., 2018b; Xiong et al., 2023; Yongo et al., 2023). Introduced from Africa and Thailand during the 1960s–1980s (De Silva et al., 2004; Li, 1999; Cai, 1979), it was commercially farmed in Guangdong, Fujian, and Guangxi but later abandoned due to small body size and slow growth (1980s–1990s). Owing to its tolerance for low temperatures, high reproductive capacity, and strong environmental adaptability, *C. zillii* has successfully established populations across major river basins in China following its escape from aquaculture. Its current distribution spans the Pearl River Basin (e.g., Hongshui River, Beijiang River, Dongjiang River, and Xijiang River), southeastern coastal basins (e.g., Hanjiang River, Minjiang River, and Qiantang River), rivers of Hainan Island (e.g., Nanduijiang River, Wanquan River, and Changhua River), and the Yangtze River Basin (e.g., Jinsha River and Xiangjiang River). The species continues to exhibit a clear trend of range expansion (Xing et al., 2025). *C. zillii* is recognized as one of the most ecologically detrimental alien fish species among the 207 introduced fish species in China, posing severe impacts to aquatic ecosystems (Chen et al., 2017). *C. zillii* degrades ecosystems by overgrazing aquatic plants, preying on native fish eggs, and increasing nitrogen/phosphorus levels, contributing to eutrophication. Its invasion threatens freshwater biodiversity, fisheries stability, and local livelihoods. Recognizing these risks, China added it to the Catalog of Key Managed Invasive Alien Species on 20 December 2022 (Gu et al., 2023), reflecting heightened governmental concern.

Under economic globalization, climate change, and increased human activity, *C. zillii* will expand across China and more of global freshwaters without intervention, threatening aquatic biodiversity. Understanding its invasion mechanisms is critical for control. While invasion biology proposes hypotheses like Diversity Resistance (Elton, 1958), Vacant

Niche (Williamson and Fitter, 1996), Enemy Release (Keane and Crawley, 2002), and Inherent Superiority (Hufbauer and Torchin, 2008), endogenous mechanisms remain understudied. We argue *C. zillii*'s success stems from intrinsic advantages: cold tolerance (Lu, 1987), broad salinity tolerance (Chervinski and Hering, 1973; El-Sayed, 2006), dietary plasticity (Tang et al., 2025), and reproductive plasticity (Li et al., 2023). These traits enhance adaptability, resource use, and population growth, surpassing non-invasive species.

The phenotypic traits of organisms are determined by their genotype, with genetic changes driving alterations in biological characteristics (Benfey and Mitchell-Olds, 2008). The genome encodes all genetic information related to invasion mechanisms. Genomic sequencing of invasive species helps reveal their molecular basis, evolutionary processes, and genetic adaptations. Recent studies demonstrate how genomics explains invasion success in fish. Adrian-Kalchauer et al. (2020) found expanded gene families (cytochrome P450 enzymes, olfactory receptors, opsins, osmoregulatory and immunity genes) in the invasive round goby (*Neogobius melanostomus*), enhancing its cold and saline adaptation. Similarly, Lv et al. (2025) identified rapidly evolving collagen-related genes in the Amazon sailfin catfish (*Pterygoplichthys pardalis*), improving its ecological flexibility. Feng et al. (2023) analyzed invasive goldfish (*Carassius auratus*) in Tibet, revealing selection on embryonic development genes in diploid populations, aiding adaptation to cold (<10 °C) waters and improving larval survival. These genomic adaptations highlight how genetic changes enable invasive species to thrive in new environments.

Genome information has been published for three tilapia species: Nile tilapia (*Oreochromis niloticus*) (Conte et al., 2017; Etherington et al., 2022; Tao et al., 2021a), blue tilapia (*O. aureus*) (Bian et al., 2019; Tao et al., 2021a), and Mozambique tilapia (*O. mossambicus*) (Tao et al., 2021b), but these studies focused on sex-determination (Conte et al., 2017; Tao et al., 2021a, 2021b), genetic introgression (Etherington et al., 2022), and biomedical applications (Bian et al., 2019) rather than invasiveness. To understand *C. zillii*'s invasive success in China, we propose its strong environmental adaptation is key and present the first chromosome-level genome using Illumina, PacBio, and Hi-C technologies. Through comparative genomics and resequencing of invasive versus native populations, we identified candidate genes for rapid adaptation, including olfactory receptors for environmental sensing and immune-related genes for environmental response, along with other potentially invasion-relevant gene families. Understanding invasive populations' genetic background is crucial for tracking invasion pathways and predicting expansion (Cristescu, 2015; Estoup and Guillemaud, 2010), as population genomics effectively reconstructs invasion histories (Coutts et al., 2018; Puckett and Munshi-South, 2019; Rouget et al., 2016; Ryan et al., 2019). Historical records indicate that *C. zillii* was introduced to China through multiple independent events during the 1960s–1980s, with reported origins including South Africa (via Taiwan, China), Thailand, and Africa (De Silva et al., 2004; Li, 1999; Cai, 1979). However, detailed genetic tracing of these introductions has been lacking, and the exact sources remain unclear due to incomplete native-range sampling and potential intermediary transfers. Here, we analyzed whole-genome resequencing data from three native and five invasive populations to examine genomic divergence and elucidate the origin, genetic structure, demographic history, and dispersal routes of invasive *C. zillii* in China.

2. Materials and methods

2.1. Sample preparation

On 26 September 2022, a 46.1 g male redbelly tilapia (standard length: 10.4 cm) was collected from Lake Qiandaohu (119.08° E, 29.58° N; 30.44 m average depth) in Zhejiang, China, with muscle tissue stored at −80 °C for Hi-C and PacBio HiFi sequencing. For population resequencing, 79 specimens were collected (Oct–Dec 2022) from three

Ethiopian lakes [native range: Lake Hora (EHA), Lake Bishoftu (EBS), Lake Ziway (EZW)] and five Chinese river systems [invasive range: Lake Qiandaohu (QDH), Shuikou Reservoir (SKSK), Letan Reservoir (XJ), Dongjiang River (DJ), Nanduijiang River (NDJ)], representing their distribution in Africa and China (Table S1). All muscle samples were preserved at -80°C for whole-genome resequencing.

2.2. Genomic DNA extraction

Genomic DNA was extracted with Qiagen DNA purification kit (Qiagen, Valencia, CA, USA), then quantified (NanoDrop, 2000; NanoDrop Technologies, Wilmington, DE, USA), assessed for purity (Qubit dsDNA HS Assay Kit, Life Technologies, Carlsbad, CA, USA), and checked for integrity (0.8% agarose gel).

2.3. Genome survey and sequencing

2.3.1. Genome size estimation

To estimate redbelly tilapia's genome size, a 300 bp insert-size Illumina library was paired-end sequenced on the HiSeq X Ten platform (Illumina, Inc., San Diego, CA, USA), with raw data filtered using Trimmomatic v0.36 (Bolger et al., 2014). K-mer analysis ($k=23$) was performed via Jellyfish v2.2.6 (Marçais and Kingsford, 2011), with genome size, heterozygosity and repeat content estimated using GenomeScope 2.0 (Ranallo-Benavidez et al., 2020).

2.3.2. Illumina and PacBio library construction and sequencing

We conducted short-read sequencing using a 300 bp paired-end library on Illumina's HiSeq X Ten platform, and long-read sequencing with a 20 kb Single Molecule Real Time (SMRT) bell library in Circular Consensus Sequencing (CCS) mode on the PacBio Sequel platform (Pacific Biosciences, Menlo Park, CA, USA). Using CCS v3.0.0 software, we sequenced 2 cells, generating 73.09 Gb of data. Illumina reads were filtered by removing adapters, reads with $>10\%$ unknown nucleotides (N), and those with $\geq 50\%$ bases having Q-values ≤ 10 . PacBio subreads underwent quality control with standard default settings, ensuring high-quality data for downstream analyses.

2.3.3. Hi-C library construction and sequencing

Following Rao et al. (2015), we crosslinked muscle tissue with 2% formaldehyde, digested DNA with *Mbo*I, filled 5' overhangs with biotinylated nucleotides, performed in situ ligation, sonicated to 300–700 bp fragments, enriched with streptavidin beads, and conducted Illumina 150 bp paired-end sequencing.

2.4. Genome assembly and annotation

2.4.1. Genome assembly

To generate a high-quality chromosome-level genome assembly of redbelly tilapia, we employed an integrated approach combining PacBio HiFi, Illumina, and Hi-C sequencing technologies. First, PacBio HiFi long reads were assembled using Hifiasm v0.16.0 (Cheng et al., 2021) to generate high-quality primary contigs. This approach avoids potential quality degradation that can occur when polishing HiFi assemblies with NGS reads. The initial assembly resulted in contigs with high continuity. Next, Hi-C sequencing data were processed to scaffold the contigs into chromosome-level assemblies. Hi-C clean reads were mapped to the draft contigs using Bowtie2 v2.2.5 (Langmead and Salzberg, 2012) with parameters optimized for Hi-C data alignment. Valid chromatin interaction pairs were identified using HiC-Pro v2.8.1 (Servant et al., 2015). The contigs were then ordered, oriented, and clustered into chromosomes using 3D-DNA v3.0 (Dudchenko et al., 2017) with subsequent manual curation in Juicebox v1.11.08 (Durand et al., 2016). This hierarchical approach allowed us to anchor contigs onto pseudochromosomes based on chromatin contact frequency patterns. The final chromosome-level assembly comprised 22 pseudochromosomes,

consistent with the known karyotype of cichlid fishes.

2.4.2. Evaluation of genome assembly

The quality and completeness of the redbelly tilapia genome assembly were rigorously assessed using multiple complementary approaches. We employed Merquy v1.3 (Rhie et al., 2020) to evaluate assembly quality using both Illumina short reads and PacBio HiFi reads as independent references. Merquy calculates k -mer based quality metrics including consensus quality value (QV) and k -mer completeness. The analysis used a k -mer size of 21 and compared k -mer spectra between the assembly and sequencing data. Assembly quality demonstrated high base-level accuracy and comprehensive sequence representation, with QV of 66 (PacBio HiFi reads) and 54 (Illumina short reads), and k -mer completeness rates of 97.03% and 97.32%, respectively. Genome completeness was evaluated using BUSCO v6.0.0 (Simao et al., 2015) against the latest Actinopterygii ortholog database (actinopterygii_odb12). Genome assembly accuracy was assessed by aligning short-insert paired-end reads to the redbelly tilapia genome using BWA-MEM (Li and Durbin, 2009). Chromosomal features, including gene density, GC content, SNP density, DNA transposon and retrotransposon densities, were visualized using Circos v0.69-9 (Krzywinski et al., 2009) to provide a comprehensive assessment of the genome assembly.

2.4.3. Transposable element annotation

For transposable element (TE) annotation, we combined homology-based and ab initio approaches. De novo repeat libraries were constructed using RepeatModeler v2.0.6 (Flynn et al., 2020), RepeatScout v1.0.7 (Price et al., 2005), PILER v1.0 (Edgar and Myers, 2005), and LTR FINDER v1.05 (Xu and Wang, 2007). These libraries along with Repbase v31.04 were then analyzed through RepeatMasker v4.2.4 (<http://www.repeatmasker.org/>) to identify repetitive sequences. Tandem repeats were detected using Tandem Repeat Finder (TRF; v4.07b). This integrated methodology combining multiple computational tools enabled comprehensive annotation of repetitive elements in the redbelly tilapia genome.

2.4.4. Protein-coding gene annotation

A comprehensive annotation of protein-coding genes in the *C. zillii* genome was performed by integrating multiple lines of evidence, including homology-based prediction, transcriptome data, and ab initio methods. For homology-based prediction, we employed two complementary strategies to maximize sensitivity. First, GeMoMa v1.6.4 (Keilwagen et al., 2019) was used to predict gene structures by leveraging conserved synteny and intron positions from ten phylogenetically close species: *Oreochromis aureus*, *O. niloticus*, *Sarotherodon melanotheron*, *Archocentrus centrarchus*, *Haplochromis burtoni*, *Etheostoma cragini*, *Perca flavescens*, *P. fluviatilis*, *Danio rerio*, and *Oryzias latipes*. Second, a custom pipeline using BLAT v319 (Kent, 2002) and Genewise v2.4.1 (Birney et al., 2004) was implemented to align protein sequences from the same set of species directly to the *C. zillii* genome, providing an additional layer of homology-based evidence. For ab initio prediction, Augustus v3.2.1 (Stanke et al., 2006) was employed, utilizing statistical models of gene structure trained on the target genome to identify potential coding regions. To incorporate transcriptome evidence, RNA-Seq data from *C. zillii* were used. Reads were aligned to the genome using HISAT2 v2.2.1 (Kim et al., 2015), followed by transcript assembly and quantification with StringTie v2.1.4 (Pertea et al., 2015). Additionally, full-length Iso-Seq data were aligned using GMAP v2023-05-15 (Wu and Watanabe, 2005). Open reading frames (ORFs) within these assembled transcripts were predicted using TransDecoder v5.5.0. PASA v2.5.3 (Haas et al., 2008) was used to generate a comprehensive set of transcript-based gene models and to annotate untranslated regions (UTRs) and alternative splicing events. All evidence lines—homology-based predictions, ab initio predictions, and transcriptome-based models—were integrated, weighted, and consolidated into a non-redundant gene set using EvidenceModeler

(EVM) v2.1.0 (Haas et al., 2008). Following the EVM integration, a small number of high-confidence GeMoMa-predicted genes that were absent from the consensus set were added back as a final refinement step, using a custom script. The final gene set underwent rigorous filtering to remove low-confidence predictions, such as those encoding proteins shorter than 50 amino acids or models supported solely by ab initio evidence without transcriptional or homology support. This integrated approach ensured the generation of a high-confidence, comprehensive annotation of protein-coding genes in the *C. zillii* genome.

2.4.5. Functional annotation of protein-coding genes

To characterize protein-coding genes functionally, we annotated the final gene set against multiple databases: NCBI non-redundant protein database (NR) (downloaded on 2023-10-08), SwissProt (downloaded on 2023-10-08, UniProt, 2023_04), Eukaryotic Orthologous Groups (KOG) (original release 2003), TrEMBL (downloaded on 2023-10-08, UniProt, 2023_04), InterPro, Gene Ontology (GO) (version 2023-10-09), and the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Release 108.0). GO annotations were obtained via InterProScan v5.64-96.0 (synchronized with InterPro 96.0, released 26 September 2023) (Jones et al., 2014) by screening protein sequences against InterPro for conserved domains. For other databases, we conducted BLASTP v2.12.0+ searches with a 1×10^{-5} E-value cutoff.

2.5. Comparative genomics

2.5.1. OrthoMCL analysis

Orthologous gene detection was essential for cross-genome comparisons. Using OrthoMCL v2.0.9 (Li et al., 2003), we clustered orthologs and identified paralogs across 17 teleost species: 11 cichlids [*C. zillii*, *Oreochromis niloticus* (GCF_001858045.2), *O. aureus* (GCF_013358895.1), *O. mossambicus* (Tao et al., 2021b), *Sarotherodon melanotheron*, *Archocentrus centrarchus* (GCF_007364275.1), *Haplochromis burtoni* (GCF_018398535.1), *Astatotilapia calliptera* (GCF_900246225.1), *Pundamilia nyererei* (GCF_000239375.1), *Maylandia zebra* (GCF_000238955.4), and *Neolamprologus brichardi* (GCF_000239395.1)] and 6 representative species [*Takifugu rubripes* (GCF_901000725.2), *Xiphophorus maculatus* (GCF_002775205.1), *Gasterosteus aculeatus* (GCF_016920845.1), *Astyanax mexicanus* (GCF_023375975.1), *Danio rerio* (GCF_000002035.6), and *Oryzias latipes* (GCF_002234675.1)]. Genomic data for 15 species came from NCBI RefSeq or literature, except *S. melanotheron*, provided by Southwest University's Key Laboratory of Freshwater Fish Reproduction and Development (Chongqing, China). For genes with multiple isoforms, only the longest transcript was analyzed. OrthoMCL parameters included a 50% match percentage cutoff and 1×10^{-5} E-value threshold, with MCL clustering at an inflation parameter of 1.5. Post-clustering, an in-house Perl script extracted single-copy orthologs, universal orthologs, and other gene categories.

2.5.2. Phylogenetic analysis of redbelly tilapia with other fish species

To reveal the phylogenetic relationship between redbelly tilapia and other fish species, a phylogenetic species tree was reconstructed from redbelly tilapia and 16 other fishes, specifically *O. niloticus*, *O. aureus*, *O. mossambicus*, *S. melanotheron*, *A. centrarchus*, *H. burtoni*, *A. calliptera*, *P. nyererei*, *M. zebra*, *N. brichardi*, *T. rubripes*, *X. maculatus*, *G. aculeatus*, *A. mexicanus*, *D. rerio*, and *O. latipes*. The maximum-likelihood species tree was reconstructed using RAxML v8.2.10 (Stamatakis, 2014) under the GTR + GAMMA model with 1000 bootstrap replicates, based on codon alignments of single-copy orthologous CDS sequences. Divergence times were estimated with MCMCtree program in the Phylogenetic Analysis by Maximum Likelihood (PAML) package (Yang, 2007) using GTR + GAMMA and seven TimeTree-derived (<http://www.timetree.org>) calibrations: 180.0–251.5 million years ago (Mya) (*D. rerio*-*O. niloticus*), 75.5–124.1 Mya (*O. latipes*-*X. maculatus*), 83.2–150.9 Mya (*G. aculeatus*-*T. rubripes*), 132.0–170.0 Mya (*A. mexicanus*-*D. rerio*),

100.0–130.0 Mya (*T. rubripes*-*O. aureus*), 81.2–95.6 Mya (*O. latipes*-*O. niloticus*), and 14.1–30.1 Mya (*O. niloticus*-*H. burtoni*). MCMCtree analysis used 5000 burn-in iterations, 20,000 sampling cycles, and sampling frequency of 50.

2.5.3. Gene family expansion and contraction analysis

To control for technical biases from varying genome assembly and annotation quality across species, we analyzed gene family evolution using CAFE 5 (Mendes et al., 2020). This version of CAFE integrates an error model that estimates per-species gene count variance, reducing false positives in expansion/contraction inferences. Inputs included a gene count file and an ultrametric species tree with branch lengths. The error model was estimated directly from the data (-e flag), and the global birth-death rate (λ) was optimized automatically while allowing lineage-specific rate variation. Families with conditional $P < 0.05$ were classified as rapidly evolving, providing a robust assessment of gene family dynamics.

To investigate the ecological adaptation strategies of *C. zillii*, we conducted comparative genomics analyses on three evolutionarily significant gene families [olfactory receptor (OR), C-type lectin receptor (CLR), and B7 co-stimulatory molecule (B7)] across six species: three reference models (*D. rerio*, *O. latipes*, and *G. aculeatus*) and two resilient tilapiine congeners (*O. niloticus* and *O. aureus*). For OR genes, we first retrieved *O. niloticus* reference sequences from NCBI GenBank, then performed BLASTP searches (e-value cutoff 1×10^{-5}) against all six species' genomes. Using HMMER 3.0 (<http://www.hmmerr.org/>), we built Hidden Markov Models (HMMs) for each gene family member and conducted genome-wide hmmsearch analyses (e-value cutoff 1×10^{-5}) to identify candidate genes. For CLR and B7 analyses, we obtained reference sequences from three model species (*O. niloticus*, *O. latipes*, and *G. aculeatus*) and manually constructed HMMs to screen all six genomes (e-value cutoff 1×10^{-5}). Additionally, we employed existing Pfam HMM profiles (CLR: PF00059; B7: PF08205; <http://pfam.xfam.org/>) to validate identified sequences through similarity searches (e-value cutoff 1×10^{-5}), ultimately obtaining the predicted gene families for subsequent evolutionary analyses.

Protein sequences were aligned using MAFFT v7.450 (Katoh et al., 2019), then refined with Gblocks v0.91b (Talavera and Castresana, 2007) to remove poorly conserved regions. The filtered alignments were analyzed with ModelFinder (Kalyaanamoorthy et al., 2017) to select optimal substitution models. Maximum-likelihood phylogenetic trees were reconstructed in IQ-TREE v1.6.6 (Nguyen et al., 2015), using 1000 ultrafast bootstrap replicates. Final trees were visualized and annotated through the interactive Tree of Life (iTOL) web platform (Letunic and Bork, 2021).

To investigate lineage-specific evolution of the OR, CLR, and B7 gene families, we performed gene tree-species tree reconciliation using Notung v2.9.1.5 (Stolzer et al., 2012). For each family, the maximum-likelihood gene tree was reconciled with the reference species tree to infer gene gain (duplication) and loss events along each phylogenetic branch. Gene tree rooting was performed using the outgroup method, with *D. rerio* sequences designated as the outgroup. Based on the reconciled trees, we quantified duplication and loss events occurring on the branch leading to *C. zillii* and other key branches (e.g., the ancestral tilapiine branch) to identify lineage-specific expansions or contractions and elucidate the timing of gene family evolution.

2.5.4. Positive selection analysis

To accurately detect genes under positive selection in *C. zillii* while accounting for potential confounding effects such as multinucleotide substitutions (MNS) and synonymous rate variation (SRV), we employed the BUSTED + MH model as implemented in the HyPhy package v2.5.42. This model extends the standard Branch-Site Unrestricted Statistical Test for Episodic Diversification (BUSTED) framework by incorporating both site-to-site synonymous rate variation (+S) and instantaneous multinucleotide (double- and triple-hit) substitutions

(+MH), thereby reducing false-positive inferences caused by unmodeled evolutionary complexities (Lucaci et al., 2023). We used the protein sequences of single-copy orthologous genes from 17 teleost species (including *C. zillii* and 16 others listed in Section 2.5.2). Multiple sequence alignments were generated at the codon level using PRANK (Löytynoja and Goldman, 2010) and were refined with Gblocks v0.91b (Talavera and Castresana, 2007) to remove poorly aligned regions. For each alignment, we performed the BUSTED + MH analysis under the “+Double + Triple” mode to allow for both double- and triple-nucleotide instantaneous substitutions. The model tests the null hypothesis that all branches and sites evolved under purifying or neutral selection ($\omega \leq 1$) against the alternative hypothesis that a proportion of branch-site combinations evolved under positive selection ($\omega > 1$). Statistical significance was assessed using a likelihood ratio test (LRT) with critical values derived from a 50:50 mixture of χ^2_0 and χ^2_1 distributions, as recommended by Murrell et al. (2015) and Wisotsky et al. (2020). To control for multiple testing, *P*-values were adjusted using the false discovery rate (FDR) method of Benjamini and Hochberg (1995). Genes with an FDR-adjusted *P*-value < 0.05 were considered to have experienced significant episodic diversifying selection in *C. zillii*. We used the species tree for positive selection gene analysis with BUSTED + MH. All analyses were conducted using the HyPhy batch file available at <https://github.com/veg/hyphy/blob/master/res/TemplateBatchFiles/SelectionAnalyses/BUSTED-MH.bf>, using default parameters unless otherwise noted.

2.6. Population genomics analysis

2.6.1. Genome resequencing

A total of 79 redbelly tilapia were collected from Ethiopian lakes (10 in EHA, 10 in EBS, and 9 in EZW) and Chinese waters (10 in QDH, 10 in SKSK, 10 in DJ, 10 in XJ, and 10 in NDJ). Genomic DNA was extracted and sequencing libraries prepared using the TruSeq Nano DNA HT Kit (Illumina, Inc., San Diego, CA, USA). Library size distribution was verified with an Agilent 2100 Bioanalyzer (Agilent Technologies, Inc., Santa Clara, CA, USA), followed by 150 bp paired-end sequencing on Illumina HiSeq X Ten.

2.6.2. Single-nucleotide polymorphism (SNP) calling

To obtain high-quality SNPs for population genomic analyses, we followed a standardized variant calling pipeline with stringent filtering. Raw paired-end sequencing reads from all 79 individuals were first processed with Trimmomatic v0.36 (Bolger et al., 2014) to remove adapters and low-quality bases. Clean reads were then aligned to the *C. zillii* chromosome-level reference genome using the Sentieon Package v202112.03 (Aldana and Freed, 2022). Following alignment, the resulting BAM files were sorted and duplicates were marked to prepare for variant analysis. For each individual sample, variant calling was performed independently with the Sentieon haplotyper module, generating a sample-specific genomic Variant Call Format (gVCF) file. To produce a comprehensive population-level dataset, joint genotyping was subsequently conducted on all sample gVCF files using the Sentieon GVCFTyper module (Aldana and Freed, 2022), yielding a single, unified VCF file for downstream analyses. Next, SNPs were extracted from the joint-genotyped VCF file using GATK v4.2.2's SelectVariants tool (McKenna et al., 2010). Subsequent hard-filtering was applied to ensure high-confidence variant calls. We used GATK's VariantFiltration with the following thresholds, which were selected based on common practices in whole-genome population studies (e.g., Liu et al., 2022; Cui et al., 2025) and preliminary data inspection: QD < 2.0 (Quality by Depth), FS > 60.0 (Fisher Strand bias), MQ < 40.0 (Mapping Quality), MQRankSum < -12.5 , ReadPosRankSum < -8.0 . These cutoffs are conservative and help remove false positives arising from sequencing artifacts, alignment errors, or strand bias. After GATK filtering, we further applied population-level filters using VCFtools v0.1.15 (Danecek

et al., 2011): only biallelic SNPs with missing genotype rate $< 5\%$ and minor allele frequency (MAF) > 0.01 were retained. Additionally, we excluded genomic regions with sequencing depth outside the 95% confidence interval of the genome-wide distribution to mitigate potential biases from paralogous regions or misalignments.

2.6.3. Population genetic diversity statistics

To estimate population genetic diversity and neutrality, genome-wide nucleotide diversity (π), Tajima's *D*, and pairwise F_{ST} values (Weir and Cockerham, 1984) were calculated using VCFtools v0.1.15 (Danecek et al., 2011) in 10-kb sliding windows with a 2-kb step size. The calculation of π relies on accurately identifying both variant and invariant sites; therefore, we utilized the comprehensive site information preserved in the joint-genotyped VCF file generated via the Sentieon gVCF-based pipeline (as described in Section 2.6.2). This approach ensures that regions with no reads (missing data) are distinguished from sites that are genuinely identical to the reference, thereby preventing inflation of diversity estimates. To further enhance reliability, we applied a uniform filter to exclude genomic windows where more than 20% of sites had missing genotype calls across all individuals. Additional diversity statistics—including observed heterozygosity (H_0), polymorphism information content (PIC), minor allele frequency (MAF), and inbreeding coefficients (*F*)—were computed using PLINK v2.0 (Chang et al., 2015).

2.6.4. Population genetic structure

Population structure was analyzed using ADMIXTURE v1.0 (Alexander et al., 2009) with *K* values ranging from 2 to 8, where the optimal *K* was selected by minimizing cross-validation error. Genomic data was formatted using PLINK v2.0 (Chang et al., 2015) prior to analysis. To visualize genetic relationships among individuals, we constructed an identity-by-state (IBS) distance matrix based on the full set of high-quality SNPs identified across the 79 individuals using PLINK v2.0 (Chang et al., 2015). A Neighbor-Joining (NJ) tree was then generated from this distance matrix using MEGA7 (Kumar et al., 2016) with *O. niloticus* as the outgroup. The resulting dendrogram was visualized using the interactive Tree of Life (iTOL) platform (Letunic and Bork, 2021). For population clustering, principal component analysis (PCA) was performed using PLINK v2.0 (Chang et al., 2015) on genotype data from filtered variants on 79 individuals.

2.6.5. Linkage-disequilibrium analysis

Linkage disequilibrium (LD) decay was analyzed with PopLDdecay v3.40 (Zhang et al., 2019) using standard parameters, calculating mean r^2 per distance interval and visualizing patterns across eight *C. zillii* populations via R script.

2.6.6. Inference of recent demographic history using GONE2

We used the linkage-disequilibrium-based method in GONE2 v2 (Santiago et al., 2020) to estimate recent fluctuations in effective population size (N_e) over the last approximately 200 generations. Because no species-specific or chromosome-scale genetic map was available for *C. zillii*, the analysis was conducted using the default uniform recombination rate of 1.0 cM/Mb, a common and practical approach in population genomics for non-model organisms lacking a high-resolution recombination landscape. GONE2 was run separately for each of the eight populations, with 100 bootstrap replicates per population by randomly resampling individuals to ensure robustness.

2.6.7. Genome-wide scan for signals of selective sweeps

Based on population structure, invasive and native *C. zillii* populations were categorized into four groups: Group 1 (ETH), comprising the three native populations (EHA, EBS, EZW); Group 2 (QDH); Group 3 (SKSK); and Group 4 (DXN), comprising the three invasive populations (DJ, XJ, NDJ). To detect signatures of local adaptation, we employed four complementary selection scan methods: pairwise F_{ST} (Weir and

Cockerham, 1984), the ratio of nucleotide diversity (π ratio), Tajima's D (Tajima, 1989), and the cross-population composite likelihood ratio test (XP-CLR; Chen et al., 2010). For the first three methods (F_{ST} , π ratio, and Tajima's D), statistics were calculated in 10-kb sliding windows with a 2-kb step size using VCFtools v0.1.15 (Danecek et al., 2011). XP-CLR analysis was performed using XP-CLR v1.1.2 (Chen et al., 2010) under a uniform recombination rate of 1.0 cM/Mb due to the absence of a species-specific genetic map for *C. zillii*. Outlier windows were initially defined as those in the top 5% of the genome-wide distribution for each statistic. To ensure robustness, candidate selective sweep regions were identified only when they were simultaneously detected as outliers by all four methods. This stringent multi-method intersection significantly reduces false-positive signals and yields high-confidence genomic intervals under strong selection. Genes overlapping these consensus outlier regions were considered candidate selected genes. To visually confirm that selection signals were localized to the candidate genes themselves, we extracted and plotted the values of all four statistics across a zoomed-in genomic region encompassing each candidate gene and its flanking sequences.

2.6.8. Functional enrichment analysis of candidate genes

We performed KEGG and GO enrichment analyses using DAVID v6.8 (Dennis Jr et al., 2003) (<https://david.ncifcrf.gov/tools.jsp>). Statistical significance was assessed using a hypergeometric Fisher exact test through OmicShare platform (www.omicshare.com/tools), with multiple-testing correction applied via FDR. KEGG pathways showing FDR-adjusted p -values below 0.05 were identified as significantly enriched. Visualization of KEGG and GO enrichment results were conducted using the ggplot2 package (version 3.4.4) in R v4.3.2.

3. Results

3.1. Genome sequencing and chromosome-level genome assembly

3.1.1. Genome size evaluation

K-mer analysis of *C. zillii* using Illumina sequencing data generated a total data volume of 128.19 Gb with 109,390,852,608 k -mers (Fig. S1). Genome size was estimated at 935.64 Mb (Table S2), with 0.2% heterozygosity and 26.76% repeats. The absence of heterozygous peaks suggests a simple genomic structure favorable for assembly (Fig. S1).

3.1.2. Illumina and PacBio sequencing and genome assembly

We sequenced the genome using both PacBio Sequel II and Illumina platforms, yielding 73.09 Gb (78.09 $\times$) of clean PacBio subreads with an average read length of 20.51 kb (Table S3). Additionally, 128.19 Gb (135 $\times$) of Illumina sequences were obtained (Table S3). The assembled genome produced a final draft of 979.98 Mb, comprising 195 contigs with an N50 of 34.72 Mb, where the longest contig reached 65.29 Mb (Table S4).

3.1.3. Chromosome sequence assembly

To achieve a chromosome-level assembly, 272.71 Gb of clean Hi-C data (280 $\times$) was used to scaffold the assembled genome (Fig. S2). After refinement, 966.59 Mb (98.63%) were anchored onto 22 pseudochromosomes (Fig. 1), with individual pseudochromosome lengths ranging from 31.73 Mb to 111.61 Mb and an average length of 43.94 Mb (Table S5). The final 980.04 Mb chromosome-level genome assembly contained 84 scaffolds and 197 contigs with a scaffold N50 of 40.09 Mb (Table 1, Table 2). BUSCO assessment demonstrated 99.72% completeness of conserved genes (Table S6).

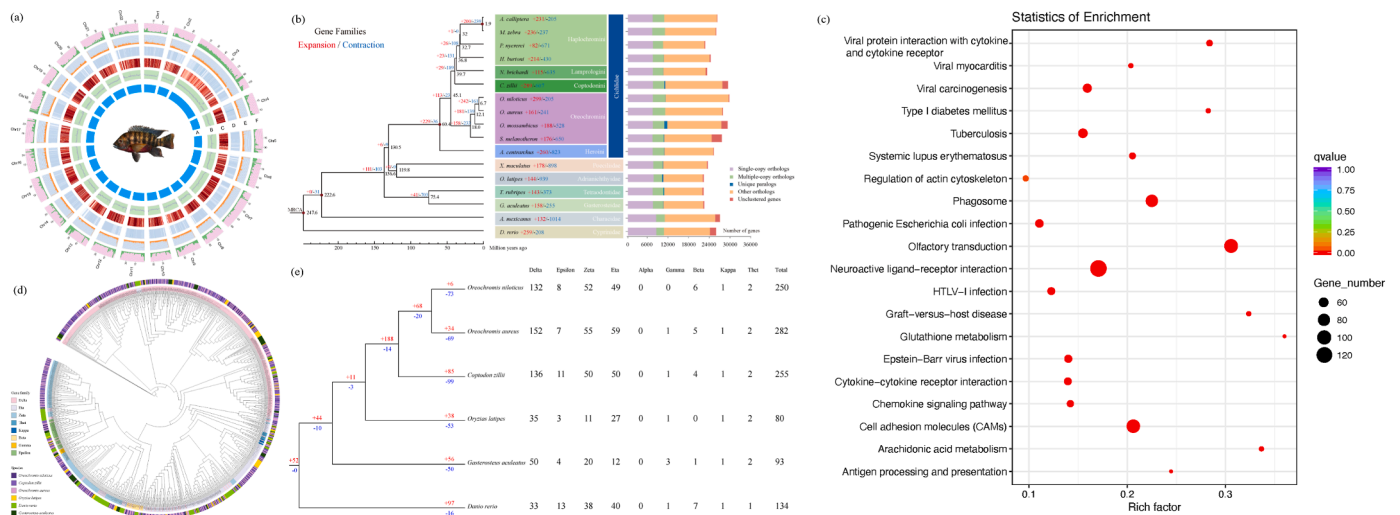


Fig. 1. Genomic characteristics, evolutionary history, and adaptive gene families of the redbelly tilapia (*Coptodon zillii*). (a) Circular plot of genome features. From inner to outer tracks: A, chromosome ideograms; B, GC content; C, density of protein-coding genes; D, density of genome-wide SNPs; E, density of DNA transposons; F, density of long terminal repeat retrotransposons (LTR-RTs). (b) Phylogenetic relationship and ortholog distribution. The maximum-likelihood species tree was inferred from 5473 strict single-copy orthologs across *C. zillii* and 16 other fish species. Numbers of expanded (red) and contracted (blue) gene families are shown on branches; node numbers indicate estimated divergence times (million years ago, Ma). MRCA is the abbreviation of most recent common ancestor. The gene family distribution for each species is displayed alongside: light purple, single-copy families; light green, multicopy families; light blue, species-specific families; light orange, other families; light red, unclustered genes. (c) Enriched KEGG pathways of rapidly expanded gene families. Significantly enriched pathways (q -value < 0.05) for gene families identified as rapidly evolving (P < 0.05). (d) Phylogeny of olfactory receptor (OR) proteins. Unrooted maximum-likelihood tree of ORs from *C. zillii*, *Oreochromis aureus*, *O. niloticus*, *Gasterosteus aculeatus*, *Danio rerio*, and *Oryzias latipes*, constructed with IQ-TREE under the JTT + G + I + F model (1000 bootstrap replicates). (e) Loss and gain events of olfactory receptor (OR) genes in *C. zillii* and related species. The species tree is shown with branch lengths proportional to divergence times. Numbers above each branch indicate inferred gene gains (red, "+") and losses (blue, "-") based on gene tree-species tree reconciliation using Notung.

Table 1
Comparison of genome assemblies in four tilapia fishes at chromosome level.

Features	<i>Coptodon zillii</i>	<i>Oreochromis niloticus</i>	<i>Oreochromis niloticus</i>	<i>Oreochromis niloticus</i>	<i>Oreochromis niloticus</i>	<i>Oreochromis niloticus</i>	<i>Oreochromis niloticus</i>	<i>Oreochromis aureus</i>	<i>Oreochromis aureus</i>	<i>Oreochromis mossambicus</i>
Assembly name	NA	O. niloticus_UMBU	XX_nile	Oren11.1	O. niloticus_UMD1	ORES2315963_Elv1.0	XX_ONJp	ZZ_OA	NA	NA
Genome size (Mb)	980.04	1006	993.5	927.7	1009	1070	993.5	1005	920	1007
Number of contigs	197	3009	1460	77,754	2566	NA	1201	805	NA	NA
Number of scaffolds	84	2459	403	5909	NA	364	403	303	NA	451
Number of assembled chromosomes	22	22	22	22	22	22	22	22	22	22
Genome assembly quality										
Contig N50 (Mb)	34.72	2.9	2.5	0.0295	3.3	24.5	2.6	4.4	0.0532	2.78
Scaffold N50 (Mb)	40.09	38.8	40.3	2.8	NA	26	40.3	40.7	1.10	11.38
BUSCO genes (%)	99.72	99.67	99.60	99.10	79.66	95.90	97.3	97.50	NA	97.71
Genomic features										
G + C (%)	40.8	40.5	40.5	40.5	NA	NA	40.4	39.4	NA	42.28
Gene annotation										
Number of genes	29033	29,550	25,264	30,978	38,412	45,826	25,264	25,467	23,117	28,902
References	This study	GCA_001858045.3	GCA_013350305.1	GCA_000188235.2	Conte et al. (2017)	Etherington et al. (2022)	Tao et al. (2021a)	Tao et al. (2021a)	Bian et al. (2019)	Tao et al. (2021b)

Notes: NA indicates that no data was available.

Table 2Summary statistics of chromosome-level genome assembly of *Coptodon zillii*.

Item	Scaffold length (bp)	Scaffold number	Contig length (bp)	Contig number
Max length	111,608,768	-	65,293,448	-
N10	111,608,768	1	47,926,456	2
N20	48,863,300	3	40,768,379	4
N30	41,199,818	5	39,488,533	7
N40	40,714,204	7	38,474,974	9
N50	40,092,963	10	34,723,038	12
N60	39,985,982	12	31,513,700	15
N70	39,467,494	15	26,186,227	19
N80	39,261,804	17	19,968,472	23
N90	34,409,495	20	7,922,129	31
Total length	980,035,945	-	979,979,445	-
Total number	-	84	-	197
GC Ratio	0.41	-	0.41	-

3.2. Genome annotation

3.2.1. Repeat annotation

Repeat annotation revealed that the *C. zillii* genome contained 423.64 Mb of repetitive sequences (43.23% of the assembled genome), with transposable elements (TEs) accounting for 42.82% (Table S7). Among TEs, long terminal repeat retrotransposons (LTR-RTs) (29.29%), DNA transposons (10.54%), and long interspersed nuclear elements (LINEs) (8.27%) were the three most abundant repeat types (Table S7, Fig. S3).

3.2.2. Gene prediction and functional annotation

Using an integrated approach, we identified 29,033 putative protein-coding genes in the *C. zillii* genome. The average gene and coding sequence (CDS) lengths were 14,514.73 bp and 1670.86 bp, respectively (Table S8, Fig. S4), with an average of 9.33 exons per gene and an average intron length of 1542.61 bp (Table S8, Fig. S4). The gene set showed 96.06% completeness (BUSCO assessment, Table S9). Among the predicted genes, 28,123 (96.87%) were functionally annotated against at least one of seven public databases (GO, KEGG, KOG, SwissProt, TrEMBL, InterPro, NR) (Table S10).

3.3. Genome evolution

3.3.1. Orthology assignment and phylogenomic analysis

Gene family clustering analysis of 17 species using OrthoMCL v2.0.9 identified 25,874 gene families, including 5473 single-copy orthologs (Fig. 1, Table S11). To determine the phylogenetic relationship of *C. zillii* with other fish species, we reconstructed a maximum-likelihood species tree using RAXML based on 5473 strictly single-copy orthologs. The results showed that *C. zillii* clustered with nine cichlid species (*A. calliptera*, *M. zebra*, *P. nyererei*, *H. burtoni*, *N. brichardi*, *O. aureus*, *O. niloticus*, *O. mossambicus*, and *S. melanotheron*), exhibiting the closest genetic distance to *N. brichardi* (Fig. 1). Divergence time estimation revealed that within Cichlidae, the four Oreochromini tribe species (*O. aureus*, *O. niloticus*, *O. mossambicus*, and *S. melanotheron*) diverged from the Coptodonini tribe member *C. zillii* approximately 45.1 million years ago (Mya) (Fig. 1).

3.3.2. Gene family expansion and contraction analysis

To investigate evolutionary changes in *C. zillii*'s gene families, CAFE analysis was performed, showing the most recent common ancestor of the 17 species possessed 17,716 gene families. In *C. zillii*, we detected 300 expanded (2114 genes) and 662 contracted families ($P < 0.05$) (Fig. 1, Table S12). KEGG analysis revealed expanded families were significantly enriched in olfactory transduction (90 olfactory receptor

genes), phagosome, antigen processing and presentation, cell adhesion molecules (CAMs), and neuroactive ligand-receptor interaction pathways (Table S13, Fig. 1). These families also participated in metabolic processes including glutathione metabolism, cytochrome P450-mediated drug metabolism, and arachidonic acid metabolism.

3.4. Positively selected and rapidly expanded genes promoting *C. zillii* adaptations to environmental stress

3.4.1. Positively selected genes associated with adaptation and invasiveness

Positive selection analysis of single-copy orthologs across 17 species identified 34 positively selected genes (PSGs) in *C. zillii* (Table S14), including key immune-related genes under strong selection (i.e., galectin-related protein B and T-cell surface glycoprotein CD4 isoform X1), suggesting adaptive evolution to environmental pressures.

3.4.2. Rapidly expanded gene families associated with adaptation and invasiveness

3.4.2.1. Evolution of olfactory receptors (ORs) gene families. The *C. zillii* genome contains 255 OR genes, the second highest among the six species examined (Supplementary Note 1). Following Liu et al. (2021), ORs from all species clustered into nine subfamilies (α , β , γ , δ , ϵ , ζ , η , θ , and κ) (Table S15). Phylogenetic reconstruction using 1094 complete OR protein sequences (>250 amino acids) under the optimal JTT + G + I + F model demonstrated that OR genes predominantly organized into evolutionarily conserved tandem arrays across species (Fig. 1). *Coptodon zillii* exhibited pronounced expansions in δ (136 genes, >50% of its OR repertoire), ζ (50), and η (50) subfamilies (Table S15). Gene tree-species tree reconciliation revealed that most OR duplications occurred in the

common tilapiine ancestor rather than on the terminal *C. zillii* branch (Fig. 1). The ancestral tilapiine branch gained 65 δ , 52 ζ , and 56 η genes. Following divergence from *Oreochromis*, the *C. zillii* branch acquired modest additional duplications (+10 δ , +2 ζ , +1 η) with few losses. Thus, while *C. zillii* shows limited lineage-specific expansion, the major OR amplification occurred earlier in tilapiine evolution, suggesting *C. zillii* inherited a substantially expanded olfactory repertoire that may contribute to its ecological adaptation.

3.4.2.2. Evolution of C-type lectin receptors (CTLRs) gene families. *Coptodon zillii* possessed the largest CTLR repertoire ($n = 62$) among the six species, more than double that of *O. latipes* ($n = 25$) and nearly triple that of *G. aculeatus* ($n = 23$) (Table S16). Following the classification system of Zhang et al. (2015), phylogenetic analysis of 225 CTLRs using the WAG + G + I + F model revealed *C. zillii* CTLRs distributed across all five subfamilies, with the CLEC4 (C-type lectin domain family 4) subfamily showing the most pronounced expansion ($n = 56$) (Table S16, Fig. S5, Supplementary Note 2). Gene tree-species tree reconciliation demonstrated lineage-specific expansion along the *C. zillii* branch (Fig. S5), with 55 net CLEC4 gains and minimal losses (7 losses). In contrast, *O. niloticus* and *O. aureus* showed modest changes (+17 and +16 gains, respectively). This indicates recent and substantial CTLR diversification in *C. zillii* following divergence from the *Oreochromis* lineage.

3.4.2.3. Evolution of B7 family of costimulatory molecules. *Coptodon zillii* contains 58 B7 genes, fewer than *O. niloticus* (74) and *O. aureus* (78) but markedly more than *O. latipes* (16), *G. aculeatus* (5), and *D. rerio* (9) (Table S17). Following the classification system of Hansen et al. (2009; seven subfamilies: B7-1, B7-2, B7-H1, B7-H2, B7-H3, B7-H4, B7-DC), phylogenetic analysis of 240 B7 sequences using the WAG + G + I + F

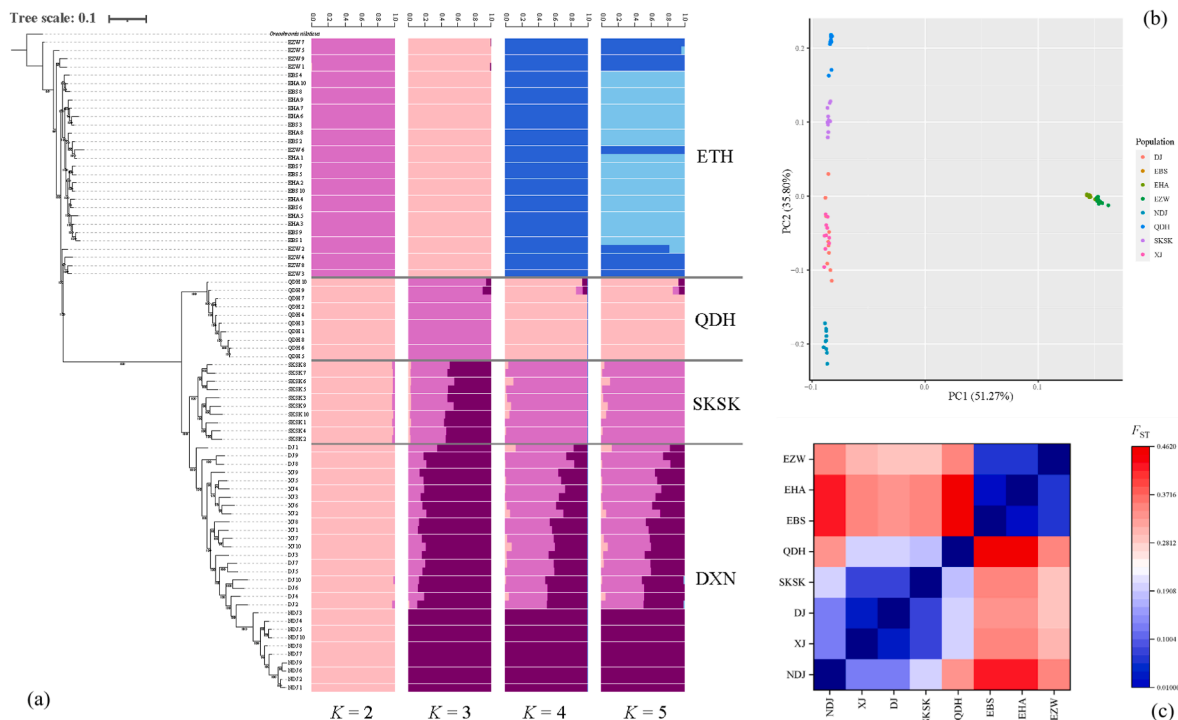


Fig. 2. Population genomic analysis of redbelly tilapia in the native and invasive regions. (a) Population structure and Neighbor-Joining dendrogram of *Coptodon zillii* populations from eight geographic sites based on genome-wide SNP data. The dendrogram was constructed using the identity-by-state (IBS) distance matrix with 1000 bootstrap replicates. Population structure inferred using ADMIXTURE analysis from $K = 2$ to $K = 5$. Clusters were noted by different colors, and each bar represented an individual. Numbers at tree nodes represented the bootstrap supports (1000 replications). (b) Principal component analysis (PCA) of resequencing individuals based on the first two principal components. Each dot represents one individual, with different colors representing different populations. (c) Heatmap of genetic differentiation index (F_{ST}) between pairwise populations. The module colors represent the F_{ST} value—darker reds indicate higher pairwise F_{ST} values and darker blues indicate lower pairwise F_{ST} values. See Table S1 for the abbreviations for sampling sites.

model showed *C. zillii* B7 genes distributed across five subfamilies, with B7-H1 being most expanded ($n = 35$) (Fig. S6, Supplementary Note 3). Gene tree-species tree reconciliation revealed that the common tilapiine ancestor experienced major expansion, particularly in B7-H1 (+44 gains). Subsequently, the *C. zillii* branch acquired moderate duplications (+10), yielding net increases of +9 B7-H1 and +1 B7-H3 (Fig. S6). *Oreochromis niloticus* and *O. aureus* each experienced additional terminal duplications (+23 and +16, respectively). Thus, while *C. zillii* showed limited lineage-specific expansion, the major B7 diversification occurred earlier in tilapiine evolution.

3.5. Invasion and spread pattern of *C. zillii* revealed by population genomics analysis

3.5.1. Population resequencing data and variant characteristics

Whole-genome resequencing of 79 individuals from eight populations produced 1428.86 Gb of high-quality clean reads (average 18.09 Gb per individual) after quality filtering (Table S18). Using the 980.04 Mb *C. zillii* genome as reference, sequencing depth ranged from $9.69\times$ to $21.62\times$ with a mean of $(15.85 \pm 2.47)\times$ (Table S19), while mapping rates varied between 99.38% and 99.84%. Variant calling

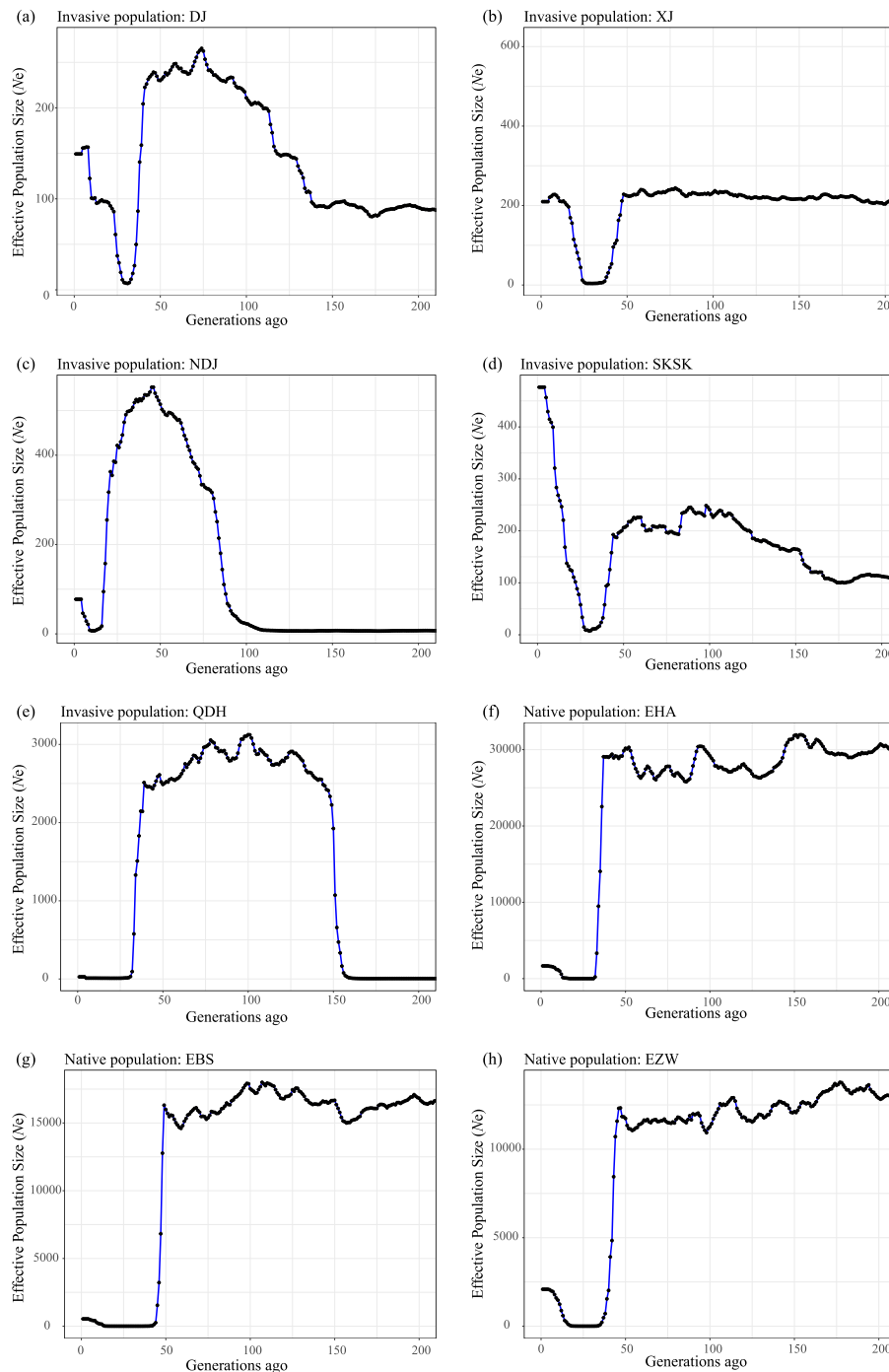


Fig. 3. Recent demographic history inferred from GONE2 analyses. Effective population sizes (N_e) over the past 200 generations are shown for eight populations of *Coptodon zillii*: (a) DJ, (b) XJ, (c) NDJ, (d) SKSK, (e) QDH, (f) EHA, (g) EBS, (h) EZW. Populations a–e are from the invasive range (China), and populations f–h are from the native range (Ethiopia). See Table S1 for the abbreviations for sampling sites.

identified 15,693,070 high-quality SNPs (1.60% genomic coverage, 1 SNP per 62 bp), with population-specific averages ranging from 3,243,181 to 7,851,847 SNPs (Table S20).

3.5.2. Genome-wide genetic diversity estimates in native and invasive range

The QDH population showed the lowest nucleotide diversity ($\pi = 1.29 \times 10^{-3}$), while EZW had the highest ($\pi = 2.62 \times 10^{-3}$) (Table S21). On average, invasive populations ($\pi = 2.13 \times 10^{-3}$) showed slightly higher nucleotide diversity than native ones ($\pi = 2.07 \times 10^{-3}$), though this difference was not statistically significant ($P = 1.0000$). NDJ exhibited maximum heterozygosity ($H_0 = 1.18 \times 10^{-2}$), contrasting with EZW's minimum ($H_0 = 0.53 \times 10^{-2}$) (Table S21). Genome-wide, invasive populations displayed higher mean heterozygosity ($H_0 = 0.93 \times 10^{-2}$) than native groups ($H_0 = 0.91 \times 10^{-2}$; $P = 0.6583$).

3.5.3. Genetic relationships among native and invasive populations

The Neighbor-Joining dendrogram based on genome-wide SNP data grouped the 79 *C. zillii* individuals into four major genetic clusters: Group 1 (EHA, EBS, EZW), Group 2 (SKSK), Group 3 (QDH), and Group 4 (DJ, XJ, NDJ) (Fig. 2). Principal component analysis confirmed this structure, with PC1 (51.27% variation) separating native (EHA, EBS, EZW) from invasive populations, and PC2 (35.80% variation) subdividing invasive populations into four groups: Group 1 (SKSK), Group 2 (QDH), Group 3 (DJ and XJ), and Group 4 (NDJ) (Fig. 2). ADMIXTURE analysis ($K = 4$, lowest CV error) identified four ancestral clusters: ETH (EHA, EBS, EZW), SKSK, QDH, and DXN (DJ, XJ, NDJ) (Fig. S7). Native populations and NDJ formed pure subpopulations, while DJ and XJ showed significant admixture from three ancestral sources, with minor hybridization in SKSK and QDH (Fig. 2). The genetic clustering indicated that Chinese invasive populations stem from three independent introductions (QDH, SKSK, and NDJ lineages). Although the clear separation from native Ethiopian populations suggests limited recent gene flow, strong founder effects and genetic drift during invasion may have amplified differentiation and obscured potential contributions from unsampled native-range sources.

3.5.4. Genetic differentiation among populations

Pairwise F_{ST} analysis of eight populations and four groups (ETH, SKSK, QDH, DXN) revealed EBS-NDJ had the highest differentiation among populations, while EBS-EHA showed the lowest (Table S22, Fig. 2). Among groups, ETH-QDH exhibited maximum F_{ST} , whereas DXN-SKSK showed significantly lower differentiation than other pairs (Wilcoxon rank-sum test, $P < 0.05$) (Table S23, Fig. S8).

3.5.5. Demographic history

To assess recent demographic histories, we estimated effective population size (N_e) over the past 200 generations using GONE2 (Fig. 3). The three native populations showed similar patterns: stable large sizes (N_e about 10,000–30,000) until approximately 50 generations ago, followed by sharp declines and a gradual recovery from approximately 12 generations onward. Among the five invasive populations in China, four (i.e., SKSK, DJ, XJ, NDJ) shared a U-shaped trajectory: starting at low N_e (200–500) around 50 generations ago, they underwent a severe bottleneck (about 30–50 generations ago) and then recovered to present N_e values of about 80–475. In contrast, the QDH population experienced an extreme, prolonged bottleneck beginning approximately 50 generations ago, with N_e dropping near zero by approximately 30 generations ago and showing no recovery. These distinct demographic signatures align with the species' introduction history into China since the 1960s, reflecting varied establishment scenarios and founder effects. Invasive populations exhibited reduced minor allele frequencies (MAF) and positive Tajima's D values (Table S21), indicating allele frequency shifts from bottlenecks. These genetic signatures collectively reveal invasion-associated population contraction in China, though recent rapid

expansion hasn't yet left detectable genomic traces.

3.5.6. Linkage disequilibrium analysis

Linkage disequilibrium (LD) decayed rapidly within <20 kb across all populations (Fig. S9). Among them, QDH showed the slowest LD decay rate while DJ demonstrated the fastest within 0–20 kb. Native populations (EHA, EBS, and EZW) maintained slower decay than invasive populations (NDJ, XJ, SKSK, and DJ). At the 20 kb distance threshold, QDH exhibited the highest LD degree (r^2 value) and DJ showed the lowest LD maintenance (Fig. S9).

3.6. Strong selection signatures contributing to new environmental adaptation of invasive *C. zillii* populations in China

We analyzed local adaptation in *C. zillii* by grouping invasive and native populations into four clusters based on population structure, identifying 27 candidate adaptation genes from overlapping genomic regions that simultaneously met the top 5% thresholds for four selection criteria (F_{ST} , π ratio, XP-CLR, and Tajima's D) (Table S24). The QDH-DXN comparison revealed 54 selected genes in the QDH population ($F_{ST} > 0.43$, $\pi[\text{QDH}]/\pi[\text{DXN}] > 31.07$, XP-CLR > 104.74, and Tajima's $D < -2.53$) (Fig. 4, Table S25), including thirteen immune-related genes (e.g., CD276 antigen-like) with strong signals (Fig. 4, Table S25). The SKSK-DXN comparison showed 85 selected genes in the SKSK population ($F_{ST} > 0.26$, $\pi[\text{SKSK}]/\pi[\text{DXN}] > 2.34$, XP-CLR > 10.17, and Tajima's $D < -1.16$) featuring seven immune-related genes (e.g., complement C3-like), four chemoreception genes [e.g., voltage-dependent L-type calcium channel subunit alpha-1D (*CACNA1F*), voltage-gated potassium channel subunit beta-2 (*KCNAB2*) and potassium voltage-gated channel subfamily A member 1 (*KCNA5*)], and two stress-response genes (e.g., cAMP-specific 3',5'-cyclic phosphodiesterase 4D) (Fig. S10, Table S26). The DXN-ETH comparison identified one selected gene in the DXN group (N(G), N(G)-dimethylarginine dimethylaminohydrolase 1), which is associated with antioxidant defence and stress response ($F_{ST} > 0.48$, $\pi[\text{DXN}]/\pi[\text{ETH}] > 13.41$, XP-CLR > 87.15, and Tajima's $D < -1.08$) (Fig. S11, Table S27). These findings demonstrate distinct adaptation signatures across populations or groups, with immune function being a major selection target.

4. Discussion

4.1. Comparative analysis and evaluation of genome assembly and annotation quality between redbelly tilapia and other tilapia species

Genomic analysis is essential for understanding redbelly tilapia invasion mechanisms, including population dynamics and adaptation. Since the first tilapia genome (*O. niloticus*) was published in 2014 (i.e., Brawand et al., 2014), several tilapia genomes (e.g., blue tilapia, Mozambique tilapia) have become publicly available. We present the first chromosome-level redbelly tilapia genome, assembled using PacBio long reads and Hi-C technology, validated through comparative analysis with published tilapia genomes. Our assembly (980.04 Mb) closely matches the Nile tilapia "XX_ONJp" genome (Tao et al., 2021a) and falls within the known tilapia genome size range (927.7–1070 Mb). At the contig level, our assembly achieves superior continuity ($N50 = 34.7$ Mb), surpassing other high-quality tilapia genomes (Table 1; Conte et al., 2017, 2019; Etherington et al., 2022; Tao et al., 2021a, 2021b). Additionally, the BUSCO assessment revealed that the genome assembly of *C. zillii* achieved a completeness of 99.72%, surpassing the BUSCO scores reported for most other published tilapia genome assemblies. This high-quality chromosome-level assembly provides crucial genomic resources for investigating redbelly tilapia's invasive success.

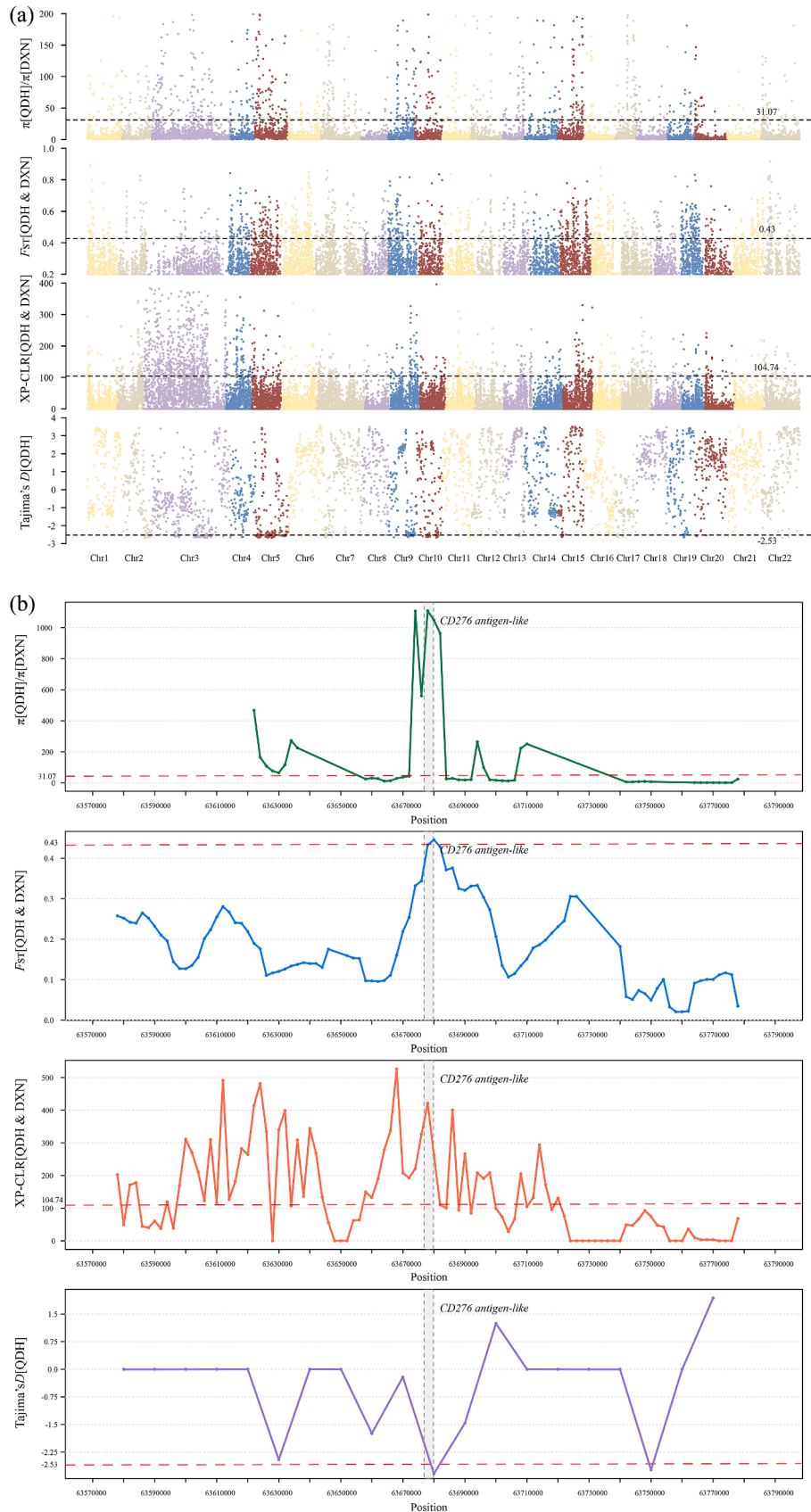


Fig. 4. Genome-wide scan signals of selective sweeps in QDH population. **(a)** Manhattan plot showing the distribution of the π ratios ($\pi[QDH]/\pi[DXN]$, top 5%), F_{ST} values (QDH and DXN, top 5%), XP-CLR values (QDH and DXN, top 5%), and Tajima's D values (QDH, top 5%) in nonoverlapped 10 kb slide windows along chromosomes. Horizontal dotted lines represent the cutoff fulfilling the requirement for the selected regions. The intersection of these candidate regions were defined as the selective sweep regions. **(b)** Distribution of π ratios, F_{ST} , XP-CLR, and Tajima's D values around the genomic regions that were positively selected in the QDH population. The red dotted line represents the outlier cutoff of π ratios, F_{ST} , XP-CLR, and Tajima's D values, respectively. The peak region in light grey background corresponds to a gene (*CD276* antigen-like) under selection in QDH population.

4.2. Genomic insights into the remarkable ecological adaptability of invasive redbelly tilapia

4.2.1. Olfactory receptor gene expansion and selection in signaling genes

Invasive fish rely on olfactory perception for detecting food, predators, and conspecifics (Hara, 1975). Teleost olfactory receptors (ORs) are divided into nine subfamilies (Liu et al., 2021), with expansions linked to invasion success in species such as round goby (*Neogobius melanostomus*) (Adrian-Kalchauer et al., 2020) and walking catfish (*Clarias batrachus*) (Li et al., 2018). *Coptodon zillii* possesses 255 predicted OR genes, ranking second among six compared species, with notable expansions in the aquatic-specific δ ($n = 136$), ζ ($n = 50$), and η ($n = 50$) subfamilies. These subfamilies are involved in detecting water-soluble odorants (Liu et al., 2024) and may facilitate environmental perception during invasion. Selective sweep analysis identified three olfactory signaling-related genes under selection in the SKSK population: *CACNA1F*, *KCNAB2* and *KCNA5*. L-type calcium channels, such as those encoded by *CACNA1F*, play a role in olfactory transduction and memory formation in vertebrates (Restrepo and Teeter, 1990; Ghosh et al., 2017). Although functional evidence in fish is limited, selection on these genes may enhance olfactory signal processing, contributing to behavioral adaptation to and exploitation of new environments.

4.2.2. Expansion and selection of innate and adaptive immune genes

Invasive species face novel pathogens, and immune gene expansions have been implicated in successful colonization, such as for round goby (Adrian-Kalchauer et al., 2020). *Coptodon zillii* shows a notable expansion of C-type lectin receptors (CTLRs), with 62 members, including 56 in the CLEC4 subfamily. CTLRs function as pattern recognition receptors in innate immunity (Ito et al., 2025; Sousa et al., 2024), and their expansion may enhance pathogen recognition in novel habitats. Additionally, selective sweep analyses revealed innate immune genes under selection in QDH [e.g., C-type lectin domain family 11 member A (*CLEC11A*)] and SKSK populations. For adaptive immunity, the B7 costimulatory molecule family is expanded in *C. zillii* (58 members), particularly the B7-H1 subfamily (35 genes). B7 molecules regulate T cell activation (Nagai and Azuma, 2019; Getu et al., 2023), and their expansion may facilitate immune regulation during pathogen exposure. Selective sweep analysis also identified adaptive immune genes under selection, including CD276 antigen gene (also known as B7-H3) in QDH, which modulates T cell responses (Sugamata et al., 2009). These findings suggest that both innate and adaptive immune gene families may contribute to environmental adaptation in invasive populations.

4.3. Sources of redbelly tilapia populations introduced to China

Reconstructing invasion routes is crucial for both understanding the evolutionary drivers of biological invasions and informing effective management strategies (Bock et al., 2015). While historical records of introduction pathways are often limited or unavailable, population genetics offers a powerful alternative by analyzing genetic structures to identify source populations and trace invasion routes. This approach enables hypothesis testing about invasion mechanisms while supporting evidence-based policy development for invasive species control.

Neighbor-Joining dendrogram, principal component, and Bayesian clustering analyses consistently showed that Ethiopia's native populations were not the sources of China's five invasive redbelly tilapia populations. Instead, the invasive populations clustered into three distinct ancestral components, suggesting three separate introduction events—corresponding to Lake Qiandao, Shuikou Reservoir, and Nanduijiang River. Published records show that China had three documented redbelly tilapia introductions between the 1960s and 1980s. The first occurred in 1963 when agricultural researchers in Taiwan introduced it from South Africa (Li, 1999), the second took place in 1978 when a Guangdong food company imported it from Thailand (Cai,

1979), and the third happened in 1981 with an introduction from Africa (De Silva et al., 2004). While our findings align with these records, the first two introductions were indirect—the South African source's origin is unspecified, and the Thai stock traced back to an undocumented 1949 Malaysian transfer (De Silva et al., 2004). Due to incomplete native-range sampling and the convoluted introduction pathways (involving multiple indirect transfers), neither our genetic data nor existing literature can definitively pinpoint the exact origins of China's invasive populations. It is important to note that strong genetic drift and bottleneck effects during colonization can diminish detectable signals from source populations, potentially causing invasive groups to appear as independent lineages even if they originated from known native regions. While our data provide strong evidence for the genetic makeup and demographic history of the Chinese invasion and do not support a direct contribution from the Ethiopian populations sampled here, they do not preclude origins from other unsampled native areas or complex admixture histories obscured by founder events. Therefore, broader conclusions about invasion pathways would benefit from more extensive native-range sampling in future studies.

4.4. Limitations of using population genomics to infer invasion mechanisms

While our study provides valuable insights into the potential genetic basis of invasion success in *C. zillii*, several limitations should be considered when interpreting the results. First, the native Ethiopian populations used in this study may not fully represent the ancestral gene pools from which the invasive Chinese populations were derived. Historical records indicate multiple and often indirect introduction pathways (e.g., via Thailand or South Africa), and our sampling within Africa was limited to three Ethiopian lakes. Genetic differences observed between Ethiopian and Chinese populations could therefore reflect either adaptive changes during invasion or pre-existing variation among geographically distinct African populations. Future studies incorporating broader native-range sampling across Africa and the Middle East, as well as genomic data from intermediate introduced populations (e.g., in Thailand), would help disentangle ancestral variation from invasion-specific selection. Second, our selective sweep analyses identified candidate genes associated with local adaptation in China, but these signals alone cannot confirm functional causation. Environmental association analyses or common-garden experiments would be needed to validate whether these genomic regions indeed contribute to phenotypic traits relevant to invasion. Third, our XP-CLR analysis relied on a default recombination rate (1 cM/Mb) due to the absence of a species-specific fine-scale recombination map for *C. zillii*. This simplification may bias the detection of selection signatures, as XP-CLR is known to be sensitive to recombination rate variation, and applying an inappropriate or uniform rate across the genome can substantially influence the identification of outlier windows under selection (Chen et al., 2010). Fourth, demographic processes such as bottlenecks, founder effects, and admixture can mimic or obscure signals of selection, complicating the identification of truly adaptive loci (REF) (Apat and Pfeifer, 2020). Finally, while we focused on olfactory and immune-related genes, other traits (e.g., thermal tolerance, reproductive behavior, or metabolic plasticity) likely also contribute to invasion success and warrant further investigation.

Despite these limitations, our chromosome-level genome and population genomic dataset provide a foundational resource for future functional and comparative studies, enabling more precise dissection of the molecular mechanisms underlying *C. zillii*'s invasive success.

5. Conclusions

This study presents a chromosome-level genome of *C. zillii* and analyzes 79 resequenced individuals from invasive populations in five Chinese water bodies and native populations in three Ethiopian lakes.

Results show significant expansion of olfaction and immune-related gene families and positive selection of immune response-associated genes in *C. zillii*, likely enhancing invasiveness. Population genomics traced China's invasive populations to three inferred sources, while selective sweep analysis identified strong selection in genes linked to environmental adaptation and stress resistance. These findings provide a genomic resource for invasive tilapia studies and advance understanding of adaptive evolution and invasion mechanisms in fishes.

Data availability statement

The data that support the findings of this study have been deposited into CNGB Sequence Archive (CNSA) (Guo et al., 2020) of China National GeneBank DataBase (CNGBdb) (Chen et al., 2010) with accession number CNP0007147 (<https://db.cngb.org/cnsa/>). All custom code and pipelines used for data analysis are available at https://github.com/zyllei/Coptodon_zillii_genome_code.

Ethics statement

The experimental protocol was approved by the Institutional Animal Care and Use Committee (IACUC) of Shanghai Ocean University (permit number: SHOU2024085), and complies with the Guidelines on Ethical Treatment of Experimental Animals established by the Ministry of Science and Technology, China.

Funding

This study was supported by the National Key Research and Development Program of China (2022YFC2601302).

CRediT authorship contribution statement

Shoujie Tang: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft. **Zhiyi Yan:** Data curation, Methodology. **Ying Xing:** Data curation, Methodology. **Keyan Dong:** Data curation. **Yuxuan Liang:** Data curation. **Chaoyi Wang:** Data curation. **Jiaying Zheng:** Data curation. **Temesgen Tola Geletu:** Resources. **Chenhong Li:** Methodology. **Yan Zhao:** Methodology. **Xiaowu Chen:** Methodology. **Yaolei Zhang:** Data curation, Methodology, Supervision. **Jinliang Zhao:** Funding acquisition, Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The affiliations with BGI-Research (including its Qingdao and Chongqing branches) are purely academic and research-oriented for this study, and none of the funding or institutional support received creates a conflict of interest.

Acknowledgements

We are grateful to Mr. Weimin Xiang (Department of Agriculture and Rural Affairs of Chun'an County, Zhejiang Province, China), Mr. Honggui Huang (Fisheries Research Institute of Fujian, China), Mr. Qiubiao Huang (Fisheries Research Institute of Huizhou City, Guangdong Province, China), Mr. Zhongbao Guo (Fisheries Science Research Institute of Guangxi Zhuang Autonomous Region, China), and Mr. Bing Wang (Guangdong HAID Group Co., Ltd.) for their selfless assistance during the sample collection.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.watbs.2026.100695>.

References

- Abdulmalik-Labe, O.P., Quilang, J.P., 2019. DNA barcoding of fishes from Lake Lanao, Philippines. *Mitochondrial DNA B* 4, 1890–1894. <https://doi.org/10.1080/23802359.2019.1614890>.
- Adrian-Kalchauer, I., Blomberg, A., Larsson, T., Musilova, Z., Peart, C.R., Pippel, M., Solbakken, M.H., Suurväli, J., Walser, J.-C., Wilson, J.Y., et al., 2020. The round goby genome provides insights into mechanisms that may facilitate biological invasions. *BMC Biol.* 18, 11. <https://doi.org/10.1186/s12915-019-0731-8>.
- Al-Zaidy, K.J., 2013. First record of *Tilapia zillii* (GEWAIS, 1848) in AL-Delmj marsh west AL-Diwania city middle of Iraq. *Diyala Agricultural Sciences Journal* 5, 9–16.
- Aldana, R., Freed, D., 2022. Data processing and germline variant calling with the Sentieon pipeline. In: Ng, C., Pisuogio, S. (Eds.), *Variant Calling. Methods in Molecular Biology*, 2493. Humana, New York, NY, USA, pp. 1–19. https://doi.org/10.1007/978-1-0716-2293-3_1.
- Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. <http://www.genome.org/cgi/doi/10.1101/gr.094052.109>.
- Andreu-Soler, A., Ruiz-Campos, G., 2013. Effects of exotic fishes on the somatic condition of the endangered killifish *Fundulus lima* (Teleostei: Fundulidae) in oases of Baja California Sur, Mexico. *Southwest. Nat.* 58, 192–201. <https://doi.org/10.1894/0038-4909-58.2.192>.
- Apata, M., Pfeifer, S.P., 2020. Recent population genomic insights into the genetic basis of arsenic tolerance in humans: the difficulties of identifying positively selected loci in strongly bottlenecked populations. *Heredity* 124, 253–262. <https://doi.org/10.1038/s41437-019-0285-0>.
- Bavali, S., Haghi, M., Zakeri, M., Kochanian, P., 2022. Feeding and reproduction ecological patterns of *Coptodon zillii* in Shadegan wetland: an aggressive potential species for local aquaculture. *Int. Aquat. Res.* 14, 71–79.
- Benfey, P.N., Mitchell-Olds, T., 2008. From genotype to phenotype: systems biology meets natural variation. *Science* 320, 495–497. <https://www.science.org/doi/10.1126/science.1153716>.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B* 57, 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
- Bernery, C., Bellard, C., Courchamp, F., Brosse, S., Gozlan, R.E., Jarić, I., Teletchea, F., Leroy, B., 2022. Freshwater fish invasions: a comprehensive review. *Annu. Rev. Ecol. Syst.* 53, 427–456. <https://doi.org/10.1146/annurev-ecolsys-032522-015551>.
- Bian, C., Li, J., Lin, X., Chen, X., Yi, Y., You, X., Zhang, Y., Lv, Y., Shi, Q., 2019. Whole genome sequencing of the blue tilapia (*Oreochromis aureus*) provides a valuable genetic resource for biomedical research on tilapias. *Mar. Drugs* 17, 386. <https://doi.org/10.3390/md17070386>.
- Birney, E., Clamp, M., Durbin, R., 2004. GeneWise and genomewise. *Genome Res.* 14, 988–995. <https://doi.org/10.1101/gr.1865504>.
- Bock, D.G., Caseys, C., Cousens, R.D., Hahn, M.A., Heredia, S.M., Hübner, S., Turner, K. G., Whitney, K.D., Rieseberg, L.H., 2015. What we still don't know about invasion genetics. *Mol. Ecol.* 24, 2277–2297. <https://doi.org/10.1111/mec.13032>.
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Brawand, D., Wagner, C.E., Li, Y.I., Malinsky, M., Keller, I., Fan, S., Simakov, O., Ng, A. Y., Lim, Z.W., Bezault, E., et al., 2014. The genomic substrate for adaptive radiation in African cichlid fish. *Nature* 513, 375–381. <https://doi.org/10.1038/nature13726>.
- Cai, Q.M., 1979. Freshwater fish from Thailand. *Freshw. Fish.* 9, 28 [in Chinese].
- Canonica, G.C., Arthington, A., McCrary, J.K., Thieme, M.L., 2005. The effects of introduced tilapias on native biodiversity. *Aquat. Conserv.* 15, 463–483. <https://doi.org/10.1002/aqc.699>.
- Carpenter, S.R., Stanley, E.H., Zanden, M.J.V., 2011. State of the world's freshwater ecosystems: physical, chemical, and biological changes. *Annu. Rev. Environ. Resour.* 36, 75–99. <https://doi.org/10.1146/annurev-environ-021810-094524>.
- Casal, C.M.V., 2006. Global documentation of fish introductions: the growing crisis and recommendations for action. *Biol. Invasions* 8, 3–11. <https://doi.org/10.1007/s10530-005-0231-3>.
- Casemiro, F.A.S., Bailly, D., da Graça, W.J., Agostinho, A.A., 2018. The invasive potential of tilapias (Osteichthyes, Cichlidae) in the Americas. *Hydrobiologia* 817, 133–154. <https://doi.org/10.1007/s10750-017-3471-1>.
- Chang, C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4, 7. <https://doi.org/10.1186/s13742-015-0047-8>.
- Chen, H., Patterson, N., Reich, D., 2010. Population differentiation as a test for selective sweeps. *Genome Res.* 20, 393–402. <http://www.genome.org/cgi/doi/10.1101/g.r.100545.109>.
- Chen, Y.Y., Sun, C.S., Zhan, A.B., 2017. Biological invasions in aquatic ecosystems in China. *Aquat. Ecosyst. Health* 20, 402–412. <https://doi.org/10.1080/14634988.2017.1403268>.

- Cheng, H.Y., Concepcion, G.T., Feng, X.W., Zhang, H.W., Li, H., 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* 18, 170–175. <https://doi.org/10.1038/s41592-020-01056-5>.
- Chervinski, J., Hering, E., 1973. *Tilapia zillii* (Gervais) (Pisces, Cichlidae) and its adaptability to various saline conditions. *Aquaculture* 2, 23–29. [https://doi.org/10.1016/0044-8486\(73\)90122-1](https://doi.org/10.1016/0044-8486(73)90122-1).
- Cishahayo, L., Yongo, E., Mutethya, E., Waithaka, E., Ndayishimiye, R., 2022. Length–weight relationship, condition factor, sex ratio and size at first maturity of the blue-spotted tilapia (*Oreochromis leucostictus*) in Lake Naivasha, Kenya. *Lakes Reservoirs Res. Manag.* 27, e12417. <https://doi.org/10.1111/lre.12417>.
- Conte, M.A., Gammerding, W.J., Bartie, K.L., Penman, D.J., Kocher, T.D., 2017. A high quality assembly of the Nile tilapia (*Oreochromis niloticus*) genome reveals the structure of two sex determination regions. *BMC Genom.* 18, 341. <https://doi.org/10.1186/s12864-017-3723-5>.
- Conte, M.A., Joshi, R., Moore, E.C., Nandamuri, S.P., Gammerding, W.J., Roberts, R.B., Carleton, K.L., Lien, S., Kocher, T.D., 2019. Chromosome-Scale assemblies reveal the structural evolution of African cichlid genomes. *GigaScience* 8, giz030. <https://doi.org/10.1093/gigascience/giz030>.
- Corrales, X., Katsanevakis, S., Coll, M., Heymans, J.J., Piroddi, C., Ofir, E., Gal, G., 2020. Advances and challenges in modelling the impacts of invasive alien species on aquatic ecosystems. *Biol. Invasions* 22, 907–934. <https://doi.org/10.1007/s10530-019-02160-0>.
- Costa-Pierce, B.A., 2003. Rapid evolution of an established feral tilapia (*Oreochromis spp.*): the need to incorporate invasion science into regulatory structures. *Biol. Invasions* 5, 71–84. <https://doi.org/10.1023/A:1024094606326>.
- Courtenay, W.R., 1997. Tilapias as non-indigenous species in the Americas: environmental, regulatory and legal issues. In: Costa-Pierce, B.A., Rakocy, J.E. (Eds.), *Tilapia Aquaculture in the Americas*. World Aquaculture Society, Baton Rouge, pp. 18–33.
- Coutts, S.R., Helmstedt, K.J., Bennett, J.R., 2018. Invasion lags: the stories we tell ourselves and our inability to infer process from pattern. *Divers. Distrib.* 24, 244–251. <https://doi.org/10.1111/ddi.12669>.
- Cristescu, M.E., 2015. Genetic reconstructions of invasion history. *Mol. Ecol.* 24, 2212–2225. <https://doi.org/10.1111/mec.13117>.
- Crutchfield, J.U., 1995. Establishment and expansion of redbelly tilapia and blue tilapia in a power plant cooling reservoir. *Am. Fish. Soc. Symp.* 15, 452–461.
- Crutchfield, J.U., Schiller, D.D., Herlong, D.D., Mallin, M.A., 1992. Establishment and impact of redbelly tilapia in a vegetated cooling reservoir. *J. Aquat. Plant Manag.* 30, 28–35.
- Cui, Y., Yuan, Y.B., Wang, B., Wu, B.L., Cai, M.Y., Fang, M., Shen, W.L., Wu, X.F., Han, F., Wang, Z.Y., 2025. T2T genomes unveil centromere architecture and adaptive divergence in large yellow croaker (*Larimichthys crocea*). *Adv. Sci.* 12, e06374. <https://doi.org/10.1002/adv.202506374>.
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 1000 Genomes Project Analysis Group, 2011. The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>.
- De Silva, S.S., Subasinghe, R.P., Bartley, D.M., Lowther, A., 2004. Tilapias as Alien Aquatics in Asia and the Pacific: a Review. FAO, Rome, p. 65p. FAO Fisheries Technical Paper No. 453. <https://www.fao.org/fishery/docs/CDrom/aquaculture/a0805e/documents/Y5728e02.pdf>.
- Deines, A.M., Wittmann, M.E., Deines, J.M., Lodge, D.M., 2016. Tradeoffs among ecosystem services associated with global tilapia introductions. *Rev. Fish. Sci. Aquacult.* 24, 178–191. <https://doi.org/10.1080/23308249.2015.1115466>.
- Dennis Jr, G., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., Lempicki, R.A., 2003. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* 4, R60. <https://doi.org/10.1186/gb-2003-4-9-r60>.
- Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., Aiden, E.L., 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* 356, 92–95. <https://www.science.org/doi/10.1126/science.aal3327>.
- Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., Aiden, E.L., 2016. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* 3, 99–101. <https://doi.org/10.1016/j.cels.2015.07.012>.
- Edgar, R.C., Myers, E.W., 2005. PILER: identification and classification of genomic repeats. *Bioinformatics* 21, i152–i158. <https://doi.org/10.1093/bioinformatics/bti1003>.
- El-Sayed, A.M., 2006. *Tilapia Culture*. CABI Publishing, Cambridge.
- Elton, C.S., 1958. *The Ecology of Invasion by Animals and Plants*. The University of Chicago Press, Chicago.
- Estoup, A., Guillemaud, T., 2010. Reconstructing routes of invasion using genetic data: why, how and so what? *Mol. Ecol.* 19, 4113–4130. <https://doi.org/10.1111/j.1365-294X.2010.04773.x>.
- Etherington, G.J., Nash, W., Ciezarek, A., Mehta, T.K., Barria, A., Peñaloza, C., Khan, M. G.Q., Durrant, A., Forrester, N., Fraser, F., et al., 2022. Chromosome-level genome sequence of the Genetically Improved Farmed Tilapia (GIFT, *Oreochromis niloticus*) highlights regions of introgression with *O. mossambicus*. *BMC Genom.* 23, 832. <https://doi.org/10.1186/s12864-022-09065-8>.
- Fatsi, P.S.K., Hashem, S., Kodama, A., Appiah, E.K., Saito, H., Kawai, K., 2020. Population genetics and taxonomic signatures of wild Tilapia in Japan based on mitochondrial DNA control region analysis. *Hydrobiologia* 847, 1491–1504. <https://doi.org/10.1007/s10750-020-04203-3>.
- Feng, X., Liu, S.L., Sui, X.Y., Chen, Y.F., Zhu, R., Jia, Y.T., Tong, J.G., Yu, X.M., Liu, C.L., Hansen, M.M., 2023. Genetic responses in sexual diploid and unisexual triploid goldfish (*Carassius auratus*) introduced into a high-altitude environment. *Mol. Ecol.* 32, 1955–1971. <https://doi.org/10.1111/mec.16864>.
- Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., Smit, A.F., 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. USA* 117 (17), 9451–9457. <https://doi.org/10.1073/pnas.1921046117>.
- Gallardo, D., Clavero, M., Sánchez, M.I., Vilà, M., 2016. Global ecological impacts of invasive species in aquatic ecosystems. *Glob. Change Biol.* 22, 151–163. <https://doi.org/10.1111/gcb.13004>.
- Geletu, T.T., Tang, S.J., Xing, Y., Zhao, J.L., 2024. Ecological niche and life-history traits of redbelly tilapia (*Coptodon zillii*, Gervais 1848) in its native and introduced ranges. *Aquat. Living Resour.* 37, 2. <https://doi.org/10.1051/alr/2023030>.
- Getu, A.A., Tigabu, A., Zhou, M., Lu, J.R., Fodstad, Ø., Tan, M., 2023. New frontiers in immune checkpoint B7-H3 (CD276) research and drug development. *Mol. Cancer* 22, 43. <https://doi.org/10.1186/s12943-023-01751-9>.
- Ghosh, A., Carew, S.J., Chen, X.H., Yuan, Q., 2017. The role of L-type calcium channels in olfactory learning and its modulation by norepinephrine. *Front. Cell. Neurosci.* 11, 394. <https://doi.org/10.3389/fncel.2017.00394>.
- Gozlan, T.E., 2008. Introduction of non-native freshwater fish: is it all bad? *Fish. Fish.* 9, 106–115. <https://doi.org/10.1111/j.1467-2979.2007.00267.x>.
- Gu, D.E., Hu, Y.C., Xu, M., Wei, H., Luo, D., Yang, Y.X., Yu, F.D., Mu, X.D., 2018a. Fish invasion in the river systems of Guangdong Province, South China: possible indicators of their success. *Fish. Manag. Ecol.* 25, 44–53. <https://doi.org/10.1111/fme.12265>.
- Gu, D.E., Luo, G., Huang, H.K., Jia, T., Zhang, C., Shen, Y.X., Chen, X.X., Xu, M., Wei, H., Fang, M., et al., 2023. The list of key invasive alien species under key management released by the ministry of agriculture and rural affairs and six other departments. *China Fish.* 1, 8–11 [in Chinese].
- Gu, D.E., Mu, X.D., Xu, M., Luo, D., Wei, H., Li, Y.Y., Zhu, Y.J., Luo, J.R., Hu, Y.C., 2016. Identification of wild tilapia species in the main Rivers of south China using mitochondrial control region sequence and morphology. *Biochem. Systemat. Ecol.* 65, 100–107. <https://doi.org/10.1016/j.bse.2016.02.007>.
- Gu, D.E., Yu, F.D., Xu, M., Wei, H., Mu, X.D., Luo, D., Yang, Y.X., Pan, Z., Hu, Y.C., 2018b. Temperature effects on the distribution of two invasive tilapia species (*Tilapia zillii* and *Oreochromis niloticus*) in the rivers of South China. *J. Freshw. Ecol.* 33, 511–524. <https://doi.org/10.1080/02705060.2018.1517698>.
- Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., Wortman, J.R., 2008. Automated eukaryotic gene structure annotation using EvidenceModeler and the program to assemble spliced alignments. *Genome Biol.* 9, R7. <https://doi.org/10.1186/gb-2008-9-1-r7>.
- Hansen, J.D., Pasquier, L.D., Lefranc, M.P., Lopez, V., Benmansour, A., Boudinot, P., 2009. The B7 family of immunoregulatory receptors: a comparative and evolutionary perspective. *Mol. Immunol.* 46, 457–472. <https://doi.org/10.1016/j.molimm.2008.10.007>.
- Hara, T.J., 1975. Olfaction in fish. *Prog. Neurobiol.* 5, 271–335. [https://doi.org/10.1016/0301-0082\(75\)90014-3](https://doi.org/10.1016/0301-0082(75)90014-3).
- Hoese, D.F., Bray, D.J., Paxton, J.R., Allen, G.R., 2006. *Fishes*. In: Beasley, O.L., Wells, A. (Eds.), *Zoological Catalogue of Australia*, 35. ABRS and CSIRO Publishing, Victoria, Australia, p. 1472, 2.
- Hufbauer, R.A., Torchin, M.E., 2008. Integrating ecological and evolutionary theory of biological invasions. In: Nentwig, W. (Ed.), *Biological Invasions*. Springer, Berlin, Germany, pp. 79–96.
- Innal, D., Giannetto, D., 2017. Age structure and length-weight relationship of non-native redbelly tilapia *Coptodon zillii* (Gervais, 1848) (Cichlidae) in the Pinarbasi Spring Creek (Burdur, Turkey). *Acta Zool. Bulg.* 9, 111–116.
- Ishikawa, T., Tachiara, K., 2008. Age, growth and maturation of the redbelly tilapia *Tilapia zillii* introduced into the Haeburu Reservoir on Okinawa-jima Island. *Fish. Sci.* 74, 527–532. <https://doi.org/10.1111/j.1444-2906.2008.01555.x>.
- Ito, T., Guenther, C., Ishikawa, E., Yabuki, T., Naga, M., Nakatani, Y., Yamasaki, S., 2025. Phylogenetic and structural insights into the origin of C-type lacinia in vertebrates. *Immunogenetics* 77, 18. <https://doi.org/10.1007/s00251-025-01375-x>.
- Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al., 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>.
- Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., Haeseler, A.V., Jermini, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. <https://doi.org/10.1038/nmeth.4285>.
- Katoh, K., Rozewicki, J., Yamada, K.D., 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings Bioinf.* 20, 1160–1166. <https://doi.org/10.1093/bib/bbx108>.
- Keane, R.M., Crawley, M.J., 2002. Exotic plant invasions and the enemy release hypothesis. *Trends Ecol. Evol.* 17, 164–170. [https://doi.org/10.1016/S0169-5347\(02\)02499-0](https://doi.org/10.1016/S0169-5347(02)02499-0).
- Keilwagen, J., Hartung, F., Grau, J., 2019. GeMoMa: Homology-based gene prediction utilizing intron position conservation and RNA-Seq data. In: Kollmar, M. (Ed.), *Gene Prediction. Methods in Molecular Biology*, 1962. Humana, New York, NY, USA. <https://doi.org/10.1007/978-1-4939-9173-0-9>.
- Kent, W.J., 2002. BLAT—The BLAST-like alignment tool. *Genome Res.* 12, 656–664. <https://doi.org/10.1101/gr.229202>.
- Kim, D., Langmead, B., Salzberg, S.L., 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. <https://doi.org/10.1038/nmeth.3317>.
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., Marra, M.A., 2009. Circos: an information aesthetic for comparative genomics.

- Genome Res. 19, 1639–1645. <http://www.genome.org/cgi/doi/10.1101/gr.092759.109>.
- Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. <https://doi.org/10.1093/molbev/msw054>.
- Langmead, B., Salzberg, S., 2012. Fast gapped-read alignment with bowtie 2. *Nat. Methods* 9, 357–359. <https://doi.org/10.1038/nmeth.1923>.
- Leprieux, F., Beauchard, O., Huguency, B., Grenouillet, G., Brosse, S., 2008. Null model of biotic homogenization: a test with the European freshwater fish fauna. *Divers. Distrib.* 14, 291–300. <https://doi.org/10.1111/j.1472-4642.2007.00409.x>.
- Letunic, I., Bork, P., 2021. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. <https://doi.org/10.1093/nar/gkab301>.
- Li, J.H., 1999. Contributions of Taiwan's agricultural varietal improvement to farming. *Taiwan Agricultural Research* 2, 16–18 [in Chinese].
- Li, J.F., Wang, J.B., Liu, Z.Y., Liu, Q., Guo, C., Liao, C.S., Li, W., Guo, C.B., Zhang, T.L., Liu, J.S., 2023. Comparative study on reproductive biology of two co-invasion tilapia species in the Shanmei reservoir. *Acta Hydrobiol. Sin.* 47, 657–665 [in Chinese with English abstract].
- Li, L., Stoeckert Jr., C.J., Roos, D.S., 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. <http://www.genome.org/cgi/doi/10.1101/gr.1224503>.
- Li, N., Bao, L., Zhou, T., Yuan, Z.H., Liu, S.K., Dunham, R., Li, Y.N., Wang, K., Xu, X.Y., Jin, Y.L., et al., 2018. Genome sequence of walking catfish (*Clarias batrachus*) provides insights into terrestrial adaptation. *BMC Genom.* 19, 952. <https://doi.org/10.1186/s12864-018-5355-9>.
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
- Li, Y.M., Yu, F.H., 2023. Managing the risk of biological invasions. *iScience* 26, 108221. <https://doi.org/10.1016/j.isci.2023.108221>.
- Liu, H., Chen, C.H., Lv, M.L., Liu, N., Hu, Y.F., Zhang, H.L., Enbody, E.D., Gao, Z.X., Andersson, L., Wang, W.M., 2021. A chromosome-level assembly of blunt snout bream (*Megalobrama amblycephala*) genome reveals an expansion of olfactory receptor genes in freshwater fish. *Mol. Biol. Evol.* 38, 4238–4251. <https://doi.org/10.1093/molbev/msab152>.
- Liu, N., Guan, S.H., Wang, W.M., Liu, H., 2024. Development of olfactory organs and gene expression patterns of representative olfactory receptors in blunt snout bream (*Megalobrama amblycephala*). *J. Fish. China* 48, 049109 [in Chinese with English abstract].
- Liu, Z.D., Xing, L.S., Huang, W.L., Liu, B., Wan, F.H., Raffa, K.F., Hofstetter, R.W., Qian, W.Q., Sun, J.H., 2022. Chromosome-level genome assembly and population genomic analyses provide insights into adaptive evolution of the red turnstone beetle, *Dendroctonus valens*. *BMC Biol.* 20, 190. <https://doi.org/10.1186/s12915-022-01388-y>.
- Lowe-McConnell, R.H., 2006. The Tilapia Trail: the Life Story of a Fish Biologist. MPM Publishing, Ascot.
- Lu, Z.K., 1987. Characteristics of tilapia culture species and their necessary environmental conditions. *Fishery Information & Strategy* 3, 21–23 [in Chinese].
- Lucaci, A.G., Zehr, J.D., Enard, D., Thornton, J.W., Kosakovsky Pond, S.L., 2023. Evolutionary shortcuts via multinucleotide substitutions and their impact on natural selection analyses. *Mol. Biol. Evol.* 40. <https://doi.org/10.1093/molbev/msad150>.
- Lv, Y.Y., Li, Y.P., Fang, M., Liu, Y., Wang, Y.Y., Yang, Y.X., Zou, Y.C., Shi, Q., Mu, X.D., 2025. Chromosome-level genome assembly reveals adaptive evolution of the invasive Amazon sailfin catfish (*Pterygoplichthys pardalis*). *Commun. Biol.* 8, 616. <https://doi.org/10.1038/s42003-025-08029-4>.
- Löytynoja, A., Goldman, N., 2010. webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. *BMC Bioinf.* 11, 579. <https://doi.org/10.1186/1471-2105-11-579>.
- Marcais, G., Kingsford, C., 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27, 764–770. <https://doi.org/10.1093/bioinformatics/btr011>.
- Martin, C.W., Valentine, M.M., Valentine, J.F., 2010. Competitive interactions between invasive Nile tilapia and native fish: the potential for altered trophic exchange and modification of food webs. *PLoS One* 5, e14395. <https://doi.org/10.1371/journal.pone.0014395>.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. <https://doi.org/10.1101/gr.107524.110>.
- Mendes, F.K., Vanderpool, D., Fulton, B., Hahn, M.W., 2020. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* 36, 5516–5518. <https://doi.org/10.1093/bioinformatics/btaa1022>.
- Mohamed, A.R.M., Al-Wan, S.M., 2020. Evaluation of biological characters of the invasive species, *Coptodon zillii* in the Garmat Ali River, Basrah, Iraq. *Int. J. Fish. Aquat. Stud.* 8, 176–185.
- Murrell, B., Weaver, S., Smith, M.D., Wertheim, J.O., Murrell, S., Aylward, A., Eren, K., Pollner, T., Martin, D.P., Smith, D.M., Scheffler, K., Kosakovsky Pond, S.L., 2015. Gene-wide identification of episodic selection. *Mol. Biol. Evol.* 32, 1365–1371. <https://doi.org/10.1093/molbev/msv035>.
- Nagai, S., Azuma, M., 2019. The CD28-B7 family of Co-signaling molecules. In: Azuma, M., Yagita, H. (Eds.), *Co-signal Molecules in T Cell Activation*, Advances in Experimental Medicine and Biology, 1189. Springer, Singapore. https://doi.org/10.1007/978-981-32-9717-3_2.
- Ng, P.K.L., Chou, L., Lam, T., 1993. The status and impact of introduced freshwater animals in Singapore. *Biol. Conserv.* 64, 19–24. [https://doi.org/10.1016/0006-3207\(93\)90379-F](https://doi.org/10.1016/0006-3207(93)90379-F).
- Nguyen, L.T., Schmidt, H.A., Haeseler, A.V., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. <https://doi.org/10.1093/molbev/msu300>.
- Panprommin, D., Soontornprasit, K., Pangson, T., 2018. Comparison of three molecular methods for species identification of the family Cichlidae in Kwan Phayao, Thailand. *Mitochondrial DNA* 30, 184–190. <https://doi.org/10.1080/24701394.2018.1472248>.
- Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., Salzberg, S.L., 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. <https://doi.org/10.1038/nbt.3122>.
- Peterson, M.S., Slack, W.T., Woodley, C.M., 2005. The occurrence of non-indigenous Nile Tilapia, *Oreochromis niloticus* (Linnaeus) in coastal Mississippi, USA: ties to aquaculture and thermal effluent. *Wetlands* 25, 112–121. [https://doi.org/10.1672/0277-5212.2005.025\[0112:TOONNT\]2.0.CO;2](https://doi.org/10.1672/0277-5212.2005.025[0112:TOONNT]2.0.CO;2).
- Philippart, J.-C.L., Ruwet, J.-C.L., 1982. Ecology and distribution of tilapias. In: Pullin, R. S.V., Lowe-McConnell, R.H. (Eds.), *The Biology and Culture of Tilapias*. ICLARM Conference Proceedings 7, 432. International Center for Living Aquatic Resources Management, Manila, Philippines, pp. 15–59. <http://pubs.iclarm.net/libinfo/Pdf/Pu/b%20C6%207.pdf>.
- Price, A.L., Jones, N.C., Pevzner, P.A., 2005. De novo identification of repeat families in large genomes. *Bioinformatics* 21, i351–i358. <https://doi.org/10.1093/bioinformatics/bti1018>.
- Puckett, E.E., Munshi-South, J., 2019. Brown rat demography reveals pre-commensal structure in eastern Asia before expansion into Southeast Asia. *Genome Res.* 29, 762–770. <https://doi.org/10.1101/gr.235754.118>.
- Ranallo-Benavidez, T.R., Jaron, K.S., Schatz, M.C., 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* 11, 1432. <https://doi.org/10.1038/s41467-020-14998-3>.
- Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., Aiden, E.L., 2015. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 162, 687–688. <https://doi.org/10.1016/j.cell.2015.07.024>.
- Restrepo, D., Teeter, J.H., 1990. Olfactory neurons exhibit heterogeneity in depolarization-induced calcium changes. *Am. J. Physiol.-Cell. Ph.* 258, C1051–C1061. <https://doi.org/10.1152/ajpcell.1990.258.6.C1051>.
- Rhie, A., Walenz, B.P., Koren, S., Phillippy, A.M., 2020. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* 21, 245. <https://doi.org/10.1186/s13059-020-02134-9>.
- Ricciardi, A., MacIsaac, H.J., 2011. Impacts of biological invasions on freshwater ecosystems. In: Richardson, D.M. (Ed.), *Fifty Years of Invasion Ecology: the Legacy of Charles Elton*. Blackwell Publishing Ltd, New York, USA, pp. 211–223.
- Robins, C.R., Bailey, R.M., Bond, C.E., Brooker, J.R., Lachner, E.A., Lea, R.N., Scott, W.B., 1991. Common and scientific names of fishes from the Americas and Canada. *Am. Fish. Soc. Spec. Publ.* 20, 183.
- Roosbehfar, R., Dehestani-Esfandabadi, M., Rooshbehfar, S., 2014. First record of the redbelly tilapia, (*Tilapia zillii* Gervais, 1848), in Iran. *J. Appl. Ichthyol.* 30, 1045–1046. <https://doi.org/10.1111/jai.12469>.
- Rouget, M., Robertson, M.P., Wilson, J.R.U., Hui, C., Essl, F., Renteria, J.L., Richardson, D.M., 2016. Invasion debt-quantifying future biological invasions. *Divers. Distrib.* 22, 445–456. <https://doi.org/10.1111/ddi.12408>.
- Ryan, S.F., Lombaert, E., Espeset, A., Vilá, R., Talavera, G., Dinca, V., Doelman, M.M., Renshaw, M.A., Eng, M.W., Horne, E.A., et al., 2019. Global invasion history of the agricultural pest butterfly *Pieris rapae* revealed with genomics and citizen science. *P. Natl. Acad. Sci. USA*. 116, 20015–20024. <https://doi.org/10.1073/pnas.1907492116>.
- Santiago, E., Novo, I., Pardiñas, A.F., Saura, M., Wang, J.L., Caballero, A., 2020. Recent demographic history inferred by high-resolution analysis of linkage disequilibrium. *Mol. Biol. Evol.* 37, 3642–3653. <https://doi.org/10.1093/molbev/msaa169>.
- Seebens, H., Blackburn, T., Dyer, E., Genovesi, P., Hulme, P.E., Jeschke, J.M., Pagad, S., Pyšek, P., Winter, M., Arianoutsou, M., et al., 2017. No saturation in the accumulation of alien species worldwide. *Nat. Commun.* 8, 14435. <https://doi.org/10.1038/s41467-017-04435-5>.
- Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.J., Vert, J.P., Heard, E., Dekker, J., Barillot, E., 2015. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* 16, 259. <https://doi.org/10.1186/s13059-015-0831-x>.
- Simao, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.
- Sousa, C.R., Yamasaki, S., Brown, G.D., 2024. Myeloid C-type lectin receptors in innate immune recognition. *Immunity* 57, 700–717. <https://doi.org/10.1016/j.immuni.2024.03.005>.
- Stamatakis, A., 2014. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.
- Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., Morgenstern, B., 2006. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* 34, W435–W439. <https://doi.org/10.1093/nar/gkl1200>.
- Stolzer, M., Lai, H., Xu, M.L., Sathaye, D., Vernot, B., Durand, D., 2012. Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. *Bioinformatics* 28, i409–i415. <https://doi.org/10.1093/bioinformatics/bts386>.

- Sugamata, R., Suetake, H., Kikuchi, K., Suzuki, Y., 2009. Teleost B7 expressed on monocytes regulates T cell responses. *J. Immunol.* 182, 6799–6806. <https://doi.org/10.4049/jimmunol.0803371>.
- Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595. <https://doi.org/10.1093/genetics/123.3.585>.
- Talavera, G., Castresana, J., 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* 56, 564–577. <https://doi.org/10.1080/10635150701472164>.
- Tang, S.J., Xing, Y., Geletu, T.T., Zhao, J.L., 2025. Trophic plasticity of the invasive redbelly tilapia (*Coptodon zillii*) in China inferred from DNA metabarcoding analysis. *Ecol. Evol.* 15, e71118. <https://doi.org/10.1002/ece3.71118>.
- Tao, W., Cao, J., Xiao, H., Zhu, X., Dong, J., Kocher, T.D., Lu, M., Wang, D., 2021b. A chromosome-level genome assembly of Mozambique tilapia (*Oreochromis mossambicus*) reveals the structure of sex determining regions. *Front. Genet.* 12, 796211. <https://doi.org/10.3389/fgene.2021.796211>.
- Tao, W.J., Xu, L.H., Zhao, L., Zhu, Z.X., Wu, X., Min, Q.W., Wang, D.S., Zhou, Q., 2021a. High-quality chromosome-level genomes of two tilapia species reveal their evolution of repeat sequences and sex chromosomes. *Mol. Ecol. Resour.* 21, 543–560. <https://doi.org/10.1111/1755-0998.13273>.
- Tarkan, S.A., 2022. *Tilapia zillii* (redbelly tilapia). CABI Compendium. <https://doi.org/10.1079/cabicompendium.61147>.
- Varela-Romero, A., Ruiz-Campos, G., Yépiz-Velázquez, L.M., Alaníz-García, J., 2002. Distribution, habitat and conservation status of desert pupfish (*Cyprinodon macularius*) in the Lower Colorado River Basin, Mexico. *Rev. Fish Biol. Fish.* 12, 157–165. <https://doi.org/10.1023/A:1025006920052>.
- Vences, M., Stützer, D., Rasoamampionona Raminosoa, N., Ziegler, T., 2022. Towards a DNA barcode library for Madagascar's threatened ichthyofauna. *PLoS One* 17, e0271400. <https://doi.org/10.1371/journal.pone.0271400>.
- Weir, B.S., Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38, 1358–1370. <https://doi.org/10.2307/2408641>.
- Williamson, M.H., Fitter, A., 1996. The characters of successful invaders. *Biol. Conserv.* 78, 163–170. [https://doi.org/10.1016/0006-3207\(96\)00025-0](https://doi.org/10.1016/0006-3207(96)00025-0).
- Wisotsky, S.R., Kosakovsky, S.L., Shank, S.D., Muse, S.V., 2020. Synonymous site-to-site substitution rate variation dramatically inflates false positive rates of selection analyses: ignore at your own peril. *Mol. Biol. Evol.* 37, 2430–2439. <https://doi.org/10.1093/molbev/msaa037>.
- Wu, T.D., Watanabe, C.K., 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21, 1859–1875. <https://doi.org/10.1093/bioinformatics/bti310>.
- Xing, Y., Tang, S.J., Geletu, T.T., Zhao, J.L., 2025. Invasion and management implications of *Coptodon zillii* in China. *Chin. J. Ecol.* 44, 645–653 [in Chinese with English abstract].
- Xiong, W., Guo, C.B., Gozlan, R.E., Liu, J.S., 2023. Tilapia introduction in China: economic boom in aquaculture versus ecological threats to ecosystems. *Rev. Aquacult.* 15, 179–197. <https://doi.org/10.1111/raq.12710>.
- Xu, Z., Wang, H., 2007. LTR FINDER: an efficient tool for the prediction of full length LTR retrotransposons. *Nucleic Acids Res.* 35, W265–W268. <https://doi.org/10.1093/nar/gkm286>.
- Yang, Z., 2007. Paml 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. <https://doi.org/10.1093/molbev/msm088>.
- Yongo, E., Agembe, S.W., Manyala, J.O., Waithaka, E., 2022. Aspects of the biology and population structure of *Oreochromis niloticus*, *Coptodon zillii* and *Oreochromis leucostictus* tilapia in Lake Naivasha, Kenya. *Lakes Reservoirs Res. Manag.* 27, e12398. <https://doi.org/10.1111/lre.12398>.
- Yongo, E., Zhang, P., Mutethya, E., Zhao, T., Guo, Z., 2023. The invasion of tilapia in South China freshwater systems: a review. *Lakes Reservoirs Res. Manag.* 28, e12429. <https://doi.org/10.1111/lre.12429>.
- Zhang, C., Dong, S., Xu, J., He, W., Yang, T., 2019. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35, 1786–1788. <https://doi.org/10.1093/bioinformatics/bty875>.
- Zhang, X.H., Shi, Y.H., Chen, J., 2015. Molecular characterization of a transmembrane C-type lectin receptor gene from ayu (*Plecoglossus altivelis*) and its effect on the recognition of different bacteria by monocytes/macrophages. *Mol. Immunol.* 66, 439–450. <https://doi.org/10.1016/j.molimm.2015.05.009>.