Aquaculture 595 (2025) 741630

Contents lists available at ScienceDirect

Aquaculture

journal homepage: www.elsevier.com/locate/aquaculture

Transcriptome analysis reveals genetic regulation of growth diversity in F_2 intergeneric hybrids of *Megalobrama amblycephala* and *Culter alburnus*

Mengxue Luo [a,1], Yakui Tai [a,1], Mengdan Li [a], Yiyan Zeng [a], Chang Wu [a], Ling Liu [a], Hong Zhang [a], Li Ren [a,b,*], Shaojun Liu [a,b,*]

[a] State Key Laboratory of Developmental Biology of Freshwater Fish, Engineering Research Center of Polyploid Fish Reproduction and Breeding of the State Education Ministry, College of Life Sciences, Hunan Normal University, Changsha 410081, China
[b] Hunan Yuelu Mountain Scienece and Technology Co.Ltd. for Aquatic Breeding, Changsha 410081, China

ARTICLE INFO

Keywords:
Intergeneric hybridization
Allele-specific expression
Growth

ABSTRACT

Hybridization between different genera can greatly shape the genotype and phenotype of individuals. It is possible to sporadically obtain individuals or populations with heterosis through conducting multiple rounds and stages of hybridization, backcrossing, and self-crossing. Investigating the regulatory relationship between allelic combinations from different species and specific economic traits in hybrid individuals will help us understand the genetic mechanism of heterosis. In this study, we conducted a transcriptomic analysis of 25 individuals from the F_2 population obtained from the intergeneric hybridization of blunt snout bream (*Megalobrama amblycephala*, BSB) and topmouth culter (*Culter alburnus*, TC). We identified 94,897 SNPs on the subgenome BSB (originating from species BSB) and 26,372 SNPs on the subgenome TC (originating from species TC). By utilizing WGCNA to construct a co-expression network, we successfully identified 54 key genes involved in growth regulation. Of these, 8 genes from subgenome BSB show 7 positively influencing and 1 negatively influencing growth rate. Additionally, 46 genes in subgenome TC include 24 with positive and 22 with negative correlations to growth rate. Using eQTL analysis, we assessed associations between SNPs and gene expression levels for 3736 *cis*- and 7683 *trans*-regulatory events. In comparative analyses of eQTL and WGCNA, we detected 2 *cis*- and 16 *trans*-regulatory genes shared in them, indicating their potential regulatory network in growth diversity in the hybrid. These results help us understand the complex regulatory network of growth traits in intergeneric hybrids. It represents a significant advancement in molecular-assisted breeding, offering valuable insights for developing effective hybrid breeding strategies and achieving heterosis in populations or individuals.

1. Introduction

Megalobrama amblycephala, commonly known as blunt snout bream (BSB; 2n = 2× = 48), and *Culter alburnus*, known as topmouth culter (TC; 2n = 2× = 48), belong to the family Cyprinidae and subfamily Cultrinae. They are economically significant freshwater fish species in China (Zhou et al., 2008). These two species exhibit distinct feeding habits and morphological characteristics (Ren et al., 2019). The BSB is herbivorous and possesses a higher dorsal fin and a shorter body compared to the carnivorous TC. The wild populations of BSB and TC have been severely depleted due to overfishing, and artificial culture has become the primary way of production for these species. However, in recent years, their growth rates and meat quality have declined during cultivation, posing challenges to sustainable aquaculture practices. It is crucial to genetically improve the existing BSB and TC populations to obtain superior strains with desirable traits, which will contribute to the healthy development of aquaculture in China.

Hybridization offers a rapid way to generate diverse genotypes and

Abbreviations: BSB, *M. amblycephala*; TC, *C. alburnus*; WGCNA, weighted correlation network analysis; SNP, single nucleotide polymorphism; BL, body length; BH, body height; BW, body weight; TPM, transcripts per million; AGP, allelic gene pair; PCC, pearson correlation coefficient; ASE, allelic-specific expression; GWAS, genome-wide association study; eQTL, expression quantitative trait loci; MB, megabase; TOM, topological overlap matrix; GB, gigabase..

* Corresponding authors at: State Key Laboratory of Developmental Biology of Freshwater Fish, Engineering Research Center of Polyploid Fish Reproduction and Breeding of the State Education Ministry, College of Life Sciences, Hunan Normal University, Changsha 410081, China.
E-mail addresses: renli_333@163.com (L. Ren), lsj@hunnu.edu.cn (S. Liu).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.aquaculture.2024.741630
Received 16 February 2024; Received in revised form 12 June 2024; Accepted 13 September 2024
Available online 14 September 2024

phenotypes, presenting ample opportunities to investigate the influence of genetic regulation on phenotypic traits. Interspecific hybridization has been extensively employed in certain plant species, such as *Triticum aestivum* × *Secale cereal* (Degraeve et al., 2021) and *Brassica nigra* × *B. rapa* (Paritosh et al., 2014), aiming to develop varieties with superior economic traits. Similarly, in the field of fish breeding, interspecific hybridization involving different genera or subfamilies has been observed in carps (Liu et al., 2016; Qin et al., 2014), salmonid fish (Hórreo et al., 2011), and cichlid fish (Selz and Seehausen, 2019). Of particular interest, the intergeneric hybrids between BSB and TC demonstrate a range of phenotypic variations, including intermediate body shapes (Rapp et al., 2009a) and diverse growth patterns. These hybrids provide valuable material for studying the genetic basis underlying the observed phenotypic diversity.

The study of genome-wide allelic expression is an important way to find out about the different kinds of quantitative traits that are linked to changes in allelic expression (Ren et al., 2019). This type of research is extensively conducted on various plant species, including cotton (Rapp et al., 2009b), rice (Song et al., 2013) and maize (Springer and Stupar, 2007). Such changes arise from hybridization and can lead to allelic expression dominance (Yoo et al., 2013). Asymmetric expression patterns are often linked to imbalanced expression of alleles in hybrids, resulting from *cis*-regulatory changes that affect transcription initiation, transcription rates, and/or transcript stability in alleles. Additionally, *trans*-regulatory changes can impact the efficiency of *cis-trans* interactions by modifying the activity or expression of transcription factors (Maheshwari and Barbash, 2012; McManus et al., 2010).

To investigate the growth diversity within these hybrids, we conducted genome-wide allelic expression analysis using the weighted gene co-expression network analysis (WGCNA) method. By employing this approach, we identified key growth-regulated genes. In addition, we found single nucleotide polymorphisms (SNPs) in the coding sequences of the alleles and looked for links between these SNPs and the growth traits in the hybrid population. Our findings highlight the importance of allelic expression in regulating growth diversity, offering valuable insights for the study and application of genetic breeding in aquaculture.

2. Material and methods

2.1. Samples

The fish used in this study involved the F_2 hybrid population, which was obtained from the hybridization of male *C. alburnus* (topmouth culter, TC) and female *M. amblycephala* (blunt snout bream, BSB), and the subsequent self-crossing of the F_1 hybrid population. These fish were fed in a pool with suitable environmental conditions involving water temperatures, oxygen content, forage, breeding density, *etc.* These pools were located in the drainage area of Dongting Lake, Hunan, China (29° 11′ 51“ N, 112° 35’ 50” E). Twenty-five healthy individuals (twelve months after hatching) were collected for this study. The growth traits, including body length (BL), body height (BH), and body weight (BW), were detected for each individual. These fish were deeply anesthetized with 300 mg/L Tricaine Methanesulfonate (Sigma-Aldrich, St. Louis, MO, USA) for 10 min (25 °C) in a separate tank. After confirming death, they were collected to extract the muscle tissue, which has significant implications for growth traits. The DNA content of the erythrocytes of the hybrids was measured using flow cytometry (Cell Counter Analyzer, Partec, Germany) (Xiao et al., 2014). The transcriptome data of muscle tissue in 25 hybrid individuals was obtained using DNA nanoball (DNBSEQ-T7) technology according to the standard method (Patterson et al., 2019).

2.2. RNA isolation and mRNA-seq

Total RNA of the 25 F_2 hybrid in muscle tissue was isolated and purified according to the TRIzol extraction method, respectively (Rio et al., 2010). The RNA concentration was measured using NanoDrop technology. Total RNA samples were treated with DNase I (Invitrogen) to remove any contaminating genomic DNA. The purified RNA was quantified using an Agilent 2100 system. Isolated mRNA was fragmented with a fragmentation buffer. The resulting short fragments were reverse transcribed and amplified to produce cDNA. The mRNA-seq libraries of the 25 samples were prepared according to the standard high-throughput method. The quality of the cDNA library was assessed by the Agilent 2100 system.

2.3. Detection of gene expression based on mRNA-seq

The sequencing adaptors, duplicated read pairs, and low-quality reads of RNA-seq were removed with Fastp (v. 0.21.0) (Chen et al., 2018). After quality control, the clean reads of RNA-seq data were aligned to the combined genomes of *M. amblycephala* and *C. alburnus* using bowtie2 (v. 2.3.5.1) (Langmead et al., 2009) with default parameters. Then the alignment result in SAM format was converted to BAM format and sorted using SAMtools (v. 1.10) (Li et al., 2009), the unique mapped reads were tabulated using htseq-count (v. 0.12.4) (Srinivasan et al., 2020). The gene expression values of 25 samples were normalized based on the ratio of the number of mapped reads for each gene to the total number of mapped reads for the entire genome. The gene with an average mapped reads <5 across all samples was filtered out, and we finally obtained 23,438 expressed genes along with their expression levels. The transcripts per million (TPM) values were calculated based on the normalized data and used for WGCNA and correlation analysis.

The allelic gene pair (AGP) between the subgenome BSB (originating from *M. amblycephala*) and subgenome TC (originating from *C. alburnus*) in the hybrid offspring was obtained using the all-against-all reciprocal BLASTP (v. 2.8.1) with an e-value threshold of exp.(−6) based on protein sequences. These AGPs were then selected based on the 23,438 expressed genes identified in the above step, resulting in 6252 AGPs. Then we obtained 6252 non-AGPs through the staggered pairing method. Furthermore, we conducted Pearson correlation coefficient (PCC) analysis on the AGPs and non-AGPs using the cor.test function in R.

To investigate allele-specific expression (ASE) regulating growth diversity in the hybrid population, we calculated bias values by using the $ASE_{bias\ value} = C_{BSB}/(C_{BSB} + C_{TC})$ -0.5 formula, where the “C” represents the read count of AGP, and used the POISSON.DIST function to compute the *p*-value. If the *p*-value was less than or equal to 0.05 ($p \leq 0.05$) and the bias value was greater than zero (bias value >0), the AGP was considered BSB allelic expression bias. Conversely, if the *p*-value met the same threshold ($p \leq 0.05$) but the bias value was less than (bias value <0), the AGP was considered TC allelic expression bias. Then we used the cor.test function to perform PCC analysis between the bias values and body weight and identified AGPs by applying a threshold of *p*-value ≤0.05 and $|PCC| \geq 0.4$.

2.4. SNP detection and eQTL analysis

SNPs were extracted using GATK (v. 4.0.4.0). First, the Picard toolkit was used to reorder the BAM file obtained in the alignment step, add the RG information in the header, remove the duplicate sequence generated by PCR, and then use the ‘HaplotypeCaller’ function to call SNPs (Gao et al., 2021). As there were multiple samples, the intermediate gVCF file needed for subsequent analysis was generated for each sample. Then, the ‘CombineGVCFs’ function was used to integrate the multiple gVCF files into a combined VCF file. The final filtered VCF file was obtained by the ‘VariantFiltration’ function with hard filtering thresholds (Van der Auwera et al., 2018). In the end, we obtained 121,269 SNPs for further analysis.

We used the PLINK (v. 1.90b6.10) (Purcell et al., 2007) to perform genome-wide association study (GWAS) analysis, with body weight as

the phenotype. To express the results more intuitively, we converted the *p*-value to -$\log_{10}$(*p*-value). Then we applied a threshold of -$\log_{10}$(*p*-value) $\geq$ 5 to identify the SNPs significantly associated with body weight.

Then R-package Matrix eQTL (v. 2.3) (Shabalin, 2012) was used to detect expression quantitative trait loci (eQTL) and calculate the relationship between SNPs and gene expression. First, the threshold cisDist was set at 1 megabase (MB), which represents the maximum distance between SNP and gene association. The *cis*-eQTL mapping window encompassed a region spanning 1 Mb upstream of the transcription start site to 1 Mb downstream of the gene end. SNP-gene pairs falling within this region were classified as *cis*-associated, whereas those outside were classified as *trans*-associated (Liu et al., 2020). For *cis*-eQTL and *trans*-eQTL, we applied the adjusted thresholds of *p*-value $\leq$ exp.(−3) and *p*-value $\leq$ exp.(−6), respectively.

2.5. WGCNA analysis

To identify specific modules associated with three phenotypes, a weighted co-expression network of all expressed genes was constructed using the R package WGCNA (v. 1.71) (Langfelder and Horvath, 2008). In this study, we chose the soft-threshold power of 14 for automatic network construction with the blockwiseModules function. Subsequently, we transformed the adjacency matrix into a topological overlap matrix (TOM). The TOM is used for hierarchical clustering, which groups genes with similar expression profiles into modules, each module represents a cluster of co-expressed genes. Each module was required to contain at least 30 genes, based on our chosen criteria. Then, the modules strongly correlated with traits (*p*-value $\leq$0.05) are potential candidates for further biological investigation. The genes within the mentioned modules are sorted out and filtered out based on the threshold of |PCC| $\geq$ 0.5 between genes and modules, as well as the threshold of |PCC| $\geq$ 0.5 between genes and three phenotypes. Finally, based on the threshold of |PCC| $\geq$ 0.5 between genes and the three phenotypes (BW, BL, and BH), these hub genes are classified into three types.

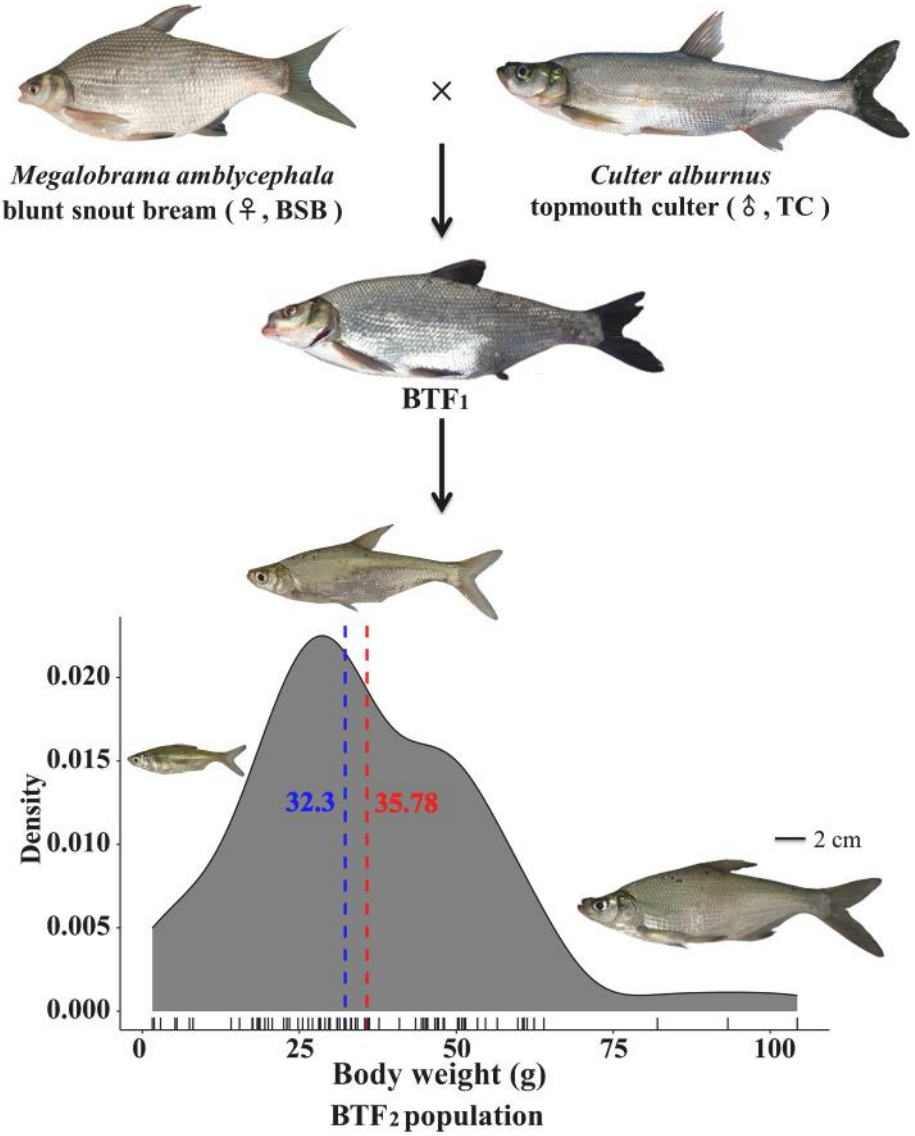

Fig. 1. Appearance of hybrid lineage of *M. amblycephala* and *C. alburnus*, body weight of BTF_2 population. Appearance of BSB, TC, and their hybrid progenies of BTF_1 (two-year-old individual), and BTF_2 (one-year-old individual). Blue dotted line represents the median value of body weight (50 individuals) in the BTF_2 population, red dotted line represents the mean value. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Results

3.1. Growth diversity

The 25 F_2 hybrid individuals were obtained from the intergeneric hybridization of female *M. amblycephala* (blunt snout bream, BSB) and male *C. alburnus* (topmouth culter, TC) and subsequent self-crossing of the F_1 hybrid population, providing key genetic material for investigating the genetic mechanisms of growth diversity in hybrids. We randomly collected 25 individuals (body weight ranged from 1.6 g to 104.3 g) from this population, which was bred in a pond with suitable environments (Fig. 1 and Table B.1). The values of body length *vs.* body height ranged from 3.56 to 4.34, while distinct ones were observed between inbred paternal BSB and TC (Table B.1) (Xiao et al., 2014). In conclusion, the F_2 hybrid resulting from the hybridization of BSB and TC provides valuable insights into the interaction of alleles from both species in regulating the diversity of growth traits.

3.2. Allele-specific expression

To further investigate ASE in the F_2 hybrid population, we mapped the 307.52 gigabase (GB) of clean reads of the 25 transcriptomes to the combined genomes of BSB and TC (Table B.2). The 65,453,073 unique reads were utilized to assess gene expression (Table B.3). We detected 15,305–19,960 expressed genes in these hybrid individuals. Among these, 6712–10,309 (42.33 %–58.72 %) originated from subgenome BSB, while 6992–10,076 (41.28 %–57.67 %) were from subgenome TC (Table B.4). We also analyzed 6252 AGPs between BSB and TC. Among these, 709 to 2867 alleles (11.34 %–45.86 %) from TC were predominantly expressed in the 25 hybrid individuals, exhibiting higher expression levels of TC alleles compared to those of BSB. Conversely, 712 to 2336 alleles (11.39 %–37.36 %) from BSB demonstrated dominant expression (Table B.5). Additionally, we observed co-expression of alleles BSB and TC in 2215–4573 (35.43 %–73.14 %) genes within the hybrid population (Table B.5), demonstrating varied expression patterns of alleles from different species.

Focusing on the distribution of diverse ASE patterns in the 25 hybrid individuals, we found that 659–1527 (10.56 %–24.45 %) gene pairs exhibited a BSB allelic expression bias, while 2427–4155 (38.84 %–66.60 %) gene pairs showed a bias towards TC allelic expression. Moreover, 1116 to 2927 (17.91 %–46.84 %) gene pairs did not display any expression bias (Table B.6). These results suggest a stronger bias in allele expression towards the subgenome TC in the hybrid progenies (Fig. 2A). In investigating ASE's role in regulating growth diversity within the hybrid population, we selected 155 AGPs through a Pearson correlation analysis between body weight and bias values (Table B.7).

Analyzing the PCC of expression values in AGPs and non-AGPs provides insights into allele interactions. Within the 6252 AGPs examined, the average absolute value of gene expression correlation coefficient was significantly higher in AGPs (average absolute value of PCC = 0.56) compared to non-AGPs (average absolute value of PCC = 0.18) (Fig. 2B). Of these AGPs, 5224 (83.56 %) demonstrated a positive correlation ($|PCC| \geq 0.3$, p-value ≤ 0.05), with 4416 (70.63 %) showing a strong positive correlation ($|PCC| \geq 0.5$, p-value ≤ 0.05) (Table B.8). These findings suggest a regulatory interplay between BSB and TC alleles, possibly driven by their shared *trans*-regulatory factors.

3.3. Genomic variations regulate gene expression

The eQTL analysis was performed on transcriptomic data from 25 hybrid individuals, aiming to assess the associations between SNPs and gene expression levels, focusing on 3736 *cis*- and 7683 *trans*-regulatory events (Fig. 3A). Notably, the distribution of observed p-values for *trans*-eQTLs demonstrated an earlier deviation from the diagonal compared to *cis*-eQTLs. This suggests that *trans*-eQTLs are more readily detectable than *cis*-eQTLs in the transcriptome data of these hybrids (Liu et al., 2020) (Fig. A.1).

Focused on 3736 *cis*-regulatory events (p-value $\leq$ exp.(−3)), we discovered 2463 *cis*-regulatory SNPs (*cis*-SNPs) affecting 1141 genes (*cis*-genes) in the subgenome BSB, while 492 *cis*-SNPs regulating 520 genes were in the subgenome TC (Fig. A.2 A and Table B.9). Additionally, our study identified 7683 *trans*-regulatory events (p-value $\leq$ exp.(−6)), revealing 2666 *trans*-regulatory SNPs (*trans*-SNPs) affecting 2027 genes (*trans*-genes) in the subgenome BSB and 752 *trans*-SNPs regulating 1957 genes in the subgenome TC (Fig. A.2B and Table B.9). In summary, our analysis based on transcriptome data revealed a higher incidence of *trans*-regulatory events compared to *cis*-regulatory events. Moreover, the subgenome BSB of the hybrid population exhibited a greater number of regulatory events than the subgenome TC.

The 7683 *trans*-regulatory events can be categorized into four distinct patterns: Pattern 1 (3128 events, BSB-SNPs regulating TC genes), Pattern 2 (3242 events, BSB-SNPs regulating BSB genes), Pattern

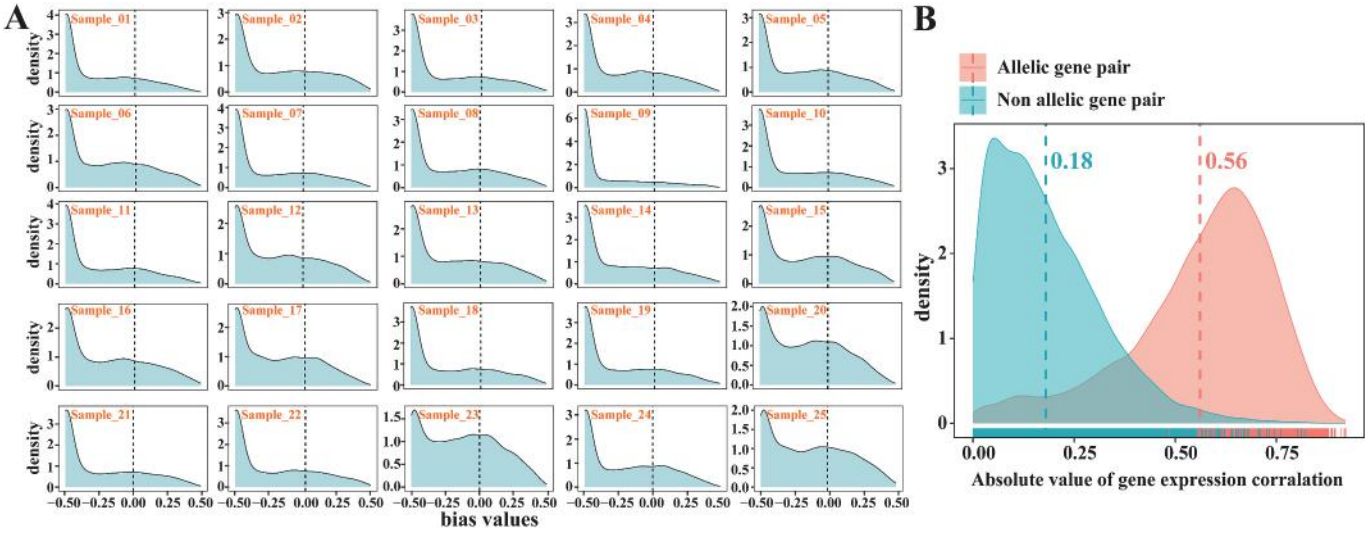

Fig. 2. Allele-specific expression and correlated analyses in AGPs and non-AGPs.
A. The gene number with TC allelic expression bias was higher than those with BSB allelic expression bias across 25 individuals. The bias values smaller than zero were considered TC allelic expression bias, while bias values bigger than zero were considered BSB allelic expression bias.
B. Expression correlational analyses of the 6252 allelic gene pairs (AGPs) and non-allelic gene pairs (non-AGPs) based on Pearson correlation coefficient (PCC) values.

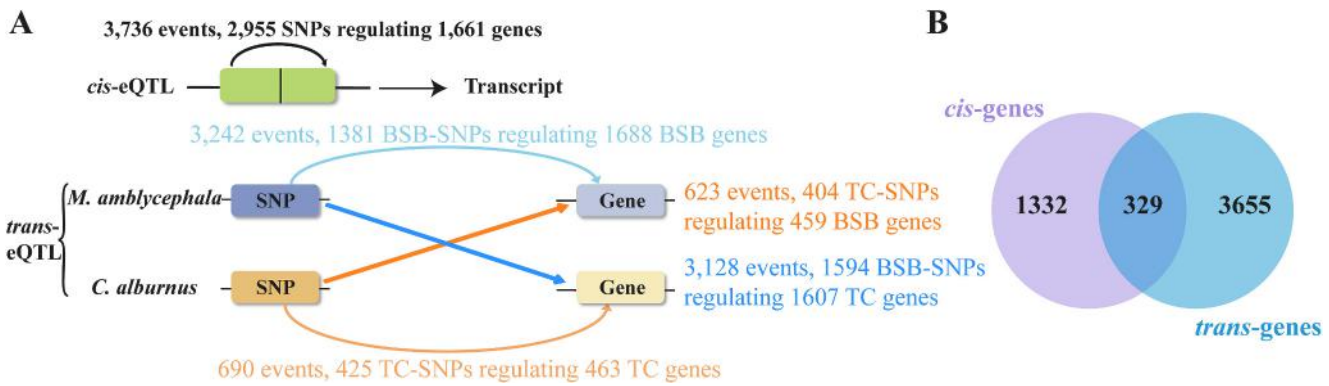

Fig. 3. Genome-wide eQTL analysis.
A. A schematic diagram of *cis*- and *trans*-regulatory events in hybrid.
B. Distribution of *cis*-genes and *trans*-genes.

3 (623 events, TC-SNPs regulating BSB genes), and Pattern 4 (690 events, TC-SNPs regulating TC genes) (Fig. A.3 and Table B.10). Our analysis reveals that SNPs on the subgenome BSB play a more prominent role in regulating gene expression within the hybrid population compared to those on the subgenome TC (Fig. A.3 and Table B.9). An intersection analysis of *cis*-genes and *trans*-genes uncovered 329 genes regulated by both *cis*- and *tran*-SNPs (Fig. 3B). Additionally, we observed that the chromosomal distribution of both *cis*-SNPs and *trans*-SNPs was predominantly concentrated on chromosomes 7 and 21. We conducted a GWAS and found that 105 SNPs across 121,269 SNPs showed a significant correlation with body weight (Fig. 4).

3.4. Growth-regulated genes predicted by WGCNA

To explore the regulation of gene expression on growth phenotypes in the hybrid population, we mapped the clean reads from 25 transcriptomes to the combined genomes of BSB and TC and calculated gene expression values. We focused on three growth phenotypes—body weight, body length, and body height—and conducted a WGCNA analysis. Out of the 23,438 total expressed genes, 54 (0.23 %) were identified as hub genes within four significant modules (MElightpink4, MEfirebrick4, MEdarkslateblue, and MEorange) (Fig. A.4). These 54 genes are considered potential regulators of growth in the hybrid population. Intriguingly, among these hub genes, 8 originated from the subgenome BSB and 46 from the subgenome TC. In the subgenome BSB, 7 out of 8 hub genes showed a positive correlation between expression and three growth phenotypes, whereas in the subgenome TC, 24 out of 46 exhibited this positive correlation (Fig. 5 and Table B.11). There are 21 unknown genes in the 54 hub genes, numbered according to the "Gene**" format (Table B.12).

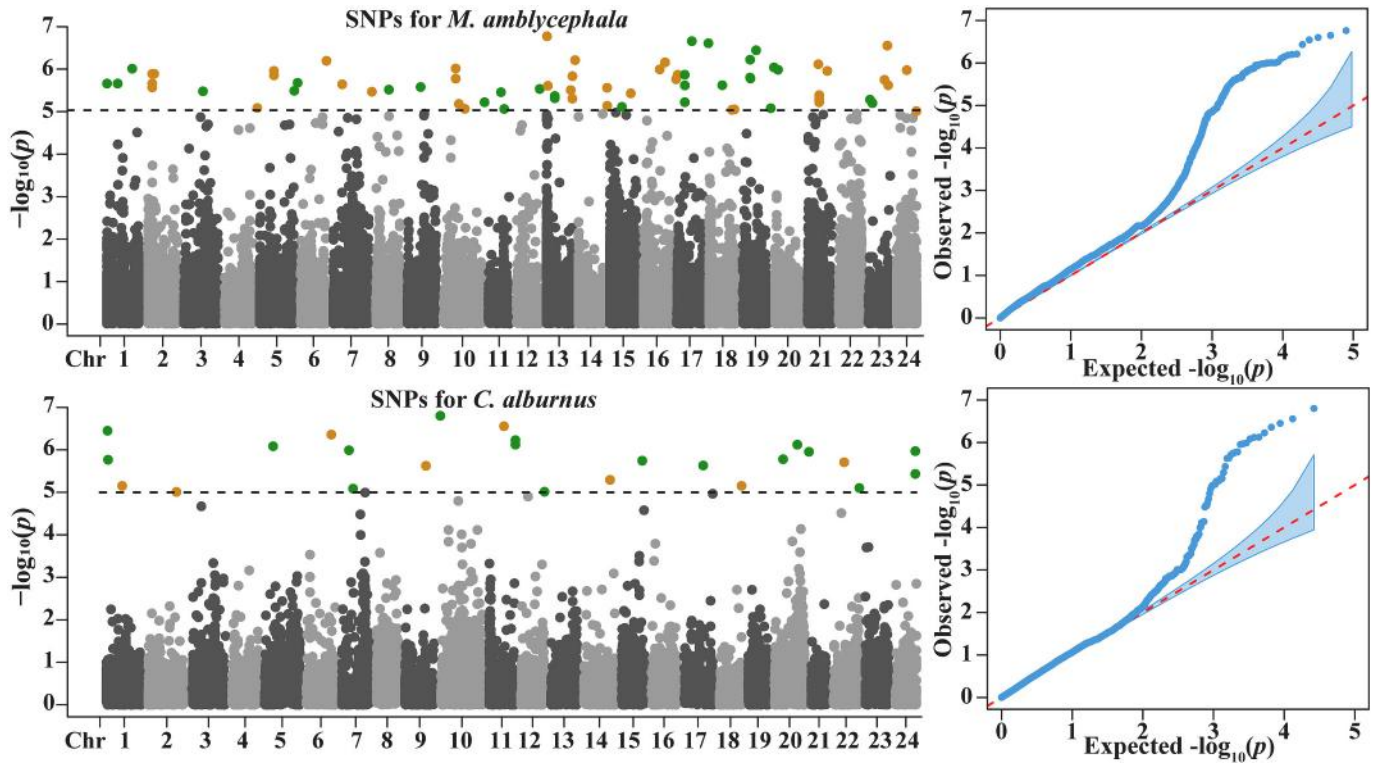

Fig. 4. Manhattan plots of the GWAS of body weight in subgenomes BSB and TC of the hybrid. The *p*-values (≤ exp.(−5)) of SNPs marked with green and orange. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

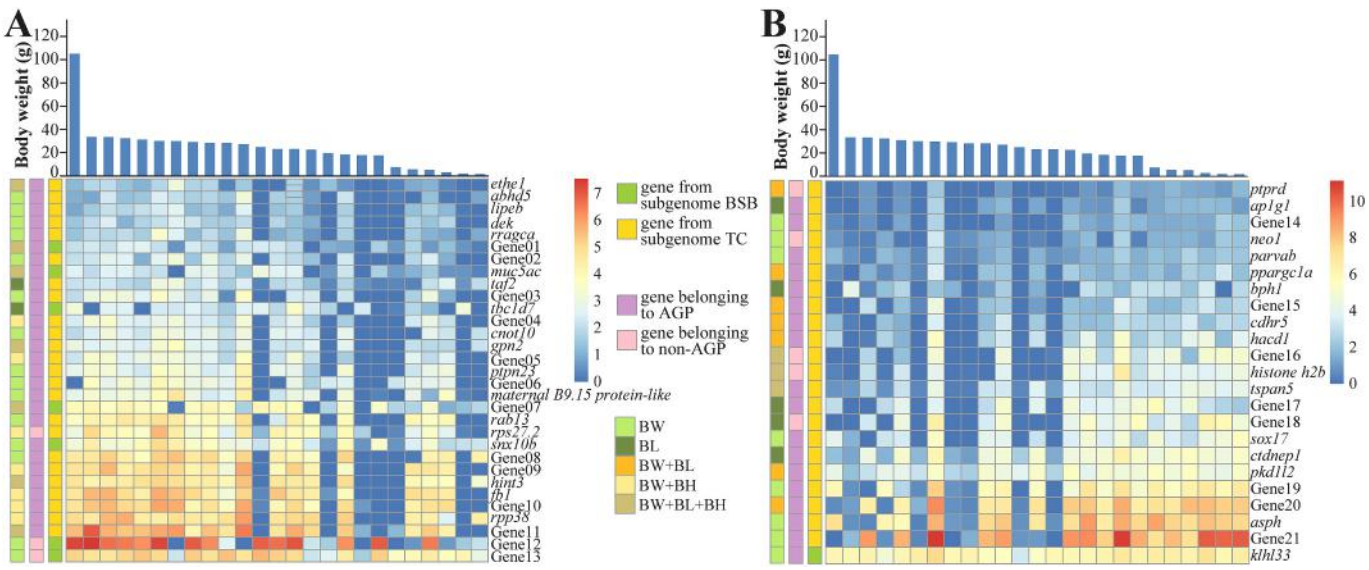

Fig. 5. Expression levels of the 54 growth-regulated genes and body weight in 25 hybrid individuals.
A. The expression values of the 31 genes exhibited a positive correlation with the values of body weight.
B. The expression values of the 23 genes exhibited a negative correlation with the values of body weight.
Note: "BW" represents the genes identified through WGCNA analyses of the body weight trait; "BL" represents the genes identified through WGCNA analyses of the body length trait; "BW + BL" represents the shared genes in WGCNA analyses of the two traits (body weight and body length); "BW + BH" represents the shared genes in WGCNA analyses of the two traits (body weight and body height); "BW + BL + BH" represents the shared genes in WGCNA analyses of the three traits (body weight, body length, and body height). The "Gene**" represents unknown gene.

3.5. Distribution of growth-regulated genes among eQTL, WGCNA, ASE, and GWAS analyses

To investigate the genetic mechanism of growth diversity in hybrids, we performed a series of analyses, including eQTL, WGCNA, and GWAS (Fig. 6A). Through the above eQTL and GWAS analyses, we detected 9 SNPs involving 9 genes shared in them, exhibiting their regulatory network in growth rate through *cis*- and *trans*-SNP-gene pairs (Fig. 6B). Meanwhile, we performed analyses on 5645 eQTL-associated genes and the 54 hub genes detected based on WGCNA analysis. The 2 *cis*-genes and 16 *trans*-genes were shared between them (Fig. 6C). Then, we performed conjoint analyses on the 54 hub growth-regulated genes and the 155 ASE-genes, and their ASE exhibited a correlation with body weight values. We found the four share genes in them, including *tspan5* (K17345, e-value = 1.1exp(−168)) (Fig. 6D). These results, including eQTL, WGCNA, ASE, and GWAS analyses, enhance my understanding of the genetic mechanisms of growth diversity in hybrids.

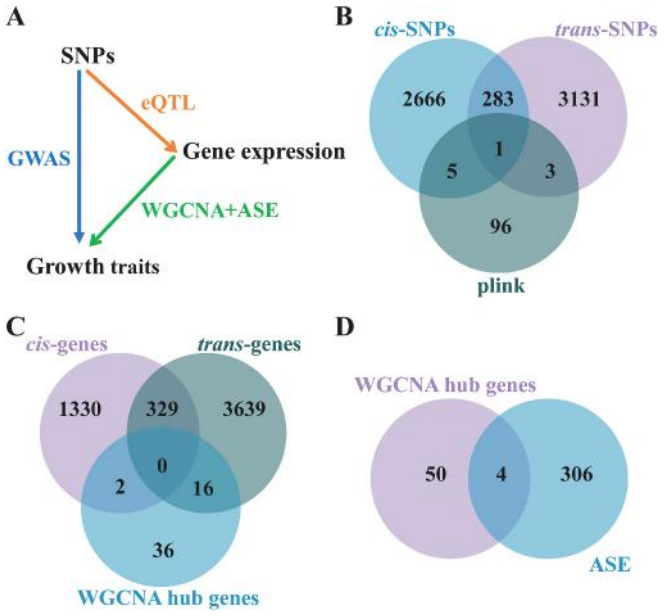

Fig. 6. Association analyses among SNPs, gene expression and growth traits, gene and SNP distributions of eQTL, WGCNA, ASE, and GWAS results.
A. Investigate the impact of SNPs on growth traits through GWAS method. Investigate the impact of SNPs on gene expression through eQTL method. Investigate the impact of gene expression on growth traits through WGCNA and ASE methods.
B. SNP distribution among *cis*- and *trans*-eQTL and GWAS.
C. Gene distribution among *cis*- and *trans*-eQTL and WGCNA.
D. Gene distribution among ASE and WGCNA.

4. Discussion

The development of the global aquaculture industry depends on superior fish varieties. In recent years, environmental changes, including pollution and habitat destruction, have led to a continuous decrease in the size and diversity of the wild freshwater farmed fish populations, thereby depleting genetic resources in aquaculture (Liu et al., 2019). As an important component of the aquaculture industry, the development of freshwater fisheries has become an important direction that we need to focus on. Many important freshwater farmed fish, such as grass carp (*Ctenopharyngodon idella*), silver carp (*Hypophthalmichthys molitrix*), and bighead carp (*Hypophthalmichthys nobilis*), belong to the family Cyprinidae (Hontela and Stacey, 2019). As the largest category of vertebrates, cyprinid fish have undergone multiple events involving hybridization, gene flow, and polyploidization during their formation. The latest studies indicate that their radiation evolution and speciation are closely associated with hybridization events (Pereira et al., 2014). Compared to other vertebrates such as birds and mammals, the gonads of cyprinid fish exhibit conservation, which provides a crucial genetic basis for the hybridization of extant cyprinid fish (Hliwa et al., 2017). In this context, carrying out genetic breeding of intergeneric and interspecific hybrid fish populations is crucial for promoting the development of freshwater fisheries.

Interspecific hybridization between BSB (blunt snout bream) and TC

(topmouth culter) offers a unique approach to studying the genetic basis of growth traits in fish. Our research utilizes the F_2 generation resulting from this hybridization to gain insights into the relationship between allelic variation and growth phenotype. The F_1 hybrid population exhibits a rich genetic tapestry due to inherent differences between parental BSB (high-backed) and TC (slender-bodied) fish. These differences stem from single nucleotide polymorphisms (SNPs) within the parental populations. Furthermore, the segregation of alleles from the F_1 gametes and their independent assortment during fertilization in the F_2 generation lead to a further diversification of genotypes (Wang and Xu, 2019). This diversification, confirmed through studies on ribosomal RNA and nuclear genes (Ren et al., 2019; Xiao et al., 2014), results in a wider range of allelic combinations influencing growth traits.

Our study, building on previous observations in plants, including cotton (Bao et al., 2019) and maize (Yang et al., 2017), and fish, including economical fish species such as swordtail fish (Powell et al., 2020), tilapia (Zhong et al., 2019) and goldfish (Ren et al., 2022), reveals that alleles from different species (allo-alleles) exert differential effects on growth rate in the F_2 hybrids. Four genes identified through ASE and WGCNA analyses, including *tspan5*, emerged as potential regulators of human body height (Yengo et al., 2022). These variations likely arise from differences in transcription sequences and translated peptide chains. These peptides can directly influence growth-related physiological pathways. The study demonstrates that allo-allele combinations of key growth-regulated genes can have both positive and negative effects on individual growth rates and body sizes.

While our F_2 population provides valuable insights, the present genotypic range is limited due to recombination events. To systematically explore the optimal combinations of allo-alleles that shape ideal growth phenotypes, further research is needed on two key fronts: 1) Building a more diverse genotypic population through targeted breeding strategies will be crucial. 2) Employing advanced genomic and transcriptomic studies is essential to identify specific genes and their regulatory mechanisms underlying the effects of heterologous alleles. Understanding the genetic basis of growth traits will empower targeted breeding strategies. Backcrossing and other selective breeding approaches can be employed to cultivate superior fish varieties with the desired body shapes and growth rates. Additionally, future research should explore the interaction between genotype and breeding environment for optimizing hybrid cultivation practices.

Animal ethics declarations

All animal experiments were approved by the Animal Care Committee of Hunan Normal University and carried out in accordance with the approved guidelines, which were approved by the Science and Technology Bureau of China.

Funding

This research was supported by the National Natural Science Foundation of China (32341057, 32293250, 32293252, and U19A2040), Hunan Provincial Natural Science Foundation (2022JJ10035), the Huxiang Young Talent Project of China (2021RC3093), National Key Research and Development Plan Program (2023YFD2401602), Special Funds for Construction of Innovative Provinces in Hunan Province (2021NK1010), Key Research and Development Program of Hunan Province of China (2023WK2001), Earmarked Fund for Agriculture Research System of China (CARS-45), and the 111 Project (D20007).

CRediT authorship contribution statement

Mengxue Luo: Writing – original draft, Data curation. **Yakui Tai:** Resources. **Mengdan Li:** Resources. **Yiyan Zeng:** Resources. **Chang Wu:** Resources. **Ling Liu:** Resources. **Hong Zhang:** Resources. **Li Ren:** Formal analysis, Writing – original draft, Writing – review and editing, Funding acquisition. **Shaojun Liu:** Writing – review & editing, Project administration, Funding acquisition.

Declaration of competing interest

The authors have declared that no competing interests exist.

Data availability

The raw reads of mRNA-seq data (25 individuals) have been deposited in National Genomics Data Center (NGDC) (https://bigd.big.ac.cn/gsub/) under the BioProject accession numbers CRA014687.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aquaculture.2024.741630.

References

Bao, Y., Hu, G., Grover, C.E., Conover, J., Yuan, D., Wendel, J.F., 2019. Unraveling cis and trans regulatory evolution during cotton domestication. Nat. Commun. 10, 5399. https://doi.org/10.1038/s41467-019-13386-w.

Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

Degraeve, S., De Baerdemaeker, N.J.F., Ameye, M., Leroux, O., Haesaert, G.J.W., Steppe, K., 2021. Acoustic vulnerability, hydraulic capacitance, and xylem anatomy determine drought response of small grain cereals. Front. Plant Sci. 12, 599824. https://doi.org/10.3389/fpls.2021.599824.

Gao, Z., Li, H., Yang, X., Yang, P., Chen, J., Shi, T., 2021. Biased allelic expression in tissues of F_1 hybrids between tropical and temperate lotus (Nelumbo nuicfera). Plant Mol. Biol. 106, 207–220. https://doi.org/10.1007/s11103-021-01138-8.

Hliwa, P., Król, J., Sikorska, J., Wolnicki, J., Dietrich, G.J., Kamiński, R., Stabińska, A., Ciereszko, A., 2017. Gonadogenesis and annual reproductive cycles of an endangered cyprinid fish, the lake minnow Eupallasella percnurus (Pallas, 1814). Anim. Reprod. Sci. 176, 40–50. https://doi.org/10.1016/j.anireprosci.2016.11.006.

Hontela, A., Stacey, N., 2019. Cyprinidae. In: Reproductive Seasonality in Teleosts. CRC Press, pp. 53–78.

Hórreo, J.L., Ayllón, F., Perez, J., Beall, E., Garcia-Vazquez, E., 2011. Interspecific hybridization, a matter of pioneering? Insights from Atlantic salmon and brown trout. J. Hered. 102, 237–242. https://doi.org/10.1093/jhered/esq130.

Langfelder, P., Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformat. 9, 559. https://doi.org/10.1186/1471-2105-9-559.

Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25. https://doi.org/10.1186/gb-2009-10-3-r25.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Subgroup, G.P.D.P., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Liu, S., Luo, J., Chai, J., Ren, L., Zhou, Y., Huang, F., Liu, X., Chen, Y., Zhang, C., Tao, M., Lu, B., Zhou, W., Lin, G., Mai, C., Yuan, S., Wang, J., Li, T., Qin, Q., Feng, H., Luo, K., Xiao, J., Zhong, H., Zhao, R., Duan, W., Song, Z., Wang, Y., Wang, J., Zhong, L., Wang, L., Ding, Z., Du, Z., Lu, X., Gao, Y., Murphy, R.W., Liu, Y., Meyer, A., Zhang, Y. P., 2016. Genomic incompatibilities in the diploid and tetraploid offspring of the goldfish x common carp cross. P. Natl. Acad. Sci. Usa. 113, 1327–1332. https://doi.org/10.1073/pnas.1512955113.

Liu, E., Wang, X., Liu, H., Liang, M., Zhu, Y., Li, Z., 2019. Chemical speciation, pollution and ecological risk of toxic metals in readily washed off road dust in a megacity (Nanjing), China. Ecotox. Environ. Safe. 173, 381–392. https://doi.org/10.1016/j.ecoenv.2019.02.019.

Liu, Y., Liu, X., Zheng, Z., Ma, T., Liu, Y., Long, H., Cheng, H., Fang, M., Gong, J., Li, X., Zhao, S., Xu, X., 2020. Genome-wide analysis of expression QTL (eQTL) and allele-specific expression (ASE) in pig muscle identifies candidate genes for meat quality traits. Genet. Sel. Evol. 52. https://doi.org/10.1186/s12711-020-00579-x.

Maheshwari, S., Barbash, D.A., 2012. Cis-by-trans regulatory divergence causes the asymmetric lethal effects of an ancestral hybrid incompatibility gene. PLoS Genet. 8, e1002597. https://doi.org/10.1371/journal.pgen.1002597.

McManus, C.J., Coolon, J.D., Duff, M.O., Eipper-Mains, J., Graveley, B.R., Wittkopp, P.J., 2010. Regulatory divergence in Drosophila revealed by mRNA-seq. Genome Res. 20, 816–825. https://doi.org/10.1101/gr.102491.109.

Paritosh, K., Gupta, V., Yadava, S.K., Singh, P., Pradhan, A.K., Pental, D., 2014. RNA-seq based SNPs for mapping in Brassica juncea (AABB): synteny analysis between the two constituent genomes a (from *B. Rapa*) and B (from *B. Nigra*) shows highly divergent gene block arrangement and unique block fragmentation patterns. BMC Genomics 15, 396. https://doi.org/10.1186/1471-2164-15-396.

Patterson, J., Carpenter, E.J., Zhu, Z., An, D., Liang, X., Geng, C., Drmanac, R., Wong, G. K., 2019. Impact of sequencing depth and technology on de novo RNA-Seq assembly. BMC Genomics 20, 604. https://doi.org/10.1186/s12864-019-5965-x.

Pereira, C.S.A., Aboim, M.A., Ráb, P., Collares-Pereira, M.J., 2014. Introgressive hybridization as a promoter of genome reshuffling in natural homoploid fish hybrids (Cyprinidae, Leuciscinae). Heredity 112, 343–350. https://doi.org/10.1038/hdy.2013.110.

Powell, D.L., Garcia-Olazabal, M., Keegan, M., Reilly, P., Du, K., Diaz-Loyo, A.P., Banerjee, S., Blakkan, D., Reich, D., Andolfatto, P., Rosenthal, G.G., Schartl, M., Schumer, M., 2020. Natural hybridization reveals incompatible alleles that cause melanoma in swordtail fish. Science 368, 731–736. https://doi.org/10.1126/science.aba5216.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575. https://doi.org/10.1086/519795.

Qin, Q., Wang, Y., Wang, J., Dai, J., Xiao, J., Hu, F., Luo, K., Tao, M., Zhang, C., Liu, Y., Liu, S., 2014. The autotetraploid fish derived from hybridization of *Carassius auratus red var.* (female) x *Megalobrama amblycephala* (male). Biol. Reprod. 91, 93. https://doi.org/10.1095/biolreprod.114.122283.

Rapp, R.A., Udall, J.A., Wendel, J.F., 2009a. Genomic expression dominance in allopolyploids. BMC Biol. 7. https://doi.org/10.1186/1741-7007-7-18.

Rapp, R.A., Udall, J.A., Wendel, J.F., 2009b. Genomic expression dominance in allopolyploids. BMC Biol. 7, 18. https://doi.org/10.1186/1741-7007-7-18.

Ren, L., Li, W., Qin, Q., Dai, H., Han, F., Xiao, J., Gao, X., Cui, J., Wu, C., Yan, X., Wang, G., Liu, G., Liu, J., Li, J., Wan, Z., Yang, C., Zhang, C., Tao, M., Wang, J., Luo, K., Wang, S., Hu, F., Zhao, R., Li, X., Liu, M., Zheng, H., Zhou, R., Shu, Y., Wang, Y., Liu, Q., Tang, C., Duan, W., Liu, S., 2019. The subgenomes show asymmetric expression of alleles in hybrid lineages of *Megalobrama amblycephala* x *Culter alburnus*. Genome Res. 29, 1805–1815. https://doi.org/10.1101/gr.249805.119.

Ren, L., Gao, X., Cui, J., Zhang, C., Dai, H., Luo, M., He, S., Qin, Q., Luo, K., Tao, M., Xiao, J., Wang, J., Zhang, H., Zhang, X., Zhou, Y., Wang, J., Zhao, X., Liu, G., Wang, G., Huo, L., Wang, S., Hu, F., Zhao, R., Zhou, R., Wang, Y., Liu, Q., Yan, X., Wu, C., Yang, C., Tang, C., Duan, W., Liu, S., 2022. Symmetric subgenomes and balanced homoeolog expression stabilize the establishment of allopolyploidy in cyprinid fish. BMC Biol. 20. https://doi.org/10.1186/s12915-022-01401-4.

Rio, D.C., Ares, M.J., Hannon, G.J., Nilsen, T.W., 2010. Purification of RNA using TRIzol (TRI reagent). Cold Spring Harb Protoc 2010. https://doi.org/10.1101/pdb.prot5439 pdb.prot5439.

Selz, O.M., Seehausen, O., 2019. Interspecific hybridization can generate functional novelty in cichlid fish. Proceed. Proc. Biol. Sci. 286, 20191621. https://doi.org/10.1098/rspb.2019.1621.

Shabalin, A.A., 2012. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics 28, 1353–1358. https://doi.org/10.1093/bioinformatics/bts163.

Song, G., Guo, Z., Liu, Z., Cheng, Q., Qu, X., Chen, R., Jiang, D., Liu, C., Wang, W., Sun, Y., Zhang, L., Zhu, Y., Yang, D., 2013. Global RNA sequencing reveals that genotype-dependent allele-specific expression contributes to differential expression in rice F_1 hybrids. BMC Plant Biol. 13, 221. https://doi.org/10.1186/1471-2229-13-221.

Springer, N.M., Stupar, R.M., 2007. Allele-specific expression patterns reveal biases and embryo-specific parent-of-origin effects in hybrid maize. Plant Cell 19, 2391–2402. https://doi.org/10.1105/tpc.107.052258.

Srinivasan, K.A., Virdee, S.K., McArthur, A.G., 2020. Strandedness during cDNA synthesis, the stranded parameter in htseq-count and analysis of RNA-Seq data. Brief. Funct. Genom. 19, 339–342. https://doi.org/10.1093/bfgp/elaa010.

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K.V., Altshuler, D., Gabriel, S., DePristo, M.A., 2018. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Curr. Protoc. Bioinformatics 43. https://doi.org/10.1002/0471250953.bi1110s43.

Wang, M., Xu, S., 2019. Statistics of Mendelian segregation—a mixture model. J. Anim. Breed. Genet. 136, 341–350. https://doi.org/10.1111/jbg.12394.

Xiao, J., Kang, X., Xie, L., Qin, Q., He, Z., Hu, F., Zhang, C., Zhao, R., Wang, J., Luo, K., Liu, Y., Liu, S., 2014. The fertility of the hybrid lineage derived from female *Megalobrama amblycephala* x male *Culter alburnus*. Anim. Reprod. Sci. 151, 61–70. https://doi.org/10.1016/j.anireprosci.2014.09.012.

Yang, J., Mezmouk, S., Baumgarten, A., Buckler, E.S., Guill, K.E., McMullen, M.D., Mumm, R.H., Ross-Ibarra, J., 2017. Incomplete dominance of deleterious alleles contributes substantially to trait variation and heterosis in maize. PLoS Genet. 13, e1007019.

Yengo, L., Vedantam, S., Marouli, E., Sidorenko, J., Bartell, E., Sakaue, S., Graff, M., Eliasen, A.U., Jiang, Y., Raghavan, S., et al., 2022. A saturated map of common genetic variants associated with human height. Nature 610, 704–712. https://doi.org/10.1038/s41586-022-05275-y.

Yoo, M.J., Szadkowski, E., Wendel, J.F., 2013. Homoeolog expression bias and expression level dominance in allopolyploid cotton. Heredity (Edinb). 110, 171–180. https://doi.org/10.1038/hdy.2012.94.

Zhong, H., Zhang, X., Xu, Q., Yan, J., Han, Z., Zheng, H., Xiao, J., Tang, Z., Wang, F., Luo, Y., 2019. Nonadditive and asymmetric allelic expression of growth hormone in hybrid tilapia. Front. Genet. 10, 961.

Zhou, Z., Ren, Z., Zeng, H., Yao, B., 2008. Apparent digestibility of various feedstuffs for bluntnose black bream *Megalobrama amblycephala Yih*. Aquac. Nutr. 14, 153–165. https://doi.org/10.1111/j.1365-2095.2007.00515.x.