Genomics, Proteomics & Bioinformatics, 2024, **22(6)**, qzae055
https://doi.org/10.1093/gpbjnl/qzae055
Advance access publication: 23 July 2024
Original Research

Variation and Interaction of Distinct Subgenomes Contribute to Growth Diversity in Intergeneric Hybrid Fish

Li Ren [1,#], Mengxue Luo [1,#], Jialin Cui [1], Xin Gao [1], Hong Zhang [1], Ping Wu [1], Zehong Wei [1], Yakui Tai [1], Mengdan Li [1], Kaikun Luo [1], Shaojun Liu [1,2,*]

[1]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
[2]Guangdong Laboratory for Lingnan Modern Agriculture, South China Agricultural University, Guangzhou 510642, China

*Corresponding author: lsj@hunnu.edu.cn (Liu S).

#Equal contribution.

Handling Editor: Yu Jiang

Abstract

Intergeneric hybridization greatly reshapes regulatory interactions among allelic and non-allelic genes. However, their effects on growth diversity remain poorly understood in animals. In this study, we conducted whole-genome sequencing and RNA sequencing analyses in diverse hybrid varieties resulting from the intergeneric hybridization of goldfish (*Carassius auratus* red var.) and common carp (*Cyprinus carpio*). These hybrid individuals were characterized by distinct mitochondrial genomes and copy number variations. Through a weighted gene correlation network analysis, we identified 3693 genes as candidate growth-regulating genes. Among them, the expression of 3672 genes in subgenome R (originating from goldfish) displayed negative correlations with body weight, whereas 20 genes in subgenome C (originating from common carp) exhibited positive correlations. Notably, we observed intriguing expression patterns of solute carrier family 2 member 12 (*slc2a12*) in subgenome C, showing opposite correlations with body weight that changed with water temperatures, suggesting differential interactions between feeding activity and weight gain in response to seasonal changes for hybrid animals. In 40.30% of alleles, we observed dominant *trans*-regulatory effects in the regulatory interactions between distinct alleles from subgenomes R and C. Integrating analyses of allele-specific expression and DNA methylation data revealed that DNA methylation on both subgenomes shaped the relative contribution of allelic expression to the growth rate. These findings provide novel insights into the interactions of distinct subgenomes that underlie heterosis in growth traits and contribute to a better understanding of multiple allelic traits in animals.

Key words: Intergeneric hybridization; Distinct allelic regulation; Growth diversity; Copy number variation; Mitochondrial regulation.

Introduction

Hybridization and polyploidization could rapidly shape various genotypes and phenotypes, providing us with abundant materials for studying the contribution of genetic regulation to phenotypes [1]. Interspecific hybridization in some plants, including *Triticum aestivum* × *Secale cereal* [2] and *Brassica nigra* × *Brassica rapa* [3], is commonly used to obtain varieties with excellent economic traits. In fish breeding, interspecific hybridization involving different genera or subfamilies has been detected in cyprinid fish [4,5], salmonid fish [6], and cichlid fish [7]. Among these fishes, *Cyprinus carpio* (common carp) and *Carassius auratus* red var. (goldfish) shared a common whole-genome duplication (WGD) event [13.75 million years ago (MYA)] as different genera of the subfamily Cyprinidae and then diverged at 10.0 MYA [8]. The specific WGD resulted in larger genome sizes and more chromosomes (2n = 100) in them than in most other carps [9]. Recent studies have shown that the high genome plasticity and diverse allelic expression shape distinct morphological characters (*e.g.*, body size and color) in some varieties of them, including goldfish and koi carp [10,11]. Meanwhile, these characteristics contribute to their adaptability in the diverse environments of slow-moving rivers, lakes, and ponds [10–13]. Interestingly, a nascent allopolyploid lineage ($4nR_2C_2$, F_3–F_{28}) was successfully established by the hybridization of female goldfish and male common carp and subsequent WGD [5,14]. Gene conversion, accompanied by allopolyploidization and multigenerational inheritance, resulted in the emergence of diverse growth phenotypes in the allopolyploid progenies [15,16].

Body growth, a classic quantitative trait that includes height in humans and weight in domestic animals [17,18], exhibits significant diversity in fishes. Although genetic variations in individual growth-regulating genes may impact the growth phenotype [18,19], the rapid genomic variation induced by hybridization and/or polyploidization is considered the most common and swift way of changing this phenotype in nature. The phenomenon known as "heterosis" occurs when hybrid offspring exhibit faster growth rates and surpass the size of their parents, a trait utilized to enhance agricultural production [20–23]. Researchers have extensively explored the genetic basis of heterosis, proposing three classic quantitative genetic hypotheses: dominance, overdominance, and epistasis. Moreover, some studies suggest that the emergence of heterosis may be linked to various molecular regulatory mechanisms, including genomic recombination [24], novel epigenetic modifications [25], and alterations in gene expression due to *trans*-regulatory factors from distinct species [26]. Additionally, research has indicated that the differential expression of alleles from different species is influenced by sequence differences in their regulatory regions [27]. In the case of orthologous genes from different genera, the greater sequence differences in their regulatory regions, as compared to intraspecific and interspecific hybridization, will influence the reshaping of allele-specific expression (ASE)

Received: 20 February 2023; **Revised:** 27 May 2024; **Accepted:** 18 June 2024.

and its impact on the growth phenotype in their hybrid offspring. This aspect promises to be an intriguing area of research.

To explore the impact of copy number variation (CNV) and mitochondrial regulation on ASE and growth traits, we collected 160 individuals representing six hybrid varieties derived from the intergeneric hybrid lineages of goldfish (2nRR) and common carp (2nCC). The integrated analyses of genomic, DNA methylation, and gene expression data provide a crucial foundation for our research. Our study will expand our understanding of gene interactions and their impact on phenotypes in animals.

Results

Determination of genotypes and growth phenotypes among hybrid varieties

Six hybrid varieties, comprising subgenomes R (originating from goldfish) and C (originating from common carp), were collected from individuals aged 8 months and 24 months after hatching, respectively (**Figure 1A**, Figure S1). To determine the ploidy levels of all hybrid individuals, flow cytometry was employed, using diploid goldfish (2n = 100) as the control group. To identify the source of mitochondrial genomes in reciprocal F_1 hybrids and allotriploid individuals with the same ploidy level, a fragment of the cytochrome b (*cytb*) gene in these hybrid individuals was obtained using Sanger sequencing. Subsequently, we compared the obtained sequences with those of goldfish and common carp to determine the origin of their mitochondrial genomes. Moreover, both whole-genome sequencing (WGS) and RNA sequencing (RNA-seq) data analyses were conducted to validate the genotypes of the six hybrid varieties. Based on the genomic data, the average depth of mapped reads of subgenomes R *vs.* C was approximately 1:1 in 2nCR, 1:2 in $3nRC_2$ and $3nC_2R$, and 2:1 in $3nCR_2$ and $3nR_2C$ (Tables S1–S3). Interestingly, we found that fish with identical subgenome ratios showed consistent distributions in the expression values of alleles R *vs.* C, as observed in the RNA-seq data (Figure 1B). Three clusters of gene expression profiles (cluster 1: 2nRC and 2nCR; cluster 2: $3nR_2C$ and $3nCR_2$; cluster 3: $3nC_2R$ and $3nRC_2$) remained stable across individuals (10–42 individuals in each variety) (Figure 1B). Similarly, analysis of the RNA-seq data clearly identified mitochondrial types (originating from goldfish or common carp) based on the different numbers of reads mapped to the two mitochondrial genomes, respectively (Tables S4 and S5). The aforementioned results can assist us in the initial identification of the genotypes among the hybrid varieties and provide insights into their breeding strategies (Figure 1A).

Analysis of body weight at 24 months after hatching showed that it was significantly higher in 2nCR compared to 2nRC (*t*-test, $P = 2.08 \times 10^{-24}$; two-tailed *t*-test, $t = 2.02$ and df = 42) (Figure 1C). Meanwhile, the allotriploid $3nC_2R$ and $3nRC_2$ (subgenomes R *vs.* C = 1:2) displayed higher body weight compared to $3nCR_2$ and $3nR_2C$ (subgenomes R *vs.* C = 2:1) (*t*-test, $P = 4.13 \times 10^{-12}$; two-tailed *t*-test, $t = 1.98$ and df = 56) (Figure 1C). Furthermore, a higher body weight was detected in $3nRC_2$ (mitochondrial genomes originating from common carp) than in $3nC_2R$ (mitochondrial genomes originating from goldfish) (*t*-test, $P = 0.0001$; two-tailed *t*-test, $t = 2.01$ and df = 46) (Figure 1C), while there is the same nuclear genome in them (Figure 1A). These findings indicate that more sets of subgenome

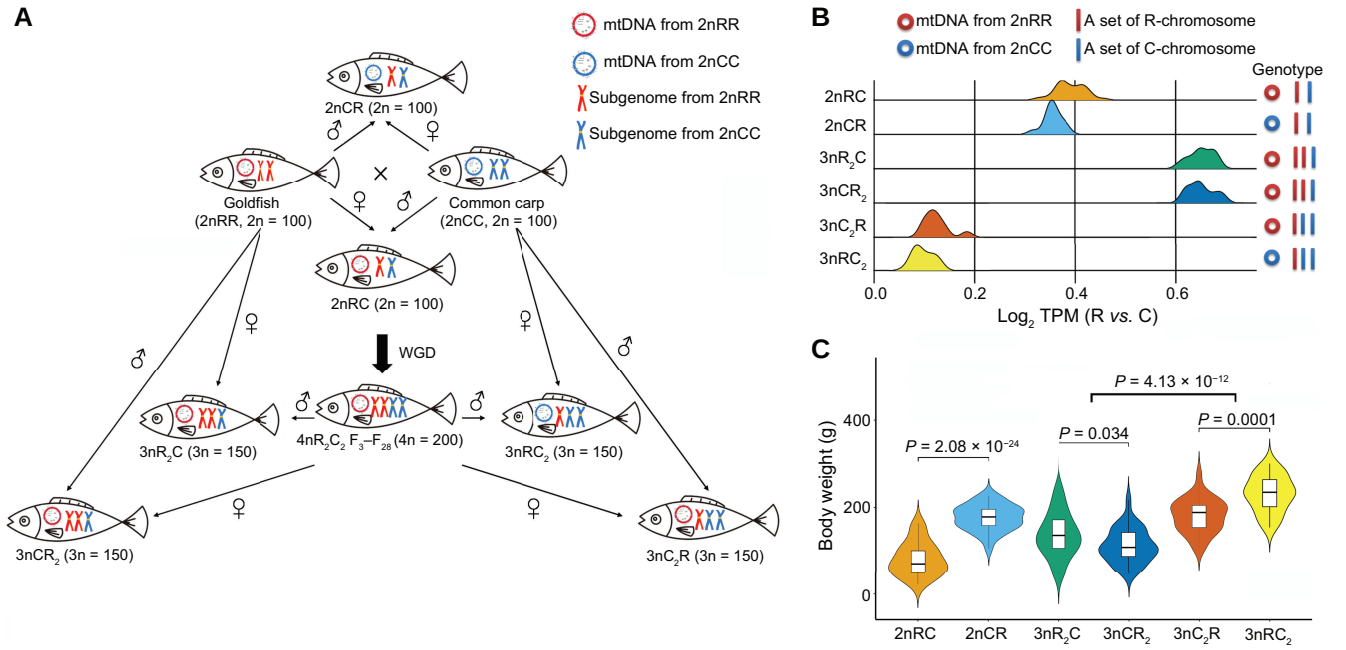


Figure 1 Generation process, genotypes, and growth phenotypes in the six hybrid varieties derived from the intergeneric hybridization of goldfish and common carp

A. The controlled genotypes of hybrid varieties obtained from the intergeneric hybridization of goldfish (*Carassius auratus* red var.) and common carp (*Cyprinus carpio*), followed by polyploidization and interploid hybridization. **B**. Genotypes of the two interspecific hybrids (2nRC and 2nCR) and the four triploid varieties ($3nR_2C$, $3nCR_2$, $3nRC_2$, and $3nC_2R$) predicted based on the mapped reads of transcriptomes. **C**. Body weight in the six hybrid varieties (24 months after hatching). mtDNA, mitochondrial DNA; WGD, whole-genome duplication; TPM, transcripts per million.

C and mitochondrial genomes originating from 2nCC contribute to the large body size among these hybrid varieties.

High CNVs in subgenome R

A previous study demonstrated that gene conversion occurred in the assembled genome of allotetraploid fish [15]. The diverse CNVs between somatic and germ cells were found in the interspecific F_1 and allotetraploid populations, revealing various mitotic and meiotic CNVs associated with gene conversion [5,15]. To investigate the presence of different CNVs in allotriploid progenies, obtained from the backcrossing of allotetraploid to diploid goldfish or common carp, we conducted WGS on 7 individuals of the F_1 hybrids and 22 individuals of triploids (8 months after hatching) (Tables S1 and S2). CNV detection for 1545–3017 genes in the 2nRC and triploid varieties revealed a higher CNV ratio in subgenome R (49.25%–83.33%) than in subgenome C (16.67%–50.75%) (*t*-test, $P = 4.02 \times 10^{-21}$; two-tailed *t*-test, $t = 2.00$ and df = 56), indicating a leading role of subgenome R in the genome plasticity of hybrid varieties (**Figure 2A**; Table S6). Focusing on 46,424 species-specific genes (SSGs) (24,283 in 2nCC and 22,141 in 2nRR) and 18,020 allelic gene pairs (AGPs), the majority of genes with CNVs (71.55%–85.55%) belonged to SSGs, and exhibited a higher ratio of genome variation in SSGs than in AGPs (*t*-test, $P = 1.80 \times 10^{-53}$; two-tailed *t*-test, $t = 2.00$ and df = 56) (Tables S6 and S7).

We investigated the distribution of CNVs in allelic genes. We observed a higher frequency of copy number increase events in allele R (52.65%–74.43%) than in allele C (25.57%–47.35%) (*t*-test, $P = 1.04 \times 10^{-43}$; two-tailed *t*-test, $t = 2.05$ and df = 28) (Table S7). In some individuals, CNV events occurred on long chromosomal segments involving contiguous genes or entire chromosomes (Figure S2). For example, CNV events occurred on the entire Chr40 in the $3nR_2C$-4 individual. In the $3nRC_2$-2 individual, CNV events occurred on the parts of Chr19 (92 genes) and Chr36 (63 genes), as well as on the entire Chr25 and Chr31. These CNV events resulted in a 1:1 ratio of gene copy numbers from the inbred parents in these allotriploid individuals (Figure 2B, Figure S2).

Furthermore, WGS data revealed allele loss caused by CNVs. Among the 22 allotriploid individuals, the number of allele loss events ranged from 2 to 223, while none were observed in the F_1 hybrids (Table S7). For instance, the loss of allele C in contiguous genes on Chr19 was observed in two individuals of $3nR_2C$ and one individual of $3nCR_2$ (Figure 2B, Figure S2). We speculated that the shared allele loss observed in different triploid progenies may be derived from the gametes of the same paternal $4nR_2C_2$ individual through interploid hybridization. Interestingly, the loss of

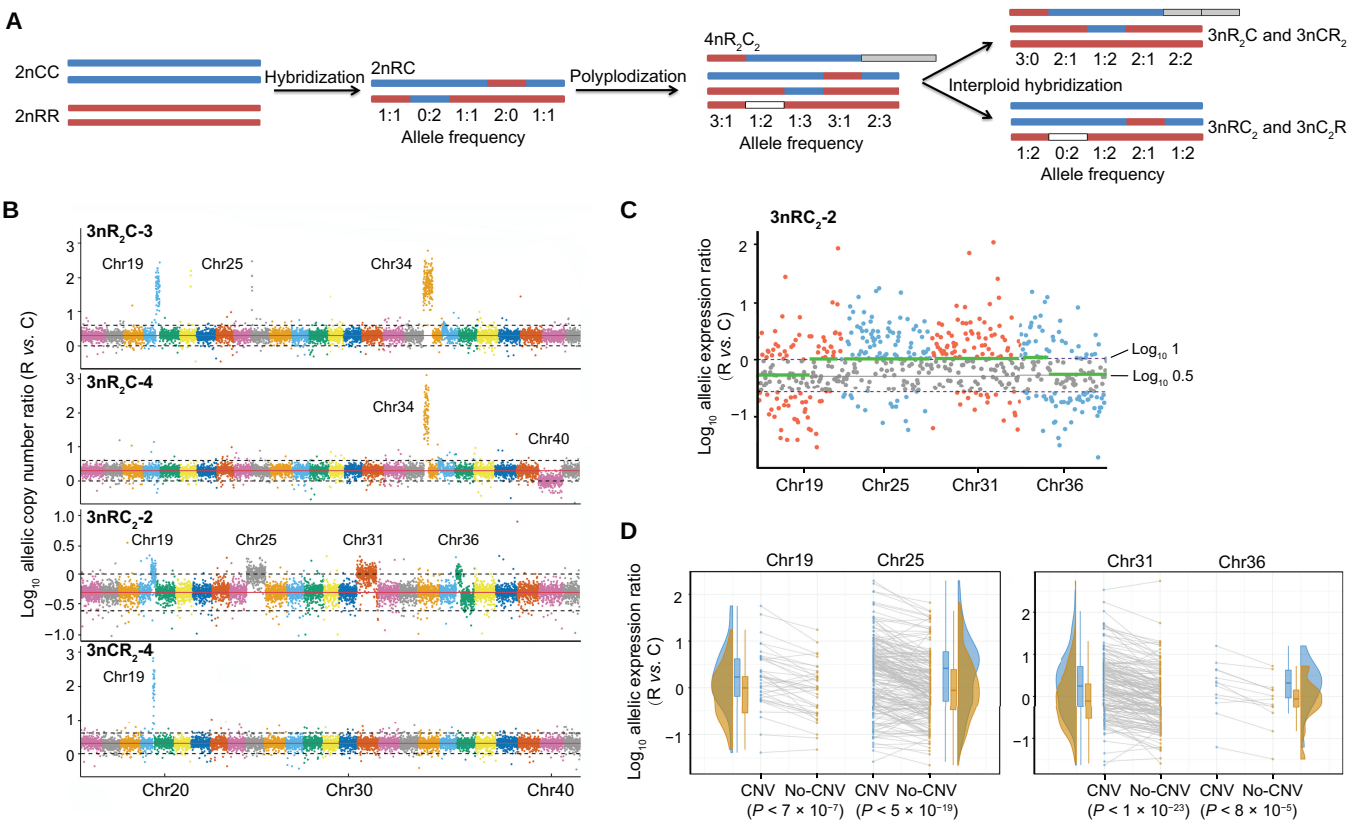


Figure 2 CNVs regulate allelic expression changes in hybrid varieties

A. Schematic diagram of CNVs accompanied by hybridization, polyploidization, and interploid hybridization. The white block represents the deletion of allele R or C, while the gray box represents the duplication of allele R or C. **B**. Allelic CNVs revealed by the ratio changes of alleles R *vs.* C. The allelic CNVs in the contiguous genes of chromosomes in four triploid individuals are shown. The loss of allele C was observed in the 92 contiguous genes on Chr19. The 1:1 ratio of allelic copy numbers was observed on Chr40 in $3nR_2C$-4 and on Chr19, Chr36, Chr25, and Chr31 in $3nRC_2$-2. In $3nR_2C$ and $3nCR_2$, the red line corresponds to $\log_{10}$ 2, while in $3nRC_2$, the red line corresponds to $\log_{10}$ 0.5. Dotted lines represent the calibration lines, which are used to determine whether CNV occurs. **C**. The changes in the expression ratio of alleles R *vs.* C on Chr19, Chr25, Chr31, and Chr36 accompanied by CNVs in $3nRC_2$-2. The green lines represent the average gene expression values in the no-CNV and CNV regions. **D**. Individuals with or without CNVs had significantly different expression ratios of alleles R *vs.* C on Chr19, Chr25, Chr31, and Chr36 in the $3nRC_2$ population (two-sided *t*-test analysis). CNV, copy number variation.

allele C was detected in the 92 contiguous genes of $3nCR_2$-4, which also exhibited the highest body weight among the $3nCR_2$ population. Gene Ontology (GO) analysis of these 92 genes identified *rfx3* and *gpat4* as being annotated to epithelial cell maturation (GO:0002071, $P = 0.001$). We speculated that the loss of allele C in these two genes may contribute to the high body weight observed in the allotriploid.

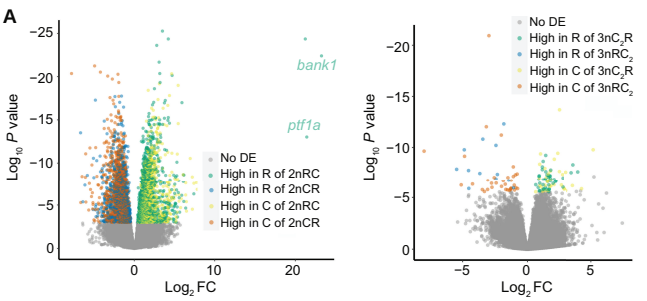
CNVs alter ASE

To investigate the impact of CNVs on allelic expression, we conducted integrated genomic and expression analyses using muscle tissue samples from 7 individuals of the F_1 hybrids and 22 individuals of the allotriploids (Table S2). Specifically, we focused on the loss of allele C in $3nR_2C$ and $3nCR_2$, by comparing the gene expression values between individuals without CNVs (no-CNV; no allelic loss) and those with CNVs (allelic loss) within the corresponding population. In the $3nRC_2$-2 individual, where CNVs led to a 1:1 allelic copy number ratio (Figure 2B), genes within the CNV regions also displayed a 1:1 allelic expression ratio (Figure 2C). Conversely, genes outside the CNV regions exhibited a 1:2 allelic expression ratio (Figure 2C). Then, differential expression analysis was performed between CNV and no-CNV individuals in the $3nRC_2$ population, and significant expression differences were detected on Chr19, Chr25, Chr31, and Chr36 (Figure 2D). These findings suggest that CNVs in the hybrids could affect gene expression, highlighting the effects of CNVs on allelic expression.

Mitochondrial regulation shapes ASE and species-specific expression

Distinct mitochondrial genomes and the same nuclear genome were in two groups (diploid group 1: 2nRC and 2nCR; triploid group 2: $3nRC_2$ and $3nC_2R$). This provided valuable insights into mitochondrial genetics and its role in regulating growth diversity through gene expression (Tables S1, S2, and S8). Comparative analyses revealed that 6516 genes (7.9%) of the total 82,464 genes (alleles R and C considered independent genes) differed in expression between 2nRC and 2nCR (13 individuals), while only 126 genes (0.15%) showed differential expression between $3nRC_2$ and $3nC_2R$ (10 individuals). The lower number of differentially expressed genes (DEGs) in triploids compared to diploids may be caused by the high frequency of CNVs which disrupted ASE regulated by mitochondrial genetics (**Figure 3A**).

In the diploid group, the analysis of differential gene expression revealed that in subgenome R, AGPs had more DEGs compared to SSGs (1834 *vs.* 1440), whereas in subgenome C, SSGs had more DEGs than AGPs (1734 *vs.* 1508) (Pearson's chi-squared test, $P = 2.04 \times 10^{-14}$, $\chi^2 = 58.497$, and df = 1) (Table S9). This suggests that orthologous genes from goldfish and SSGs from common carp are more susceptible to regulation by maternal effects, resulting in differential expression between reciprocal F_1 hybrids. Furthermore, we detected that highly expressed genes in 2nRC were predominantly in allele R (1881) rather than in allele C (1545), while the opposite trend was observed in 2nCR, where highly expressed genes were more in allele C (1697) than in allele R (1393) (Figure 3A). This finding suggests that mitochondria in hybrids may preferentially up-regulate the expression of nuclear genes from the same species. Interestingly, we noticed that allele R of *ptf1a* was silenced in 2nCR, whereas its



B

TPM in R TPM in C	Diploid	Triploid	TPM in R TPM in C	Diploid	Triploid
I	5	0	VI	507	13
II	479	5	VII	10	0
III	261	0	VIII	767	23
IV	414	11	IX	311	0
V	15,265	17,962			

Figure 3 Mitochondrial genetics regulates ASE in reciprocal hybrids
A. Differential expression analyses were performed in the diploid (2nRC and 2nCR) and triploid ($3nC_2R$ and $3nRC_2$) groups. "No DE" represents no differential expression (gray). "High in R of 2nRC" represents the genes showing higher expression in allele R than in allele C in 2nRC (1881 genes, green). There are also "High in R of 2nCR" (1393 genes, blue), "High in C of 2nRC" (1545 genes, yellow), and "High in C of 2nCR" (1697 genes, orange). The silencing of allele R was detected in the two genes (*bank1* and *ptf1a*) of 2nCR, while gene expression was observed in allele R of 2nRC. **B**. Schematic diagram of nine allelic expression patterns regulated by mitochondrial genetics. Blue dots represent the gene expression values in 2nCR and $3nRC_2$ (both mtDNA originating from 2nCC), while red dots represent the gene expression values in 2nRC and $3nC_2R$ (both mtDNA originating from 2nRR). "TPM in R" represents the gene expression value in allele R, and "TPM in C" represents the gene expression value in allele C. ASE, allele-specific expression; FC, fold change.

expression was detected in 2nRC, indicating that the mitochondrial genes originating from 2nCC inhibit the expression of allele R (Figure 3A). The significant diversity in allelic gene expression between 2nRC and 2nCR, resulting from the silencing of allele R of *ptf1a*, may be linked to their distinct growth rates [28].

To further investigate the magnitude of independence and interaction between alleles R and C, we established nine expression patterns for alleles R and C (Figure 3B). Out of these, 2754 genes (15.28%) in the diploid group (2nRC and 2nCR) and 52 genes (0.28%) in the triploid group exhibited expression changes regulated by distinct mitochondrial genetics. In the diploid group, we found that the expression of either allele R or C was altered in specific patterns (II, IV, VI, and VIII) across 2167 genes (12.02%). These patterns indicate the existence of independent regulatory networks connecting mitochondrial genes to the expression of allele R or C (Figure 3B, Figure S3). Additionally, a shared regulatory network in mitochondrial genetics regulating both alleles R and C (patterns III and IX) was observed in 572 genes (3.17%) (Figure 3B, Figure S3). Interestingly, we observed an opposite trend between the expression of alleles R and C in 15 genes (0.08%; patterns I and VII), which was likely related to an antagonistic relationship between the regulatory networks of the distinct alleles. These results showed the diversified regulatory

networks in mitochondrial genetics that regulate allelic gene expression.

Gene coexpression analyses reveal the genetic networks underlying growth rate diversity

The aforementioned results demonstrate the diversified ASE and species-specific expression in the six hybrid varieties. We then performed a weighted gene correlation network analysis (WGCNA) to detect gene coexpression networks and genetic modules in the 131 individuals (24 months after hatching). Twelve modules were identified based on the correlation of the expression profiles of the 65,495 expressed genes (31,974 genes in subgenome R and 33,521 genes in subgenome C) (Figure S4A). Among these modules, Module Eigengene 1 (ME01) exhibited the highest correlation with growth-related phenotypic values, including body weight and body length, while a significant correlation was detected between gene significance and module membership [Pearson correlation coefficient (PCC) = 0.81, $P = 1 \times 10^{-20}$] (Figure S4B). After gene filtering, we identified 3693 genes within ME01 as candidate growth-regulating genes. Within this module, the coexpression network revealed that 3672 genes in subgenome R and 21 genes in subgenome C showed significant correlations between gene expression and body weight (Figure 4A). These results indicate that CNVs in subgenome R, along with changes in gene expression within subgenome R, are primarily responsible for regulating growth diversity in these hybrid varieties.

To further identify the relationship between the expression of the 3693 growth-regulating genes and body weight, we analyzed the gene expression diversity and body weight diversity across individuals. The highest diversities of gene expression and body weight were observed in $3nR_2C$, while the lowest diversities were observed in $3nC_2R$ (Figure 4B). Importantly, both $3nR_2C$ and $3nCR_2$ (with two sets of subgenome R) exhibited higher diversities in gene expression and body weight compared to $3nRC_2$ and $3nC_2R$ (with one set of subgenome R). These findings, coupled with the enrichment of CNVs in subgenome R, indicate that the high CNV ratio in subgenome R contributes to the increased diversity of allele R expression, which subsequently leads to the diversification of growth phenotypes in the hybrid population (Figure 2; Table S6). In the diploid group, higher gene expression diversity was detected in 2nRC than in 2nCR (Figure 4B), suggesting that the maternal effects related to goldfish-originated regulation may be more beneficial to the CNV of allele R and result in gene expression diversity across individuals.

Among the 3693 growth-regulating genes, we found that the expression of 3672 genes in subgenome R and one gene

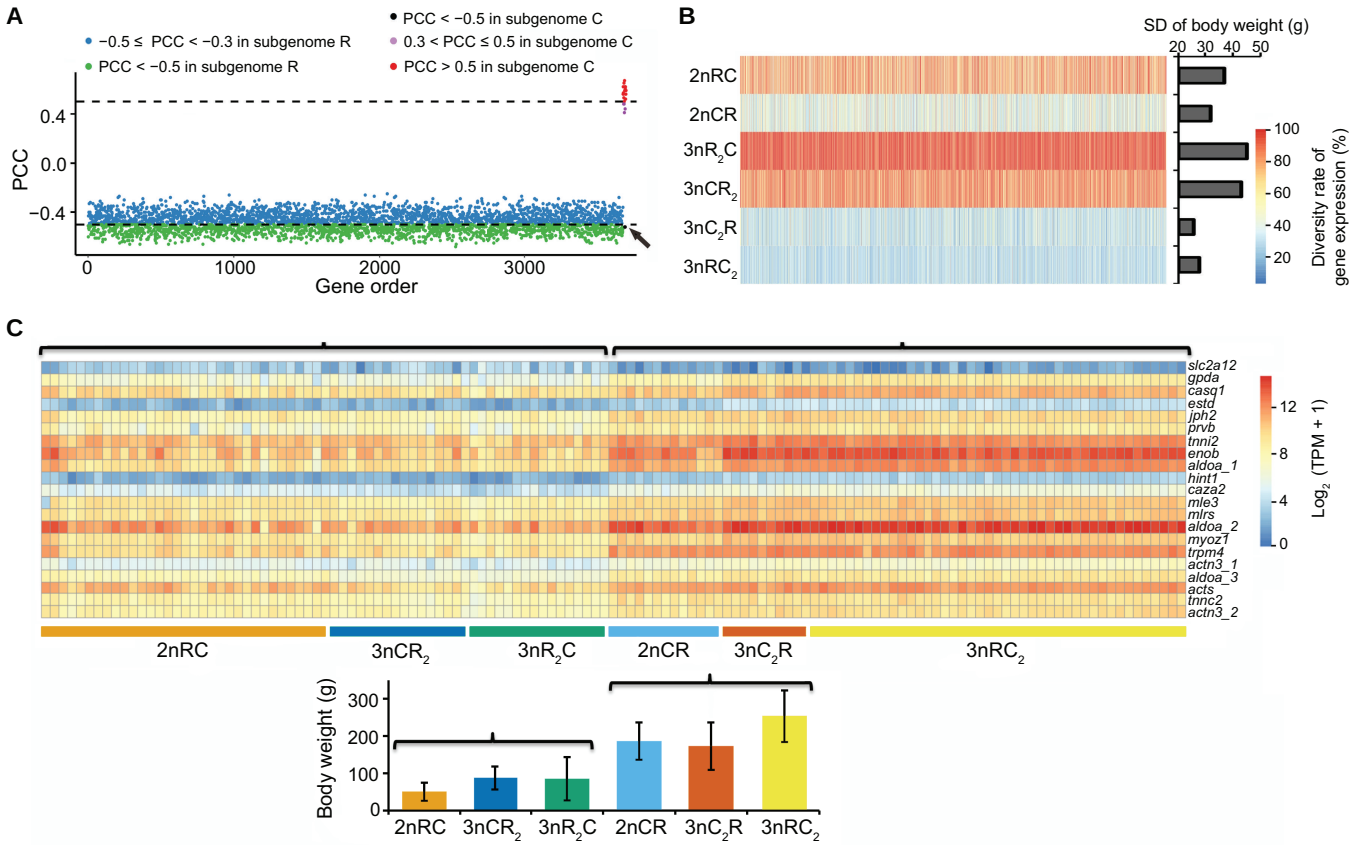


Figure 4 Expression of 3693 growth-regulating genes alters body weight in hybrid varieties

A. Pearson correlation analysis between gene expression and body weight (24 months after hatching) in 3693 predicted growth-regulating genes. PCC < −0.5 indicates strong negative correlations including 1640 genes in subgenome R (green dots) and one gene in subgenome C (*slc2a12*, black dot pointed by the arrow); −0.5 ≤ PCC < −0.3 indicates negative correlations including 2032 genes in subgenome R (blue dots); 0.3 < PCC ≤ 0.5 indicates positive correlations including 3 genes in subgenome C (pink dots); and PCC > 0.5 indicates strong positive correlations including 17 genes in subgenome C (red dots). **B**. Diversities in gene expression and body weight across the individuals of each variety. The SD of GPM values in the 3693 growth-regulating gene was performed to assess the expression diversity across individuals. The SD of body weights was calculated to assess the growth diversity across individuals. **C**. A heatmap exhibiting the expression of the 21 growth-regulating genes in subgenome C across individuals. Two groups were classified into the six hybrid varieties based on the clustering of body weight and gene expression. PCC, Pearson correlation coefficient; GPM, gene per million; SD, standard deviation; *slc2a12*, solute carrier family 2 member 12.

[solute carrier family 2 member 12 (*slc2a12*)] in subgenome C exhibited negative correlations with body weight (PCC < −0.3), while the other 20 genes in subgenome C showed positive correlations (PCC > 0.3) (Figure 4A). Additionally, the expression of these 20 genes was found to be higher in the 2nCR, $3nRC_2$, and $3nC_2R$ populations (Group 1) compared to that in the 2nRC, $3nR_2C$, and $3nCR_2$ populations (Group 2) (Figure 4C). We compared gene expression data from 24-month-old individuals (n = 131) with data from 8-month-old individuals (n = 29). Genes exhibiting strong positive correlations (PCC > 0.5) between body weight and gene expression in the larger dataset also showed strong positive correlations in the smaller dataset, with the exception of the *prvb* gene (PCC = 0.37, $P = 0.05$) (Figure S5). Interestingly, the expression of *slc2a12* in subgenome C exhibited a positive correlation with body weight at low water temperature (about 8°C, 8 months after hatching) (PCC = 0.67, $P = 0.003$) (Figure S5), while a negative correlation was detected at high water temperature (about 20°C, 24 months after hatching) (Figure 4A and C). The opposite correlations were likely related to the different strategies for rapid growth rate of these hybrid individuals in alternation with the seasons, in which the up-regulated expression of *slc2a12* in subgenome C could decrease the amount of exercise and energy consumption at low temperatures (winter), while the down-regulated expression could increase the amount of exercise for obtaining food at high temperatures (spring) [29].

Variations in gene regulatory networks and their effects on growth rate

The decreased expression of genes in subgenome R and the increased expression of genes in subgenome C both contributed to the rapid growth rate, prompting the question of how variations in gene regulatory networks regulated allelic expression and growth diversity. Among the 3693 growth-regulating genes detected in the 131 individuals of six hybrid varieties, 2094 were orthologous genes (*i.e.*, AGPs), while 1587 were SSGs in subgenome R and 12 were SSGs in subgenome C (Figure S6). The average PCC between the expression levels of alleles R and C was higher in the 2094 AGPs than that in the non-AGPs (ANOVA *F*-test, $F = 0.85$, df = 2081, $P = 0.000123$) (**Figure 5A**; see Materials and methods for details). This finding suggests that shared *trans*-regulatory factors between alleles R and C homogenize the expression levels of the two orthologous genes originating from goldfish and common carp, leading to increased synchrony in the allelic expression in hybrids. Among the 2094 AGPs, 832 (39.73%) displayed positive correlations (PCC > 0.3), with 304 (14.52%) exhibiting strong positive correlations (PCC > 0.5). In contrast, only 12 AGPs (0.57%) displayed negative correlations (PCC < −0.3), and no AGPs exhibited strong negative correlations (PCC < −0.5). These results reflect synchrony of allelic expression in 40.30% of AGPs and independence in 59.70% of AGPs. Furthermore, we analyzed nine AGPs where the expression of allele C exhibited a strong positive correlation (PCC > 0.5) with body weight (Figure 5B). Among these nine AGPs, six also showed strong positive correlations (PCC > 0.5) between the expression levels of alleles R and C (Figure 5B). Interestingly, higher PCC values of two alleles and lower body weights were observed in 2nRC than in 2nCR, and the same pattern was detected in $3nR_2C$ and $3nCR_2$ than in $3nRC_2$ and $3nC_2R$ (Figure 5C). Among hybrids with the same ploidy level, more diversified gene regulatory networks between alleles of growth-regulating genes may be associated with faster growth. This result sheds light on why heterosis often appears in hybrid F_1 generations, and with successive generations of hybrid offspring, the heterosis diminishes or decreases [20–24]. This phenomenon might be attributed to the presence of complete *trans*-regulatory factors from different species only in F_1 hybrids, leading to the maximization of differences in the gene regulatory network between allelic genes.

The PCC values between expression levels of alleles could also allow us to assess the variation of gene regulatory networks among the six hybrid varieties. The highest difference was observed between the PCC values of $3nR_2C$ and $3nRC_2$ (*t*-test, $P = 6.9 \times 10^{-152}$; two-tailed *t*-test, $t = 1.96$ and df = 2084) (Figure 5D), which had distinct mitochondrial genomes and different subgenome ratios of R *vs.* C (Figure 1A). We further investigated the differences in gene regulatory networks influenced by mitochondrial genetics in the two reciprocal F_1 hybrids (*t*-test, $P = 2.66 \times 10^{-6}$; two-tailed *t*-test, $t = 1.96$ and df = 2084) (Figure 5D), and reversed PCC values were observed in 310 growth-regulating AGPs in the two reciprocal F_1 hybrids (Table S10). Further analysis of AGPs with strong positive correlations (PCC > 0.5) between the expression levels of alleles R and C revealed that only 194 AGPs (10.35% of 1874 AGPs) were shared among the six hybrid varieties, while 438 common AGPs were observed between the two F_1 hybrids and 362 common AGPs were identified among the four triploid varieties (Figure 5E, Figure S7). Further, 2nCR and $3nRC_2$ exhibited similar PCC values in allelic gene expression (*t*-test, $P = 0.80$; two-tailed *t*-test, $t = 1.96$ and df = 2084), raising the question of why these hybrids with different ploidy levels exhibited similar gene regulatory networks. A recent study has revealed that DNA methylation inhibits the transcription of genes in the additional subgenome C of $3nRC_2$ (**Figure 6A**) [30], which may account for the similar gene regulatory networks between 2nCR and $3nRC_2$. The aforementioned results reveal a high divergence of gene regulatory networks between alleles in these hybrid varieties.

Allele-specific DNA methylation regulates growth rate diversity

Through analyzing the DNA methylation in the 3693 growth-regulating genes, we investigated whether and how DNA methylation affects growth rates via regulating ASE. We first analyzed the whole-genome bisulfite sequencing data (the muscles of 2nRR, 2nCC, $3nR_2C$, and $3nRC_2$) and growth phenotypic data (a significant difference in body weight between $3nR_2C$ and $3nRC_2$) to investigate the contribution of DNA methylation to subgenomes R and C [30]. After mapping clean reads to the combined genome of 2nRR and 2nCC, the uniquely mapped reads were used to assess the methylation levels in CpG islands. Comparative analysis revealed that the DNA methylation level in each subgenome was higher in the two triploids ($3nR_2C$ and $3nRC_2$) than in their inbred parents (2nRR and 2nCC) (Figure 6A), indicating that high DNA methylation suppresses gene expression in triploids.

We focused on the 2094 AGPs that were predicted to be involved in growth regulation, and compared the DNA methylation levels of the alleles R and C in $3nR_2C$ and $3nRC_2$. We found that allele R showed lower DNA methylation and higher expression in $3nR_2C$ than in $3nRC_2$ (Figure 6B and

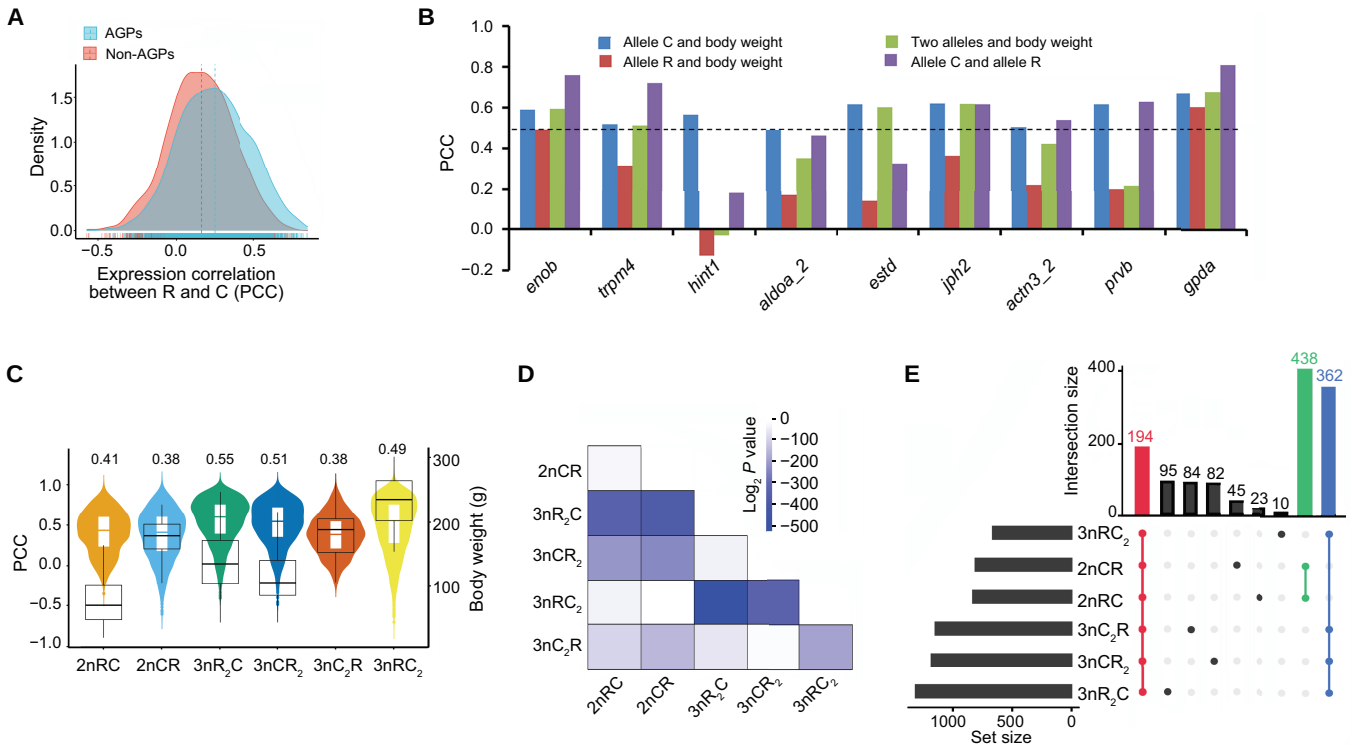


Figure 5 Comprehensive analysis of gene expression correlations and their impact on body weight across hybrid varieties
A. Expression correlation analyses between alleles R and C in the 2094 growth-regulating AGPs as well as in the non-AGPs. The selection of non-AGPs was detailed in the Materials and methods section. Red and cyan dotted lines represent average PCC values. **B**. Correlation analyses of 9 AGPs belonging to the 21 growth-regulating genes in subgenome C. **C**. The distributions of PCC values between expression of alleles R and C (violin plot) and body weights (box plot) in the six hybrids. The median PCC for each hybrid is indicated on the top. **D**. Effects of gene regulatory networks between two hybrid varieties detected by two-sided *t*-test of PCC values. **E**. The distribution of AGPs with PCC values between the expression levels of alleles R and C more than 0.5 among six hybrid varieties. AGP, allelic gene pair.

C). In contrast, allele C showed higher DNA methylation and lower expression in $3nR_2C$ than in $3nRC_2$ (Figure 6B and C). Comparing $3nR_2C$ and $3nRC_2$, we identified 84 differentially methylated genes (DMGs) (4.06%) from AGPs and 57 DMGs (3.56%) from SSGs (Table S11). The majority of DMGs (AGPs: 73 of 84; SSGs: 46 of 57) had reduced DNA methylation in $3nR_2C$ (Table S11), indicating that the lower DNA methylation in subgenome R contributes to the higher expression of allele R in $3nR_2C$ than in $3nRC_2$. Indeed, hematoxylin and eosins (H&E) staining of skeletal muscle tissues uncovered that the cross-sectional area of myofibers was larger in 2nCC than in 2nRR, which was likely related to the difference in skeletal muscle development between the two species (*t*-test, $P = 3.18 \times 10^{-16}$; two-tailed *t*-test, $t = 2.00$ and $df = 56$) (Figure 6D). We detected a larger cross-sectional area of myofibers in $3nRC_2$ than in $3nR_2C$, suggesting that the high expression of allele C benefits the growth of myofibers (*t*-test, $P = 1.40 \times 10^{-5}$; two-tailed *t*-test, $t = 2.00$ and $df = 56$) (Figure 6D). In conclusion, our findings indicate that the regulation of DNA methylation may contribute to the differential activity of ASE and growth rates between $3nRC_2$ and $3nR_2C$, possibly explaining the variations in their growth rates.

Discussion

Using different hybrid strategies that produce the controlled genotypes of hybrids, we characterized the mitochondrial and nuclear genetic variants associated with variations in gene expression and growth diversity. We detected the CNVs of subgenomes R and C across hybrid individuals, including allele loss in triploids, and found their effects on ASE and species-specific expression. By applying WGCNA, we identified 3693 genes (including 2094 AGPs and 1599 SSGs) as candidate growth-regulating genes, whose expression exhibited significant correlations with growth diversity. Using correlation analyses between the expression of distinct alleles R and C, we detected the different degrees of independence and interactions in the allelic regulatory networks, which reflected variations of gene regulatory networks among different hybrid varieties. Importantly, we found that the diversified gene regulatory networks of distinct alleles R and C in growth-regulating genes may contribute to the rapid growth rate. This result suggests that maintaining and increasing differentiation in allelic expression will result in the emergence of heterosis and be beneficial for aquaculture and animal breeding. In addition, we revealed that DNA methylation shaped allelic expression variations in subgenomes R and C.

Following their divergence 10 MYA, high genome plasticity in goldfish and common carp facilitated diverse growth phenotypes during domestication [8,31,32]. Now, we detected diverse CNVs in the nascent allotetraploid population ($4nR_2C_2$, F_3–F_{28}) [15,33], which were derived from the hybridization of goldfish and common carp [5,14], and provided abundant genetic diversities in their allotriploid progenies through different interploid crossings with their inbred parents [16]. Meanwhile, the random emergence of repair DNA damage during mitosis also resulted in CNVs, including allele loss in triploid individuals. Joint analyses of the CNV region, growth rate, and functional annotation data could

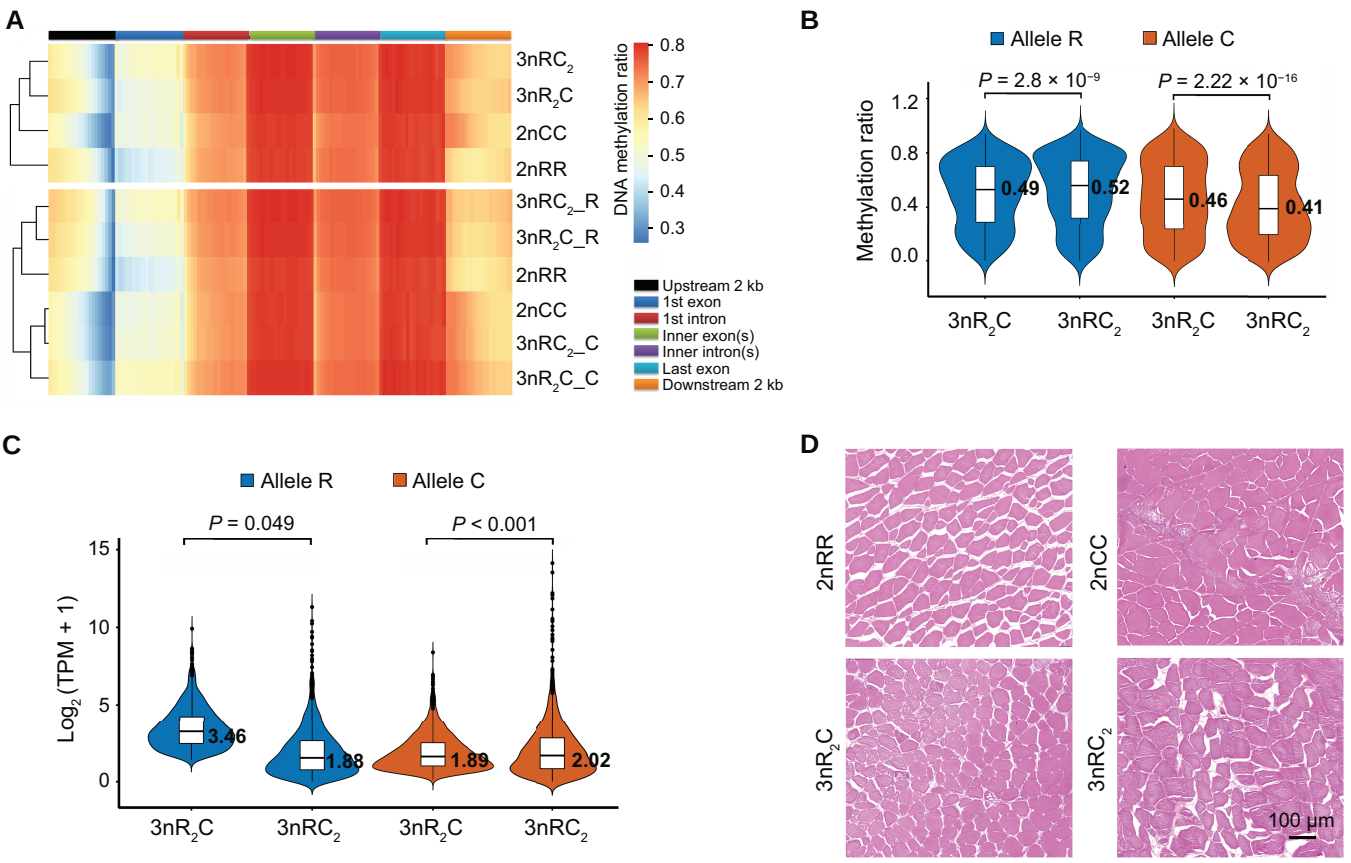


Figure 6 DNA methylation, ASE, and histological observation between hybrids with different body weights
A. Heatmap showing the DNA methylation patterns of different gene regions in the two triploids ($3nR_2C$ and $3nRC_2$) and their inbred diploid parents (2nRR and 2nCC). Each region was divided into 20 bins based on its total length. "$3nR_2C$" and "$3nRC_2$" represent the combined methylation ratios of subgenomes R and C in $3nR_2C$ and $3nRC_2$, respectively. "$3nR_2C$-R" and "$3nRC_2$-R" represent the methylation ratios in the subgenome R of $3nR_2C$ and $3nRC_2$, respectively. "$3nR_2C$-C" and "$3nRC_2$-C" represent the methylation ratios in the subgenome C of $3nR_2C$ and $3nRC_2$, respectively. **B**. Violin plot showing the DNA methylation levels of alleles R and C in the 2094 growth-regulating AGPs. **C**. Violin plot showing the expression levels of alleles R and C in the 2094 growth-regulating AGPs. **D**. Cross-section of skeletal muscles (H&E staining) showing the myofibers of the two triploids ($3nR_2C$ and $3nRC_2$) and their inbred diploid parents (2nRR and 2nCC) (n = 3 biologically independent samples). Scale bar, 100 μm. H&E, hematoxylin and eosins.

provide an effective way to identify causal genes associated with the growth variation in inter- and intra-hybrid populations. For example, we showed that the loss of allele C in *rfx3* and *gpat4* might contribute to the high growth rate. The majority of CNVs in the hybrid varieties were distributed in subgenome R, where dynamic transposition of transposable elements could result in CNVs and allelic expression variation and increase growth diversity [15].

Diverse subgenome ratios, mitochondrial genetics, and CNVs provide abundant genetic materials for investigating the relative contribution of gene regulatory networks to the variations in ASE and species-specific expression. When exposed to the common *trans*-acting regulatory factors, the distinct *cis*-regulatory elements in allelic genes may cause a target gene to interact or bind differentially with the transcription factors, thus resulting in differential expression between alleles [34,35]. We showed that the dominance of *trans*-acting regulatory effects decreased the expression diversity of alleles in most genes, while *cis*-regulatory regulatory effects increased it in a few genes [36,37]. Different *trans*-acting influences involving mitochondrial regulation in the reciprocal F_1 hybrids reflect differential expression of SSGs primarily in subgenome C, which play key roles in growth diversity. Additionally, the distinct subgenome ratios in allotriploids also shape allelic CNVs and alter the dosage of the *trans*-acting factors between alleles, although the diverse effects of DNA methylation occur in different subgenomes [26,38]. Our findings indicate that dynamic changes in variations of gene regulatory networks increase the magnitude of independence and interactions in allelic regulatory networks, resulting in a great increase in allelic expression variation and growth diversity in these intergeneric hybrid varieties.

A previous study indicated that the expression dominance in allele R was beneficial to high body height in allotriploids, while the expression dominance in allele C was beneficial to high body length [30]. We showed that the decreased expression of genes in subgenome R and the increased expression of genes in subgenome C contributed to the higher body weight. These findings suggest that the appropriate combination of ASE may contribute to the heterosis in quantitative traits. Interestingly, *slc2a12* belongs to a family of transporters that catalyze the uptake of sugars through facilitating diffusion [39]. The increased expression of *slc2a12* in winter could decrease the amount of exercise and energy consumption in low temperatures, while the decreased expression of *slc2a12* in spring could increase the amount of exercise needed to obtain

food [29]. These different strategies in alternation with the seasons will increase the growth rates of these hybrid individuals. Our results provide new viewpoints: the great diversity in *cis*-regulatory sequences between distinct alleles could decrease the synergy of ASE and increase the magnitude of heterosis in the growth rate. In summary, our findings shed light on how the regulatory network underlying distinct subgenomes regulates growth plasticity.

Conclusion

The allotriploids obtained through interploidy hybridization between allotetraploid and diploid have the advantages of ecological friendliness (sterility) and fast growth rates, which have made contributions to fish breeding for more than 20 years in China. It has been a huge challenge for us to figure out how to improve the allotriploid fish even more. Our results reveal that variations in CNVs and gene regulatory networks between alleles could shape the growth diversity in the hybrid populations. Consequently, selecting individuals with low-copy growth-regulating genes from allele R (originating from goldfish) and high-copy growth-regulating genes from allele C (originating from common carp) will contribute to the breeding of allotriploid populations exhibiting a rapid growth phenotype. This study will help us develop a novel hybrid breeding strategy through the genotype selection of parental allotetraploids.

Materials and methods

Collection and determination of samples

The fish used in this study included an F_1 diploid hybrid 2nRC obtained from hybridization between *C. auratus* red var. (goldfish, 2nRR, ♀) and *C. carpio* (common carp, 2nCC, ♂), an F_1 diploid hybrid 2nCR obtained from hybridization between 2nCC (♀) and 2nRR (♂), an allotriploid $3nR_2C$ obtained from interploid crossing of 2nRR (♀) with an allotetraploid $4nR_2C_2$ (2nRR × 2nCC, ♂), an allotriploid $3nRC_2$ obtained from interploid crossing of 2nCC (♀) with $4nR_2C_2$ (♂), an allotriploid $3nCR_2$ obtained from interploid crossing of $4nR_2C_2$ (♀) with 2nRR (♂), and an allotriploid $3nC_2R$ obtained from interploid crossing of $4nR_2C_2$ (♀) with 2nCC (♂). These hybrid varieties, including two reciprocal F_1 hybrids (2nRC and 2nCR) and four triploids ($3nR_2C$, $3nRC_2$, $3nCR_2$, and $3nC_2R$), were fed in separate pools under identical environmental conditions. These conditions included suitable water temperatures, oxygen levels, food supply, breeding density, *etc.* These pools were located in the drainage area of Dongting Lake, Hunan Province, China (N29°11′51″, E112°35′50″). Body weights were measured at multiple growth stages. Twenty-nine healthy individuals at 8 months after hatching (7 in 2nRC, 4 in $3nR_2C$, 7 in $3nRC_2$, 8 in $3nCR_2$, and 3 in $3nC_2R$) and 131 healthy individuals at 24 months after hatching were collected for this study, respectively. These hybrid varieties were deeply anesthetized with 300 mg/l tricaine methanesulfonate (Catalog No. MS-222, Sigma-Aldrich, St. Louis, MO) for 10 min at 25°C in a separation tank. After confirming the death, all samples were collected for dissection. The DNA content of erythrocytes from 2nRR, 2nCC, and the hybrids was measured using flow cytometry (Cell Counter Analyzer, Partec GmbH, Munster, Germany) for identifying chromosome number [40].

DNA isolation and WGS

High-quality genomic DNA of 2nRR, 2nCC, and the hybrids was isolated from the muscle tissues using DNeasy Blood & Tissue Kits (Catalog No. 69504, QIAGEN, Hilden, Germany). The quality of DNA was checked by a NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific, Wilmington, DE) with 260/280 and 260/230 ratios. The type of mitochondrial genome in the hybrids was identified based on a fragment of *cytb*. Then, the high-quality DNA was used to construct a paired-end library (2 × 150 bp) and sequenced by Illumina HiSeq X Ten Sequencing System (Illumina, San Diego, CA) according to standard protocol. The detail was as follows: a mixture containing equal amounts of 2nRR and 2nCC DNA, the muscle of 29 hybrid individuals. After obtaining the raw data, the sequencing adaptors were removed. fastp (v0.21.0) was used to remove duplicated read pairs and low-quality reads based on the default parameters [41].

Detection of CNVs

High-quality reads in the hybrids were mapped to the combined nuclear and mitochondrial genomes of goldfish [8] [nuclear DNA (nDNA): PRJCA001234 from the National Genomics Data Center (NGDC), Beijing Institute of Genomics (BIG), Chinese Academy of Sciences (CAS) / China National Center for Bioinformation (CNCB); mitochondrial DNA (mtDNA): AY714387.1] and common carp [32] [nDNA of Yellow River carp: PRJNA510861 from the National Center for Biotechnology Information (NCBI); mtDNA: AP009047.1] using BWA with the default parameters. Coordinate-sorted BAM output files of WGS were obtained to calculate the number of mapped reads in the coding region of each gene using htseq-count (v0.12.4) with parameters of "-m union --nonunique = none". The gene per million (GPM) is a value to measure how many reads are mapped to each gene in genomic data. It helps us understand the relative abundance of gene copies in a sample by considering the length of the gene and the total number of reads. The formula for calculating GPM is as follows:

$$\mathrm{GPM} = \mathrm{A} \times \frac{1}{\sum \mathrm{A}} \times 10^6$$

where A = Total reads mapped to gene × 10^6 / Gene length (bp).

For *in silico* F_1, we sequenced an equal mixture of 2nRR and 2nCC DNA using the same sequencing platform as other genomic data. By comparing the hybrid varieties with the *in silico* F_1, we were able to detect CNVs in all the hybrid varieties. To identify CNVs, we set thresholds based on the genotype. We calculated $\log_2 \mathrm{GPM_{hybrid}/GPM_{mixed}}$ and compared it to $\log_2 \frac{B\times 2}{C}$ and $\log_2 \frac{B\times 0.5}{C}$. If $\log_2 \mathrm{GPM_{hybrid}/GPM_{mixed}}$ was greater than $\log_2 \frac{B\times 2}{C}$ or less than $\log_2 \frac{B\times 0.5}{C}$, it was considered a CNV. In this formula, "B" represents the allelic ratio (R or C) in hybrids, and "C" represents the allelic ratio (R or C) in *in silico* F_1. For 2nRC, "B" is $\frac{1}{2}$; for allele R of $3nRC_2$ and $3nC_2R$ and allele C of $3nR_2C$ and $3nCR_2$, "B" is $\frac{1}{3}$; for allele C of $3nRC_2$ and $3nC_2R$ and allele R of $3nR_2C$ and $3nCR_2$, "B" is $\frac{2}{3}$.

AGPs between the subgenomes R (originating from 2nRR) and C (originating from 2nCC) in the hybrid varieties, including the F_1 hybrids (2nRC and 2nCR) and four triploids ($3nR_2C$, $3nRC_2$, $3nCR_2$, and $3nC_2R$), were obtained

using the all-against-all reciprocal BLASTP (v2.8.1) with an E-value of 1E−6 based on protein sequences. Then, transcripts that lacked gene annotation and were shorter than 300 bp were discarded from AGPs. GPM values in each AGP could be used to assess allelic CNVs between the subgenomes R and C in the hybrid varieties. The values of $\log_{10}(R_{GPM\ in\ hybrid}/C_{GPM\ in\ hybrid} - R_{GPM\ in\ mixed}/C_{GPM\ in\ mixed})$ could be used to assess the allelic CNVs. The detail thresholds for them were set up based on the genotype as follows: (1) $\log_{10}$ 2 and $\log_{10}$ 0.5 in F_1 hybrids (2nRC and 2nCR); (2) $\log_{10}$ 4 and $\log_{10}$ 1 in $3nR_2C$ and $3nCR_2$; (3) $\log_{10}$ 1 and $\log_{10}$ 0.25 in $3nC_2R$ and $3nRC_2$.

RNA isolation and RNA-seq

To obtain gene expression profiles of the two reciprocal F_1 hybrids and four allotriploids, total RNA of the muscle tissue was isolated and purified according to a TRIzol extraction method [42]. The RNA concentration was measured using NanoDrop technology (NanoDrop ND-1000 UV/Vis spectrophotometer, NanoDrop Technologies, Wilmington, DE). Total RNA samples were treated with DNase I (Catalog No. 18068-015, Thermo Fisher Scientific, Philadelphia, PA) to remove any contaminating genomic DNA. The purified RNA was quantified using a 2100 Bioanalyzer system (Catalog No. 5067-5576, Agilent, Santa Clara, CA). Isolated messenger RNA (mRNA) was fragmented with a fragmentation buffer. The resulting short fragments were reversely transcribed and amplified to produce complementary DNA (cDNA). Illumina RNA-seq libraries of the 29 samples were prepared according to the standard high-throughput method. The quality of the cDNA library was assessed by the 2100 Bioanalyzer system (Catalog No. 5067-5576, Agilent). The library was sequenced with a paired-end (2 × 150 bp) setting using the Illumina HiSeq X Ten Sequencing System (Illumina, San Diego, CA). The transcriptomic data of muscle tissues in 131 hybrid individuals were obtained using DNA nanoball technology using DNBSEQ-T7 (MGI, Shenzhen, China) according to the standard method [43]. Then, low-quality bases and adapters were trimmed out using fastp (v0.21.0). The high-quality reads were used in downstream analyses.

Gene expression profiling based on RNA-seq data

All RNA-seq reads of 2nRR, 2nCC, and the hybrids were mapped to the combined nuclear and mitochondrial genomes of *C. auratus* red var. [8] and *C. carpio* [32] using HISAT2 (v2.1.0) [44] with default parameters. Then, the mapped files were handled with SAMtools (v1.10) [45], while the unique mapped reads were obtained using htseq-count (v0.12.4) [46]. The expression value was 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 transcripts per million (TPM) values were calculated based on the normalized data. These reads in the two F_1 hybrids and four allotriploids were used to calculate the expression values of the genes in subgenomes R and C [47]. The genes with mapped reads in each sample < 10 and TPM < 1 were not used in our subsequent analyses. Differential expression analysis was performed using DESeq2 of the R package with the below thresholds: fold change > 3, P value < 0.001, and P adjusted < 0.001. Differential expression analysis was performed in 13 individuals of the diploid group (2nRC and 2nCR) and in 10 individuals of the allotriploid group ($3nRC_2$ and $3nC_2R$).

Both WGS and RNA-seq were performed in the 29 individuals collected at 8 months after hatching (7 in 2nRC, 4 in $3nR_2C$, 7 in $3nRC_2$, 8 in $3nCR_2$, and 3 in $3nC_2R$) for investigating the effects of CNVs on ASE and species-specific expression, which were assessed based on the aforementioned thresholds of differential expression analysis.

WGCNA and functional annotation

To investigate expression patterns across samples, we conducted a coexpression analysis based on the two F_1 hybrids and four allotriploid samples using WGCNA (v1.67). An unsupervised network on gene expression was built using the following default parameters. First, a matrix of PCC values between genes was generated based on expression values. Then, an adjacency matrix representing the connection strength among genes was established by raising the correlation matrix to a soft threshold power. Next, the adjacency matrix was used to calculate a topological overlap matrix. Genes with similar coexpression patterns were clustered using hierarchical clustering of dissimilarity. Pearson correlation analysis between the expression level of that gene and module was performed using eigengene-based connectivity, while PCC values were further calculated to measure the strength and direction of association between modules and growth traits. The coexpressed modules were determined and used in our subsequent functional analyses. The hub genes related to growth regulation were further filtered based on thresholds of module membership > 0.8 and gene significance > 0.3. Functional enrichment analyses were conducted and annotated with GO and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases. The standard deviation (SD) of body weight divided by the average body weight across individuals was used to assess the body weight diversity in each population. The SD of the TPM values across the individuals of the six hybrid varieties was used to assess the expression diversity of the predicted growth-regulating genes.

Pearson correlation analyses

Pearson correlation analyses between the expression of alleles R and C as well as between body weight and the expression of allele R or C were performed based on the below thresholds: positive correlation, PCC > 0.3; strong positive correlation, PCC > 0.5; negative correlation, PCC < −0.3; strong negative correlation, PCC < −0.5. Differential analysis was performed based on the P value using a t-test. AGPs refer to the gene pairs which maintain strict orthologous relationships between subgenomes R and C. Non-AGPs were established by pairing non-orthologous genes from the AGPs mentioned above.

Mapping of methylation sequencing data and differential methylation analysis

The whole-genome bisulfite sequencing data of 2nRR, 2nCC, $3nR_2C$, and $3nRC_2$ (muscle tissue, 2 years old, three biological replicates in each variety) were obtained from the NGDC, BIG, CAS / CNCB (BioProject: PRJCA003625). After quality checking of the methylation sequencing reads, the clean reads of 2nRR and 2nCC were mapped to the respective genomes, and the clean reads of the two triploids ($3nR_2C$ and $3nRC_2$) were mapped to the combined genome sequences of 2nRR and 2nCC [8,32]. The Bismark analysis pipeline was used to detect the methylated loci with the parameters of "-score_min L, 0, −0.2 −X 1000 -no-mixed -no-discordant" [48,49]. The

clean reads were mapped to the reference genome four times, and only the reads that mapped to the same position of the reference genome each time were retained in our subsequent analysis. A binomial distribution test was performed to identify 5-methylcytosine for each cytosine site. The potential methylation sites were then checked using the thresholds of depth > 4 and false discovery rate (FDR) < 0.05.

The average CpG methylation level was detected in different gene regions, including 1st exon, 1st intron, inner exon(s), inner intron(s), and last exon, as well as the upstream and downstream 2-kb regions (a window size of 100 bp) flanking the gene body. The average CpG methylation levels in the upstream and downstream 2-kb regions flanking the gene body were calculated and visualized using R. The regions with different methylation levels were detected using MOdel based Analysis of Bisulfite Sequencing data (MOABS) [50]. The R packages Dispersion Shrinkage for Sequencing data (DSS) and bsseq were used to call differentially methylated regions and predict DMGs based on $P < 0.01$.

H&E staining

A 10-mm trunk muscles from 2nRR, 2nCC, $3nR_2C$, and $3nRC_2$ were dissected from the region in the dorsum and fixed in Bouin's solution for 24 h, respectively. The fixed tissues were washed with distilled water for 4 h at 20°C. After dehydration in ethanol gradients and xylene, the samples were fixed in 4% paraformaldehyde and cut into serial paraffin sections (5–7 μm in thickness). Sections were processed for H&E staining. Digital images were captured with a microscope (DX8, Olympus, Tokyo, Japan). Three independent biological replicates were used to collect quantitative data on H&E staining in each hybrid variety.

Ethical statement

All procedures performed on animals in this study were approved by the Academic Committee at Hunan Normal University, China (Approval No. 2018D013).

Data availability

The raw reads of WGS and RNA-seq data have been deposited in the Genome Sequence Archive [51] at the NGDC, BIG, CAS / CNCB (GSA: CRA009160 for WGS data for 29 individuals, CRA009161 for RNA-seq data for 29 individuals, and CRA009164 for RNA-seq data for 131 individuals; BioProject: PRJCA013677), and are publicly accessible at https://ngdc.cncb.ac.cn/gsa. The raw reads of equally mixed DNA samples from goldfish and common carp have been deposited in BioSample at the NGDC, BIG, CAS / CNCB (BioSample: SAMC449140), and are publicly accessible at https://ngdc.cncb.ac.cn/biosample/browse/SAMC449140.

CRediT author statement

Li Ren: Conceptualization, Data curation, Formal analysis, Writing – original draft, Project administration, Writing – review & editing, Validation, Funding acquisition. **Mengxue Luo:** Data curation, Writing – review & editing, Validation. **Jialin Cui:** Investigation. **Xin Gao:** Data curation, Investigation. **Hong Zhang:** Investigation. **Ping Wu:** Investigation. **Zehong Wei:** Investigation. **Yakui Tai:** Investigation. **Mengdan Li:** Investigation. **Kaikun Luo:** Investigation. **Shaojun Liu:** Conceptualization, Project administration, Validation, Funding acquisition. All authors have read and approved the final manuscript.

Supplementary material

Supplementary material is available at *Genomics, Proteomics & Bioinformatics* online (https://doi.org/10.1093/gpbjnl/qzae055).

Competing interests

The authors have declared no competing interests.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant Nos. 32293252, 32341057, U19A2040, and 32273111), the Natural Science Foundation of Hunan Province (Grant No. 2022JJ10035), the Huxiang Young Talent Project of China (Grant No. 2021RC3093), the National Key R&D Program of China (Grant No. 2023YFD2401602), the Laboratory of Lingnan Modern Agriculture Project (Grant No. NT2021008), the Special Funds for Construction of Innovative Provinces in Hunan Province (Grant No. 2021NK1010), the Earmarked fund for China Agriculture Research System (Grant No. CARS-45), and the 111 Project (Grant No. D20007), China. We thank Dr. Zhenzhen Dong (Hunan Normal University, China) for professional language editing.

ORCID

0000-0003-4613-0380 (Li Ren)
0009-0003-5224-1738 (Mengxue Luo)
0000-0001-5901-8107 (Jialin Cui)
0000-0001-7162-4177 (Xin Gao)
0009-0001-7442-3079 (Hong Zhang)
0009-0004-6829-9800 (Ping Wu)
0000-0003-0688-1032 (Zehong Wei)
0009-0004-4525-5813 (Yakui Tai)
0009-0003-5896-7564 (Mengdan Li)
0000-0002-5735-1202 (Kaikun Luo)
0000-0001-5761-8571 (Shaojun Liu)

References

[1] Mallet J. Hybrid speciation. Nature 2007;446:279–83.
[2] Degraeve S, De Baerdemaeker NJF, Ameye M, Leroux O, Haesaert GJW, Steppe K. Acoustic vulnerability, hydraulic capacitance, and xylem anatomy determine drought response of small grain cereals. Front Plant Sci 2021;12:599824.
[3] Paritosh K, Gupta V, Yadava SK, Singh P, Pradhan AK, Pental D. 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 2014;15:396.
[4] Qin Q, Wang Y, Wang J, Dai J, Xiao J, Hu F, et al. The autotetraploid fish derived from hybridization of *Carassius auratus* red var. (female) × *Megalobrama amblycephala* (male). Biol Reprod 2014;91:93.
[5] Liu S, Luo J, Chai J, Ren L, Zhou Y, Huang F, et al. Genomic incompatibilities in the diploid and tetraploid offspring of the

goldfish × common carp cross. Proc Natl Acad Sci U S A 2016; 113:1327–32.
[6] Hórreo JL, Ayllón F, Perez J, Beall E, Garcia-Vazquez E. Interspecific hybridization, a matter of pioneering? Insights from Atlantic salmon and brown trout. J Hered 2011;102:237–42.
[7] Selz OM, Seehausen O. Interspecific hybridization can generate functional novelty in cichlid fish. Proc Biol Sci 2019; 286:20191621.
[8] Luo J, Chai J, Wen Y, Tao M, Lin G, Liu X, et al. From asymmetrical to balanced genomic diversification during rediploidization: subgenomic evolution in allotetraploid fish. Sci Adv 2020; 6:eaaz7677.
[9] Song C, Liu S, Xiao J, He W, Zhou Y, Qin Q, et al. Polyploid organisms. Sci China Life Sci 2012;55:301–11.
[10] Chen D, Zhang Q, Tang W, Huang Z, Wang G, Wang Y, et al. The evolutionary origin and domestication history of goldfish (*Carassius auratus*). Proc Natl Acad Sci U S A 2020; 117:29775–85.
[11] Xu P, Zhang X, Wang X, Li J, Liu G, Kuang Y, et al. Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. Nat Genet 2014;46:1212–9.
[12] Zhang Y, Liu J, Fu W, Xu W, Zhang H, Chen S, et al. Comparative transcriptome and DNA methylation analyses of the molecular mechanisms underlying skin color variations in *Crucian carp* (*Carassius carassius* L.). BMC Genet 2017;18:95.
[13] Chen Z, Omori Y, Koren S, Shirokiya T, Kuroda T, Miyamoto A, et al. *De novo* assembly of the goldfish (*Carassius auratus*) genome and the evolution of genes after whole-genome duplication. Sci Adv 2019;5:eaav0547.
[14] Liu S, Liu Y, Zhou G, Zhang X, Luo C, Feng H, et al. The formation of tetraploid stocks of red crucian carp × common carp hybrids as an effect of interspecific hybridization. Aquaculture 2001;192:171–86.
[15] Ren L, Gao X, Cui J, Zhang C, Dai H, Luo M, et al. Symmetric subgenomes and balanced homoeolog expression stabilize the establishment of allopolyploidy in cyprinid fish. BMC Biol 2022; 20:200.
[16] Ren L, Zhang X, Li J, Yan X, Gao X, Cui J, et al. Diverse transcriptional patterns of homoeologous recombinant transcripts in triploid fish (Cyprinidae). Sci China Life Sci 2021;64:1491–501.
[17] Kemper KE, Visscher PM, Goddard ME. Genetic architecture of body size in mammals. Genome Biol 2012;13:244.
[18] Visscher PM, Macgregor S, Benyamin B, Zhu G, Gordon S, Medland S, et al. Genome partitioning of genetic variation for height from 11,214 sibling pairs. Am J Hum Genet 2007; 81:1104–10.
[19] Romano C, Koot MB, Kogan I, Brayard A, Minikh AV, Brinkmann W, et al. Permian–Triassic Osteichthyes (bony fishes): diversity dynamics and body size evolution. Biol Rev Camb Philos Soc 2016;91:106–47.
[20] Hochholdinger F, Baldauf JA. Heterosis in plants. Curr Biol 2018;28:R1089–92.
[21] Wu X, Li R, Li Q, Bao H, Wu C. Comparative transcriptome analysis among parental inbred and crosses reveals the role of dominance gene expression in heterosis in *Drosophila melanogaster*. Sci Rep 2016;6:21124.
[22] Maynard BT, Taylor RS, Kube PD, Cook MT, Elliott NG. Salmonid heterosis for resistance to amoebic gill disease (AGD). Aquaculture 2016;451:106–12.
[23] Clasen JB, Norberg E, Madsen P, Pedersen J, Kargo M. Estimation of genetic parameters and heterosis for longevity in crossbred Danish dairy cattle. J Dairy Sci 2017;100:6337–42.
[24] Shapira R, Levy T, Shaked S, Fridman E, David L. Extensive heterosis in growth of yeast hybrids is explained by a combination of genetic models. Heredity (Edinb) 2014;113:316–26.
[25] Yang H, Li Q. The DNA methylation level is associated with the superior growth of the hybrid crosses in the Pacific oyster *Crassostrea gigas*. Aquaculture 2022;547:737421.
[26] Ren L, Yan X, Cao L, Li J, Zhang X, Gao X, et al. Combined effects of dosage compensation and incomplete dominance on gene expression in triploid cyprinids. DNA Res 2019;26:485–94.
[27] Springer NM, Stupar RM. Allelic variation and heterosis in maize: how do two halves make more than a whole? Genome Res 2007;17:264–75.
[28] Pan FC, Bankaitis ED, Boyer D, Xu X, Van de Casteele M, Magnuson MA, et al. Spatiotemporal patterns of multipotentiality in *Ptf1a*-expressing cells during pancreas organogenesis and injury-induced facultative restoration. Development 2013;140:751–64.
[29] Xiong Y, Lei F. *SLC2A12* of SLC2 gene family in bird provides functional compensation for the loss of *SLC2A4* gene in other vertebrates. Mol Biol Evol 2021;38:1276–91.
[30] Ren L, Zhang H, Luo M, Gao X, Cui J, Zhang X, et al. Heterosis of growth trait regulated by DNA methylation and miRNA in allotriploid fish. Epigenetics Chromatin 2022;15:19.
[31] Li JT, Wang Q, Huang Yang MD, Li QS, Cui MS, Dong ZJ, et al. Parallel subgenome structure and divergent expression evolution of allo-tetraploid common carp and goldfish. Nat Genet 2021; 53:1493–503.
[32] Xu P, Xu J, Liu G, Chen L, Zhou Z, Peng W, et al. The allotetraploid origin and asymmetrical genome evolution of the common carp *Cyprinus carpio*. Nat Commun 2019;10:4625.
[33] Wang J, Ye LH, Liu QZ, Peng LY, Liu W, Yi XG, et al. Rapid genomic DNA changes in allotetraploid fish hybrids. Heredity (Edinb) 2015;114:601–9.
[34] Pastinen T. Genome-wide allele-specific analysis: insights into regulatory variation. Nat Rev Genet 2010;11:533–8.
[35] Bao Y, Hu G, Grover CE, Conover J, Yuan D, Wendel JF. Unraveling *cis* and *trans* regulatory evolution during cotton domestication. Nat Commun 2019;10:5399.
[36] He F, Wang W, Rutter WB, Jordan KW, Ren J, Taagen E, et al. Genomic variants affecting homoeologous gene expression dosage contribute to agronomic trait variation in allopolyploid wheat. Nat Commun 2022;13:826.
[37] 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* × *Culter alburnus*. Genome Res 2019;29:1805–15.
[38] Pala I, Coelho MM, Schartl M. Dosage compensation by gene-copy silencing in a triploid hybrid fish. Curr Biol 2008;18:1344–8.
[39] Rogers S, Macheda ML, Docherty SE, Carty MD, Henderson MA, Soeller WC, et al. Identification of a novel glucose transporter-like protein—GLUT-12. Am J Physiol Endocrinol Metab 2002;282:E733–8.
[40] 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* × male *Culter alburnus*. Anim Reprod Sci 2014;151:61–70.
[41] Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018;34:i884–90.
[42] Rio DC, Ares M Jr, Hannon GJ, Nilsen TW. Purification of RNA using TRIzol (TRI reagent). Cold Spring Harb Protoc 2010;2010: pdb.prot5439.
[43] 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;20:604.
[44] Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 2019;37:907–15.
[45] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078–9.
[46] 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;19:339–42.
[47] Ren L, Cui J, Wang J, Tan H, Li W, Tang C, et al. Analyzing homoeolog expression provides insights into the rediploidization event in gynogenetic hybrids of *Carassius auratus* red var. × *Cyprinus carpio*. Sci Rep 2017;7:13679.

[48] Song Q, Zhang T, Stelly DM, Chen ZJ. Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. Genome Biol 2017;18:99.
[49] Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-seq applications. Bioinformatics 2011;27:1571–2.
[50] Sun D, Xi Y, Rodriguez B, Park HJ, Tong P, Meong M, et al. MOABS: model based analysis of bisulfite sequencing data. Genome Biol 2014;15:R38.
[51] Chen T, Chen X, Zhang S, Zhu J, Tang B, Wang A, et al. The Genome Sequence Archive Family: toward explosive data growth and diverse data types. Genomics Proteomics Bioinformatics 2021;19:578–83.

Genomics, Proteomics & Bioinformatics, 2024, 22, 1–13
https://doi.org/10.1093/gpbjnl/qzae055
Original Research