DNA Research, 2019, 26(6), 485–494
doi: 10.1093/dnares/dsz026
Advance Access Publication Date: 27 December 2019
Full Paper

OXFORD

Full Paper

Combined effects of dosage compensation and incomplete dominance on gene expression in triploid cyprinids

Li Ren[1,2†], Xiaojing Yan[1,2†], Liu Cao[1,2], Jiaming Li[1,2], Xueyin Zhang[1,2], Xin Gao[1,2], Jia Liu[1,2], Jialin Cui[1,2], and Shaojun Liu[1,2*]

[1]State Key Laboratory of Developmental Biology of Freshwater Fish, Hunan Normal University, Changsha, Hunan 410081, P.R. China, and [2]College of Life Sciences, Hunan Normal University, Changsha, Hunan 410081, P.R. China

*To whom correspondence should be addressed. Tel. +86 731 8887 3074. Fax. +86 731 8887 3010.
Email: lsj@hunnu.edu.cn
[†]These authors contributed equally to this work.

Received 8 September 2019; Editorial decision 17 December 2019; Accepted 24 December 2019

Abstract

Hybridization and polyploidy are pervasive evolutionary features of flowering plants and frequent among some animal groups, such as fish. These processes always lead to novel genotypes and various phenotypes, including growth heterosis. However, its genetic basis in lower vertebrate is still poorly understood. Here, we conducted transcriptome-level analyses of the allopolyploid complex of *Carassius auratus* red var. (R) (♀) × *Cyprinus carpio* L. (C) (♂), including the allodiploid and allotetraploid with symmetric subgenomes, and the two allotriploids with asymmetric subgenomes. The gradual changes of gene silencing and novel gene expression suggested the weakening of the constraint of polymorphic expression in genotypic changes. Then, analyses of the direction and magnitude of homoeolog expression exhibited various asymmetric expression patterns, which supported that R incomplete dominance and dosage compensation were co-regulated in the two triploids. Under these effects, various magnitudes of R-homoeolog expression bias were observed in growth-regulated genes, suggesting that they might contribute to growth heterosis in the two triploids. The determination of R incomplete dominance and dosage compensation, which might be led by asymmetric subgenomes and multiple sets of homologous chromosomes, explained why various expression patterns were shaped and their potential contribution to growth heterosis in the two triploids.

Key words: dosage compensation, incomplete dominance, asymmetric subgenomes, allotriploid, homoeolog expression

1. Introduction

Dosage compensation is when the expression of sex-linked genes originating from different biological sexes is equalized.[1] Although most studies have focused on dosage compensation of sex chromosomes, the dosage compensation of sets of genes in subgenome or whole genomes has been detected in some allopolyploids including *Oryza sativa* L.,[2] *Arabidopsis*,[3] *Squalius alburnoides*,[4] and the hybrid of *Ctenopharyngodon idellus* × *Megalobrama amblycephala*.[5] The potential mechanisms of dosage compensation include chromatin remodelling[6] and miRNAs.[7] Dosage compensation is described as an important contributor to heterosis, or 'hybrid vigour' as it is commonly known, which is defined as the superior performance of

© The Author(s) 2019. Published by Oxford University Press on behalf of Kazusa DNA Research Institute.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from https://academic.oup.com/dnaresearch/article-abstract/26/6/485/5688746 by Hunan Normal University user on 23 March 2020

hybrids compared with either of their parents.[8,9] However, its genetic basis and molecular mechanism in lower vertebrates have remained obscure.

Growth heterosis is an important trait in aquaculture species. The different amounts of heterosis were observed in the two allotriploids ($3n = 150$, RC_2 and R_2C), which were obtained from backcrossing of the allotetraploid ($4n = 200$) of *Carassius auratus* red var. (♀) × *Cyprinus carpio* L. (♂) with their two inbred parents, respectively.[10,11] The allotriploid RC_2 fish [intercrossing between diploid *C. carpio* L. (♀) and allotetraploid *C. auratus* red var. × *C. carpio* L. (♂)] showed highly significant growth heterosis comparing to the allotriploid R_2C fish [intercrossing between *C. auratus* red var. (♀) and allotetraploid *C. auratus* red var. × *C. carpio* L. (♂)], suggesting the potential interaction between growth heterosis and genetic diversity, which affects various gene expression.[12,13] Heterosis is related not only to changes in the global expression levels of all alleles,[13] but also to interactions among multiple alleles originating from different species (also described as homoeologs).[14,15] Thus, focusing on both of these factors could help us investigate the causes and effects of gene expression changes in hybrids. After the merging of divergent genomes, the novel dominant–recessive relationship between alleles was rebuilt.[16] Study on the quantitative traits under incomplete dominance was complex for multiple allele control. However, a few studies have described that some trait variation and heterosis are led by the incomplete dominance of deleterious alleles in maize.[17] Both dosage effects and incomplete dominance could shape direction and magnitude of homoeolog expression, further contributing to various phenotypes in allopolyploid plant, including heterosis.[9,18,19]

In our study, we performed mRNA-seq on the two allotriploids with growth heterosis as described earlier. After a series of expression analysis, we obtained the homoeolog expression levels in liver and muscle tissues in allopolyploid complex of *C. auratus* red var. (♀) × *C. carpio* L. (♂). After genotype determination of the two triploids, comparative transcriptome analysis will help us understand the expression changes of one or two copies of genes originating from divergent genomes. In comparison of various growth features,[10,14] the homoeolog expression changes of growth-regulated genes were used to explore the genetic basis of growth heterosis in triploid fish.

2. Materials and methods

2.1. Animals

A total of 6 types of fishes were used in our experiments: diploid *C. auratus* red var. (R); diploid *C. carpio* L. (C); first generation of allodiploid (F_1) originating from outcrossing of *C. auratus* red var. (♀) and *C. carpio* L. (♂); eighteenth generation of allotetraploid of *C. auratus* red var. (♀) × *C. carpio* L. (♂) (F_{18}); triploid fish (R_2C) originating from intercrossing between female *C. auratus* red var. and male allotetraploid of *C. auratus* red var. (♀) × *C. carpio* L. (♂); and triploid fish (RC_2) originating from intercrossing between female *C. carpio* L. (♀) and male allotetraploid of *C. auratus* red var. (♀) × *C. carpio* L. (♂). All fishes were fed in a pool with the same environment including suitable illumination, water temperature, dissolved oxygen content, and adequate forage. These experiments were conducted at the Engineering Center of Polyploidy Fish Breeding of the National Education Ministry, Hunan Normal University, Hunan, China, and conformed to the National Institutes of Health Guide for Care and Use of Laboratory Animals. The animal work was approved by the academic committee in State Key Laboratory of Developmental Biology of Freshwater Fish (approval ID: 01/2018). Two-year-old male individuals of each genotype were collected with three biological replications. Fish were deeply anaesthetized with 100 mg/l MS-222 (Sigma-Aldrich, St Louis, MO, USA) before dissection. To exclude effects of different tissues on gene expression, all liver and muscle tissue samples were excised carefully to avoid contamination from the gut, and were used for total RNA extraction.

2.2. Genotype determination

To determine ploidy levels, a flow cytometer was used to measure the DNA content of erythrocytes from diploid and triploid individuals. Chromosomal preparations were made from peripheral blood cell cultures of 15-month-old fishes. Blood (0.2 ml) was collected using a syringe soaked with 0.1% sodium heparin, cultured in nutrient solution at 25.5°C and 5% CO_2 for 68–72 h, and then colchicine was added 3.5 h before harvest. Cells were harvested by centrifugation, followed by hypotonic treatment with 0.075 M KCl at 26°C for 25–30 min, and then fixed in methanol–acetic acid (3:1, v/v) with three changes.[20] The number of chromosomes was counted under a microscope.

To determine the genotypes of the two triploid fish (R_2C and RC_2), a 9,468 bp-sized DNA repeat fragment of *C. auratus* red var. was used as a probe for fluorescence *in situ* hybridization (FISH) analyses. These analyses allowed us to characterize the chromosomes originating from *C. auratus* red var.[11] The FISH probe was produced and labelled with digoxigenin-11-dUTP (ROCHE, Mannheim, Germany) of purified PCR products. A total of 150 metaphase chromosome spreads from 10 individuals were examined under an inverted microscope (CW4000, Leica, Wetzlar, Germany) with a confocal imaging system (LCS SP2, Leica). Captured images were coloured and superimposed using Adobe Photoshop CS5 software.

2.3. mRNA-seq sequencing, alignments, and differential global expression analysis

The samples used to produce the mRNA-seq libraries were stored in RNALater (AM7021, Ambion Life Technologies, Carlsbad, CA, USA) following the manufacturer's instructions. After DNase treatment, total RNA (~2 μg) was used to construct mRNA-seq libraries according to the manufacturer's instructions. For each type of fish, an mRNA-seq library was constructed with three biological replicates with paired ends (2 × 150 bp) using Illumina HiSeq × Ten (Illumina, Sad Diego, CA, USA). For the F_1, F_{18}, and their parents, we downloaded mRNA-seq data for liver tissue from the NCBI database (accession numbers: SRX668436, SRX175397, SRX668453, SRX177691, SRX671568, SRX671569, SRX668467, and SRX161099).

The initial quality control of Illumina reads was performed with FastQC software,[21] and adapters were trimmed using Trimmomatic software.[22] The mRNA-seq reads of *C. auratus* red var. and *C. carpio* L. were mapped to genome sequences of *C. auratus* red var. (NCBI accession numbers: PRJNA487739 and PRJNA481500) and *C. carpio* L.,[23] respectively. Gene annotations were obtained by BLASTX searches of the NCBI, gene ontology, and Swiss-Prot databases. Because of the lack of a reference genome, mRNA-seq reads for F_1 and F_{18}, R_2C, and RC_2 were mapped to combined genome sequences of *C. auratus* red var. and *C. carpio* L. Read mapping was conducted using STAR software with default options. Read counts of *C. auratus* red var. and *C. carpio* L. for orthologous genes were

Downloaded from https://academic.oup.com/dnaresearch/article-abstract/26/6/485/5688746 by Hunan Normal University user on 23 March 2020

summarized for each gene pair. Normalization, estimation of expression levels, and designation as differentially expressed genes (DEGs) were performed using the edgeR package in Bioconductor.[24] The DEGs between the two genotypes were detected by edgeR using threshold of $|\log_2$ fold change$| > 1\%$ and 5% false discovery rate (FDR < 0.05) with three biological replicates.

2.4. Identification of species-specific single nucleotide polymorphisms in orthologs

Orthologous gene alignments between *C. auratus* red var. and *C. carpio* L. were obtained from all-against-all reciprocal BLASTP (v 2.2.26) comparisons with the parameters of '-e 10-5 –F –v 1 –m 8' based on protein sequences. Unannotated transcripts and coding sequences shorter than 300 bp were discarded. In total, orthologous gene pairs were selected from the above best scoring match and filtering.

To investigate the expression level of parent-of-origin genes (homoeolog expression), the LASTZ pairwise alignment tool (v 1.02.00)[25] with default parameters was used to obtain the corresponding loci for orthologous gene pairs between *C. auratus* red var. and *C. carpio* L. The insertion–deletion polymorphisms loci were discarded, and only aligned loci with best scoring matches were used for further analyses. The single nucleotide polymorphisms (SNPs) were collected using SNP Calling pipeline with GATK (v 3.8) based on the results of parent transcriptome mapping to their respective genomes as described earlier.[26] According to the comparison of SNPs and other loci in orthologous gene pairs between *C. auratus* red var. and *C. carpio* L., we identified heterozygous and homozygous loci with species-specific SNPs as used elsewhere.[27,28] To remove the negative effects of sequencing and mapping, the screening of species-specific SNPs was checked for consistency in three biological replicates. Locus possessing read counts ($\geq$1) were retained for all accessions and biological replicates in each comparison.

2.5. Detection of homoeolog expression levels in hybrids

To explore homoeolog expression in hybrids, the hybrid mapping results (bam files) of the transcriptome as described earlier were used in the following analyses. The mapping files for each hybrid were divided into two categories based on the two parental genomes. The R-/C-homoeolog reads in hybrids were calculated using in-house perl scripts, based on corresponding R-/C-species-specific SNPs in their corresponding mapping files. Gene expression levels in the parents were also determined based on the respective reference genome related to species-specific SNPs. To remove the negative effects of mutation sites in hybrids with different biological replicates, if the mapping reads of the species-specific SNP did not comply with the threshold of mean $\pm$ 2 S.D. in the three biological replicates, the abnormal value was discarded when estimating the R- and C-homoeolog expression level. Then, the sum counts of R- and C-homoeologs of each gene 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.[29] The sum of R-/C-homoeolog reads in all species-specific SNPs of each gene were used to assess the R-/C-homoeolog expression levels. The average deviation (AD) values of R- and C-homoeolog expression were used to assess the magnitude of deviation.

2.6. Homoeolog expression silencing and bias

After confirming orthologous gene pairs between *C. auratus* red var. and *C. carpio* L., we detected the extent of homoeolog silencing in liver and muscle tissues of the different types of fish. The following thresholds were set for this analysis: (i) read counts $\geq$5 in both two parents in three biological replications and (ii) R- or C-homoeologous reads in hybrid = 0 in three biological replications.

To investigate the homoeolog expression bias (HEB) in hybrids, the value was set based on the formula $\left(E_{bias} = \log_2\left(\frac{E_C\ homoeolog}{E_R\ homoeolog}\right)\right)$. The distribution of E_{bias} was tested using the *t*-test in the 'ggpubr' package in R. We determined another value based on the parents' expression levels $\left(E_{in\ silico\ hybrid} = \log_2\left(\frac{E_{paternal\ C}}{E_{maternal\ R}}\right)\right)$, which was used as the reference for the beginning of changes in homoeolog expression. We calculated another two values for R_2C $\left(E_{in\ silico\ R2C} = \log_2\left(\frac{E_{paternal\ C}}{2\times E_{maternal\ R}}\right)\right)$ and $RC_2\left(E_{in\ silico\ RC2} = \log_2\left(\frac{2\times E_{paternal\ C}}{E_{maternal\ R}}\right)\right)$, based on genotypes. Statistical analyses were performed to detect correlation coefficients between R- and C-homoeolog expression levels in the hybrids in those in their respective parents using Pearson's test and Student's *t*-test in GraphPad Prism (v 7.0) software.

3. Results

3.1. Determination of genotypes of two allotriploids

To determine whether dosage compensation occurred and shaped the various expression patterns in triploid fish, the genotypes of the two triploid fish (R_2C and RC_2) originating from the backcrossing of allotetraploid of *C. auratus* red var. (♀) × *C. carpio* L. (♂) were necessary to check (Fig. 1).[30] A total of 150 chromosomes were detected by flow cytometry and microscopy in R_2C and RC_2. Two sets of *C. auratus* red var. ($2n = 100$) chromosomes and one set of *C. carpio* L. ($n = 50$) chromosomes were detected in R_2C using FISH with a specific probe,[11] while one set of *C. auratus* red var. ($n = 50$) chromosomes and two sets of *C. carpio* L. ($2n = 100$) chromosomes were detected in RC_2 (Fig. 1). The results showed the two sets of homologous chromosomes in both of the two triploids. The symmetric subgenomes at the genomic level were detected in F_1 and F_{18}.[11,30,31]

3.2. Changes of gene silencing and novel expression with genotypic changes

After initial adapter trimming and quality filtering of the mRNA-seq results, we obtained 29.6 Gb clean data for R_2C, 32.6 Gb for RC_2, 17.3 Gb for *C. auratus* red var., and 18.5 Gb for *C. carpio* L. for liver tissue; and obtained 36.1 Gb clean data for R_2C, 34.5 Gb for RC_2, 36.7 Gb for *C. auratus* red var., and 33.7 Gb for *C. carpio* L. for muscle tissue (Supplementary Table S1). Then, 35.54 million cleaned reads (73.42%) of *C. auratus* red var. and 59.21 million cleaned reads (74.69%) of *C. carpio* L. were mapped against their respective reference genomes, and 154.04 million clean reads (72.49%) of R_2C and RC_2 were mapped to the combined genome sequences of *C. auratus* red var. and *C. carpio* L. by STAR.

Detection of R- and C-homoeolog expression values can investigate gene silencing and novel gene expression accompanying genotypic changes in hybrids. The results both in non-orthologous genes and 16,558 orthologous genes showed similar increases in novel expression of R-/C-homoeologs (R-/C-NEGs) from F_1 to F_{18} in the liver (Supplementary Fig. S1). There were also increases in silencing of R-/C-homoeologs (R-/C-SEGs) from F_1 to F_{18} in the liver

Downloaded from https://academic.oup.com/dnaresearch/article-abstract/26/6/485/5688746 by Hunan Normal University user on 23 March 2020

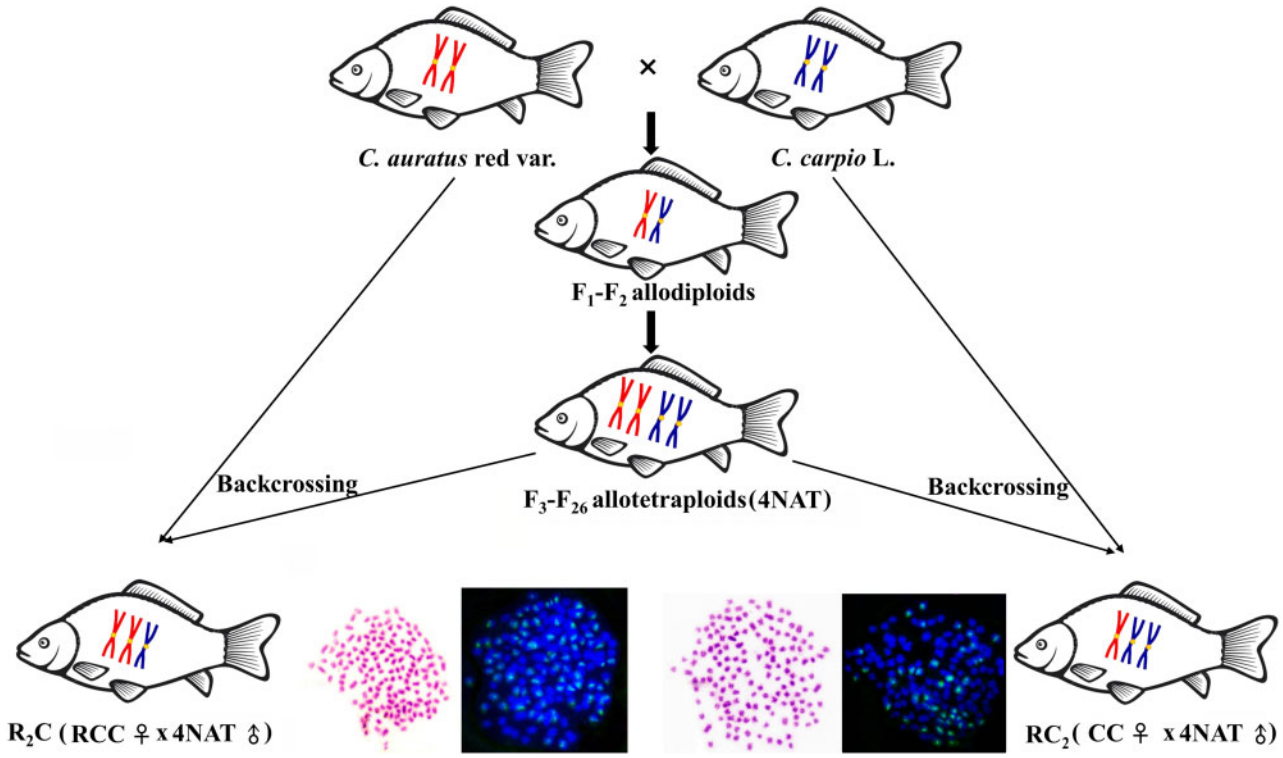

Figure 1. Genotypes of two cyprinids and their four types of hybrid offspring. R_2C and RC_2 derived from backcrossing of allotetraploid of *C. auratus* red var. (♀) × *C. carpio* L. (♂). One hundred signals were detected in R_2C (100 chromosomes from *C. auratus* red var. and other 50 from *C. carpio* L.). Fifty signals were detected in RC_2 (50 chromosomes from *C. auratus* red var. and other 100 from *C. carpio* L.). The specific-probe of *C. auratus* red var. was detected using FISH.[11] Color figures are available at *DNARES* online.

(Supplementary Fig. S1). In the backcrossing of F_{18}, the increasing number of R-/C-NEGs was observed in both non-orthologous and orthologous of the liver tissue (Supplementary Fig. S1), while the decreasing number of R-/C-SEGs was detected in them.

Next, we explored whether the various changes in gene silencing and novel gene expression were led by divergent effects of the two genotyping triploid fishes (R_2C and RC_2). Analyses of gene expression data from the liver exhibited more C-SEGs in R_2C than in RC_2 (non-orthologous genes: 81 vs. 19; orthologous genes: 40 vs. 2), and more R-NEGs in R_2C than in RC_2 (non-orthologous genes: 595 vs. 405, orthologous genes: 237 vs. 185). In contrast, there were fewer R-SEGs in R_2C than in RC_2 (non-orthologous genes: 105 vs. 161, orthologous genes: 37 vs. 45) and fewer C-NEGs in R_2C than in RC_2 (non-orthologous genes: 56 vs. 128, orthologous genes: 19 vs. 51). The similar trends were detected in the muscle tissues (Supplementary Fig. S1).

3.3. Differential global expression

Focusing on extent of changes in the polymorphic expression, we analysed the global expression of both R-/C-homoeologs in the 16,558 orthologous gene pairs. The greatest number of DEGs between the inbred maternal *C. auratus* red var. and its progeny was in the triploids (2,091 DEGs in R_2C and 2,371 DEGs in RC_2) (Fig. 2a). Meanwhile, the greatest number of DEGs between the inbred paternal *C. carpio* L. and its progeny was also in the triploids (735 DEGs in R_2C and 400 DEGs in RC_2) (Fig. 2a). The extents of differential gene expression in R_2C and RC_2 were larger than ones between the inbred parents and other hybrids (F_1 and F_{18}) (Fig. 2a), revealing that cumulative effects occurred in global expression accompanied with genotypic changes. However, in the muscle, there were fewer DEGs between R_2C and maternal R than between R_2C and paternal C, and vs. in RC_2 (Fig. 2b). These results suggested different genetic regulation between the liver and muscle.

3.4. Differential homoeolog expression analysis

To detect homoeolog expression in the hybrids, a number of 533,453 species-specific SNPs were used to detect 16,558 orthologous gene pairs of the two inbred parent's genomes using all–against–all reciprocal BLAST. We selected 4,695 orthologous gene pairs that were expressed in the liver of the two inbred parents and their hybrids. The distribution of R- and C-homoeolog expression was determined in the two inbred parents and their hybrids (Fig. 3). The expression levels of the R- and C-homoeologs differed significantly between the two inbred parents (*t*-test: $P = 0.0086$ in liver, $P = 0.0461$ in muscle), and also differed significantly among the hybrids with different genotypes ($P < 0.0001$), except in the muscle of RC_2 ($P = 0.072$) (Fig. 3). Differences in expression levels between the all two homoeologs were found in R_2C in the liver (AD = 1.87) and in the muscle (AD = 1.79). However, similar homoeolog expression levels between the two subgenomes were detected in both the liver and the muscle of RC_2, despite the imbalance of the two subgenomes.

The changes in R- and C-homoeolog expression levels led us to investigate the correlation between R-/C-homoeolog expression levels in the hybrids and those in their inbred parents. We calculated the

Downloaded from https://academic.oup.com/dnaresearch/article-abstract/26/6/485/5688746 by Hunan Normal University user on 23 March 2020

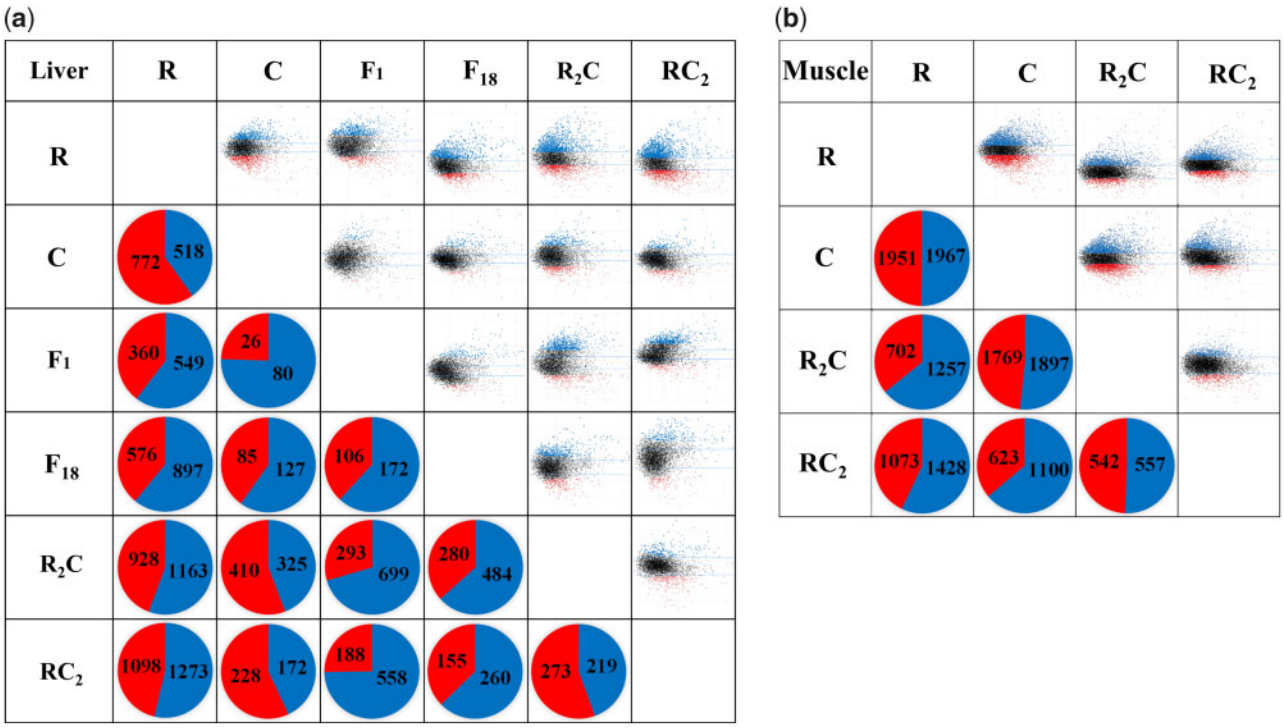

Figure 2. Differential gene expression analyses of global expression levels of the allopolyploid complex. (a) Differential expression in liver tissue [e.g. 518 up-regulated genes of R (blue) and 772 up-regulated genes of C (red) were detected between *C. auratus* red var. and *C. carpio* L.]. The blue dot in MA plot and blue colour in pie graph represent up-regulated in R, while red dot and red colour represent up-regulated in C. (b) Differential expression in muscle tissue. Color figures are available at *DNARES* online.

coefficient of determination for homoeolog expression levels in liver tissue between hybrids and their respective parents. The highest coefficient of determination for C-homoeologs was in F_1 (R^2= 0.49) and that for R-homoeologs was in F_{18} (R^2 = 0.59) (Fig. 3c). In addition, the coefficients of determination were higher for the muscle than the liver (Fig. 3c) and higher for R-homoeologs than C-homoeologs (Supplementary Tables S2 and S3).

3.5. Determination of R incomplete dominance and dosage compensation

To explore the reasons for the differences in magnitude of R- and C-homoeolog expression levels, we set two thresholds to determine the R- and C-homoeolog expression bias (R-HEB and C-HEB, respectively). The same maternal R-HEB was detected in the liver of three fish types (ratio of R-HEB: 87.45% in F_1, 73.18% in R_2C, and 73.63% in RC_2), while R- and C-HEB were balanced in parents based on $E_{in\ silico\ \mathrm{hybrid}}$ (Fig. 4a and Table 1). This result confirmed the existence of R incomplete dominance in this allopolyploid complex. Interestingly, although there was a 1:1 ratio of R:C chromosomes in F_1 and F_{18}, R-HEB and C-HEB differed markedly between them. We detected similar R-HEB and C-HEB in the two triploid fish, although they had different ratios of R:C chromosomes. These results further suggested that different regulation mechanisms might occur in these hybrids. However, the maternal R-HEB was high in R_2C and was greater in the muscle (87.17%) than in the liver (73.18%). We detected slight paternal C-HEB in the muscle of RC_2 (52.50%) (Table 1). These results further indicated that the regulation mechanisms might differ between the liver and the muscle.

To explore the potential mechanisms related to changes of homoeolog expression in the triploids, we tried to calculate the unknown effect under R incomplete dominance that had been observed in F_1 and F_{18} individuals in this allopolyploid complex. The reverse direction of the unknown effect was observed in liver of RC_2 and R_2C (Fig. 4b and c). After calculating the values of $E_{in\ silico\ \mathrm{R2C}}$ and $E_{in\ silico\ \mathrm{RC2}}$, the dosage compensation in liver of RC_2 and R_2C were determined by calculating the difference between $E_{in\ silico\ \mathrm{R2C}}$/$E_{in\ silico\ \mathrm{RC2}}$ and $E_{in\ silico\ \mathrm{hybrid}}$, relative to the expression level of parental R and C (Fig. 4b and c). Under this situation, we also tried to assess the regulation mechanism in the muscle. The same direction of HEB and a similar magnitude of expression indicated that only R incomplete dominance affected gene expression in the muscle (Fig. 4d).

3.6. Expression of growth-related genes under R incomplete dominance and dosage compensation

To investigate the potential relationship between growth heterosis and R incomplete dominance/dosage compensation, we calculated the E_{bias} and $E_{in\ silico\ \mathrm{hybrid}}$ values for the growth-regulated genes (Supplementary Table S4). In these analyses, the two triploid fishes were clustered with each other (Fig. 5a). For investigating the relationship of gene expression and growth heterosis, the expression values of the growth-regulated genes were calculated from *in silico* parents to the two triploid progenies. There were high ratios of R-HEB genes (76.74% in R_2C and 72.10% in RC_2) among the 43 genes in the liver , and high ratios of R-HEB genes (91.14% in R_2C and 55.70% in RC_2) among the 79 genes in the muscle (Fig. 5b). Further analyses showed that the most (28 in R_2C and 27 in RC_2) of

Downloaded from https://academic.oup.com/dnaresearch/article-abstract/26/6/485/5688746 by Hunan Normal University user on 23 March 2020

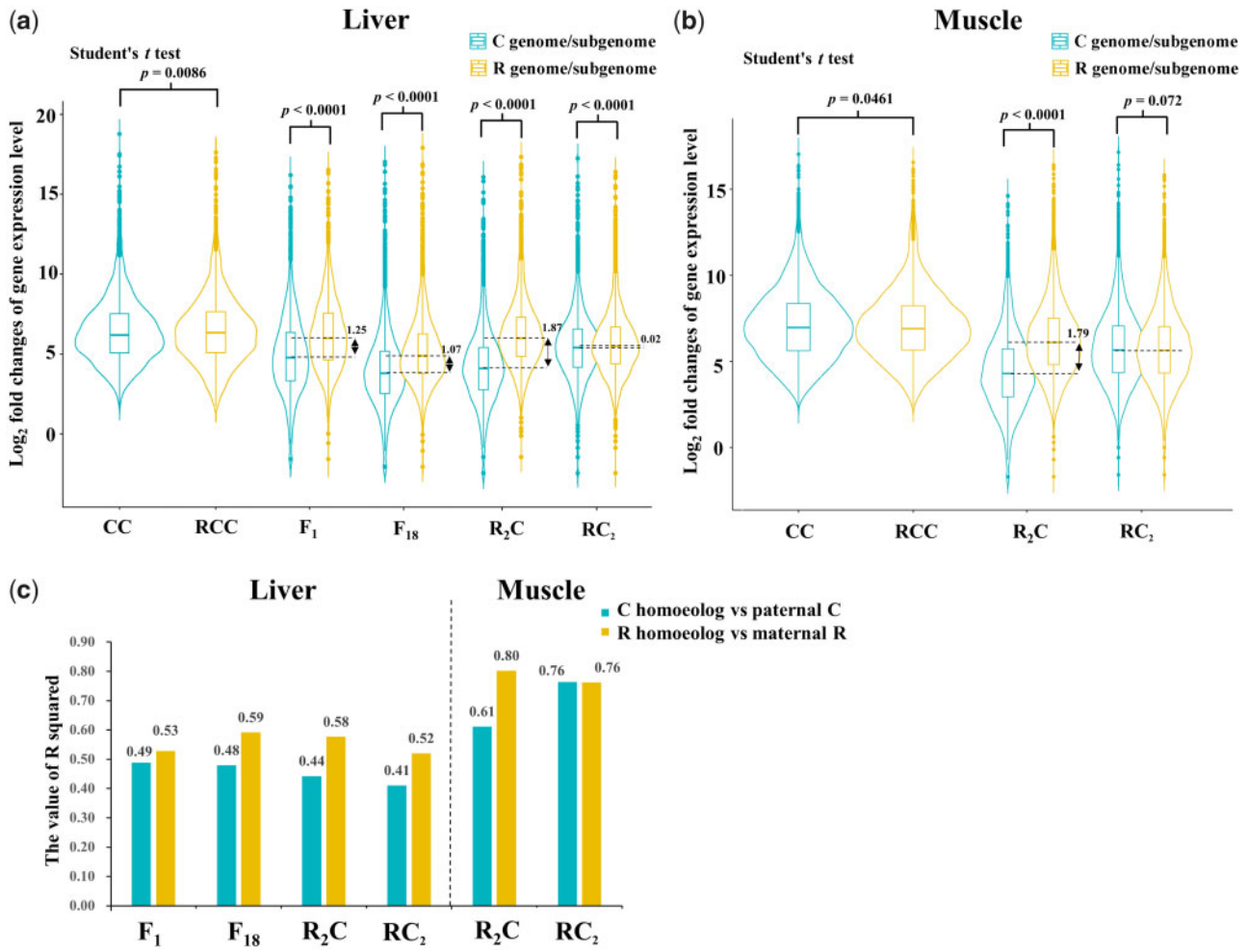

Figure 3. Distribution of R- and C-homoeolog expression values in allopolyploid complex. (a) Homoeolog expression level in liver. (b) Homoeolog expression level in muscle. The AD values was calculated between R- and C-homoeolog expression in hybrids, for example, the AD was 1.25 in liver of F_1. (c) Coefficients of determination calculated for R-/C-homoeolog expression between hybrids and their respective inbred parents ($P < 0.0001$). The values of $\log_2$ fold changes of gene expression level present the values of $\log_2$ normalized read counts.

the 43 growth-regulated genes in the liver of the allotriploids exhibited greater R-HEB than that in their parents (Fig. 5b). Sixty-six of 79 growth-regulated genes in the muscle of R_2C also had greater R-HEB than that in their parents, while only 29 of 79 genes in ones of RC_2 had greater R-HEB than that in their parents (Fig. 5b).

4. Discussion

The complex model of allopolyploid lineages provides abundant genotypes and phenotypes. Transcriptome analysis of them can be used to investigate the potential regulation mechanisms related to allopolyploidization.[32] In recent studies, dosage compensation and incomplete dominance have emerged as important processes in the formation of qualitative and quantitative traits, including heterosis.[9,33] In this study, R incomplete dominance was detected in the allotetraploid lineage including the allodiploid F_1, allotetraploid F_{18},[14] and the allotriploids R_2C and R_2C. We also studied dosage compensation in the two triploid genotypes (R_2C and R_2C) with two sets of R or C chromosomes, because integrated analyses of dosage compensation and incomplete dominance can shed light on the potential mechanisms regarding heterosis.

The successful establishment of allotetraploids of *C. auratus* red var. (♀) × *C. carpio* L. (♂) provided a chance to obtain triploids showing heterosis by backcrossing breeding.[11,12,31] Although the allotetraploids were considered to be a stable allotetraploid population based on chromosome number analysis,[11] there was lack of genotypic data of the heterozygous gamete. Therefore, the consistent genotypes in R_2C and R_2C individuals supported the stable heterozygous gametes from female allotetraploid individuals (Fig. 1). Assessment of gene silencing and novel gene expression in the allodiploid F_1 and allotetraploid F_{18} revealed increases in R-/C-NEGs and R-/C-SEGs (Supplementary Fig. S1), suggesting that the constraint of the expression of polymorphism gradually weakened through the whole genome duplication in the hybrids. The duplication of genes, chromosomes, and whole genome could contribute to expression divergence.[34,35] Then, the triploids with asymmetric subgenomes showed increased R-/C-NEGs and decreased R-/C-SEGs in the liver. This result suggested that two sets of the homologous chromosomes

Downloaded from https://academic.oup.com/dnaresearch/article-abstract/26/6/485/5688746 by Hunan Normal University user on 23 March 2020

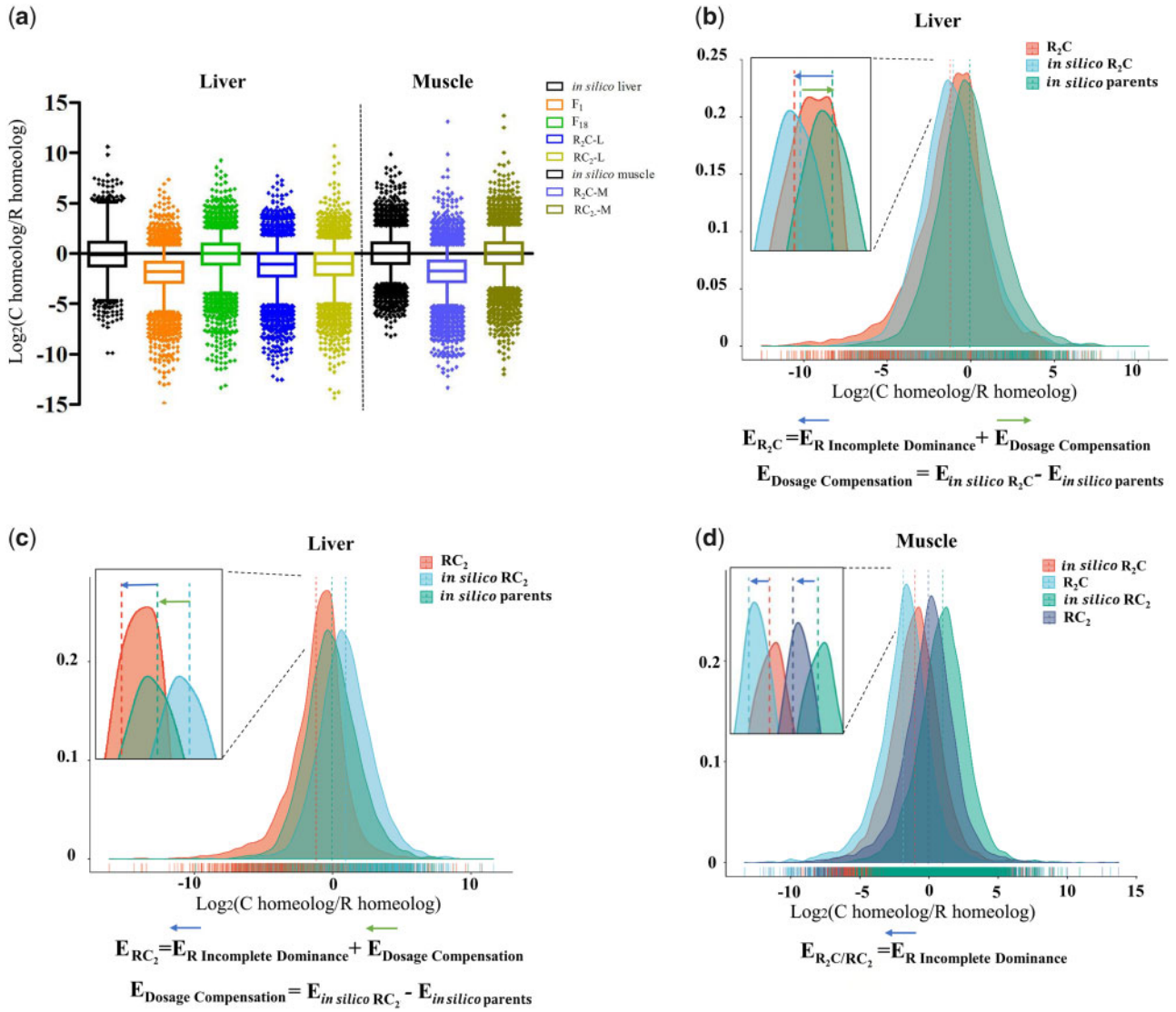

Figure 4. Effect of R incomplete dominance and dosage compensation in the allopolyploid complex. (a) Distribution of $\log_2$(C-homoeolog/R-homoeolog) values in liver and muscle of *in silico* parents and their hybrid offspring. Scatter point represents 5% highest and lowest values. (b–d) Effects of R incomplete dominance and dosage compensation in the two triploid fish. Blue arrow represents direction and magnitude of R incomplete dominance; green arrow represents direction and magnitude of dosage compensation. Color figures are available at *DNARES* online.

Table 1. Homoeologue expression bias (HEB) was calculated on the two thresholds in hybrids

	No. of genes with potential R-HEB[a]	No. of genes with potential C-HEB[a]	No. of genes with R-HEB[b]	No. of genes with C-HEB[b]	Total
Liver					
In silico parents	2,484 (52.91%)	2,211 (47.09%)	656 (13.97%)	612 (13.04%)	4,695
F_1	4,106 (87.45%)	589 (12.55%)	2,094 (44.60%)	93 (1.98%)	4,695
F_{18}	2,303 (49.05%)	2,392 (50.95%)	646 (13.76%)	409 (8.71%)	4,695
R_2C	3,436 (73.18%)	1,259 (26.82%)	1,371 (29.20%)	203 (4.32%)	4,695
RC_2	3,457 (73.63%)	1,238 (26.37%)	1,244 (26.50%)	155 (3.30%)	4,695
Muscle					
In silico parents	3,653 (49.21%)	3,770 (50.79%)	897 (12.08%)	810 (10.91%)	7,423
R_2C	6,471 (87.17%)	952 (12.83%)	3,107 (41.86%)	147 (1.98%)	7,423
RC_2	3,526 (47.50%)	3,897 (52.50%)	952 (12.83%)	815 (10.98%)	7,423

[a]If the value of E_{bias} is > 0, it is considered as potential C-biased in hybrids. If the value of E_{bias} is < 0, it is considered as potential R-biased.
[b]If the value of E_{bias} is > 2, it is considered as C-biased in hybrids. If the value of E_{bias} is < -2, it is considered as R-biased.

Downloaded from https://academic.oup.com/dnaresearch/article-abstract/26/6/485/5688746 by Hunan Normal University user on 23 March 2020

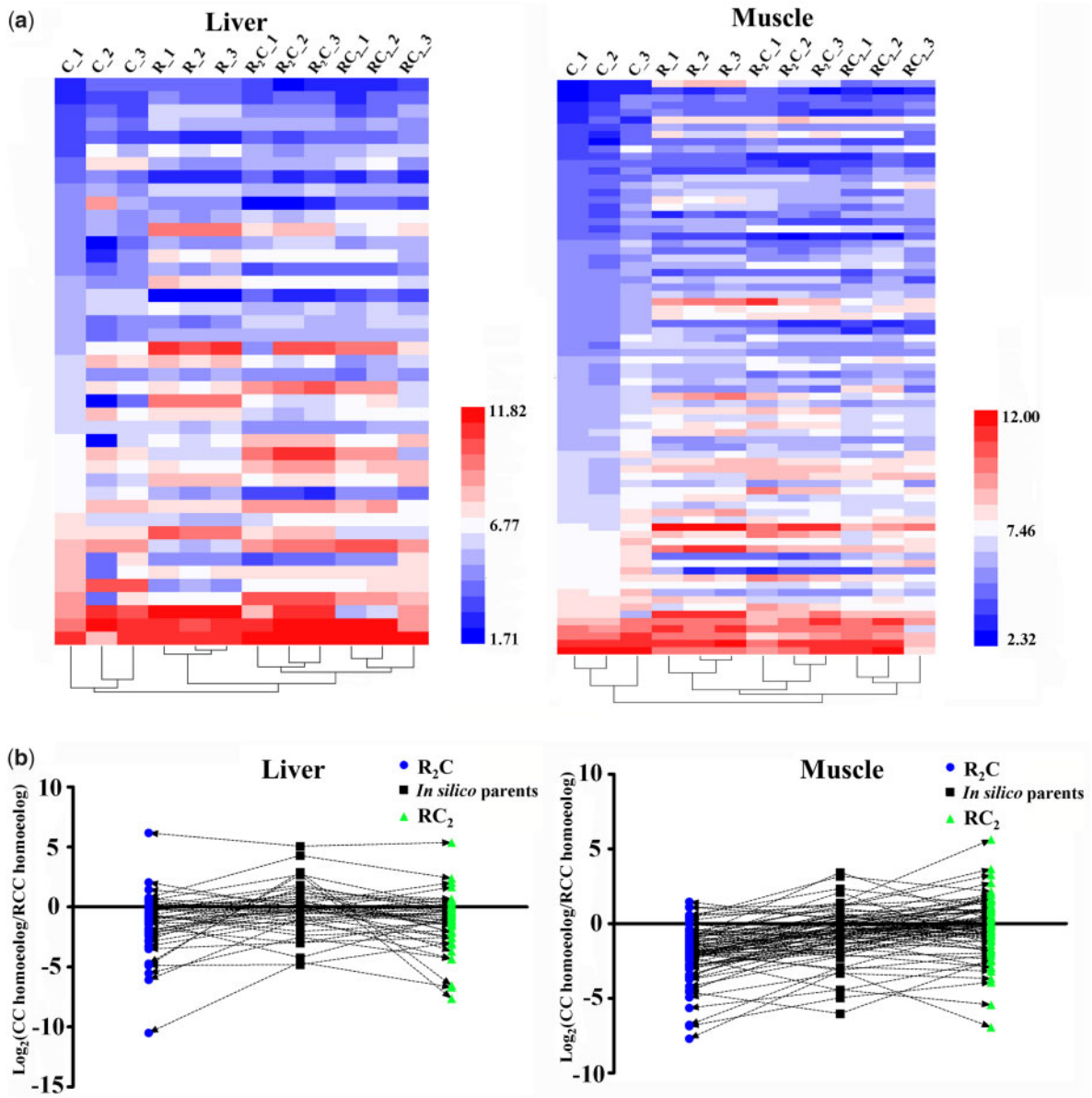

Figure 5. Expression profiles of growth-regulated genes in the two triploids and their inbred parents. (a) Values of global expression levels in liver (43 genes) and muscle (79 genes). Euclidean distance and maximum linkage clustering method were used. The ordinate of heatmap is determined by the values of $\log_2$ normalized read counts. (b) $\log_2$(C-homoeolog/R-homoeolog) values in liver and muscle of *in silico* parents and the two triploid offspring.

could inhibit silencing and promote novel expression of corresponding homoeologs based on the copy number of R-/C-homologous chromosomes. These changes will lead to the emergence of dominance of partial homoeologs and contribute to heterosis.[2] Global expression changes also indicated that constraints of gene expression in the inbred parents gradually weakened as a result of genotypic changes in hybrids (Fig. 2). Moreover, the most DEGs were detected in triploids, suggesting that the asymmetric subgenomes could rapidly and greatly speeded up this process, which may help us understand of the emergence of heterosis (Fig. 2).

Analyses of the magnitude of gene expression in the four hybrids (F_1, F_{18}, R_2C, and RC_2) exhibited that the greatest HEB was in the allotriploids R_2C (AD = 1.87 in the liver and AD = 1.79 in the muscle) (Fig. 3). This phenomenon revealed that the imbalance of the two homoeologs at the genomic level might contribute to greater HEB, although smallest HEB was found in RC_2 (Fig. 3). Analyses of homoeolog expression between inbred parents and hybrids revealed higher coefficients of determination for R-homoeologs than for C-homoeologs (Supplementary Tables S2 and S3), indicating that the R genome plays the dominant role in these hybrid fish. Therefore, evolution could be occurring rapidly with respect to C-homoeolog sequences and expression. Based on the *in silico* values of $E_{in\ silico\ hybrid}$, $E_{in\ silico\ R2C}$, and $E_{in\ silico\ RC2}$, we determined the direct effects and the magnitude of the regulation mechanisms. Both dosage compensation and the R incomplete dominance effects were detected in the liver of R_2C and RC_2. The different directions of HEB led by dosage compensation and R incomplete dominance in the liver of R_2C were partly counteracted by R-HEB. Meanwhile, the same direction

Downloaded from https://academic.oup.com/dnaresearch/article-abstract/26/6/485/5688746 by Hunan Normal University user on 23 March 2020

of HEB led by dosage compensation and R incomplete dominance together shaped the increased trend of R-HEB in the liver of RC_2 (Fig. 4). Only R incomplete dominance was detected as a regulation mechanism in the muscle of the two triploids. This may be related to epigenetic regulation such as miRNAs and RNA-directed DNA methylation in the allopolyploids.[36] In addition, the differences in the magnitude of expression between the liver and muscle may relate to mosaic dominance (Fig. 3 and Table 1).[37] Further research using more genotypes is required to explore this possibility.

Both R incomplete dominance and the dosage compensation shaped mRNA expression in these triploids. These effects made the changes in mRNA expression at global and homoeolog levels more complex, which made it difficult to understand the mechanism of regulation of biological traits in the triploids. The phenomenon of growth heterosis in these fishes is well known but still poorly understood.[10] Focussing on growth-regulated genes, comparisons between the two inbred parents and their triploid offspring at the global expression level revealed 43 genes of interest in the liver and 79 in the muscle. The R incomplete dominance of growth-regulated genes, which is related to differences in growth vigour between the two inbred parents, may contribute to growth heterosis (Fig. 5). In addition, the magnitude of R-HEB in the 79 growth-regulated genes of RC_2 muscle (55.70%) was slighter than that in R_2C (91.14%) (Fig. 5b), while individuals of RC_2 exhibited more rapid growth than R_2C. These results suggested that the appropriate balance of R-/C-HEB in the growth-regulated genes could effectively contribute to the more obvious growth heterosis in the two triploids. However, the further testing is needed to validate it. A similar phenomenon has been described in rice,[2,38] tomato,[39] and maize.[18] Although few studies have focussed on the relationship between R incomplete dominance and heterosis in animals,[40] there are some reports on this relationship in salmonids[41] and in an allotetraploid of goldfish × common carp.[14] Although dosage compensation could be related to heterosis,[5,9] our results suggested that R incomplete dominance contributes to heterosis in the muscle of triploid fishes.

Supplementary data

Supplementary data are available at *DNARES* online.

Accession numbers

All short-read RNA-seq data have been deposited in the Short Read Archive under the following accession numbers: SRS4475349, SRS4475350, SRS4475351, SRS4475352, SRS4475353, and SRS4475354.

Funding

This research was supported by the National Natural Science Foundation of China (Grant No. 31702334, 31730098, and 31430088), the Key Research and Development Program of Hunan Province (Grant No. 2018NK2072), the Earmarked Fund for China Agriculture Research System (Grant No. CARS-45), the Key Research and Development Project of Hunan Province (Grant No. 2016NK2128), the Key Research and Development Program of Hunan Province (Grant No. 2018NK2072), Hunan Provincial Natural Science and Technology Major Project (Grant No. 2017NK1031), and the Cooperative Innovation Center of Engineering and New Products for Developmental Biology of Hunan Province (Grant No. 20134486).

Conflict of interest

None declared.

References

1. Nguyen, D.K. and Disteche, C.M. 2006, Dosage compensation of the active X chromosome in mammals, *Nat. Genet.*, **38**, 47–53.
2. Birchler, J.A. 2015, Heterosis: the genetic basis of hybrid vigour, *Nat. Plants*, **1**, 15020.
3. Degenhardt, R.F. and Bonham-Smith, P.C. 2008, Transcript profiling demonstrates absence of dosage compensation in Arabidopsis following loss of a single RPL23a paralog, *Planta*, **228**, 627–40.
4. Pala, I., Coelho, M.M. and Schartl, M. 2008, Dosage compensation by gene-copy silencing in a triploid hybrid fish, *Curr. Biol.*, **18**, 1344–8.
5. Ren, L., Tang, C., Li, W., et al. 2017, Determination of dosage compensation and comparison of gene expression in a triploid hybrid fish, *BMC Genomics*, **18**, 38.
6. Lucchesi, J.C., Kelly, W.G. and Panning, B. 2005, Chromatin remodeling in dosage compensation, *Annu. Rev. Genet.*, **39**, 615–51.
7. Man-Sai, A., Francisco, S.-C. and Mora-Rodriguez, R.A. 2016, A biocomputational platform for the automated construction of large-scale mathematical models of miRNA-transcription factor networks for studies on gene dosage compensation. In: *2016 IEEE 36th Central American and Panama Convention (CONCAPAN XXXVI)*. IEEE, pp. 1–7.
8. Huang, X., Yang, S., Gong, J., et al. 2015, Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis, *Nat. Commun.*, **6**, 6258.
9. Yao, H., Gray, A.D., Auger, D.L., et al. 2013, Genomic dosage effects on heterosis in triploid maize, *Proc. Natl. Acad. Sci. USA*, **110**, 2665–9.
10. Shen, J.M., Liu, S.J., Sun, Y.D., et al. 2006, A new type of triploid crucian crap-red crucian carp (female) × allotetraploid (male), *Prog. Nat. Sci.*, **16**, 1348–52.
11. Liu, S., Luo, J., Chai, J., et al. 2016, Genomic incompatibilities in the diploid and tetraploid offspring of the goldfish × common carp cross, *Proc. Natl. Acad. Sci. USA*, **113**, 1327–32.
12. Chen, S., Wang, J., Liu, S., et al. 2009, Biological characteristics of an improved triploid crucian carp, *Sci. China C Life Sci.*, **52**, 733–8.
13. Zhong, H., Zhou, Y., Liu, S., et al. 2012, Elevated expressions of GH/IGF axis genes in triploid crucian carp, *Gen. Comp. Endocrinol.*, **178**, 291–300.
14. Ren, L., Cui, J., Wang, J., et al. 2017, Analyzing homoeolog expression provides insights into the rediploidization event in gynogenetic hybrids of *Carassius auratus* red var. × *Cyprinus carpio*, *Sci. Rep.*, **7**, 13679.
15. Ren, L., Li, W., Tao, M., et al. 2016, Homoeologue expression insights into the basis of growth heterosis at the intersection of ploidy and hybridity in Cyprinidae, *Sci. Rep.*, **6**, 27040.
16. Greig, D. 2007, A screen for recessive speciation genes expressed in the gametes of F1 hybrid yeast, *PLoS Genet.*, **3**, e21.
17. Yang, J., Mezmouk, S., Baumgarten, A., et al. 2017, Incomplete dominance of deleterious alleles contributes substantially to trait variation and heterosis in maize, *PLoS Genet.*, **13**, e1007019.
18. Hochholdinger, F. and Baldauf, J.A. 2018, Heterosis in plants, *Curr. Biol.*, **28**, R1089–92.
19. Li, A., Liu, D., Wu, J., et al. 2014, mRNA and small RNA transcriptomes reveal insights into dynamic homoeolog regulation of allopolyploid heterosis in nascent hexaploid wheat, *Plant Cell*, **26**, 1878–900.
20. Xiao, J., Kang, X., Xie, L., et al. 2014, The fertility of the hybrid lineage derived from female Megalobrama amblycephala × male Culter alburnus, *Anim. Reprod. Sci.*, **151**, 61–70.
21. Andrews, S. 2010, FastQC: a quality control tool for high throughput sequence data, *Bioinformatics*, **35**, 871–3.
22. Bolger, A.M., Lohse, M. and Usadel, B. 2014, Trimmomatic: a flexible trimmer for Illumina sequence data, *Bioinformatics*, **30**, 2114–20.
23. Xu, P., Zhang, X., Wang, X., et al. 2014, Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*, *Nat. Genet.*, **46**, 1212–9.

Downloaded from https://academic.oup.com/dnaresearch/article-abstract/26/6/485/5688746 by Hunan Normal University user on 23 March 2020

24. Robinson, M.D., McCarthy, D.J. and Smyth, G.K. 2010, edgeR: a bioconductor package for differential expression analysis of digital gene expression data, *Bioinformatics*, **26**, 139–40.
25. Harris, R.S. 2007, Improved pairwise alignment of genomic DNA. PhD Thesis, Pennsylvania State University, University Park, PA.
26. McKenna, A., Hanna, M., Banks, E., et al. 2010, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data, *Genome Res*., **20**, 1297–303.
27. Schaefke, B., Emerson, J.J., Wang, T.Y., et al. 2013, Inheritance of gene expression level and selective constraints on trans- and cis-regulatory changes in yeast, *Mol. Biol. Evol*., **30**, 2121–33.
28. McManus, C.J., Coolon, J.D., Duff, M.O., et al. 2010, Regulatory divergence in Drosophila revealed by mRNA-seq, *Genome Res*., **20**, 816–25.
29. Quackenbush, J. 2002, Microarray data normalization and transformation, *Nat. Genet*., **32**, 496–501.
30. Liu, S., Liu, Y., Zhou, G., et al. 2001, The formation of tetraploid stocks of red crucian carp × common carp hybrids as an effect of interspecific hybridization, *Aquaculture*, **192**, 171–86.
31. Liu, S. 2010, Distant hybridization leads to different ploidy fishes, *Sci. China Life Sci*., **53**, 416–25.
32. Liu, S., Qin, Q., Xiao, J., et al. 2007, The formation of the polyploid hybrids from different subfamily fish crossings and its evolutionary significance, *Genetics*, **176**, 1023–34.
33. Goff, S.A. and Zhang, Q. 2013, Heterosis in elite hybrid rice: speculation on the genetic and biochemical mechanisms, *Curr. Opin. Plant Biol*., **16**, 221–7.
34. Wang, S. and Chen, Y. 2019, Fine-tuning the expression of duplicate genes by translational regulation in arabidopsis and maize, *Front. Plant Sci*., **10**, 534.
35. Wang, Y., Wang, X. and Paterson, A.H. 2012, Genome and gene duplications and gene expression divergence: a view from plants, *Ann. N. Y. Acad. Sci*., **1256**, 1–14.
36. Ng, D.W., Lu, J. and Chen, Z.J. 2012, Big roles for small RNAs in polyploidy, hybrid vigor, and hybrid incompatibility, *Curr. Opin. Plant Biol*., **15**, 154–61.
37. Tan, C.C. 1946, Mosaic dominance in the inheritance of color patterns in the lady-bird beetle, Harmonia axyridis, *Genetics*, **31**, 195.
38. Birchler, J.A., Yao, H., Chudalayandi, S., et al. 2010, Heterosis, *Plant Cell*, **22**, 2105–12.
39. Lippman, Z.B. and Zamir, D. 2007, Heterosis: revisiting the magic, *Trends Genet*., **23**, 60–6.
40. Wu, X., Li, R., Li, Q., et al. 2016, Comparative transcriptome analysis among parental inbred and crosses reveals the role of dominance gene expression in heterosis in Drosophila melanogaster, *Sci. Rep*., **6**, 21124.
41. Maynard, B.T., Taylor, R.S., Kube, P.D., et al. 2016, Salmonid heterosis for resistance to amoebic gill disease (AGD), *Aquaculture*, **451**, 106–12.

Downloaded from https://academic.oup.com/dnaresearch/article-abstract/26/6/485/5688746 by Hunan Normal University user on 23 March 2020