BMC Genomics

RESEARCH Open Access

Gap-Free T2T assemblies of *Micropterus salmoides* identify a Y-Linked SV hotspot underlying sexual dimorphism

Shili Liu[1], Jianbo Zheng[1], Shun Cheng[1], Wenping Jiang[1], Meili Chi[1], Chao Zhu[1], Miao Peng[1] and Fei Li[1*]

Abstract

Background The largemouth bass (*Micropterus salmoides*, LMB), a globally significant aquaculture species, exhibits pronounced growth dimorphism between the sexes, yet the genetic basis of its sex determination remains unresolved. While published genomes exist, gaps, misassemblies, and unresolved repetitive regions hinder investigations into structural variations (SVs) linked to sexually dimorphic traits. High-quality telomere-to-telomere (T2T) assemblies are essential to address these limitations and advance sex determination studies.

Methodology We generated the first T2T gap-free genome assemblies for male and female LMB using Oxford Nanopore (ONT), PacBio HiFi, and Hi-C sequencing. Genome quality was validated via BUSCO, Merqury, Hi-C contact maps, and synteny comparisons. Repeat content and protein-coding genes were annotated, and male-specific SVs were identified through cross-genome alignment. Transcriptomic data from gonads and muscle were integrated to analyze differential expression in previously unresolved regions (PURs).

Results The female (871.07 Mb) and male (873.86 Mb) T2T assemblies achieved BUSCO completeness of 99.3% and QV scores > 50, resolving all gaps in prior references. Comparative analysis identified a 90-kb Y-specific SV-enriched region (SVER-Y) on chromosome 10, harboring male-specific insertions and marked by transposable element (TE) enrichment, with LTRs (74.9% higher) and LINEs (63.7% higher) relative to the chromosome-wide average. PURs (46.96 Mb), localizing to telomeres and centromeres, showed high TE content (LINEs: 29.6%, LTRs: 23.3%) and tissue-specific gene expression. Notably, 610 putative PUR-associated transcripts were differentially expressed between gonadal and somatic tissues, suggesting regulatory roles in sexual development.

Conclusion This study provides the complete T2T genomes for LMB, correcting structural errors in prior assemblies and resolving repetitive regions critical for understanding sexual dimorphism. The SVER-Y region, enriched in male-specific SVs and TEs, offers novel insights into early Y chromosome differentiation. These resources establish a foundation for deciphering sex determination mechanisms and optimizing marker-assisted breeding strategies in LMB aquaculture.

Keywords Micropterus salmoides, Largemouth bass, Telomere-to-telomere genome, Sex determination, Structural variation

*Correspondence:
Fei Li
lifeibest1022@163.com

[1]Key Laboratory of Freshwater Aquatic Animal Genetic and Breeding of Zhejiang Province, Ministry of Agriculture and Rural Affairs, Zhejiang Institute of Freshwater Fisheries (Zhejiang Freshwater Fishery Environmental Monitoring Station), Huzhou 313001, China

Background

The advancements and advantages of telomere-to-telomere (T2T) genome assemblies have revolutionized genomics in recent years, providing unprecedented completeness and precision as a vital tool for multidisciplinary research [1]. By constructing a gap-free T2T genome, the T2T-CHM13 human genome closed the last 8% of GRCh38, added 2,000 paralogous genes, and incorporated previously missing heterochromatic, centromeric and acrocentric sequences [2–4]. The T2T genome technology has been rapidly applied in aquaculture [5]. A gap-free, chromosome-level reference genome systematically resolves assembly challenges in complex repetitive regions (e.g., disease-resistant gene clusters and telomere-centromere modules) [6–11], and provides a high-precision framework for deciphering genetic regulatory networks and molecular mechanisms underlying key traits (such as stress resistance and growth performance) in aquatic species [12–16]. This breakthrough enables multi-chromosomal resolution of complex structural variations and, when integrated with multi-omics strategies (e.g., transcriptional and epigenetic profiling), substantially advances molecular breeding technologies [17, 18]. For instance, T2T-based assemblies have enhanced Marker-Assisted Selection (MAS) accuracy by 30%, doubled the efficiency of gene editing target screening, and elevated Genome-wide Selection (GS) model prediction accuracy to over 85%, significantly shortening breeding cycles [19–24]. Furthermore, comprehensive genome annotations provide molecular evidence for elucidating environmental adaptation mechanisms (e.g., salinity response and hypoxia tolerance), supporting precision gene-pyramiding breeding programs and phenotype-genotype association optimization strategies [25–27].

In 2020, the first chromosome-level genome assembly of the largemouth bass (*Micropterus salmoides*) was published, laying a crucial foundation for germplasm conservation, trait research, and genetic improvement of the species. Its assembly and annotation were widely cited by researchers as NCBI reference sequences [27]. With advancements in sequencing technologies, an updated genome with higher continuity and improved scaffolding metrics was published in 2022 [28].To date, four versions (ASM1485139v1,ASM2243578v1,CAFS_Msal_1,HNU_Msal_1.0) have been released on NCBI. However, these assemblies remain incomplete relative to T2T standards. They show inter-release size differences, widespread sequence gaps and contig misjoins, which undermine analytical reliability. This technical fragmentation, compounded by the paucity of functionally-curated annotation resources, substantially constrains their utility in comparative genomics.

Teleost fish, as the most diverse vertebrate group, exhibit extraordinary variability in their sex-determination systems [29–39]. Pronounced sexual dimorphism is observed in multiple teleost species, underpinning crucial economic traits such as differential growth rates and body size variations. The strategic development of sex-associated molecular markers and sex manipulation technologies has revolutionized aquaculture practices by enabling precise sex-specific breeding protocols that enhance both productivity and market value [40–42]. Female largemouth bass typically grow faster and live longer than males [43, 44]. Recent resequencing-based F_{ST} analyses in multiple sex-specific studies have identified male-specialized regions on chromosome 10 and developed a suite of associated molecular markers [45–47]. Nevertheless, comprehensive investigations into genome-wide structural variations, sex-determining genes, and their regulatory mechanisms remain critically lacking.

To bridge critical genomic gaps, we generated the gap-free, T2T assemblies for largemouth bass (*Micropterus salmoides*), comprising sex-specific genomes (female and male). Comparative genomics with established reference genomes and male-female divergence analyses revealed a 90 kb Y-specific SV-enriched region (SVER-Y). These assemblies exhibit unprecedented accuracy and contiguity, resolving prior inconsistencies and establishing a comprehensive genomic framework for evolutionary studies, population management, and functional research in this ecologically and economically vital species.

Methods

Sample collection and extraction

A 1-year-old female largemouth bass (body weight: 532.6 g) from Deqing Comprehensive Experimental Base of Zhejiang Institute of Freshwater Fisheries was euthanized by immersion in a 20-L aerated tank containing 500 mg/L MS-222 (CAS No.: 886-86-2, Sigma-Aldrich, USA). Immediately upon death, skeletal muscle, ovarian tissue, and visceral tissues (liver, kidney, heart, and gill) were collected for sequencing. Tissues were washed three times with phosphate-buffered saline (PBS), frozen in liquid nitrogen and stored at – 80 °C for subsequent nucleic acid extraction. Male fish of the same age (body weight: 510.1 g) were processed using the same protocol, with the exception that the sampled gonad tissue was testis (replacing ovary in females).

High-molecular-weight (HMW) genomic DNA was isolated from muscle tissue using the sodium dodecyl sulfate (SDS) lysis method. DNA integrity and concentration were evaluated by 1.0% agarose gel electrophoresis and quantified using a Qubit 4.0 Fluorometer (Invitrogen Life Technologies, Carlsbad, CA, USA). DNA purity was further assessed with a NanoDrop™ One UV-Vis spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and cross-validated by the Qubit 4.0 Fluorometer,

with absorbance ratios of OD260/280 ranging from 1.8 to 2.0 and OD260/230 between 2.0 and 2.2.

Total RNA was isolated by manual tissue homogenization followed by TRIzol™-based extraction. TRIzol™ Reagent (Cat# 15596018; Invitrogen, USA) was utilized for this procedure. RNA concentration and purity were subsequently evaluated using the same methodology as described for DNA quantification (agarose gel electrophoresis, Qubit 4.0 Fluorometer, and NanoDrop™ One spectrophotometer).x` by TRIzol™-based extraction. TRIzol™ Reagent (Cat# 15596018; Invitrogen, USA) was utilized for this procedure. RNA concentration and purity were subsequently evaluated using the same methodology as described for DNA quantification (agarose gel electrophoresis, Qubit 4.0 Fluorometer, and NanoDrop™ One spectrophotometer).

PacBio HiFi sequencing

A total of 5 μg genomic DNA was fragmented using g-TUBEs (Covaris, USA), followed by purification with 0.45×AMPure XP beads (Beckman Coulter, USA). Size selection targeting ~ 20-kb fragments was performed using the BluePippin™ System (Sage Science, USA) with a 0.75% agarose cassette. Sequencing libraries were prepared using the SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, USA), involving sequential steps of end repair, A-tailing, and blunt-end ligation of hairpin adapters to generate SMRTbell® constructs. Post-ligation, libraries were treated with the SMRTbell Enzyme Cleanup Kit (exonuclease digestion) and repurified using AMPure PB Beads (Pacific Biosciences). Target fragments were further refined via size selection on the PippinHT System (Sage Science).

Library quality was validated using the Agilent 2100 Bioanalyzer (Agilent Technologies, USA) with High Sensitivity DNA chips. Final libraries were sequenced on a PacBio Revio system (Kind Sequenon, Wuhan, China) employing the Revio Binding Kit 2.2 and Sequencing Primer V5. HiFi reads were generated using the Circular Consensus Sequencing (CCS) mode with a minimum of 3 full passes per read.

Nanopore Ultra-long sequencing

Large DNA fragments (> 100 kb) were size-selected using the Blue Pippin™ System (Sage Science, Beverly, MA) with 0.75% agarose gel cassettes. For Nanopore ultra-long-read sequencing, 10 μg genomic DNA was processed using the Ligation Sequencing Kit (SQK-LSK109, Oxford Nanopore Technologies, Oxford, UK) following the manufacturer's protocol. Sequencing libraries were constructed and loaded onto a PromethION flow cell (Oxford Nanopore Technologies) for ultra-long-read data generation (N50 > 150 kb) at Kind Sequenon (Wuhan, China). Raw signal data were basecalled in real-time using the Dorado software (v0.7.0) in "super-accurate" mode, with a minimum quality score (Q) threshold of 10.

Hi-C sequencing

1 g of muscle tissue was cross-linked with 1% (w/v) fresh formaldehyde for 10 min at room temperature, followed by quenching with 0.2 M glycine for 5 min; nuclei were isolated by cell lysis, resuspended in 150 μL of 0.1% SDS, and incubated at 65 °C for 10 min to permeabilize chromatin; SDS was neutralized by adding 120 μL H_2O and 30 μL 10% Triton X-100, followed by 15-min incubation at 37 °C; chromatin was digested overnight at 37 °C with 150 U DpnII in 1× NEB buffer; after heat inactivation of DpnII at 65 °C for 20 min, cohesive ends were filled using the Klenow fragment (37 °C, 2 h), and proximity ligation was performed with T4 DNA ligase at 16 °C for 4 h; cross-links were reversed by incubating with 200 μg/mL proteinase K (Thermo Fisher Scientific) overnight at 65 °C; DNA was purified using QIAamp DNA Mini Kits (Qiagen), sheared to 200–400 bp fragments, and ligation junctions were enriched via Dynabeads™MyOne™ Streptavidin C1 (Thermo Fisher Scientific); the Hi-C library was prepared using the MGIEasy Universal DNA Library Prep Kit v1.0 and sequenced on DNBSEQ-T7RS platform (MGI, Shenzhen, China).

MGI WGS sequencing

Genomic DNA libraries were prepared for MGI sequencing using the VAHTS Universal DNA Library Prep Kit for MGI (Cat# NDM607, Vazyme Biotech, Nanjing, China). Briefly, 1 μg of genomic DNA was fragmented using the Covaris M220 Focused-ultrasonicator (Covaris, MA, USA). Fragmented DNA was size-selected (200–400 bp average insert size) with MGIEasy DNA Clean Beads (Cat# 1000005279, MGI, Shenzhen, China). The selected DNA underwent end-repair, 3'-adenylation, and adapter ligation following the VAHTS kit protocol. Amplified libraries were purified using MGIEasy DNA Clean Beads (Cat# 1000005279, MGI). Library quality was assessed by Qubit 4.0 Fluorometer (Thermo Fisher Scientific, MA, USA) for concentration measurement and Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA) for size distribution analysis.

The library was heat-denatured and circularized using splint oligos from the MGIEasy Circularization Module (Cat# 1000005260, MGI, Shenzhen, China), generating single-strand circular DNA (ssCir DNA). Final libraries meeting quality thresholds were sequenced on the DNBSEQ-T7RS platform (MGI, Shenzhen, China).

MGI RNA-seq sequencing

2 μg total RNA underwent mRNA enrichment using VAHTS mRNA Capture Beads 2.0 (Cat# N304, Vazyme Biotech, Nanjing, China). Enzymatically fragmented

RNA was size-selected (200–400 bp average insert size) prior to library construction with the VAHTS Universal v10 RNA-seq Library Prep Kit for MGI (Vazyme Biotech). Library quality was assessed using the Qubit 4.0 Fluorometer (Thermo Fisher Scientific, MA, USA) for concentration quantification and the Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA) for size distribution profiling.

The library was heat-denatured and circularized with splint oligos from the MGIEasy Circularization Module (Cat# 1000005260, MGI, Shenzhen, China), yielding single-strand circular DNA (ssCir DNA). Final libraries meeting quality criteria were sequenced on the DNBSEQ-T7RS platform (MGI, Shenzhen, China).

Survey for genome size

The genome size of the target species was estimated based on a k-mer frequency analysis using Jellyfish (v2.3.0) and GenomeScope (v2.0) [48, 49]. High-quality MGI paired-end reads were used for this analysis. A k-mer size of 21 was chosen to balance sensitivity and computational efficiency.

K-mer counting was performed using Jellyfish with the following parameters: -m 21 -C, and the resulting histogram was generated for input into GenomeScope. The k-mer distribution was modeled using GenomeScope to estimate genome size, heterozygosity, and repetitive content.

Genome assembly

The genome assembly was performed using Hifiasm (v0.16.1) with a combination of PacBio HiFi reads and ultra-long Oxford Nanopore reads [50]. Only Nanopore reads longer than 100 kb were included in the assembly process to enhance contiguity and scaffold bridging. Hifiasm was run in the default mode for hybrid assembly using both read types.

To assess the accuracy and structural integrity of the assembly, Hi-C data were employed for validation. Potential mis-assemblies were detected by mapping Hi-C reads to the draft assembly and examining abnormal contact patterns, such as discontinuous interaction signals and off-diagonal contacts, which indicate structural inconsistencies [51].

To assess the structural continuity and detect potential assembly artifacts, we visualized the genome assembly graphs generated by the Hifiasm in GFA (Graphical Fragment Assembly) format using Bandage v0.8.1 [52].

Telomere region refinement

To improve the assembly of telomeric regions, PacBio HiFi reads were first aligned to the draft genome using Minimap2 (v2.24) with the -ax map-hifi preset [53, 54]. Reads that remained unmapped, contained canonical telomeric repeat motifs (e.g., AACCCT), or were aligned to the unassembled ends of chromosomes were extracted. These reads were then subjected to local assembly to generate contigs that extended and overlapped with the adjacent assembled sequences at chromosome termini. The resulting contigs were realigned to the draft genome, and sequences showing confident overlap with chromosome ends were used to replace the corresponding telomeric gap regions. Ultra-long Oxford Nanopore reads were also utilized to guide the assembly graph resolution of HiFi reads, providing long-range information for selecting the most consistent extension paths in telomeric regions.

Telomere and centromere analysis

Telomeric sequences in the assembled genome were identified and analyzed using TeloExplorer, a module of the quarTeT toolkit, with the parameter setting -c animal [55]. This mode was used to detect canonical animal-type telomeric repeats and evaluate the presence and completeness of telomeric regions at chromosome termini.

For centromere identification, we used Centromics (v1.0) to analyze raw ONT and PacBio HiFi reads. Centromics detects centromeric satellite arrays based on k-mer periodicity and abundance patterns. Consensus satellite sequences were extracted, clustered, and mapped across the genome to identify putative centromeric regions.

Calculation of assembly quality value

To assess the consensus accuracy of the male and female T2T genome assembly, 21-mers were first generated from MGI short reads using the meryl tool included in Merfin (v1.1) [56]. These k-mers were then used as input for Merqury (v1.3) to calculate the consensus base Quality Value (QV), providing a k-mer-based estimate of the assembly's base-level accuracy [57].

Assembly completeness and coverage assessment

Genome assembly completeness was evaluated using BUSCO (v5.4.3) with the vertebrata_odb10 lineage dataset [58, 59]. The analysis was performed in genome mode to assess the presence of conserved single-copy orthologs, providing a quantitative measure of gene space completeness.

To evaluate sequencing coverage, both Oxford Nanopore ultra-long reads and PacBio HiFi reads were aligned separately to the assembled female T2T genome using Minimap2 (v2.24) [53, 54]. The -ax map-ont preset was used for Nanopore reads, and -ax map-hifi was used for HiFi reads. Coverage depth and distribution were then calculated based on the alignment results.

Repeat annotation

Repetitive elements in the genome were annotated using a combination of homology-based and *de novo* approaches. For homology-based annotation, RepeatMasker (v4.1.5) (available at https://www.repeatmasker.org/) was used with default parameters to search against known repeat libraries and classify transposable elements based on sequence similarity. For *de novo* repeat identification, EarlGrey was employed to generate a custom repeat library and annotate novel repeat families not present in existing databases [60].

Gene prediction and functional annotation

Gene structure annotation was performed using EGAPX (available at https://github.com/ncbi/egapx) with the taxonomic identifier set to taxid = 27,706. RNA-seq reads derived from muscle and gonad tissues were provided as transcriptomic evidence to guide gene prediction.

All predicted protein-coding genes were functionally annotated by aligning their translated protein sequences against several public databases, including the NCBI non-redundant (NR) protein database (http://ftp.ncbi.nih.gov/blast/db/), Swiss-Prot (http://web.expasy.org/docs/swiss-prot_guideline.html), and the Kyoto Encyclopedia of Genes and Genomes (KEGG).

Methylation based on PacBio long reads

Methylation of cytosine residues was analyzed using PacBio HiFi long-read data. Specifically, 5-methylcytosine (5mC) sites in the CpG context were identified by first aligning the raw HiFi reads to the female T2T genome using pbmm2 (v1.13.0) (https://github.com/PacificBiosciences/pbmm2). Methylation calling was then performed using pb-CpG-tools (v1.0.0) (https://github.com/PacificBiosciences/pb-CpG-tools), which detects methylated cytosines based on kinetic signals preserved in HiFi sequencing data.

Detection and characterization of previously unresolved regions

To identify previously unresolved regions (PURs), the chromosome-level assembly of the earlier version ASM1485139v1 (ASM1485139v1.chromosomes.fasta) was aligned to the newly assembled female T2T genome using Winnowmap2 (v2.03) [61]. Regions present in female T2T but absent from the aligned chromosomes of ASM1485139v1 were defined as PURs. The non-aligned regions were extracted and summarized using BEDTools (v2.30.0) [62, 63].

To classify the PURs, we overlapped their coordinates with the gene annotation file using the bedtools intersect function. Based on their overlap with annotated genomic features, PURs were categorized into gene-associated regions and repeat-rich regions based on their overlap with annotated gene models and repetitive elements. Further characterization was conducted to evaluate repeat content and transcript evidence within each PUR category.

Transcriptome analysis and functional enrichment

RNA-seq reads from gonad and muscle tissues were aligned to the T2T reference genome using STAR (v2.7.10a) with default parameters [64]. Gene-level expression was quantified based on uniquely mapped reads. A gene was considered to have transcriptional support if it was covered by at least 10 reads in the dataset.

Due to the lack of biological replicates, differential putative gene expression analysis between gonadal and muscular tissues was performed using the edgeR package (v4.0.16) with a quasi-likelihood F-test under a no-replicate assumption [65]. Genes with an absolute log2 fold change greater than or equal to 1 and a p-value less than 0.05 were considered differentially expressed.

Significantly differentially expressed genes were subjected to functional enrichment analysis using the clusterProfiler package (v4.14.4) [66]. Gene Ontology (GO) terms and KEGG pathways were tested using the enrichGO and enrichKEGG functions, respectively, with the genome-wide annotated gene set used as the background.

Identification of sex-linked region

To identify structural variations (SVs) between the male and female genomes, we first aligned the male assembly to the female genome using Minimap2 (v2.17) with default parameters for assembly-to-assembly alignment [53, 54]. Structural variants were then called using paftools.js call from the Minimap2 toolkit. We extracted male-specific SV sequences (insertions and deletions relative to the female genome) and compared them against previously published *Micropterus salmoides* genome assemblies using BLAST.

To ensure sex-specificity, we applied the following criteria:

(1) Insertions identified in the male genome relative to the female must be absent from the female genome and present exclusively in the male genome;
(2) Deletions identified in the male genome relative to the female must be present only in the female genome and absent in the male.

Results

Female and male T2T genome assembly

We selected an adult male and female *Micropterus salmoides* from Deqing Comprehensive Experimental Base of Zhejiang Institute of Freshwater Fisheries (Fig. 1A). To assemble the female T2T genome, a total of

31.30 Gb of ultra-long ONT data (35.99×coverage), 46.25 Gb of PacBio HiFi data (53.16×coverage), and 106.64 Gb of Hi-C data were generated (Table S1).

Initially, a hybrid genome assembly was constructed by integrating Oxford Nanopore (ONT) and PacBio (PB) sequencing datasets, achieving a gap-free assembly through pre-processing. Using Hi-C data, the contigs were anchored and oriented onto 23 pseudochromosomes, thereby confirming the absence of misassembly. All autosomes were systematically named (e.g., Chr1 for *M. salmoides* chromosome 1) based on sequence similarity to the NCBI genome reference ASM1485139v1 (GCF_014851395.1).

Subsequently, manual telomere repair was performed using PacBio HiFi data. HiFi reads containing telomeric sequences that did not align to the assembly and mapping to chromosomal ends lacking assembled telomeres were extracted. These reads underwent local assembly to determine the optimal path for telomere extension, producing sequences overlapping one side of the terminal gaps. The terminal regions were updated using these newly assembled sequences (Fig. 1B). The final gap-free T2T genome assembly, T2T_MSal_F1.0, covering all the 23 chromosomes, was obtained with a total length of 871.07 Mb (Fig. 2; Table 1). The total length closely approximated the genome size estimated from NGS reads (Fig. S1), and was consistent with the ASM2243578v1 assembly, but notably different from ASM1485139v1. Using the same strategy, we obtained sequencing data from multiple platforms for the assembly of the male T2T genome (Table S1). A final gap-free T2T assembly size of 873.86 Mb was determined for the T2T_MSal_M1.0 genome (Fig. 2; Table 1).

Validation of the T2T genome assemblies

Multiple strategies were further used to assess the integrity and accuracy of T2T genomes. Benchmarking Universal Single-Copy Orthologs (BUSCO) analyses revealed identical 99.0% completeness of BUSCO genes in the T2T_MSal_F1.0 and T2T_MSal_M1.0 genomes (Fig. S2). The genome-wide average consensus quality value (QV) of the female and male genomes were estimated to be 51.36 and 50.76 using NGS data, reflecting high base accuracy (Table 1). In addition, we employed Hi-C contact matrices to further assess the structural accuracy of the T2T assemblies. The resulting interaction heatmaps exhibited uniform and continuous intra-chromosomal contact signals with clearly defined diagonal patterns, indicative of high assembly contiguity and correct chromosomal anchoring. No abnormal inter-chromosomal interactions or local disruptions were observed, suggesting the absence of large-scale misassemblies in both the T2T_MSal_F1.0 and T2T_MSal_M1.0 genomes (Fig. 1C; Fig. S3).

Alignment of HiFi and ONT raw reads to the genome assemblies resulted in high mapping efficiencies (99.91%–99.96%), while chromosome coverage and depth based on ONT data exhibited excellent uniformity (Fig. 1D; Fig. S3). Graph-based assembly structure derived from the HifiasmGFA output showed predominantly long, well-connected contigs with minimal branching, supporting the structural integrity of the T2T assemblies (Fig. S4). Both male and female *M. salmoides* T2T genomes exhibit high synteny, as evidenced by their alignment with published genomes (Fig. 3A) and illustrated in the Circos plot (Fig. 2).

Verification of telomeres (TTAGGG repeats) on each chromosome demonstrates that all 23 chromosomes have achieved gap-free T2T assembly (Table 1). Using Centromics software, centromere sequences were identified from raw ONT and PB datasets. Through pairwise comparative analysis, repetitive sequences across all 23 chromosomes were classified into four centromeric satellite units: Sat1 (723 bp), Sat2 (725 bp), Sat3 (608 bp) and Sat4 (39 bp) (Table S2 and S3), providing critical clues for subsequent validation via fluorescence in situ hybridization (FISH). These data suggest that our final T2T genomes for female and male LMB are of high quality. As the first two T2T gap-free genomes of LMB to date, these assemblies are comparable in quality to the high-quality reference genome of the model organism *Danio rerio* (Table S4) [67–70].

Genome annotation

Annotation of repeat content based on a combination of prediction and evidence-based tools in the female genome revealed 41.59% repetitive elements, including 16.22% DNA transposons, 7.27% long interspersed nuclear elements (LINEs), 0.9% short interspersed nuclear elements (SINEs), 12.42% long terminal repeats (LTRs) and 1.18% satellites (Table S5; Fig.S5A). Similar results were observed in the male genome (Table S5; Fig. S5B). The total repeat sequence content was closely aligned with ASM2243578v1 and slightly higher than the NCBI reference genome version ASM1485139v1. Although significant differences exist in repeat classification among different genome assembly versions of *Micropterus salmoides* and across perciform species, the annotated proportions of each repeat subcategory consistently show an increasing trend with improvements in genome assembly quality (Table S6) (Shao et al. 2018; Vij et al. 2016; Tine et al. 2014).

Coding gene annotation in the repeat-masked genome was conducted using the NCBI Eukaryotic Genome Annotation Pipeline (EGAP), which also supports annotations for the Human Pangenome Project and Vertebrate Genome Project [71, 72]. RNA-seq data from gonad and muscle tissues were utilized for transcriptome-based

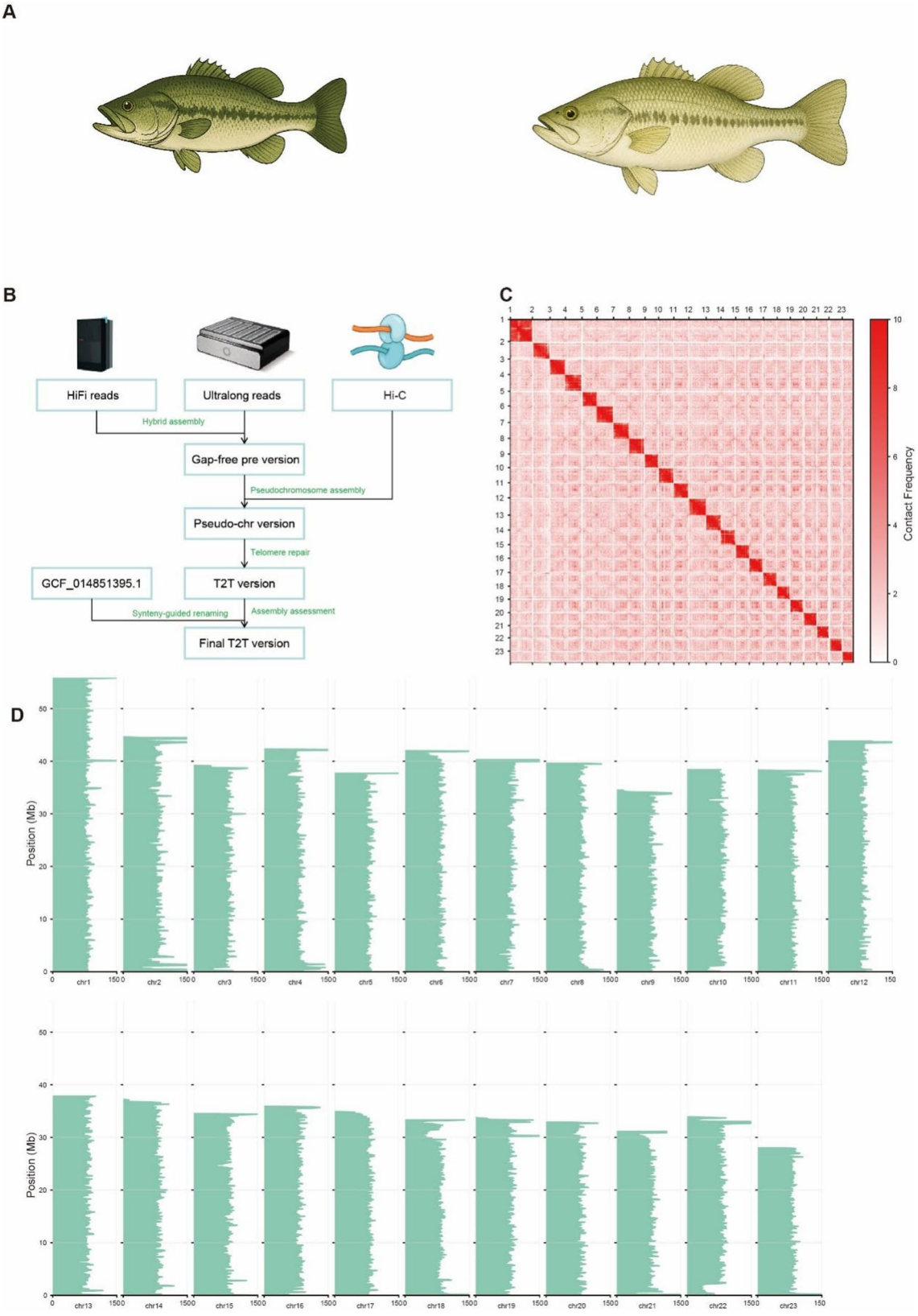

Fig. 1 (See legend on next page.)

(See figure on previous page.)

Fig. 1 Genome assembly and evaluation of largemouth bass (*Micropterus salmoides*). **A**, Illustration of largemouth bass. **B**, Workflow of the genome assembly process. HiFi reads, ultra-long reads, and Hi-C data were integrated to generate a hybrid assembly. A gap-free pre-assembly was produced, followed by pseudomolecule construction, telomere repair, and generation of a T2T assembly. The final T2T version was obtained after synteny-guided renaming and assembly evaluation. **C**, Hi-C contact map of the female individual showing chromatin interaction frequencies across the 23 chromosomes. D, Read coverage depth along each chromosome

training (Table S1). A total of 22,411 high-confidence protein-coding genes were finally identified in the female genome, whereas 22,527 were annotated in the male genome (Fig. S5C and S5D; Table S7). Of the total predicted genes in *M. salmoides*, 95.98% (21,512/22,411) in the female and 96.97% (21,845/22,527) in the male genome were functionally annotated, demonstrating high coverage of coding region characterization (Fig. S5E and S5F).

Improvement of T2T_MSal_F1.0

A high degree of collinearity was observed between T2T_MSal_F1.0 and the NCBI LMB genome reference ASM1485139v1, indicating the high quality and reliability of T2T_MSal_F1.0 (Fig. 3A). The assembly T2T_MSal_F1.0 closed all the 4729 gaps in ASM1485139v1 and resolved the telomeric regions. T2T_MSal_F1.0 corrected numerous structural errors in the ASM1485139v1, for example, the 57.92 ~ 67.34 Mb region of chromosome 1 in the ASM1485139v1 aligns to the proximal end of chromosome 12 in the T2T_MSal_F1.0, while the first 10 Mb of chromosome 2 aligns to the distal end of chromosome 22 in the T2T assembly. Large-scale inversions (> 5 Mb) are present at the distal end of Chr2 and the proximal regions of Chr5 and Chr17(Fig. 3A). These are unlikely to represent chromosomal rearrangements or fusions, as collinearity analysis in the ASM2243578v1 genome assembly supports the structural accuracy of the T2T organization (Fig.S6A and S6B). A substantial number of assembly gaps are observed on Chr1 of the ASM1485139v1 genome, with prominent clustering near chromosomal termini (potentially telomeric or centromeric domains) characterized by dense repetitive elements (Fig. 3D). The misassembly of chromosome 1 in the ASM1485139v1 genome was observed, attributable to its highly repetitive sequences, as evidenced by abrupt spikes in sequencing depth within IGV (Fig. 3E). These two contigs were annotated as telomeric regions in our genome assembly. The elevated error rate of PB CLR sequencing and the complexity of these sequences are more plausible explanations.

The cultured populations of largemouth bass exhibit significant strain uniformity and low genetic diversity [73]. The T2T_MSal_F1.0 genome exhibits exceptionally high collinearity with published genomic references, making it a reliable reference for analysis of previously unresolved regions (PURs). PURs, with a total size of 46.96 Mb, were identified by comparing the full sequences of T2T_MSal_F1.0 and ASM1485139v1 to infer misassembled or newly assembled regions (Fig. 4a). These regions are predominantly composed of repetitive sequences, with Long Interspersed Nuclear Elements (LINEs) representing the highest proportion at 29.61% (13.90 Mb), followed by Long Terminal Repeats (LTRs) at 23.34% (10.96 Mb) and Satellite DNA at 22.11% (10.38 Mb) (Fig. 4B). We labeled the PUR regions on the chromosomes and observed that the majority of these regions were located at chromosomal ends, with a small fraction positioned in central regions, consistent with the telomere and centromere regions we previously identified (Fig. 4A).

Through *de novo* prediction, 704 putative novel transcripts were identified within the PURs, among which 610 showed detectable preliminary expression based on transcriptomic data. Differential expression analysis between gonad and muscle tissues revealed that 435 and 436 of these transcripts were differentially expressed in two respective comparisons, with 223/212 and 199/237 transcripts being upregulated and downregulated, respectively (Fig. 4C). Although these transcripts were not significantly enriched in known GO terms or KEGG pathways, their consistent differential expression between gonadal and muscular tissues suggests potential biological relevance. This lack of enrichment may stem from incomplete or underdeveloped functional annotations in current databases, particularly for non-model teleosts. As such, PUR-associated transcripts represent promising candidates for investigation into tissue-specific regulatory mechanisms that remain poorly understood.

Identification of Y-specific SV-enriched region (SVER-Y)

Sex chromosomes (e.g., the mammalian XY or avian ZW systems) gradually lose their recombination ability during evolution, leading to structural differentiation (such as the degeneration of Y/W chromosomes) [74, 75]. Recent studies have revealed that sex-specific mutation accumulation and the activity of transposable elements (TEs) are core driving factors in this process [76–78]. The largemouth bass (*Micropterus salmoides*) exhibits a monofactorial male heterogametic sex determination system (XX/XY), accompanied by undifferentiated sex chromosomes. Previous studies have demonstrated that sex-specific SNPs are significantly enriched on Chr10 in largemouth bass (LMB). The male-specific markers on Chr7 identified in Du et al.'s study were found to align to Chr10 (ASM1485139v1)

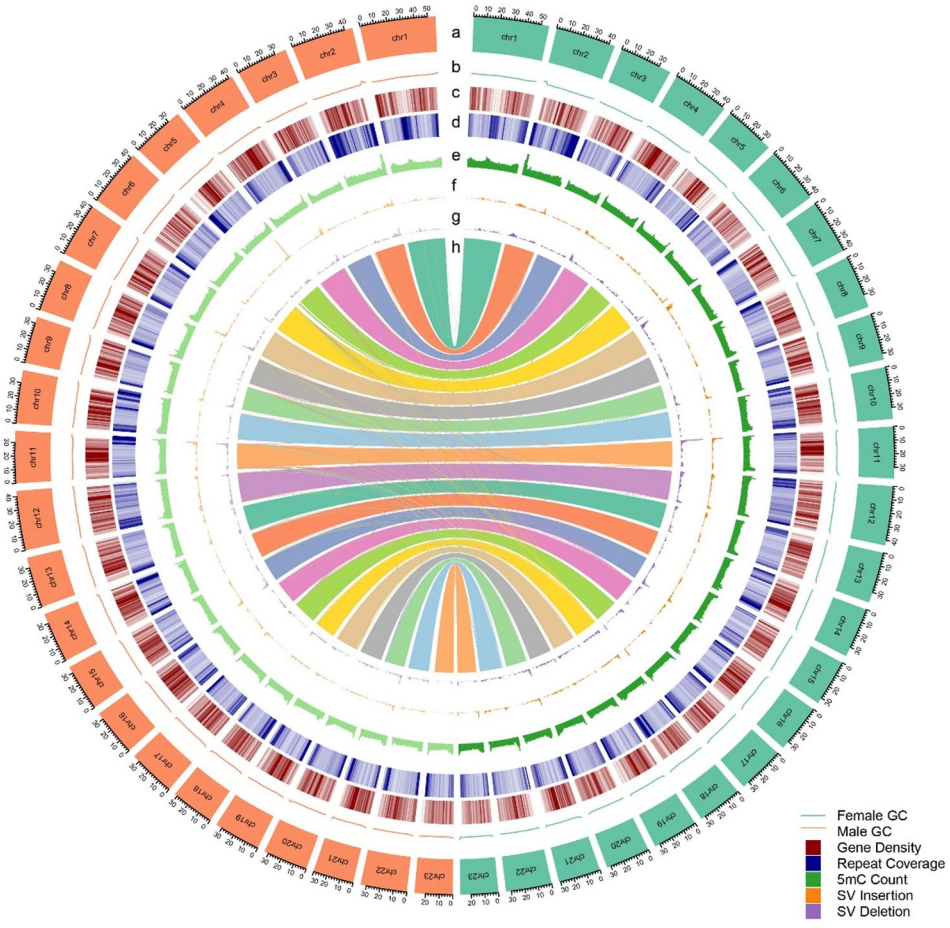

Fig. 2 Comparative genomic features and synteny between female and male largemouth bass genomes.
a, Chromosomal coordinates of the female (right) and male (left) genomes
b, GC content
c, Gene density
d, Repeat coverage
e, Distribution of 5-methylcytosine (5mC)
f, Structural variant (SV) insertions
g, Structural variant (SV) deletions
h, Collinearity between female and male chromosomes

upon re-analysis, likely due to differences in the reference sequences used by the authors. This further underscores the critical role of a high-quality genome assembly in subsequent gene mapping and functional studies. A 0.17 Mb sequence segment (HNU_Msal_1.0,Chr7:3,338,087–3,510,683) from the Du et al.'s study was extracted and precisely mapped to the T2T_MSal_F1.0 reference genome (Chr10: 35,986,998–38,350,439) through collinear analysis. Similarly, a 6.62 Mb segment on chromosome 10 (ASM1485139v1,Chr10:30.37-36.99 Mb) from He et al.'s work was aligned to T2T_MSal_F1.0(Chr10:31547266–38351426) using synteny-based localization (FigS7). The sex-specific SNP-dense chromosomal interval in the latter study completely encompasses the homologous region identified in the former.

Benefiting from our male and female genome assemblies, as well as the four published versions of the largemouth bass (LMB) genome, we can systematically identify sex-specific regions at the structural variation (SV) level. Through collinear comparative analysis between T2T_MSal_F1.0 and T2T_MSal_M1.0, we identified a total of 24,177 structural variations, specifically insertion and deletion (InDel) events with sizes exceeding 50 bp, which showed no significant variation in chromosomal distribution (Fig. 5A). To further minimize individual variation, we validated the structural variations (SVs) across six assemblies, ultimately identifying 36 male-specific insertions unique to Chr10 (Fig. 5B). Notably, the largest SV (13 kb insertion) ranked first and was located within a 90-kb genomic region harboring 6 shared SVs (Fig. 5C and d), which drew our attention.

Table 1 Summary of the present and four published Micropterus salmoides genome assemblies

Assembly feature	ASM1485139v1ASM1485139v1	ASM2243578v1	CAFS_Msal_1	HNU_Msal-1.0	T2T_MSal_F1.0	T2T_MSal_M1.0
Size of assembly	963,613,914	844,893,128	875,695,184	877,669,248	871,075,730	873,866,909
Contig N50	1,227,323	15,296,922	14,257,025	5,637,587	37,869,003	38,378,056
Scaffold N50	36,481,429	35,769,121	35,933,337	37,181,689	37,869,003	38,378,056
Anchored chromosome size	862,264,699	827,394,836	834,598,459	861,610,724	871,075,730	873,866,909
Anchored chromosome percent (%)	89.48	98.03	95.30	98.17	100.00	100.00
Gap number	4,729	241	449	246	0	0
Single telomeric chromosome	0	0	0	0	0	0
Double telomeric chromosome	0	0	0	0	23	23
Annotated Protein-coding gene	23,701	26,370	-	-	22,411	22,527
Complete BUSCOs (%)	97.20	97.49	99.20	99.20	99.30	99.30
Merqury	-	-	-	-	51.36	50.76

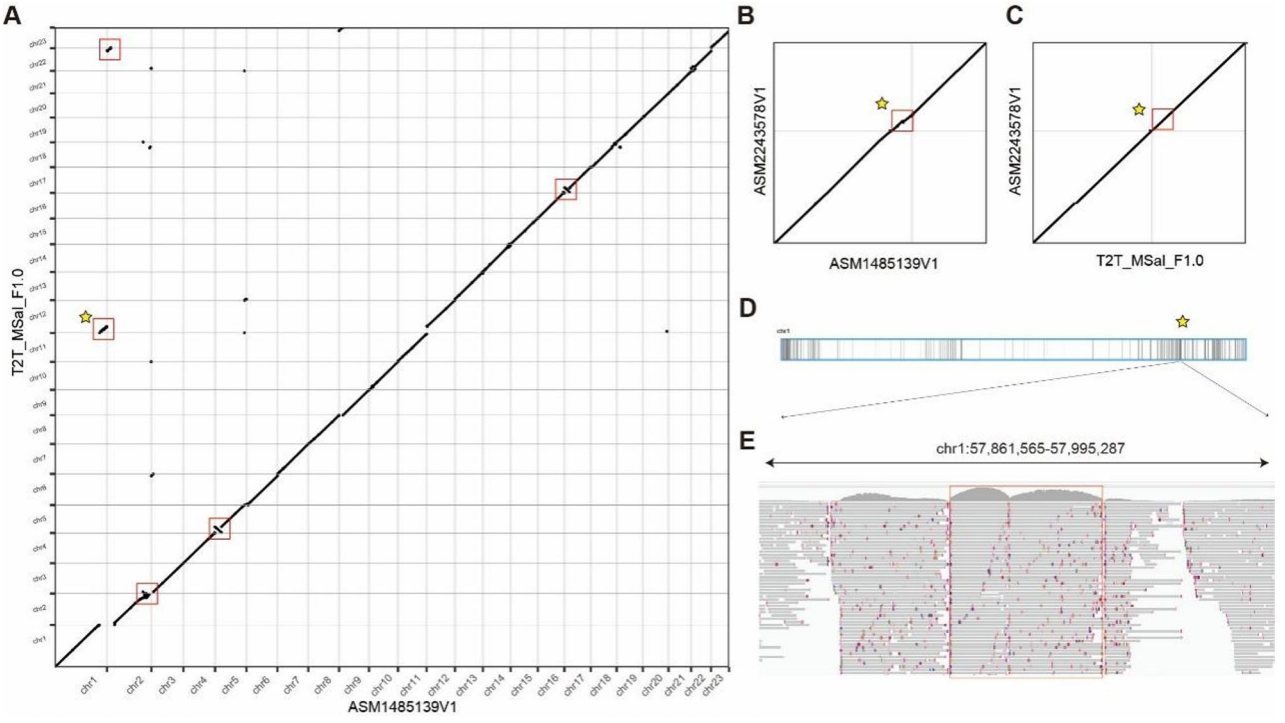

Fig. 3 Comparison and validation of genome assembly quality. **A**, Whole-genome synteny comparison between the current assembly (T2T_MSal_F1.0) and the previously published reference (ASM1485139V1). Red boxes indicate potential misassemblies, and the yellow star marks a selected region for further validation. **B**, Synteny comparison between ASM1485139V1 and another published assembly (ASM2243578V1). **C**, Synteny comparison between T2T_MSal_F1.0 and ASM2243578V1. **D**, Gap distribution in the current assembly. The selected region (marked by yellow star) shows a cluster of gaps. **E**, Read coverage of the selected region (chr1:57,861,565–57,995,287) shows abnormal depth, supporting the presence of assembly errors

This region at Chr10:35,942,121 – 36,902,615 in the T2T_MSal_F1.0 assembly, overlaps a previously reported sex-specific SNP-dense chromosomal interval. Its presence was further validated in TGS (Fig. S8). We designated this 90-kb region as the Y-specific SV-enriched region (SVER-Y). Based on gene structure prediction, seven open reading frames (ORFs) were identified, among which only G1 and G7 exhibited transcriptional support

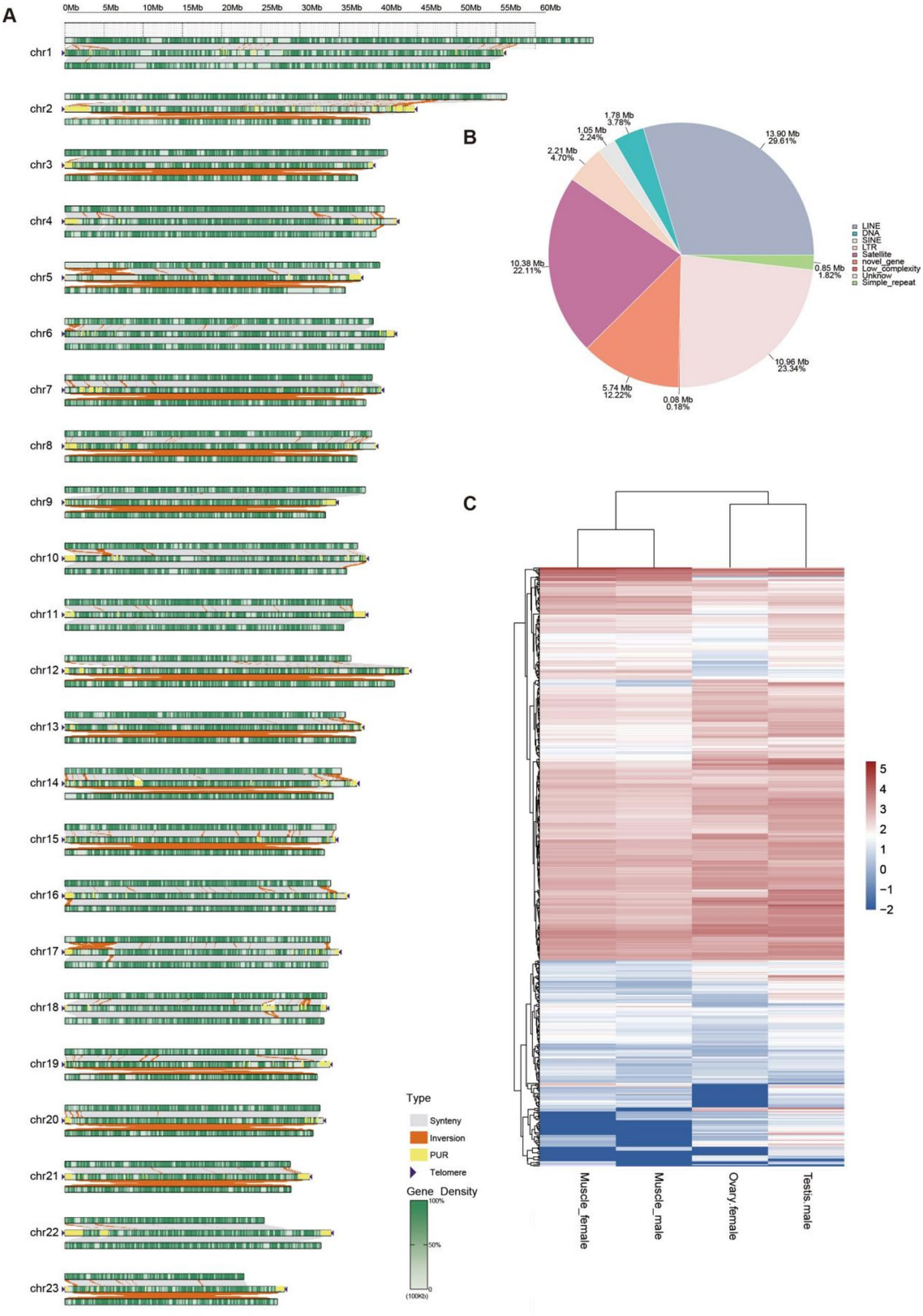

Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 Identification and functional analysis of positively selected and structurally variable regions. **A**, Whole-genome synteny comparison across three assemblies: ASM1485139v1 (top), T2T_female1.0 (middle), and ASM2243578v1 (bottom). Syntenic blocks, inversions, and positively selected regions (PURs) are shown along each chromosome, with gene density indicated in green. Telomeric regions are marked with arrowheads. **B**, Annotation of sequences located within PURs, classified by repeat and sequence types. **C**, Expression heatmap of genes located in PURs across different tissues (muscle and gonads of male and female). Expression values are normalized and clustered by gene

from RNA-seq reads (Fig. 5C). Domain analysis of the corresponding seven protein sequences using InterProScan revealed that all genes except G2 possessed distinct domain or gene family annotations, implying potential functional relevance (Table S9). All structural variations (SVs) except for those in the Sex-Linked Region were confirmed manually as false positives through IGV visualization, primarily due to genomic heterozygosity.

We also observed significant enrichment of transposable elements (TEs) in regions adjacent to this SVER-Y. Compared to the entire chromosome 10, the SVER-Y adjacent region exhibited the most pronounced increase in LTR proportion (74.92%), followed by LINEs (63.73%), SINEs (50.40%), and DNA transposons (40.33%). This enrichment pattern suggests transposable elements are involved in the structural evolution of SVER-Y (Fig. 5E).

A recent study has reported characteristics of this region, but our findings based on comparative genomic analyses revealed subtle discrepancies [79]. Specifically, we identified six insertions within the SVER-Y region and the discovery of a novel MSR region (SVER-6) (Table S8). To validate the accuracy of these regions, we analyzed resequencing data from female and male LMB downloaded from public databases, confirming their accuracy (Fig. S9). Additionally, collinearity analysis of genes uncovered extremely high variability in the arrangement of the gene cluster at the terminal end of chr10, which is speculated to be associated with the evolution of sex chromosomes (Fig. S10).

These findings suggest that, unlike the well-differentiated sex chromosomes of the Qingjiang medaka (*Oryzias latipes*), the Y chromosome of *Micropterus salmoides* remains in an early-stage differentiation state.

Discussion

With the continuous advancement of sequencing technologies, an increasing number of *de novo* assembled genomes have been generated. While aquatic animal genome sequencing has experienced rapid expansion, the accelerated iteration of sequencing platforms has simultaneously introduced challenges in genome utility evaluation. The completion of the human telomere-to-telomere genome has marked a milestone achievement in *de novo* genome assembly research. *Micropterus salmoides* is an economically important aquaculture species. Here, we present the gap-free T2T genome assemblies covering both sexes, establishing a comprehensive genomic foundation for subsequent genetic and molecular investigations.

Through comparative analysis of current reference genomes and our T2T assemblies, we have rectified assembly errors attributable to technological limitations and characterized previously unresolved regions (PURs). These sequences, predominantly encompassing telomeres and centromeres, represent persistent challenges in genome assembly. Furthermore, we annotated 610 novel genes localized within PURs. However, although more than half of PUR-associated genes showed tissue-specific differential expression, our analyses lacked biological replicates and were not enrichment in known functional pathways. These genes may represent uncharacterized functional and regulatory elements, and deserve further exploration through the design of transcriptomic experiments or targeted gene function studies. To advance this work, future efforts will focus on refining the genome through fluorescence in situ hybridization (FISH)-based validation of telomeres/centromeres and pangenome sequencing of representative breeding populations.

The sex determination mechanisms of teleosts remain a key research frontier. In the largemouth bass (*Micropterus salmoides*) — a sexually dimorphic species where females exhibit higher commercial value than males — previous resequencing-based F_{ST} analyses have localized Y-specific SNP-dense chromosomal interval. However, comprehensive investigations into genomic architecture divergence and the regulatory machinery of sex determination genes remain lacking. By integrating gap-free genomes of both sexes with published genomic datasets, we identified a 90-kb male-specific structural variation (SV)-enriched region (SVER-Y) on chromosome 10 (Chr10) and utilized multiple types of data evidence (including resequencing data) to precisely define the boundaries. This male-specific region was not fully resolved in previous studies, likely due to limitations in genome continuity and the inability to accurately assemble repeat-rich sequences. In contrast, our telomere-to-telomere assemblies enabled comprehensive reconstruction of this region, uncovering previously undetected structural features. Outside this region, other segments of Chr10 exhibit extremely high similarity between sexes.

TEs can promote sex chromosome differentiation by mediating chromosomal rearrangements and regulating sex-related genes [80, 81]. An explosion of transposable elements (TEs) and simple repeat accumulation has been

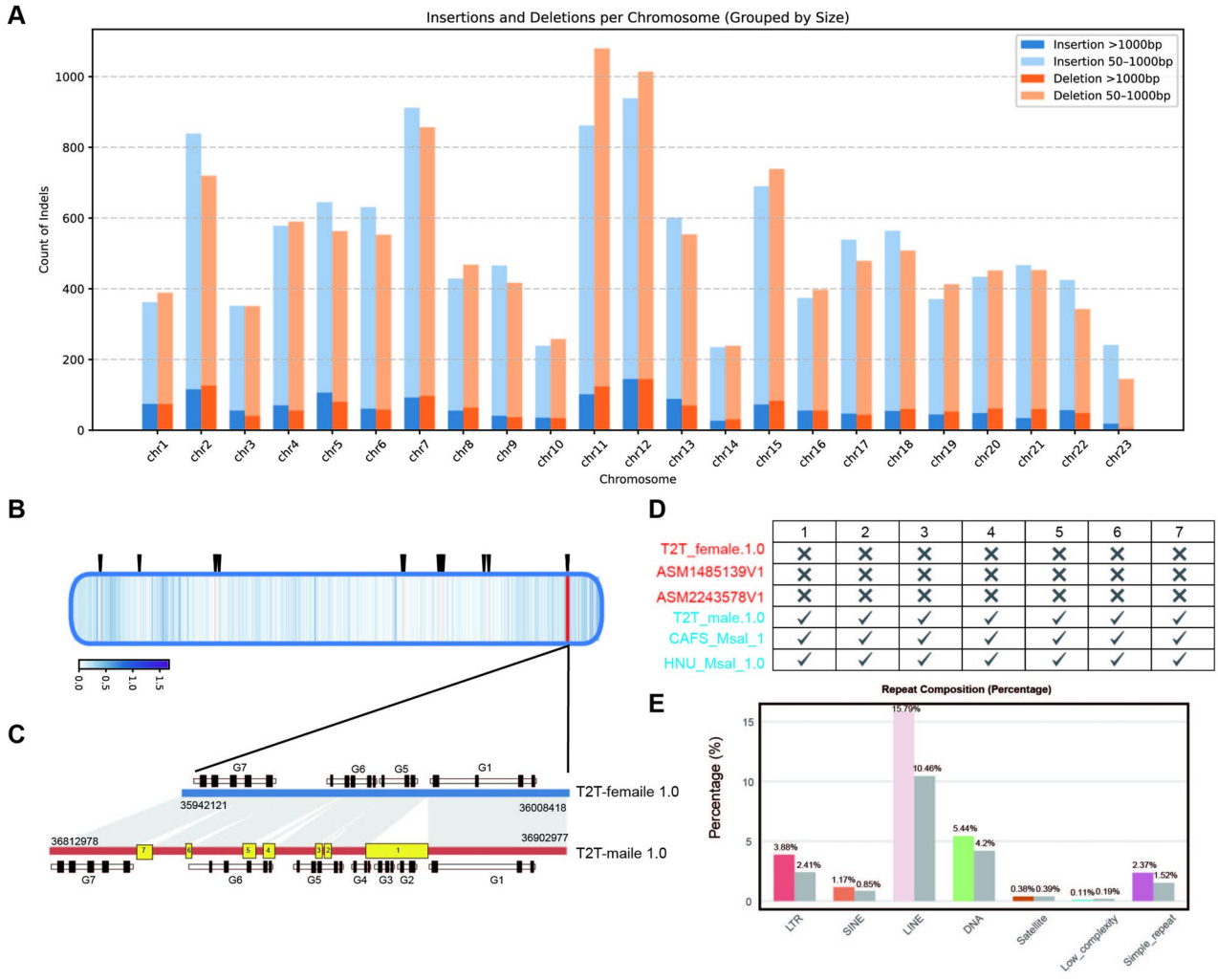

Fig. 5 Characterization of male-specific structural variations and repetitive elements. **A**, Structural variations (SVs) between female and male genomes across all chromosomes, grouped by size (insertions and deletions of 50–1,000 bp and >1,000 bp). **B**, Distribution of repeats along chromosome 10 in the male genome. The color gradient indicates repeat density, and the red bar marks a region enriched for male-specific sequences. **C**, Synteny comparison of the SV enrichment region between female and male assemblies. Yellow boxes indicate SVER loci. **D**, Summary of the presence or absence of the six SVER loci across different published genome assemblies. **E**, Composition and length of transposable elements (TEs) families within the SVER region in the male genome

observed around the young sex-determining loci in fish [82]. Analysis of repetitive sequences in the SVER-Y and its adjacent regions revealed significant enrichment of transposable elements (TEs). These results indicate that TEs are involved in the structural evolution of SVER-Y and that the Y chromosome of *Micropterus salmoides* remains in an early stage of differentiation. However, the molecular identity of sex-determining genes and their regulatory networks still require in-depth investigation. However, the molecular identity of sex-determining genes and their regulatory networks still require further in-depth investigation. This will be a key direction for future research, requiring experiments that combine RNA-seq with ATAC-seq and ChIP-seq, along with functional validation assays, to systematically clarify the regulatory role of TEs in sex chromosome differentiation in *Micropterus salmoides*.

Conclusions

This study presents the first gap-free, telomere-to-telomere genome assemblies for male and female *Micropterus salmoides*. These assemblies resolve thousands of gaps and correct major mis-assemblies across all 23 chromosomes. The complete references uncover 46.96 Mb of previously unresolved regions (PURs) enriched in telomeric and centromeric repeats, which contain 610 newly annotated gonad-biased genes—suggesting novel roles in growth and sexual development. Within this refined genomic framework, we also identify a 90 kb Y-linked hotspot on chromosome 10 (SVER-Y), characterized

by male-specific insertions and transposable elements, thereby elucidating the earliest steps of Y-chromosome differentiation. Collectively, these high-accuracy assemblies, the PUR gene catalogue, and the precisely delineated SVER-Y region not only provide a definitive, high-resolution reference sequence for future basic research but also establish a foundation for deciphering sex-determination pathways in largemouth bass and developing robust genome-enabled breeding tools for aquaculture.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-12468-y.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements
Not applicable.

Authors' contributions
Liu S., Zheng J., and Li F., conceived of and designed the research.Liu S., Cheng S.and Jiang W. collected the biological sample; Liu S.,ChiM., Zhu C. and Peng M.performed laboratory procedures; Liu S. and Li F. led the writing of the paper; All authors reviewed and approved the final manuscript.

Funding
This work was financially supported by Basic Public Welfare Research Program of Zhejiang Province (LGN22C190024), Major Science and Technology Programs for Agriculture (Aquaculture) New Variety Selection of Zhejiang Province (2021C02069-2) and Zhejiang Provincial Financial Special Project (2025CZZX02).

Data availability
The raw sequencing data and assembled genomes for both male and female Micropterus salmoides individuals are available in NCBI under BioProject accessions PRJNA1280199 (male) and PRJNA1280200 (female). Genome assemblies and annotations are also archived on Zenodo at https://doi.org/10.5281/zenodo.15708890for long-term access. The female WGS datasets include SRR26340594, SRR26340595, and SRR26340596, while the male datasets include SRR26340608, SRR26340609, and SRR26340610.

Declarations

Ethics approval and consent to participate
Animal care and use for this research was approved by the Institutional Animal Care and Use Committee (IACUC) of the Zhejiang Institute of Freshwater Fisheries (ZIFF20241006).Consent for publication.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 11 July 2025 / Accepted: 17 December 2025
Published online: 23 December 2025

References

1. Li H, Durbin R. Genome assembly in the telomere-to-telomere era. Nat Rev Genet. 2024;25(9):658–70.
2. Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A et al. The complete sequence of a human genome. Science. 2022; Available from: https://www.science.org/doi/10.1126/science.abj6987 . [cited 2025 Jun 21].
3. Rhie A, Nurk S, Cechova M, Hoyt SJ, Taylor DJ, Altemose N, et al. The complete sequence of a human Y chromosome. Nature. 2023;621(7978):344–54.
4. Completing human genomes. Nat Methods. 2022;19(6):629.
5. Zhou Q, Wang J, Li J, Chen Z, Wang N, Li M, et al. Decoding the fish genome opens a new era in important trait research and molecular breeding in China. Sci China Life Sci. 2024;67(10):2064–83.
6. Zhu Lyun, Nie L, Zhu G, Xiang L, xin, Shao J. zhong. Advances in research of fish immune-relevant genes: a comparative overview of innate and adaptive immunity in teleosts. Dev Comp Immunol. 2013;39(1–2):39–62.
7. Fu X, Chen Y, Wang L, Zhou Q, Li M, Song Y, et al. Identification and functional analysis of the perforin-1 like gene in disease resistance in half smooth tongue sole (*Cynoglossus semilaevis*). Dev Comp Immunol. 2021;122:104135.
8. Wang Y, Zhang H, Xian W, Iwasaki W. Chromosome genome assembly and annotation of the spiny red Gurnard (*Chelidonichthys spinosus*). Sci Data. 2023;10(1):443.
9. Xue L, Gao Y, Wu M, Tian T, Fan H, Huang Y, et al. Telomere-to-telomere assembly of a fish Y chromosome reveals the origin of a young sex chromosome pair. Genome Biol. 2021;22(1):203.
10. Sun Z, Li S, Liu Y, Li W, Liu K, Cao X, et al. Telomere-to-telomere gapless genome assembly of the Chinese sea bass (*Lateolabrax maculatus*). Sci Data. 2024;11(1):175.
11. Zhang K, Guo S, Yang S, Zhou W, Wu J, Zhang X, et al. A telomere-to-telomere genome assembly of the protandrous hermaphrodite Blackhead seabream, *Acanthopagrus schlegelii*. Sci Data. 2025;12(1):350.
12. Ao J, Mu Y, Xiang LX, Fan D, Feng M, Zhang S, et al. Genome sequencing of the perciform fish *Larimichthys Crocea* provides insights into molecular and genetic mechanisms of stress adaptation. PLoS Genet. 2015;11(4):e1005118.
13. Xu XW, Zheng W, Meng Z, Xu W, Liu Y, Chen S. Identification of stress-related genes by co-expression network analysis based on the improved turbot genome. Sci Data. 2022;9(1):374.
14. Xu XW, Zheng W, Yang Y, Hou J, Chen S. High-quality Japanese flounder genome aids in identifying stress-related genes using gene coexpression network. Sci Data. 2022;9(1):705.
15. Yang Y, Wang T, Chen J, Wu L, Wu X, Zhang W, et al. Whole-genome sequencing of brown-marbled grouper (*Epinephelus fuscoguttatus*) provides insights into adaptive evolution and growth differences. Mol Ecol Resour. 2022;22(2):711–23.
16. Zhou Q, Gao H, Zhang Y, Fan G, Xu H, Zhai J, et al. A chromosome-level genome assembly of the giant grouper (*Epinephelus lanceolatus*) provides insights into its innate immunity and rapid growth. Mol Ecol Resour. 2019;19(5):1322–32.
17. Shao C, Li Q, Chen S, Zhang P, Lian J, Hu Q, et al. Epigenetic modification and inheritance in sexual reversal of fish. Genome Res. 2014;24(4):604–15.
18. Piferrer F. Epigenetics of Sex Determination and Differentiation in Fish. In: Sex Control in Aquaculture. JOHN WILEY & SONS, LTD; 2014:65–83. Available from: https://www.sciencegate.app/document/10.1002/9781119127291.ch3 [cited 2025 Jun 21].
19. Lu J, Fang W, Huang J, Li S. The application of genome editing technology in fish. Mar Life Sci Technol. 2021;3(3):326–46.
20. Wang C, Xu J, Kocher TD, Li M, Wang D. CRISPR knockouts of *Pmela* and *Pmelb* engineered a golden tilapia by regulating relative pigment cell abundance. J Hered. 2022;113(4):398–413.
21. Zhao J, Zhou T, Bai H, Ke Q, Li B, Bai M, et al. Genome-wide association analysis reveals the genetic architecture of parasite (*Cryptocaryon irritans*) resistance in large yellow croaker (*Larimichthys crocea*). Mar Biotechnol N Y N. 2021;23(2):242–54.
22. Zhao J, Bai H, Ke Q, Li B, Zhou Z, Wang H, et al. Genomic selection for parasitic ciliate *Cryptocaryon irritans* resistance in large yellow croaker. Aquaculture. 2021;531:735786.
23. Meuwissen TH. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 2001;157(4):1819–29.
24. Zhou Z, Shao G, Shen Y, He F, Tu X, Ji J, et al. Extreme-phenotype genome-wide association analysis for growth traits in spotted sea bass (*Lateolabrax maculatus*) using whole-genome resequencing. Anim Open Access J MDPI. 2024;14(20):2995.
25. Prchal M, D'Ambrosio J, Lagarde H, Lallias D, Patrice P, François Y, et al. Genome-wide association study and genomic prediction of tolerance to acute hypoxia in rainbow trout. Aquaculture. 2023;565:739068.

26. Shan X, Xu T, Ma Z, Zhang X, Ruan Z, Chen J, et al. Genome-wide association improves genomic selection for ammonia tolerance in the orange-spotted grouper (*Epinephelus coioides*). Aquaculture. 2021;533:736214.
27. Sun C, Li J, Dong J, Niu Y, Hu J, Lian J, et al. Chromosome-level genome assembly for the largemouth bass *Micropterus salmoides* provides insights into adaptation to fresh and brackish water. Mol Ecol Resour. 2021;21(1):301–15.
28. He K, Zhao L, Yuan Z, Canario A, Liu Q, Chen S, et al. Chromosome-level genome assembly of largemouth bass (*Micropterus salmoides*) using PacBio and Hi-C technologies. Sci Data. 2022;9(1):482.
29. Li XY, Mei J, Ge CT, Liu XL, Gui JF. Sex determination mechanisms and sex control approaches in aquaculture animals. Sci China Life Sci. 2022;65(6):1091–122.
30. Kobayashi Y, Nagahama Y, Nakamura M. Diversity and plasticity of sex determination and differentiation in fishes. Sex Dev Genet Mol Biol Evol Endocrinol Embryol Pathol Sex Determ Differ. 2013;7(1–3):115–25.
31. Mei J, Gui JF. Genetic basis and biotechnological manipulation of sexual dimorphism and sex determination in fish. Sci China Life Sci. 2015;58(2):124–36.
32. Nanda I, Kondo M, Hornung U, Asakawa S, Winkler C, Shimizu A, et al. A duplicated copy of DMRT1 in the sex-determining region of the Y chromosome of the medaka, *Oryzias latipes*. Proc Natl Acad Sci U S A. 2002;99(18):11778–83.
33. Li M, Sun Y, Zhao J, Shi H, Zeng S, Ye K, et al. A tandem duplicate of Anti-Müllerian hormone with a missense SNP on the Y chromosome is essential for male sex determination in nile Tilapia, *Oreochromis niloticus*. PLoS Genet. 2015;11(11):e1005678.
34. Peichel CL, McCann SR, Ross JA, Naftaly AFS, Urton JR, Cech JN, et al. Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. Genome Biol. 2020;21(1):177.
35. Feron R, Zahm M, Cabau C, Klopp C, Roques C, Bouchez O, et al. Characterization of a Y-specific duplication/insertion of the anti-Mullerian hormone type II receptor gene based on a chromosome-scale genome assembly of yellow perch, *Perca flavescens*. Mol Ecol Resour. 2020;20(2):531–43.
36. Nakamoto M, Uchino T, Koshimizu E, Kuchiishi Y, Sekiguchi R, Wang L et al. C <>Moens editor 2021 A Y-linked anti-Müllerian hormone type-II receptor is the sex-determining gene in ayu, plecoglossus altivelis. PLOS Genet 17 8 e1009705.
37. Dan C, Lin Q, Gong G, Yang T, Xiong S, Xiong Y, et al. A novel PDZ domain-containing gene is essential for male sex differentiation and maintenance in yellow catfish (*Pelteobagrus fulvidraco*). Sci Bull. 2018;63(21):1420–30.
38. Bao L, Tian C, Liu S, Zhang Y, Elaswad A, Yuan Z, et al. The Y chromosome sequence of the channel catfish suggests novel sex determination mechanisms in teleost fish. BMC Biol. 2019;17(1):6.
39. Chen S, Zhang G, Shao C, Huang Q, Liu G, Zhang P, et al. Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. Nat Genet. 2014;46(3):253–60.
40. Liang H, Wang L, Sha H, Zou G. Development and validation of sex-specific markers in *Pelodiscus sinensis* using restriction site-associated DNA sequencing. Genes. 2019;10(4):302.
41. Casalini M, Agbali M, Reichard M, Konecná M, Bryjová A, Smith C. Male dominance, female mate choice, and intersexual conflict in the Rose Bitterling (*Rhodeus ocellatus*). Evol Int J Org Evol. 2009;63(2):366–76.
42. de Oliveira JJ, Silva PF, Lopes F, de Luchiari A. Individual differences guide mate choice in the fighting fish (*Betta splendens*). J Comp Psychol Wash DC 1983. 2021;135(3):338–48.
43. Yokogawa K. Morphological variations in the largemouth bass *Micropterus salmoides* with particular emphasison growth-related changes. Aquac Sci. 2014;62(4):361–74.
44. Wang Q, Lu S, Hua J, Li Y, Tao Y, Jiang B et al. Transcriptome profiling reveals that differences in regulatory patterns of gonad development between male and female largemouth bass (*Micropterus salmoides*) contribute to dimorphic weight gain. Research Square. 2024. Available from: https://www.researchsquare.com/article/rs-4293130/v1 [cited 2025 Jun 21].
45. Wen M, Zhang Y, Wang S, Hu F, Tang C, Li Q, et al. Sex locus and sex markers identification using whole genome pool-sequencing approach in the largemouth bass (*Micropterus salmoides L*). Aquaculture. 2022;559:738375.
46. Du J, Zhou J, Li S, Shao J, Jiang P, Dong C, et al. A PCR-based method for genetic sex identification and evidence of the XX/XY sex determination system in largemouth bass (*Micropterus salmoides* L). Aquaculture. 2021;545:737220.
47. He Q, Ye K, Han W, Yekefenhazi D, Sun S, Xu X, et al. Mapping sex-determination region and screening DNA markers for genetic sex identification in largemouth bass (*Micropterus salmoides*). Aquaculture. 2022;559:738450.
48. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinforma Oxf Engl. 2011;27(6):764–70.
49. Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, et al. GenomeScope: fast reference-free genome profiling from short reads. Bioinforma Oxf Engl. 2017;33(14):2202–4.
50. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved *de Novo* assembly using phased assembly graphs with hifiasm. Nat Methods. 2021;18(2):170–5.
51. Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 2016;3(1):95–8.
52. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of *de Novo* genome assemblies. Bioinforma Oxf Engl. 2015;31(20):3350–2.
53. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinforma Oxf Engl. 2018;34(18):3094–100.
54. Li H. New strategies to improve minimap2 alignment accuracy. Bioinforma Oxf Engl. 2021;37(23):4572–4.
55. Lin Y, Ye C, Li X, Chen Q, Wu Y, Zhang F, et al. QuarTeT: a telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. Hortic Res. 2023;10(8):uhad127.
56. Formenti G, Rhie A, Walenz BP, Thibaud-Nissen F, Shafin K, Koren S, et al. Merfin: improved variant filtering, assembly evaluation and Polishing via k-mer validation. Nat Methods. 2022;19(6):696–704.
57. Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol. 2020;21(1):245.
58. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol. 2021;38(10):4647–54.
59. Manni M, Berkeley MR, Seppey M, Zdobnov EM. BUSCO: assessing genomic data quality and beyond. Curr Protoc. 2021;1(12):e323.
60. Baril T, Galbraith J, Hayward A. Earl grey: A fully automated user-friendly transposable element annotation and analysis pipeline. Arkhipova I, editor. Mol Biol Evol. 2024;41(4):msae068.
61. Jain C, Rhie A, Hansen NF, Koren S, Phillippy AM. Long-read mapping to repetitive reference sequences using Winnowmap2. Nat Methods. 2022;19(6):705–10.
62. Dale RK, Pedersen BS, Quinlan AR. Pybedtools: a flexible python library for manipulating genomic datasets and annotations. Bioinforma Oxf Engl. 2011;27(24):3423–4.
63. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinforma Oxf Engl. 2010;26(6):841–2.
64. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinforma Oxf Engl. 2013;29(1):15–21.
65. Robinson MD, McCarthy DJ, Smyth GK. EdgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinforma Oxf Engl. 2010;26(1):139–40.
66. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. ClusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innov Camb Mass. 2021;2(3):100141.
67. Shao C, Li C, Wang N, Qin Y, Xu W, Liu Q, et al. Chromosome-level genome assembly of the spotted sea bass, *Lateolabrax maculatus*. GigaScience. 2018;7(11):giy114.
68. Vij S, Kuhl H, Kuznetsova IS, Komissarov A, Yurchenko AA, Van Heusden P, et al. Chromosomal-level assembly of the Asian Seabass genome using long sequence reads and multi-layered scaffolding. PLoS Genet. 2016;12(4):e1005954.
69. Tine M, Kuhl H, Gagnaire PA, Louro B, Desmarais E, Martins RST, et al. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. Nat Commun. 2014;5:5770.
70. Lawson ND. Decoding the zebrafish genome. Nat Genet. 2022;54(7):917–9.
71. Wang T, Antonacci-Fulton L, Howe K, Lawson HA, Lucas JK, Phillippy AM, et al. The human pangenome project: a global resource to map genomic diversity. Nature. 2022;604(7906):437–46.
72. Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, et al. Towards complete and error-free genome assemblies of all vertebrate species. Nature. 2021;592(7856):737–46.

73. Yu P, Chen H, Liu M, Zhong H, Wang X, Wu Y, et al. Current status and application of largemouth bass (*Micropterus salmoides*) germplasm resources. Reprod Breed. 2024;4(2):73–82.
74. Sinclair AH, Berta P, Palmer MS, Hawkins JR, Griffiths BL, Smith MJ, et al. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature. 1990;346(6281):240–4.
75. Smith CA, Roeszler KN, Ohnesorg T, Cummins DM, Farlie PG, Doran TJ, et al. The avian Z-linked gene DMRT1 is required for male sex determination in the chicken. Nature. 2009;461(7261):267–71.
76. Charlesworth D, Charlesworth B, Marais G. Steps in the evolution of heteromorphic sex chromosomes. Heredity. 2005;95(2):118–28.
77. Zhou Y, Zhan X, Jin J, Zhou L, Bergman J, Li X, et al. Eighty million years of rapid evolution of the primate Y chromosome. Nat Ecol Evol. 2023;7(7):1114–30.
78. Wang Y, Gasser RB, Charlesworth D, Zhou Q. Evolution of sexual systems, sex chromosomes and sex-linked gene transcription in flatworms and roundworms. Nat Commun. 2022;13(1):3239.
79. Zhang X, Dong J, Ruan Z, Gao F, Zhou W, You X et al. Identification and characterization of a male-specific region in largemouth bass (*Micropterus Salmoides*) by whole-genome sequencing, resequencing and genomics comparison. Front Mar Sci. 2025;12. Available from: https://www.frontiersin.org/journals/marine-science/articles/10.3389/fmars.2025.1586534/full . [cited 2025 Jun 21].
80. Chalopin D, Volff JN, Galiana D, Anderson JL, Schartl M. Transposable elements and early evolution of sex chromosomes in fish. Chromosome Res. 2015;23(3):545–60.
81. Stévant I, Gonen N, Poulat F. Transposable elements acquire time- and sex-specific transcriptional and epigenetic signatures along mouse fetal gonad development. Front Cell Dev Biol. 2024;11:1327410.
82. Chiang VS, DeRosa H, Park JH, Hunter RG. The role of transposable elements in sexual development. Front Behav Neurosci. 2022;16:923732.

Publisher's Note