BMC Biology

RESEARCH

Open Access

Chromatin-mediated mechanisms of hybrid fertility in F_1 male hybrids of *Culter alburnus* and *Megalobrama amblycephala*

Li Ren[1,2,3*†], Kai Yang[1†], Yiyan Zeng[1], Ruyi Zhang[1], Ling Liu[1], Huiya Guan[1], Jinhui Zhang[1], Xiaohuan Han[1] and Shaojun Liu[1,2,3*]

Abstract

Background Interspecific hybridization is a fundamental evolutionary process, yet it often results in male sterility due to genomic incompatibilities and disrupted epigenetic regulation. While fertility variation among hybrids is increasingly acknowledged, the precise mechanisms linking these molecular disruptions to reproductive outcomes in vertebrates remain underexplored.

Results To address this, we investigated testicular chromatin dynamics and gene expression in 19 mature male F_1 hybrids derived from a paternal bluntnose black bream (*Megalobrama amblycephala*, BSB) and a maternal topmouth culter (*Culter alburnus*, TC). These two species diverged approximately 12.74 million years ago, and their F_1 hybrid offspring display a wide range of fertility. Our integrative multi-omics analysis revealed several insights into the molecular basis of hybrid fertility. Genes inherited from the BSB parent exert a disproportionate influence on reproductive outcomes, indicating an asymmetric parental contribution. Furthermore, chromatin architecture analysis showed that BSB-derived enhancer-promoter networks are characterized by longer interaction distances and greater regulatory strength, suggesting distinct subgenome-specific topologies. Then, our analysis identified the BSB-derived gene *LOC125267388* as a critical regulator of fertility. Homologous to a retrotransposon esterase and containing conserved teleost motifs, this gene likely contributes to the epigenetic modulation of hybrid fertility.

Conclusions These findings provide insights into the chromatin-mediated mechanisms underpinning reproductive barriers in vertebrate hybrids, contributing to a deeper understanding of evolutionary divergence and hybrid incompatibility.

Keywords Hybrid fish, Male fertility, Chromatin accessibility, Multi-omics

Background

Hybridization is a key driver of evolution, contributing to phenotypic diversification, rapid adaptation, and speciation in both plants and animals [1, 2]. Recent studies have shown that hybridization reshapes genomes, epigenetic landscapes, and gene expression, leading to rapid phenotypic changes [3–5]. However, despite these evolutionary benefits, interspecific hybridization frequently results in reproductive barriers, a central theme in evolutionary biology.

†Li Ren and Kai Yang contributed equally to this work.

*Correspondence:
Li Ren
rl@hunnu.edu.cn
Shaojun Liu
lsj@hunnu.edu.cn
[1] Engineering Research Center of Polyploid Fish Reproduction and Breeding, College of Life Sciences, Ministry of Education, Hunan Normal University, Changsha, Hunan 410081, China
[2] Yuelushan Laboratory, Changsha, Hunan 410128, China
[3] Hunan Yuelu Mountain Science and Technology Co., Ltd. for Aquatic Breeding, Changsha, Hunan 410081, China

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Hybrid sterility, especially male sterility, is a common consequence of interspecific crosses, as predicted by Haldane's rule, which states that sterility is more pronounced in the heterogametic sex [6, 7]. This sterility often arises from genomic conflicts between divergent parental genomes, which disrupt key reproductive processes such as gonad development and gamete formation without affecting somatic cell proliferation [8, 9]. Recent research has identified specific regulatory genes linked to impaired gonadogenesis or gametogenesis in hybrids [10–13]. Yet, the complex molecular mechanisms that underlie differential fertility in viable hybrids remain largely underexplored in vertebrates. Our research addresses this knowledge gap using hybrid lineages from two ecologically distinct and naturally isolated fish species, *Megalobrama amblycephala* (BSB) and *Culter alburnus* (TC). These species show significant differences in diet and body morphology and are reproductively isolated in the wild due to ecological barriers [14, 15]. Through laboratory hybridization, we generated reciprocal hybrid lineages spanning multiple generations. The differential fertility in male F_1 individuals makes them an ideal model for exploring the genetic and molecular basis of hybrid sterility. Ultimately, this work could offer new insights into the conserved epigenetic drivers of speciation in teleosts [16, 17].

Chromatin remodeling, the dynamic restructuring of chromatin to control DNA accessibility, is a fundamental epigenetic mechanism for regulating gene expression [18]. In reproductive biology, it ensures precise gene regulation during critical stages of testicular development and spermatogenesis, including spermatogonial differentiation and meiosis [19]. In fish, these processes are essential for proper gonadal differentiation and germ cell maturation [20]. While disruptions in chromatin regulation are known to cause male infertility and testicular dysfunction, the specific role of chromatin regulation in mediating hybrid fertility, particularly in teleosts, is a significant knowledge gap. Exploring this link could elucidate the fundamental mechanisms behind hybrid sterility and reproductive isolation.

Given the high number of gametes produced by large-bodied fish, their hybrid populations serve as an excellent model for studying the intricate processes of male fertility. This study investigates the chromatin structural changes and gene expression in a population of male F_1 hybrids from crosses between two ecologically isolated species, *M. amblycephala* (male) and *C. alburnus* (female), which diverged approximately 12.74 million years ago [21, 22]. These hybrids display differentiated fertility phenotypes. By leveraging an integrated multi-omics approach to explore the link between differential fertility and chromatin remodeling, our study aims to provide novel insights into the epigenetic mechanisms of hybrid fertility and, more broadly, vertebrate reproductive barriers.

Results

Evaluation of male reproductive traits

The reproductive capacity of 19 healthy, sexually mature male F_1 hybrid individuals (derived from crossing a male *M. amblycephala* and a female *C. alburnus*) was evaluated. These fish were reared for 24 months and showed no significant differences in growth phenotypes. Semen volume was obtained by abdominal compression, and testicular weight was measured post-dissection. The sperm content index was calculated as the ratio of sperm-occupied area to the total testis area in hematoxylin–eosin (HE)-stained sections (Additional file 1: Fig. S1). A strong positive correlation ($R > 0.7$, $p < 0.001$) was observed among the three phenotypes: sperm content index, semen volume, and testicular weight (Fig. 1A).

Differential parental contributions to genes associated with male reproductive phenotypes

To elucidate the genetic mechanisms underlying the male reproductive phenotype differences among these individuals, we conducted testis transcriptome sequencing and subsequent analysis of 19 hybrid individuals. Alignment of the reads to a merged reference genome of the parental BSB and TC resulted in an average mapping rate exceeding 99% (Additional file 2: Table S1). After filtering out genes with an average read count below 10 across the 19 samples, we obtained a final set of 44,998 expressed genes (Additional file 1: Fig. S2). To delve into how parental genome components influence gene expression and function in the hybrids, we first identified and compared the orthologous genes between BSB and TC. This allowed us to obtain 13,068 paired gene copies originating from both parents in the hybrid progeny (Additional file 2: Table S2), which serve as alleles in these individuals.

Following the assessment of male hybrid fertility and the acquisition of testis gene expression data, we performed a WGCNA. The analysis clustered 33,570 genes (18,166 from TC, 15,404 from BSB) into distinct co-expression modules. Our focus was on six modules significantly correlated with the reproductive phenotypes ($p < 0.05$): four negatively correlated (indianred4, plum1, red, and blue) and two positively correlated (salmon, palevioletred2) (Additional file 1: Fig. S2). We identified a total of 593 genes significantly associated with reproductive phenotypes from these modules. These genes were defined by a significant association with the reproductive phenotypes ($p < 0.05$) and high module membership (> 0.5). A closer examination of their parental origin revealed that 367 genes were derived from BSB, while 226

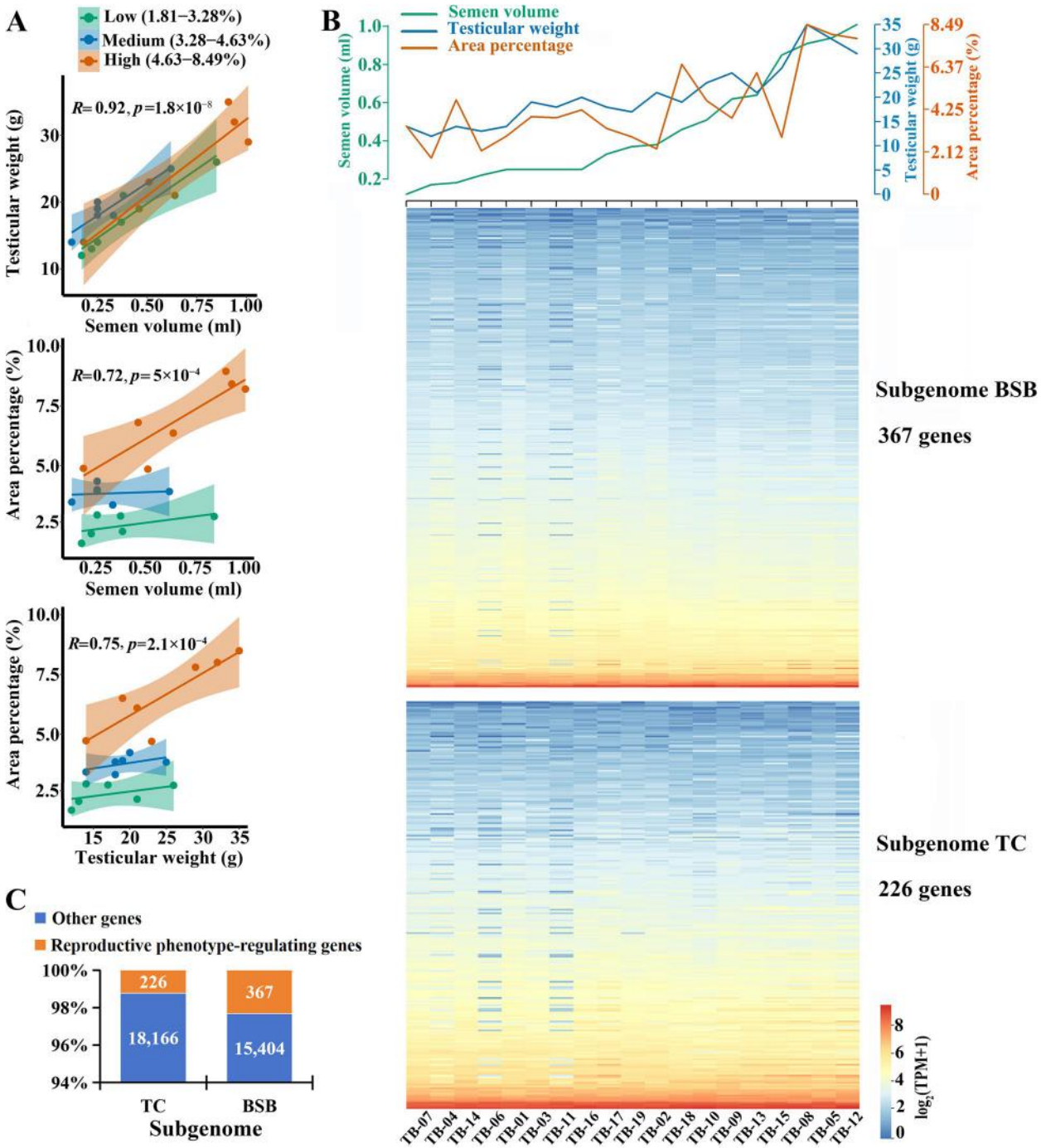


Fig. 1 Reproductive phenotypes and parental gene contributions in F_1 hybrid males. **A** Correlation of reproductive phenotypes in 19 sexually mature F_1 hybrid males, measured by semen volume, testicular weight, and sperm content index ($R > 0.7$, $p < 0.001$). **B** Weighted gene co-expression network analysis (WGCNA) results, with a line plot (top) showing variation in 3 reproductive phenotypes across 19 individuals and a heatmap (bottom) depicting expression levels of 593 fertility-associated genes. **C** Distribution of 33,570 genes from BSB (15,404) and TC (18,166) subgenomes, highlighting 593 fertility-associated genes (367 BSB, 226 TC) identified by WGCNA

were from TC. Further allelic analysis indicated that 179 of the BSB-derived genes were allelic, compared to 128 of the TC-derived genes (Fig. 1B). These results suggest that genes related to male hybrid fertility are disproportionately contributed by the paternal BSB. To statistically validate this observation, we conducted a chi-square test

on the ratio of parental gene contributions. The test confirmed a highly significant difference in the proportion of reproductive phenotype-regulating genes from BSB *vs.* TC ($\chi^2 = 20.31$, $df = 1$, $p = 6.58 \times 10^{-6}$) (Fig. 1C). This finding supports the observation that BSB provides a greater proportion of phenotype-associated genes.

Chromatin accessibility analysis shows high regulatory similarity

To investigate the regulatory mechanisms underlying the observed gene expression patterns, we performed ATAC-seq on the same 19 testis samples. We aligned the resulting 402.47 Gb of ATAC-seq data to the merged reference genome of BSB and TC (Additional file 2: Table S3). With an average mapping rate exceeding 97%, we identified chromatin accessibility at 29,892 potential promoters and 392,444 potential enhancers, based on genome annotation and the shared chromatin accessibility across all 19 samples. The principal component analysis (PCA) of the chromatin accessibility data showed no distinct clustering of individuals, indicating a high degree of regulatory similarity among the hybrids (Additional file 1: Fig. S3). Likewise, a clustering analysis based on sub-genomic distribution also showed no clear separation among individuals (Additional file 1: Fig. S4).

Subgenome-specific differences in enhancer–promoter network topology

To map the active regulatory network, we integrated Hi-C and ATAC-seq data from the hybrid individuals to identify enhancer–promoter interactions. To investigate the relationship between active enhancers and their target genes in hybrid individuals, we integrated 905.44 Gb of Hi-C data with ATAC-seq data from the same set of male individuals. Following stringent quality control, we obtained a total of 276,013,146 nonredundant valid interactions (Additional file 2: Table S4), which were evenly distributed across the BSB and TC subgenomes (Additional file 1: Fig. S5). By cross-referencing the Hi-C interaction data with the potential enhancer regions identified by ATAC-seq, we successfully pinpointed 33,493 high-confidence enhancer–promoter pairs ($q < 0.01$). Of these, the BSB subgenome comprised 17,096 pairs, while the TC subgenome had 16,397 pairs (Fig. 2A). These pairs

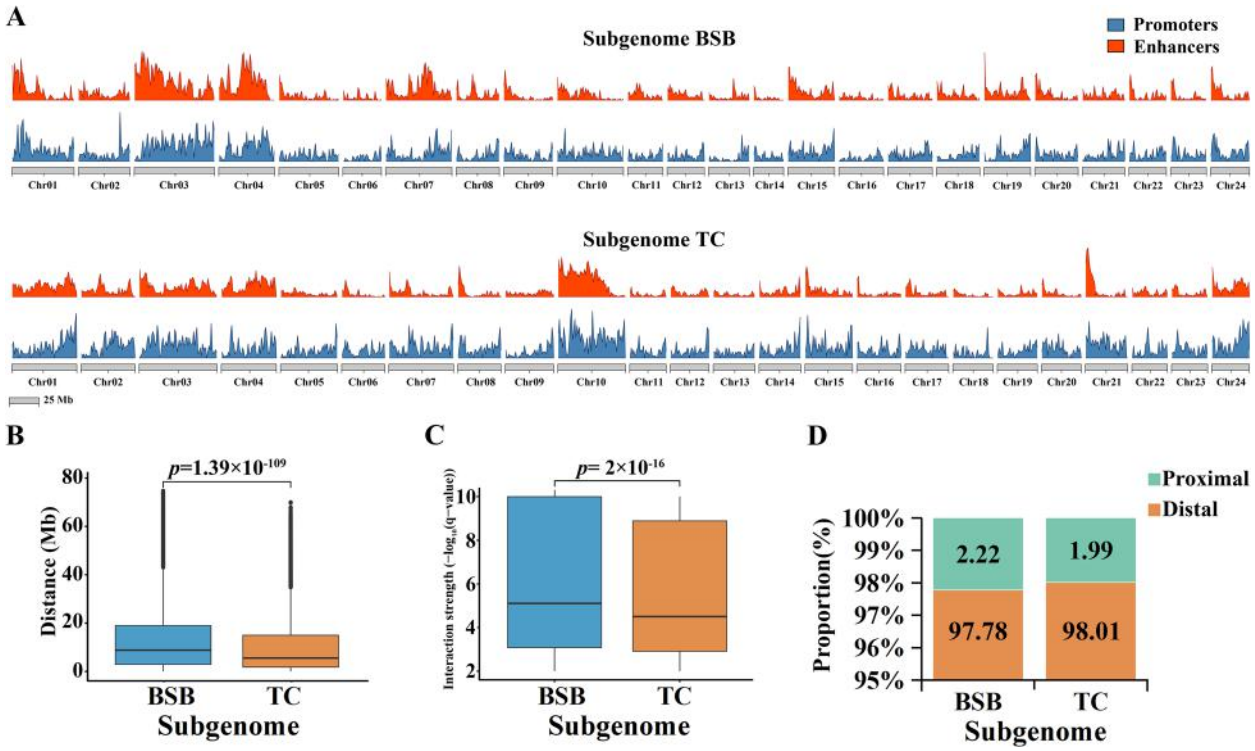


Fig. 2 Identification and characterization of enhancer-promoter interactions in F_1 hybrid testes. **A** Genome-wide distribution of 33,493 high-confidence enhancer–promoter (E–P) pairs (17,096 BSB, 16,397 TC) identified by integrating Hi-C and ATAC-seq data. **B** Comparison of E–P interaction distances between BSB and TC subgenomes (Wilcoxon test), showing that the BSB subgenome tends to form longer-range regulatory contacts. **C** Comparison of E–P interaction strengths between BSB and TC subgenomes (Wilcoxon test), indicating stronger enhancer–promoter connectivity in BSB. **D** Proportion of distal (> 10 kb) and proximal (≤ 10 kb) enhancers by subgenome. Distal enhancers predominate in both BSB and TC, suggesting that long-range chromatin interactions are a key regulatory feature in hybrid testes

collectively involved 21,411 enhancers and 10,747 promoters. Specifically, the BSB subgenome contained 10,749 enhancers and 5201 promoters, whereas the TC subgenome had 10,603 enhancers and 5496 promoters (Fig. 2A). This regulatory network likely represents the active enhancer–promoter relationships in the testis that drive the expression of genes associated with reproduction and testis development.

Analysis of the 33,493 enhancer–promoter pairs revealed that both the interaction distance (Wilcoxon test, $p = 1.39 \times 10^{-109}$, Fig. 2B) and interaction strength (Wilcoxon test, $p = 5.36 \times 10^{-28}$, Fig. 2C) were significantly higher in the BSB subgenome compared with the TC subgenome, indicating a clear difference in the topological structure of their regulatory networks. In the BSB subgenome, distal enhancers (>10 kb) accounted for 97.8% (16,716), while proximal enhancers (≤10 kb) made up 2.22% (380) (Fig. 2D). For the TC subgenome, distal enhancers constituted 98.0% (16,133), and proximal enhancers made up 1.99% (328) (Fig. 2D). These topological differences in the regulatory networks of the parental subgenomes are a potential reason for the fertility differences observed in the F_1 hybrids.

Subgenome-specific regulatory associations from multi-omics integration

We performed a correlation analysis between the chromatin accessibility of promoter regions and gene expression levels for 21,403 genes (11,160 from the TC subgenome and 10,262 from the BSB subgenome). This analysis revealed a significant correlation in 541 genes (220 from the TC subgenome and 321 from the BSB subgenome). Significant correlation was defined by a q-value < 0.05 and an absolute Spearman's correlation coefficient ≥ 0.3. A chi-square test revealed a highly significant difference in the proportion of genes with a significant promoter-expression correlation between the two subgenomes ($\chi^2 = 28.59$, $df = 1$, $p < 0.001$). Specifically, the number of significantly correlated genes in the BSB subgenome (321 genes, 3.13%) was significantly higher than in the TC subgenome (220 genes, 1.97%) (Fig. 3A). Simultaneously, we conducted a correlation analysis between the chromatin accessibility of 61,721 enhancers (32,822 from the TC subgenome and 28,899 from the BSB subgenome) and gene expression levels. A significant correlation was found in 895 of these enhancers (324 in the TC subgenome and 571 in the BSB subgenome) (Fig. 3B). The difference in the proportion of enhancers with significant correlations was also statistically significant ($p < 0.0001$), with a higher proportion in the BSB subgenome (1.98%) compared to the TC subgenome (0.99%).

Finally, we identified a set of core regulatory networks where enhancer accessibility, promoter accessibility, and gene expression were all significantly correlated. These networks involve 62 genes (20 from the TC subgenome and 42 from the BSB subgenome) and correspond to 175 enhancer regions (Fig. 3C). These core networks offer a valuable resource for future investigation into the epigenetic regulation of fertility.

Identification of a gene in the core regulatory network associated with male fertility

To investigate the association between the core regulatory network and male fertility phenotypes in hybrid fish, we performed an intersection analysis of genes within the network and 593 fertility-related genes identified through WGCNA. This analysis highlighted *LOC125267388* from *M. amblycephala* as a key candidate gene (Fig. 4A). Although its specific function is currently unknown, its expression is significantly associated with 21 enhancer elements, with 10 showing a strong correlation ($p < 0.05$) (Fig. 4B, C). Furthermore, the expression of *LOC125267388* also significantly correlates with three distinct fertility-related phenotypes (Fig. 4D), underscoring its potential role in reproductive regulation.

To determine if this unknown sequence was species-specific to *M. amblycephala*, we performed a BLAST analysis against the NCBI database. We identified 51 homologous sequences in other teleost species, with 6 amino acid sites universally conserved and 43 sites conserved in over 80% of the sequences (Fig. 4E, Additional file 1: Fig. S6, and Additional file 2: Table S5). Intriguingly, a specific segment of *LOC125267388* showed high similarity to the esterase domain of the zebrafish ZfL2-1 retrotransposon ORF1 protein (PDB: 4C1B_A, NCBI) (Fig. 4F). To further trace its evolutionary lineage, we constructed a phylogenetic tree, which provided additional insights into its evolutionary process (Additional file 1: Fig. S7).

To validate the importance of this gene in male fertility, we analyzed its expression patterns across six tissues (liver, muscle, brain, intestine, kidney, and testis) using previously published transcriptomic data from three hybrid progeny (NGDC: CRA017468). This analysis revealed that *LOC125267388* exhibits its highest expression in the testis (Fig. 4G, Additional file 2: Table S6), a finding that strongly supports its direct involvement in male fertility regulation.

Discussion

Hybridization is a potent evolutionary force, driving both diversity and adaptation [1, 2]. However, this process often leads to hybrid sterility, a key reproductive barrier that can be explained by Haldane's rule, which posits that

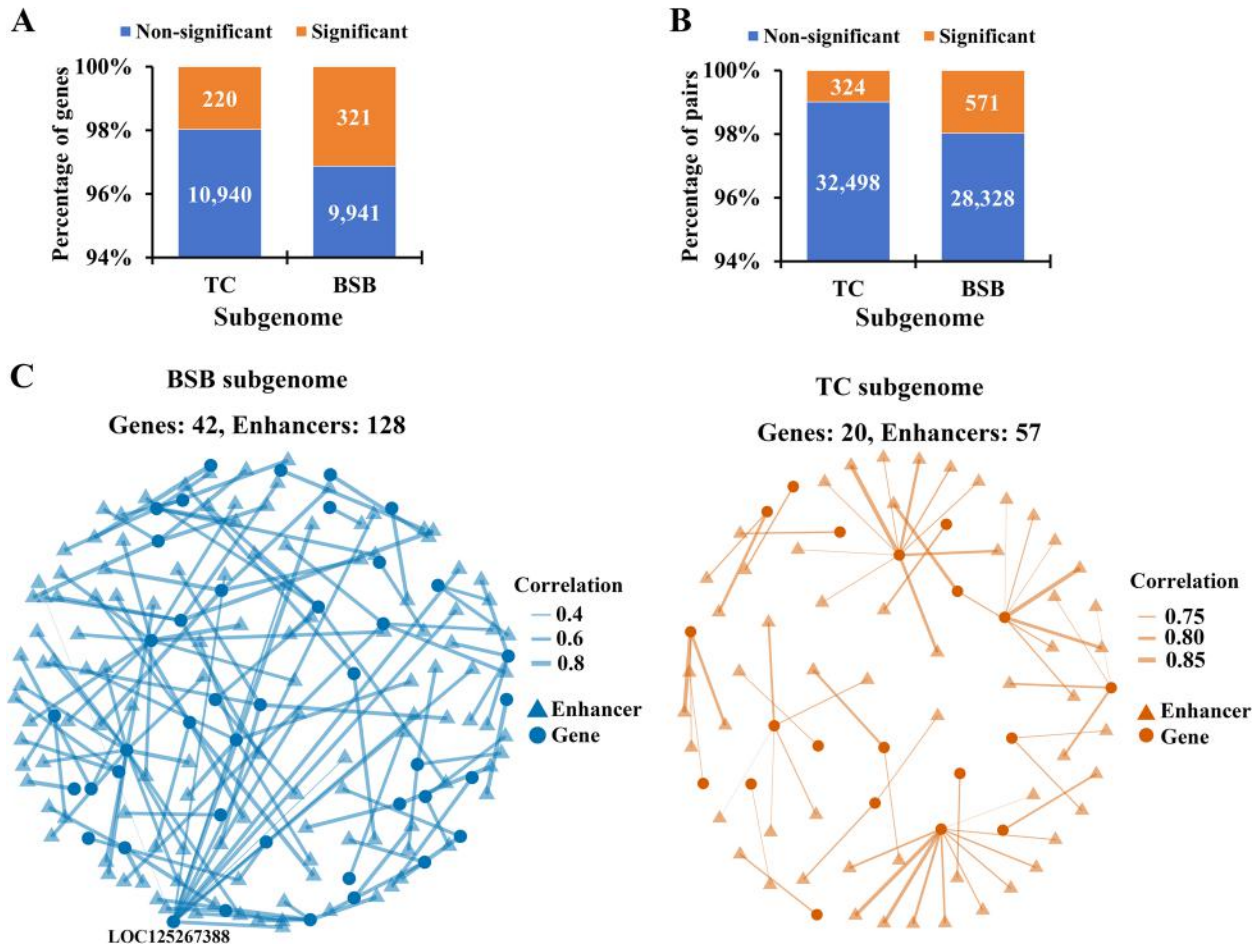


Fig. 3 Subgenome-specific regulatory contributions in F_1 hybrid testes. **A** proportion of genes showing significant correlations between promoter accessibility and gene expression in each subgenome, revealing a stronger promoter–expression association in BSB. **B** Proportion of enhancers showing significant correlations between chromatin accessibility and target gene expression, indicating higher enhancer regulatory activity in the BSB subgenome. **C** Integrated core regulatory networks linking enhancer accessibility, promoter accessibility, and gene expression, highlighting subgenome-specific regulatory modules associated with fertility variation

sterility is more pronounced in the heterogametic sex [6–9]. This sterility is often attributed to genomic conflicts between divergent parental genomes, which disrupt critical reproductive processes. While previous studies have identified specific regulatory genes linked to this disruption, the precise mechanisms linking these conflicts to differential fertility in viable hybrids remain underexplored. Our work addresses this gap by showing how chromatin-mediated mechanisms provide a new layer of explanation.

Specifically, we found that shared orthologous genes exhibit expression stability, suggesting a buffering mechanism against the genomic shock of hybridization [23, 24]. In contrast, species-specific genes showed significant variability, with the paternal BSB subgenome exhibiting a larger contribution to fertility-related genes. This asymmetry is rooted in the distinct regulatory topologies of the two parental subgenomes [24–26]. We observed that BSB's enhancer–promoter networks are characterized by longer interaction distances and greater regulatory strength, a chromatin-driven divergence that provides a robust mechanism to explain the observed fertility variation among hybrids [18–20, 27]. Our findings thus complement existing theories on genomic conflict by revealing the epigenetic basis of asymmetric parental contributions to hybrid fertility.

Using an integrated multi-omics approach, we identified *LOC125267388*—a BSB-derived gene with no known function—as a key gene for hybrid fertility. This gene's high activity in the testis and its tight control by 10 enhancers underscore its pivotal role in reproductive processes. Phylogenetic analysis further shows its sequence is highly conserved across 51 fish species, preserving key amino acid patterns that align with species divergence.

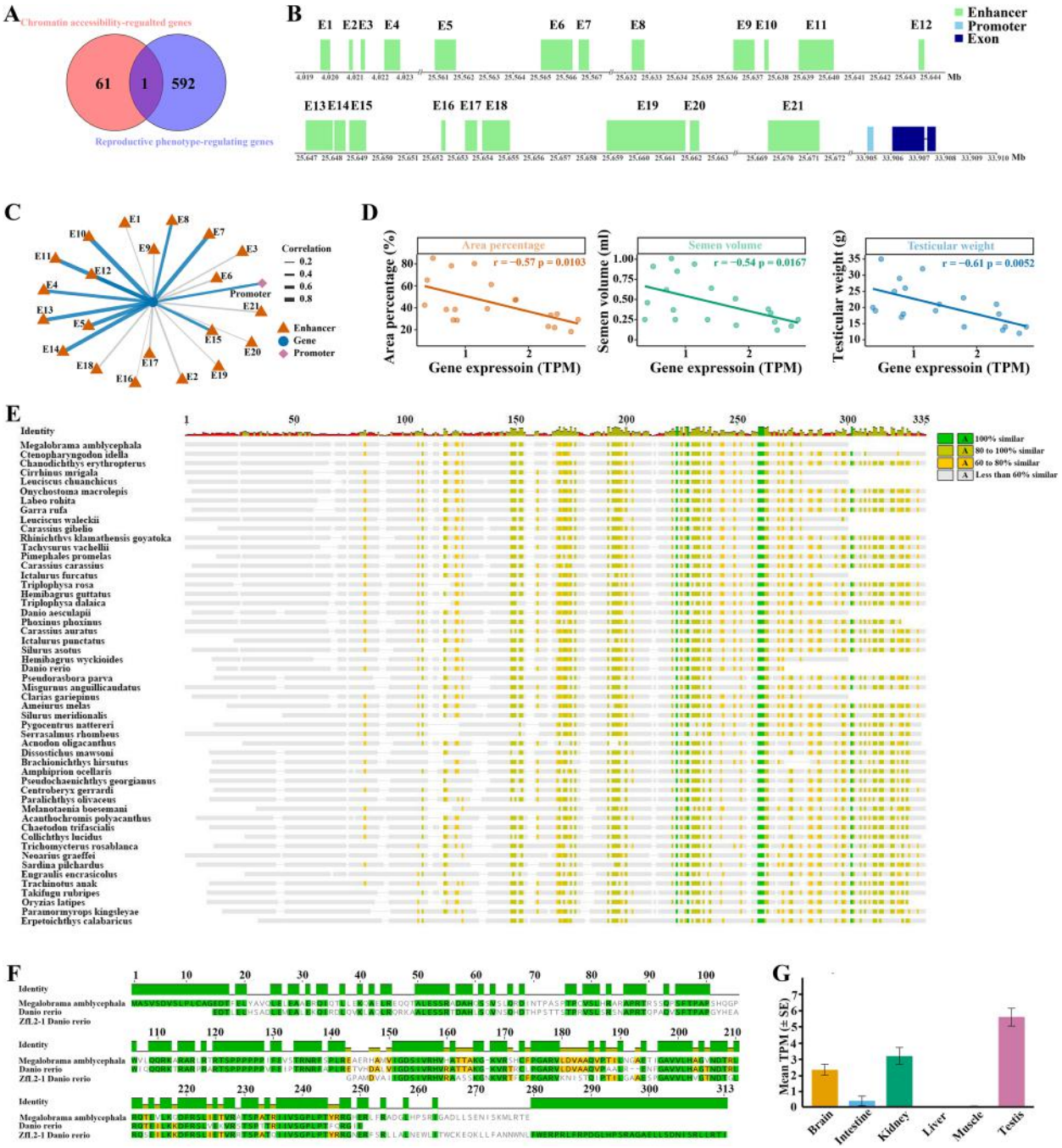


Fig. 4 Identification and characterization of a key fertility-related gene in $\mathbf{F}_1$ hybrid testes. **A** Overlap between 62 core regulatory network genes and 593 fertility-associated genes identified by WGCNA, pinpointing *LOC125267388* as a candidate fertility regulator. **B** Genomic distribution of 21 enhancers associated with *LOC125267388*. **C** Correlation of 10 significant enhancers with *LOC125267388* expression (Spearman, $p < 0.05$). **D** Expression of *LOC125267388* shows a significant Pearson correlation with all 3 reproductive phenotypes (semen volume, testis weight, and sperm content index) across the 19 hybrid individuals ($p < 0.05$). **E** Sequence alignment of *LOC125267388* protein with 51 teleost homologs, indicating **6** universally conserved and 43 highly conserved (> 80%) amino acid sites. **F** Alignment of a *LOC125267388* segment with its zebrafish homolog and the esterase domain of a retrotransposon ORF1 protein, suggesting possible esterase-like or retrotransposon-related function. **G** Expression profile of *LOC125267388* across six tissues, showing predominant expression in the testis, consistent with a role in spermatogenesis (*t*-test, $p < 0.01$)

The gene's partial similarity to the esterase domain of a zebrafish retrotransposon protein [28] leads us to hypothesize two potential, nonexclusive functions. First, it may contribute to genomic stability by regulating retrotransposon activity, a critical process during spermatogenesis that is often disrupted in hybrids [29, 30]. Second, its esterase-like function could involve lipid metabolism, aiding in the crucial membrane remodeling required for sperm formation [29, 31]. This proposed dual function offers a plausible explanation for the observed fertility differences and highlights *LOC125267388* as a valuable candidate for future mechanistic studies on hybrid reproductive barriers.

Our multi-omics analysis identified *LOC125267388*, a previously uncharacterized gene originating from BSB, as a key determinant of hybrid fertility. Its high expression in the testis and stringent regulation by 10 distinct enhancers strongly suggests a critical, core function in the reproductive pathway. This functional importance is further supported by phylogenetic data showing remarkable conservation of its sequence, including key amino acid patterns, across 51 fish species, correlating with species divergence. Based on the partial similarity of *LOC125267388* to the esterase-like domain of the zebrafish retrotransposon ORF1 protein [28], we propose two mechanistic hypotheses. First, the gene may maintain genomic stability by regulating transposable element (TE) activity, a process vital for spermatogenesis but often compromised in hybrids [29, 30]. Specifically, *LOC125267388* could modulate the stability and processing of TE transcripts, perhaps by associating with retrotransposon RNP complexes or factors in the piRNA pathway. The known consequence of disrupting piRNA-mediated silencing—the derepression of LINE/LTR elements leading to spermatogenic defects and infertility—provides a clear precedent for TE control affecting fertility [32]. Furthermore, this regulatory role could extend to influencing the epigenetic status of TE loci, thereby establishing the appropriate chromatin environment for TE repression and normal meiosis. Second, the predicted esterase-like function might participate in lipid metabolism, supporting crucial membrane remodeling events necessary for proper sperm formation and maturation [29, 31].

It is also important to acknowledge that the chromatin states and subgenome-specific topologies observed in the F_1 hybrids may partly reflect inheritance from the parental species. Because the original parental individuals cannot be resampled under identical environmental and developmental conditions, we cannot completely disentangle inherited chromatin features from those that emerged through hybridization. Nevertheless, our results emphasize that these asymmetric regulatory architectures acquire functional significance only within the hybrid genomic context, where divergent parental chromatin landscapes coexist and interact. This interaction itself constitutes the mechanistic basis for fertility variation. From an evolutionary perspective, parental divergence and hybridization are inseparable processes—the inherited differences between parental genomes set the stage for hybrid genomic interactions and potential incompatibilities [33].

These findings elucidate how epigenetic changes drive hybrid male fertility, linking genome conflicts, chromatin restructuring, and retrotransposon control. Our study is limited by its focus on one hybrid cross and by the unclear function of *LOC125267388*, which calls for further experiments such as CRISPR editing. Future work could explore how mechanisms like piwi-interacting RNA (piRNA) suppress retrotransposons in these hybrids [34, 35] to better understand reproductive barriers. In summary, our work provides a molecular framework for understanding how parental genome conflicts translate into phenotypic outcomes in hybrids, with implications for selective breeding in aquaculture.

Conclusions

This multi-omics study reveals the epigenetic basis of hybrid sterility:

asymmetric chromatin regulatory topologies between parental subgenomes determine hybrid male fertility differences. We conclude that the paternal BSB subgenome, utilizing stronger long-range enhancer networks, dictates the variation in the expression of fertility genes. This finding offers a new mechanism for Haldane's rule, linking genomic conflict to phenotypic outcomes via chromatin restructuring. Interesting, we identified the BSB-derived gene *LOC125267388* as a key factor, hypothesized to modulate spermatogenesis through retrotransposon control or lipid metabolism. Our work establishes a molecular framework for understanding hybrid reproductive barriers, providing vital targets and theoretical guidance for optimizing breeding strategies in aquaculture.

Methods

Sample collection and preparation

F_1 hybrid offspring (BSB × TC) were produced by crossing a sexually mature male *M. amblycephala* (BSB) and a female *C. alburnus* (TC). The hybrids were reared in experimental ponds at the aquaculture base in Wangcheng District, Changsha, China (28°22′25″ N, 112°47′26″ E). Throughout the rearing period, optimal environmental conditions were maintained, including stable water temperature, pH, and dissolved oxygen levels, with an ad libitum food supply.

Prior to tissue collection, the ploidy level of each fish was determined using flow cytometry (Cell Counter Analyzer, Partec, Germany). Blood samples (0.5–1 mL) were collected from the caudal vein with a syringe containing 100–150 units of sodium heparin, and the DNA content of erythrocytes was measured following Xiao et al. [36], using BSB and TC as controls.

At 24 months post-hatching, 19 healthy male fish showing no significant differences in growth phenotypes were randomly selected for sampling. In compliance with ethical standards and veterinary best practices, the fish were euthanized by immersion in an overdose of tricaine methanesulfonate (300 mg/L; Sigma-Aldrich, USA) for 30 min. Death was confirmed by the permanent cessation of opercular movement and loss of vestibulo-ocular reflex for at least 10 min prior to dissection.

Semen volume was assessed by gentle abdominal massage to stimulate milt release, which was collected in sterile containers. Subsequently, the fish were weighed and dissected. The left testis of each individual was harvested, weighed, and partitioned for multiple downstream analyses, including histological examination via hematoxylin–eosin (HE) staining, messenger RNA sequencing (mRNA-seq), and assay for transposase-accessible chromatin using sequencing (ATAC-seq).

Histological analysis with hematoxylin–eosin stain

To assess reproductive capacity, we evaluated residual sperm in the seminiferous tubules of the testes. Testis Sects. (10-mm thick) were dissected from each of the 19 hybrid individuals and fixed in Bouin's solution for 24 h, followed by a 4-h wash in distilled water. The tissues were then dehydrated through a graded ethanol series (70%, 80%, 90%, and 100%) and embedded in paraffin. The paraffin blocks were sectioned into 10-μm slices using a microtome, and the sections were stained with hematoxylin–eosin (HE), following the manufacturer's protocol (HE staining kit). Digital images of the stained sections were captured using a DX8 microscope (Olympus, Tokyo, Japan). Sperm content was quantified using ImageJ software by calculating the sperm content index, defined as the ratio of sperm-occupied area to the total testis area in each section.

RNA isolation and transcriptome library preparation

Total RNA was extracted from the testis tissues of the 19 individuals using the TRIzol method [37]. RNA concentration was measured with a NanoDrop, and genomic DNA contamination was removed using DNase I (Invitrogen). RNA quality was assessed using a 2100 Bioanalyzer (Agilent, USA). For transcriptome sequencing, mRNA was fragmented using a fragmentation buffer, reverse transcribed, and amplified into cDNA. Transcriptome data were generated for all 19 samples using DNBSEQ-T7 DNA nanoball technology, following the standard protocol [38]. The library preparation included the following steps: (1) Single-stranded circular library preparation using MGI Library Prep Kits. (2) combinatorial probe anchor sequencing, involving DNA anchor hybridization and attachment of a fluorescent probe to the DNA nanoball; (3) high-resolution imaging to capture the fluorescent signal; and (4) digital processing and sequencing to produce high-quality sequencing data. Low-quality bases and adapter sequences were trimmed using Fastp (v0.21.0) [39].

Gene expression analysis

Clean reads from 19 samples (1 per individual) and 3 additional individuals (each with 6 tissues: liver, muscle, brain, intestine, kidney, and testis) were aligned to the combined reference genomes of BSB (NCBI BioProject: PRJNA343584) [40, 41] and TC (National Genomics Data Center, accession number GWHDOEX00000000) [42, 43] using Bowtie2 (v 2.4.1) with default parameters [44]. Mapped reads were processed with SAMtools (v1.10) [45] to obtain uniquely mapped reads, which were then quantified using htseq-count (v0.12.4) [46]. Gene expression values for mRNA-seq were normalized and calculated as transcripts per million. Genes with fewer than five mapped reads in any sample were excluded from further analysis. Gene Ontology enrichment analysis was performed with a significance threshold of Benjamini–Hochberg false discovery rate (FDR) < 0.05.

Allo-allele detection

In F_1 hybrids from BSB and TC crosses, alleles were derived from parental orthologous genes. Allelic sequences for B (BSB-derived) and T (TC-derived) subgenomes were identified using reciprocal all-against-all BLASTX (v2.8.1) with an *e*-value threshold of $1e^{-6}$ and a minimum of 70% protein sequence alignment. Transcripts < 300 bp were excluded. Chromosome collinearity was analyzed with MCScanX, identifying 13,068 homologous chromosome pairs based on orthologous relationships.

Weighted gene correlation network analysis

Using WGCNA (v1.67), we analyzed gene expression data from 19 samples to identify key genes in testicular development and spermatogenesis. A Pearson correlation matrix was transformed into an adjacency matrix using a soft-threshold power, followed by the topological overlap matrix calculation to cluster co-expressed genes into modules via hierarchical clustering. Module eigengenes were correlated with semen volume, testicular weight, and the sperm content index using the Pearson

correlation. Key genes were selected from modules significantly correlated with the three reproductive phenotypes ($p < 0.05$), based on their strong association with these phenotypes ($p < 0.05$) and a high module membership (> 0.5).

ATAC-seq data processing and peak identification

ATAC-seq data were generated from 19 hybrid fish testis samples using the BGI DNBSEQ-T7 platform, following standard protocols [38]. Raw FASTQ files were quality controlled with fastp (v0.23.2) to remove adapters and low-quality reads. High-quality reads were aligned to a combined reference of the paternal genomes (BSB and TC) using Bowtie2 (v2.3.5) [44] with parameters –very-sensitive and –end-to-end. Aligned BAM files were sorted and deduplicated using samtools (v1.9), retaining only unique reads. Peaks were called with MACS2 (v2.2.7.1) [47] using parameters –nomodel –shift −100 –extsize 200, producing narrowPeak files. To ensure peak reliability, narrowPeak files were filtered for each sample at a q-value < 0.05 to exclude low-quality peaks. Peak signals were converted to BigWig format using bamCoverage (deepTools v3.5.1) and normalized to reads per million to account for sequencing depth differences across samples.

NarrowPeak files from the 19 samples were merged using bedtools (v2.29.2) with a q-value < 0.05 filter, generating a unified peak set. Based on gene annotations, transcription start sites (TSS ± 1 kb) were defined, and peaks in non-promoter regions were designated as candidate enhancers, creating separate promoter and enhancer peak sets. Signal intensities in promoter regions were extracted from BigWig files using bigWigAverageOverBed (UCSC tools), yielding per-sample mean signal data. Additionally, signal intensities from candidate enhancer regions were extracted for subsequent analyses.

Hi-C data processing and interaction identification

High-throughput chromosome conformation capture (Hi-C) data were derived from two mature testicular samples from the same hybrid population, sequenced as part of a previous study [24]. The raw sequencing data were quality-controlled using fastp (v0.23.2) with default parameters to remove adapters and low-quality reads (Phred score < 20, length < 50 bp). Processed reads were aligned to a combined reference of the paternal genomes (BSB and TC) using Bowtie2 (v2.3.5) with the –very-sensitive-local option to ensure accurate mapping. HiC-Pro (v3.1.0) was employed to filter invalid pairs, deduplicate reads, and generate contact matrices at 10-kb resolution. Significant chromatin interactions were identified using FitHiC (v2.0.7) with a q-value threshold of < 0.01 to control for false discovery rate (FDR).

To specifically identify enhancer-promoter (E-P) interactions, we filtered the significant chromatin interactions from FitHiC. This was done by intersecting them with predefined enhancer and promoter regions. Enhancer candidates were obtained from ATAC-seq peak calling, as described in the ATAC-seq section of the manuscript. Promoter regions were defined as the genomic regions within ± 1 kb of a TSS, based on the GFF annotation. The intersection was performed using the intersect command in pybedtools (v0.9.0), which retained only the overlapping regions. The resulting interactions were further filtered and manipulated using pandas (v2.0.3) to ensure consistent coordinate systems and to specifically select interactions with a q-value < 0.01. Finally, we extracted ATAC-seq signals for both the enhancers and promoters involved in the identified E-P interactions. This was accomplished using the bigWigAverageOverBed tool from the UCSC tools suite, which generated a per-sample signal file based on the identified E–P interactions.

Statistical analysis and network construction

All statistical analyses were conducted in R (v4.2.0) using the tidyverse and broom packages for efficient data manipulation, analysis, and output tidying. We integrated three datasets—gene expression, promoter chromatin accessibility, and enhancer chromatin accessibility—by mapping sample and gene identifiers to ensure consistency. To investigate regulatory relationships, we performed Spearman's rank (ρ) and Pearson's (r) correlation analyses between gene expression and three key traits (semen volume, testicular weight, and area percentage), as well as between regulatory elements and their associated genes. For each correlation, we calculated a coefficient and a p-value. p-values were then adjusted using the Benjamini–Hochberg procedure to control the false discovery rate (FDR), yielding q-values. We defined a significant association by a threshold of q-value < 0.05 and an absolute correlation coefficient ≥ 0.3. Genes with both a significant promoter-expression correlation and a significant enhancer-expression correlation were identified as the core regulatory network.

Abbreviations

ATAC-seq	Assay for transposase-accessible chromatin using sequencing
BSB	*Megalobrama amblycephala*
E–P	Enhancer–promoter
FDR	False discovery rate
HE	Hematoxylin–eosin
Hi-C	High-throughput chromosome conformation capture
mRNA-seq	Messenger RNA sequencing
NCBI	National Center for Biotechnology Information
NGDC	National Genomics Data Center
PCA	Principal component analysis
piRNA	PIWI-interacting RNA
TC	*Culter alburnus*
TE	Transposable element
TSS	Transcription start site
WGCNA	Weighted gene co-expression network analysis

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12915-026-02523-9.

Additional File 1: Figures S1-S7. Fig. S1. Comprehensive analysis of testicular phenotypes in 19 TB individuals. Fig. S2. PCA plot of gene expression in testis transcriptomes. Fig. S3. PCA of chromatin accessibility. Fig. S4. Subgenome distribution of chromatin accessibility in testis samples. Fig. S5. Distribution of valid non-redundant HI-C interactions. Fig. S6. Multiple sequence alignment of amino acid sequences. Fig. S7. Phylogenetic analysis of *LOC125267388* across teleost species.

Additional File 2: Tables S1-S6. Table S1. Summary of transcriptome alignment to the genome. Table S2. Orthologous genes of *Megalobrama amblycephala* and *Culter alburnus*. Table S3. Overview of ATAC-seq data quality and alignment statistics for 19 testis samples. Table S4. Summary of Hi-C data. Table S5. Homologous sequences of uncharacterized gene *LOC125267388* from NCBI database. Table S6. Alignment information of transcriptome data for six tissues.

Acknowledgements
Not applicable.

Authors' contributions
L.R. conceived the study, acquired funding, handled project administration and supervision, developed the methodology, performed formal analysis and investigation, conducted validation, and wrote the original draft. S.L. contributed to conceptualization, funding acquisition, project administration, and supervision. Y.Z., K.Y., R.Z., and L.L. performed validation and developed the methodology. H.G. contributed to the methodology. J.Z. and X.H. performed validation. All authors read and approved the final manuscript.

Funding
This research was supported by the National Natural Science Foundation of China (32341057 and 32293252), National Key Research and Development Plan Program (2023YFD2401602), Special Funds for Construction of Innovative Provinces in Hunan Province (2021NK1010), Earmarked Fund for China Agriculture Research System (CARS-45), 111 Project (D20007), and the world-class cultivation discipline (biology) of Hunan Province.

Data availability
The raw mRNA-seq data have been deposited in the National Center for Biotechnology Information (NCBI) under SRR29824779–SRR29824869 and in the National Genomics Data Center (NGDC) under CRA017468 and CRA017690. The raw ATAC-seq data have been deposited in the National Genomics Data Center (NGDC) under CRA029187 (https://ngdc.cncb.ac.cn/gsa/search?searchTerm=CRA029187).

Declarations

Ethics approval and consent to participate
All experimental procedures involving animals in this study were conducted in strict accordance with the guidelines for the care and use of laboratory animals. The study protocol was formally reviewed and approved by the Academic Committee of Hunan Normal University (Approval No. 2025003), Changsha, China.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 10 October 2025 Accepted: 16 January 2026
Published online: 23 January 2026

References

1. Tigano A, Friesen VL. Genomics of local adaptation with gene flow. Mol Ecol. 2016;25(10):2144–64.
2. Keller I, Wagner C, Greuter L, Mwaiko S, Selz O, Sivasundar A, et al. Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. Mol Ecol. 2013;22(11):2848–63.
3. Shen H, He H, Li J, Chen W, Wang X, Guo L, et al. Genome-wide analysis of DNA methylation and gene expression changes in two Arabidopsis ecotypes and their reciprocal hybrids. Plant Cell. 2012;24(3):875–92.
4. Runemark A, Moore EC, Larson EL. Hybridization and gene expression: beyond differentially expressed genes [J]. Mol Ecol. 2025;34(15):e17303.
5. Madlung A, Tyagi AP, Watson B, Jiang H, Kagochi T, Doerge RW, et al. Genomic changes in synthetic Arabidopsis polyploids. Plant J. 2005;41(2):221–30.
6. McDermott SR, Noor MA. The role of meiotic drive in hybrid male sterility. Philos Trans R Soc Lond B Biol Sci. 2010;365(1544):1265–72.
7. Hurst LD, Pomiankowski A. Causes of sex ratio bias may account for unisexual sterility in hybrids: a new explanation of Haldane's rule and related phenomena. Genetics. 1991;128(4):841–58.
8. Maheshwari S, Barbash DA. The genetics of hybrid incompatibilities. Annu Rev Genet. 2011;45(1):331–55.
9. Yoshikawa H, Xu D, Ino Y, Yoshino T, Hayashida T, Wang J, et al. Hybrid sterility in fish caused by mitotic arrest of primordial germ cells. Genetics. 2018;209(2):507–21.
10. Niayale R, Cui Y, Adzitey F. Male hybrid sterility in the cattle-yak and other bovines: a review. Biol Reprod. 2021;104(3):495–507.
11. Good JM, Dean MD, Nachman MW. A complex genetic basis to X-linked hybrid male sterility between two species of house mice. Genetics. 2008;179(4):2213–28.
12. Phadnis N, Orr HA. A single gene causes both male sterility and segregation distortion in Drosophila hybrids. Science. 2009;323(5912):376–9.
13. Ponjarat J, Singchat W, Monkheang P, Suntronpong A, Tawichasri P, Sillapaprayoon S, et al. Evidence of dramatic sterility in F1 male hybrid catfish [male *Clarias gariepinus* (Burchell, 1822)× female *C. macrocephalus* (Günther, 1864)] resulting from the failure of homologous chromosome pairing in meiosis I. Aquaculture. 2019;505:84–91.
14. Chen J, Liu H, Gooneratne R, Wang Y, Wang W. Population genomics of Megalobrama provides insights into evolutionary history and dietary adaptation. Front Genet. 2020;11:710.
15. Sun N, Zhu DM, Li Q, Wang GY, Chen J, Zheng F, et al. Genetic diversity analysis of topmouth culter (*Culter alburnus*) based on microsatellites and D-loop sequences. Environ Biol Fishes. 2021;104(1):111–20.
16. Vernaz G, Hudson AG, Santos ME, Fischer B, Carruthers M, Shechonge AH, et al. Epigenetic divergence during early stages of speciation in an African crater lake cichlid fish. Nat Ecol Evol. 2022;6(12):1940–51.
17. Piferrer F, Miska EA, Anastasiadi D. Epigenetics in fish evolution. In: Guerrero-Bosagna C, editor. On Epigenetics and Evolution. United States: Academic Press; 2024. p. 283–306.
18. Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. Nat Rev Genet. 2019;20(4):207–20.
19. Murat F, Mbengue N, Winge SB, Trefzer T, Leushkin E, Sepp M, et al. The molecular evolution of spermatogenesis across mammals. Nature. 2023;613(7943):308–16.
20. Hou M, Wang Q, Zhao R, Cao Y, Zhang J, Sun X, et al. Analysis of chromatin accessibility and DNA methylation to reveal the functions of epigenetic modifications in *Cyprinus carpio* gonads. Int J Mol Sci. 2024;25(1):321.
21. Ren L, Li W, Qin Q, Dai H, Han F, Xiao J, et al. The subgenomes show asymmetric expression of alleles in hybrid lineages of *Megalobrama amblycephala* x *Culter alburnus*. Genome Res. 2019;29(11):1805–15.
22. Xiao J, Hu F, Luo K, Li W, Liu S. Unique nucleolar dominance patterns in distant hybrid lineage derived from *Megalobrama amblycephala* × *Culter alburnus*. BMC Genet. 2016;17(1):131.
23. Luo M, Tai Y, Li M, Zeng Y, Wu C, Liu L, et al. Transcriptome analysis reveals genetic regulation of growth diversity in F2 intergeneric hybrids of *Megalobrama amblycephala* and *Culter alburnus*. Aquaculture. 2025;595:741630.
24. Ren L, Zeng Y, Liu Q, Tu X, Chen F, Wu H, et al. Genomic and chromosomal architectures underlying fertility maintenance in the testes of

intergeneric homoploid hybrids. Sci China Life Sci. 2025. https://doi.org/10.1007/s11427-024-2868-y.
25. Zhang Y, Xiang Y, Yin Q, Du Z, Peng X, Wang Q, et al. Dynamic epigenomic landscapes during early lineage specification in mouse embryos. Nat Genet. 2018;50(1):96–105.
26. Mack KL, Nachman MW. Gene regulation and speciation. Trends Genet. 2017;33(1):68–80.
27. Schneider AM, Schmidt S, Jonas S, Vollmer B, Khazina E, Weichenrieder O. Structure and properties of the esterase from non-LTR retrotransposons suggest a role for lipids in retrotransposition. Nucleic Acids Res. 2013;41(22):10563–72.
28. Chuma S, Nakano T. piRNA and spermatogenesis in mice. Philos Trans R Soc Lond B Biol Sci. 2013;368(1609):20110338.
29. Tam PLF, Leung D. The molecular impacts of retrotransposons in development and diseases. Int J Mol Sci. 2023;24(22):16418.
30. Bertucci A, Pierron F, Thébault J, Klopp C, Bellec J, Gonzalez P, et al. Transcriptomic responses of the endangered freshwater mussel *Margaritifera margaritifera* to trace metal contamination in the Dronne River, France. Environ Sci Pollut Res. 2017;24(35):27145–59.
31. Yang F, Wang PJ. Multiple LINEs of retrotransposon silencing mechanisms in the mammalian germline. Semin Cell Dev Biol. 2016;59:118–25.
32. Runemark A, Vallejo-Marin M, Meier JI. Eukaryote hybrid genomes. PLoS Genet. 2019;15(11):e1008404.
33. Romero-Soriano V, Modolo L, Lopez-Maestre H, Mugat B, Pessia E, Chambeyron S, et al. Transposable element misregulation is linked to the divergence between parental piRNA pathways in Drosophila hybrids. Genome Biol Evol. 2017;9(6):1450–70.
34. Watanabe T, Cheng EC, Zhong M, Lin H. Retrotransposons and pseudogenes regulate mRNAs and lncRNAs via the piRNA pathway in the germline. Genome Res. 2015;25(3):368–80.
35. Xiao J, Kang X, Xie L, Qin Q, He Z, Hu F, et al. The fertility of the hybrid lineage derived from female *Megalobrama amblycephala* x male *Culter alburnus*. Anim Reprod Sci. 2014;151(1–2):61–70.
36. Rio DC, Ares M Jr, Hannon GJ, Nilsen TW. Purification of RNA using TRIzol (TRI reagent). Cold Spring Harb Protoc. 2010;2010(6):pdb.prot5439.
37. Patterson J, Carpenter EJ, Zhu Z, An D, Liang X, Geng C, et al. Impact of sequencing depth and technology on de novo RNA-seq assembly. BMC Genomics. 2019;20(1):604.
38. Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884–90.
39. Liu H, Chen C, Lv M, Liu N, Hu Y, Zhang H, et al. A chromosome-level assembly of blunt snout bream (*Megalobrama amblycephala*) genome reveals an expansion of olfactory receptor genes in freshwater fish. Mol Biol Evol. 2021;38(10):4238–51.
40. Liu H, Chen C, et al. *Megalobrama amblycephala* genome sequencing and assembly. NCBI BioProject. 2021. https://identifiers.org/ncbi/bioproject:PRJNA343584.
41. Ren L, Tu X, Luo M, Liu Q, Cui J, Gao X, et al. Genomes reveal pervasive distant hybridization in nature among cyprinid fishes. GigaScience. 2025;14:giaf001.
42. Ren L, Tu X, Luo M, Liu Q, Cui J, Gao X, et al. *Culter alburnus* genome assembly. Genome Warehouse in National Genomics Data Center. 2025. https://identifiers.org/cncb.ngdc.gwh:GWHDOEX00000000.
43. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.
44. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.
45. Srinivasan KA, Virdee SK, McArthur AG. Strandedness during cDNA synthesis, the stranded parameter in htseq-count and analysis of RNA-seq data. Brief Funct Genomics. 2020;19(5–6):320–9.
46. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31(2):166–9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.