SCIENCE CHINA
Life Sciences

•RESEARCH PAPER•

https://doi.org/10.1007/s11427-020-1749-9

Diverse transcriptional patterns of homoeologous recombinant transcripts in triploid fish (Cyprinidae)

Li Ren[1,2†], Xueyin Zhang[1,2†], Jiaming Li[1,2†], Xiaojing Yan[1,2], Xin Gao[1,2], Jialin Cui[1,2], Chenchen Tang[1,2] & Shaojun Liu[1,2*]

[1]*State Key Laboratory of Developmental Biology of Freshwater Fish, Hunan Normal University, Changsha 410081, China;*
[2]*College of Life Sciences, Hunan Normal University, Changsha 410081, China*

Received July 17, 2020; accepted August 13, 2020; published online January 6, 2021

Homoeologous recombination (HR), the exchange of homoeologous chromosomes, contributes to subgenome adaptation to diverse environments by producing various phenotypes. However, the potential relevance of HR and innate immunity is rarely described in triploid cyprinid fish species. In our study, two allotriploid genotypes (R_2C and RC_2), whose innate immunity was stronger than their inbred parents (*Carassius auratus* red var. and *Cyprinus carpio* L.), were obtained from backcrossing between male allotetraploids of *C. auratus* red var.×*C. carpio* L. and females of their two inbred parents, respectively. The work detected 140 HRs shared between the two triploids at the genomic level. Further, transcriptions of 54 homoeologous recombinant genes (HRGs) in R_2C and 65 HRGs in RC_2 were detected using both Illumina and PacBio data. Finally, by comparing expressed recombinant reads to total expressed reads in each of the genes, a range of 0.1%–10% was observed in most of the 99–193 HRGs, of which six recombinant genes were classified as "response to stimulus". These results not only provide a novel way to predict HRs in allopolyploids based on cross prediction at both genomic and transcriptional levels, but also insight into the potential relationship between HRs related to innate immunity and adaptation of the triploids and allotetraploids.

homoeologous recombination, recombinant transcripts, triploid, innate immunity

Citation: Ren, L., Zhang, X., Li, J., Yan, X., Gao, X., Cui, J., Tang, C., and Liu, S. (2021). Diverse transcriptional patterns of homoeologous recombinant transcripts in triploid fish (Cyprinidae). Sci China Life Sci 64, https://doi.org/10.1007/s11427-020-1749-9

INTRODUCTION

Polyploidy and hybridization are two important processes that drive the formation of novel genotypes and phenotypes by reshaping genomes (Soltis and Soltis, 1999; Mallet, 2007). Polyploidy and hybridization are common and provide evolutionary advantages for many plants (Ni et al., 2009) and some animals (Dickerson, 1973; Bayne et al., 1999; Zhou and Gui, 2017). In hybridization, the occurrence of polyploidy events helps enhance resistance to deleterious effects caused by the coexistence of two or more different subgenomes (Liu et al., 2016). Additionally, hybridization can accelerate speciation if it results in reproductive isolation from inbred diploid parents (Soltis et al., 2009). Moreover, the coexistence of two or more subgenomes can lead to genomic changes, including genome rearrangements, which lead to recombination between homoeologous genes of different subgenomes (Gaeta and Chris Pires, 2010).

Homoeologous recombination (HR), the exchange of genetic information between homoeologs of different subgenomes, plays a role in generating of species diversity during reproduction. HR has been observed in many plants and some animals, including sunflower (Rieseberg et al., 1995), water frog (Christiansen and Reyer, 2009), and allotetraploid cyprinid fish (Liu et al., 2016). Additionally, HR may play essential roles in preventing conflicts in homo-

†Contributed equally to this work
*Corresponding author (email: lsj@hunnu.edu.cn)

eolog chromosome pairing during meiosis, and lead to the fusion of subgenomes originating from different species (Gaeta and Chris Pires, 2010). However, potential mechanisms of regulation of HR have been rarely addressed in non-model species, apart from two studies that have reported that the recombinases *rad51* and *dmc1* have essential roles in fusing homologous DNA molecules (Shinohara et al., 2000; Masson and West, 2001).

Similar to gene duplication, sequence divergence and HR are also major contributors to gene evolution (Eichler, 2001), and can probably produce new genes from existing sequences. If gene recombination of homoeologs occurs in non-coding sequence regions, it can lead to regulatory module exchange in the subgenomes. Consequently, the expression of the two homoeologous genes will change to a novel state under the control of the *cis*-regulatory sequences (Gaeta and Chris Pires, 2010; Starcevic et al., 2011). If HR occurs in coding sequences it can cause assembly of a recombined gene; the expression of recombined mRNAs and proteins may lead to new functions and alter the corresponding phenotype (Qi et al., 2007). While the disease resistance and stress resistance in allotetraploid lineages of *C. auratus* red var. (RCC)×*C. carpio* L. (CC) have been shown to be stronger than that of their inbred parents (Chen et al., 2009; Xiao et al., 2018), the genetic basis of innate immunity and adaptation is unknown (Zhang and Gui, 2018); however, some genes that were differentially expressed among the diploid parents, triploid progeny, and allotetraploid progeny were identified based on their expression profiles (Xiao et al., 2018). HR was observed in the mitochondria of triploids and potentially contributes to the production of phenotypic diversity (Guo et al., 2006). Identifying HR in immune-regulated genes may help elucidate the mechanism underlying the strong innate immune response of hybrids.

Recently, high-throughput sequencing has been used to screen HRs on a genome-wide scale (Schumer et al., 2018). Because of challenges due to sequence similarity between the homoeologous genes from the two subgenomes analyzed in this study, some homoeologs were distinguished using species-specific single nucleotide polymorphisms (SNPs). In our study, genomic and transcriptional strategies were used to assess HRs in the two triploids (R_2C and RC_2), which were obtained from the backcrossing of allotetraploid *C. auratus* red var.×*C. carpio* L. to the two diploid inbred parents (*C. auratus* red var. and *C. carpio* L). Genome-wide analysis of shared HRs of the two triploids could be used to assess conserved HRs in allotetraploid populations, whereas unique HRs in different triploid individuals could reflect different distributions of HRs in their paternal allotetraploid progenitors. Furthermore, HR detection at the transcriptional level can provide insight into the various gene expression patterns caused by HR and potential functional effects on resistance to disease and stress.

RESULTS

Distribution of genomic homoeologous recombinant genes in two triploids

We obtained two types of triploid fish by backcrossing allotetraploid *C. auratus* red var.×*C. carpio* L. (4NAT) with the two inbred parents (Figure 1 and Table S1 in Supporting Information). The genomic DNA changes that occurred in triploids were mostly from the paternal allotetraploids. A series of Illumina paired-end genomic sequencing procedures were used to detect HR occurrence (Figure 2A and Table S2 in Supporting Information). In total, 21,062 and 8,809 recombination events between the two subgenomes were observed in the R_2C and RC_2, respectively (Figure S1 in Supporting Information). Among these, most of the recombination events were unique in the two triploids, whose male parents were different 4NAT individuals (Figure S1 in Supporting Information). This diversified distribution of recombination events indicated that HR was not consistent among different individuals. After filtering, only 834 recombination events were shared between the two triploids, which indicates that the genomic recombination originated from their shared paternal allotetraploids (Figure S1 in Supporting Information). In addition, 760 shared recombination events, including those in intergenic regions and gene regions, were found in 50 homoeologous chromosome pairs, whereas only 140 recombination events of 111 genes were determined to be HR among 13,008 orthologous pairs (Figure 3 and Table S3 in Supporting Information). The gene ontology (GO) analysis indicated that 17 of these genes were classified as “response to stimulus” (GO: 0050896) (Figure S2 in Supporting Information).

Transcripts of homoeologous recombination predicted by Illumina data

To further investigate the transcription of recombinant genes in the two triploids, mRNA of their liver and muscle tissues was sequenced using Illumina sequencing. In total, 29.6 Gb and 32.6 Gb of data were obtained from R_2C and RC_2 liver, whereas 36.1 Gb and 34.5 Gb of data were obtained from

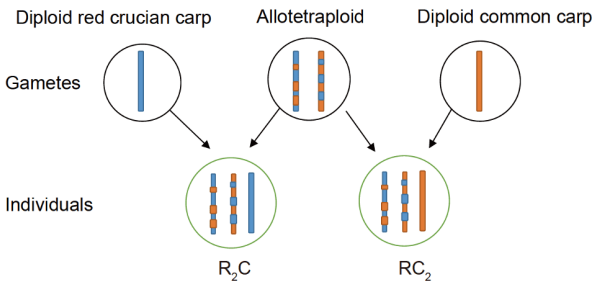

Figure 1 Procedure for creating the two types of triploid fish. The recombinant DNA sequences of triploid progeny originated from the paternal allotetraploids.

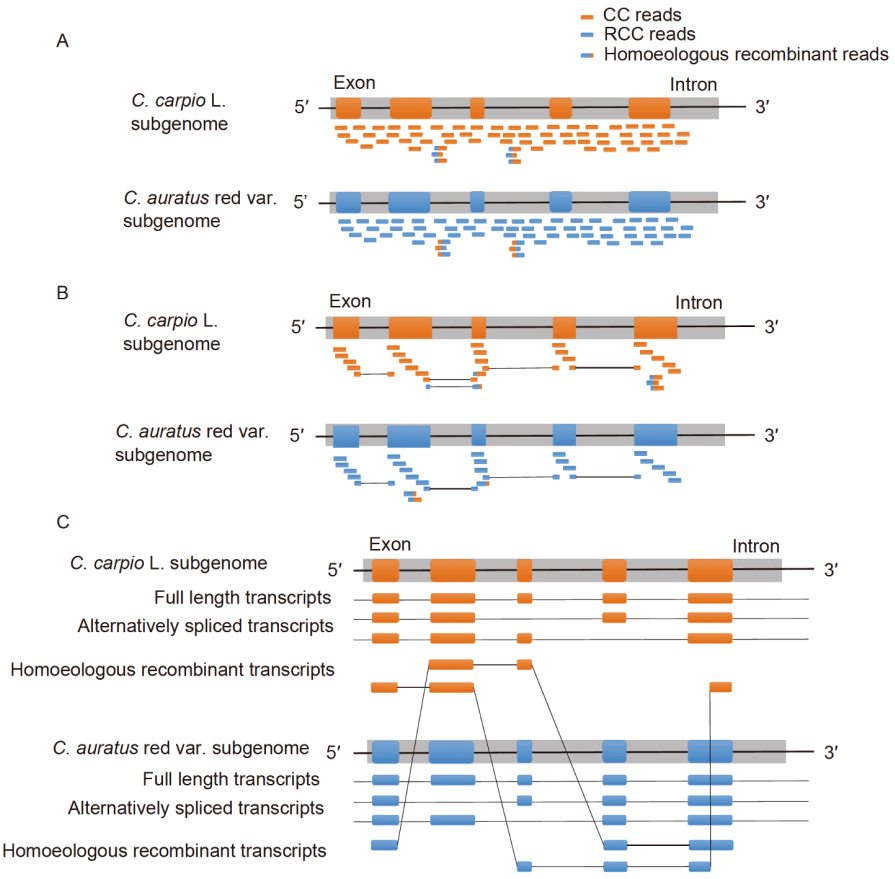

Figure 2 Schematic diagrams for the detection of homoeologous recombination based on Illumina and PacBio data. A, Homoeologous recombinant reads from Illumina data detected at the genomic level. B, Homoeologous recombinant reads from Illumina data detected at the transcriptional level. C, Homoeologous recombinant reads from PacBio data detected at the transcriptional level.

R_2C and RC_2 muscle, respectively (Table S4 in Supporting Information). After selecting suitable thresholds, the comparison of RCC and CC subgenomes described 28,148–33,215 expressed recombinant reads in liver of R_2C and 35,040–39,187 reads in RC_2, whereas 34,879–44,277 and 33,517–54,896 were detected in muscle of R_2C and RC_2, respectively (Figure 2B and Table S5 in Supporting Information). The average ratios of recombinant transcripts in each gene (0.0693%–0.0909%) were calculated from the total mapped reads. We detected 19,440–37,039 recombination events in short-length reads of the two tissues of R_2C and RC_2 based on paired-end sequencing (Table S6 in Supporting Information).

Subsequently, we analyzed the recombination events in 13,008 orthologous pairs of RCC and CC. Widespread recombinant reads were observed in 4,139–6,055 genes of RCC homoeologs and 4,442–7,004 CC homoeologous genes (Table S6 in Supporting Information). However, only 260 homoeologs shared in all samples exhibited inconsistency among different individuals and different genotypes (Figure S3 in Supporting Information). In addition, we found on average a higher number of recombination events in RC_2 than R_2C (liver: 21,642 vs. 27,616; muscle: 25,465 vs. 29,169) (Table S6 in Supporting Information).

Distribution of expressed HRGs

To better understand the HRs in the two triploid subgenomes, some of the recombination events at the chromosome level were examined, with these being determined by the incomplete genomes of the two inbred parents. In total, 254 HR gene pairs were observed in R_2C liver based on a threshold of >4× read coverage in each gene, whereas 318 were found in R_2C muscle (Figure S4 in Supporting Information). Moreover, 284 gene pairs of HRs were detected in RC_2 liver based on the same thresholds, while 312 were found in RC_2 muscle (Figure S4 in Supporting Information). The accurate prediction of averages recombinant reads in each gene pair was 41 and 44 in liver of R_2C and RC_2, 56, and 56 in muscle of R_2C and RC_2, respectively. However, most HR gene pairs (175 in liver and 247 in muscle) were detected in both R_2C and RC_2 (Figure S4 in Supporting Information). Additionally, groups of HR gene pairs were detected in liver (79 in R_2C, 109 in RC_2) and muscle (71 in R_2C,

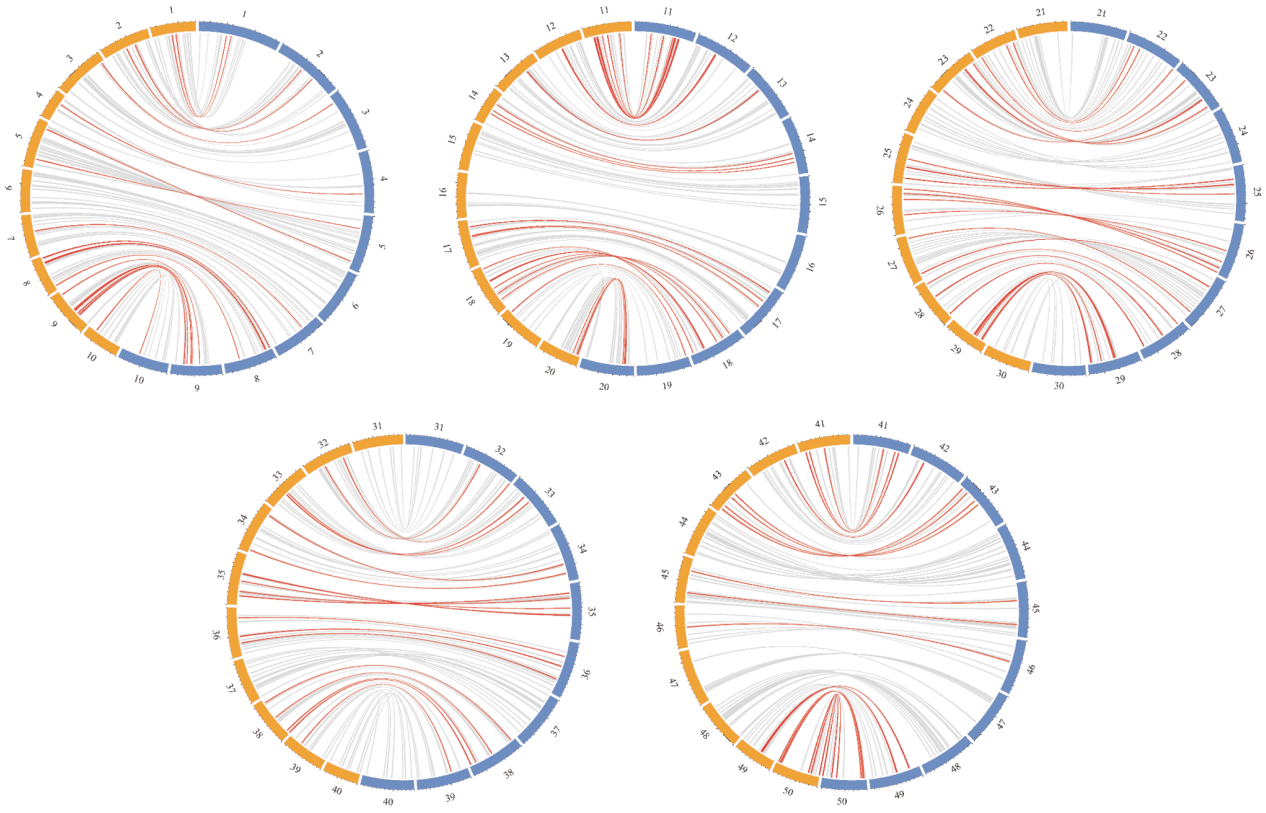

Figure 3 Shared recombination in the two triploids. The gray line represents recombination that occurred in non-syntenic blocks of the homoeologous chromosomes between the *C. auratus* red var. and *C. carpio* L. subgenomes. The red line represents HR between the two subgenomes. Outer ring, chromosomes; inner ring, line that denotes the position of recombination.

65 in RC_2) (Figure S4 in Supporting Information).

Expression level of homoeologous recombinant transcripts (HRTs)

To further investigate the effects of recombination events, we mainly studied the ratio of the shared HRTs based on the numbers of recombinant and non-recombinant reads. For the *C. auratus* red var. subgenome, a total of 152 and 193 genes shared in muscle and liver were found in R_2C and RC_2, respectively. For the *C. carpio* L. subgenome, a total of 99 and 114 genes shared in liver and muscle were found in R_2C and RC_2, respectively. (Figure 4A). Furthermore, we calculated the expressed recombinant ratio by dividing the number of recombinant reads by the total number of reads in each gene. These ratios in different samples were mainly distributed in a range of 0.1%–10% (Figure 4B–E). No differences were observed between the ratios of R_2C and RC_2 for RCC subgenomes in liver (P=0.2020) or muscle (P=0.1635), nor for CC subgenomes in liver (P=0.9045); a significant difference was, however, detected between the ratios of R_2C and RC_2 for CC subgenomes in muscle (P=0.0040) (Figure 4B–E). Additionally, significant differences were observed between R_2C liver and muscle tissues for both the RCC and CC subgenomes (P=0.0015, and P=0.0282, respectively). In contrast, while no significant difference was found between RC_2 liver and muscle tissues for the CC subgenome (P=0.9240, CC), a significant difference was observed between those of the RCC subgenome (P=0.0490) (Figure S5 in Supporting Information).

Detection of expressed HRs using PacBio sequencing

Unlike partial detection of the recombination events in fractured sequences based on paired-end sequencing, the full-length recombinant transcripts were observed in PacBio sequencing data (Figure 2C). In total, 655,012 isotypes, which included 86.78% full-length transcripts, were obtained from R_2C, while 561,453 isotypes, which included 86.42% full-length transcripts, were sequenced in RC_2 samples (Table S7 in Supporting Information). After screening, 3,049 and 3,181 recombinant transcripts were obtained from R_2C and RC_2, respectively. Among these, 1,731 recombination events in R_2C were found in coding sequences between RCC and CC, whereas 133 were distributed in non-coding sequences (Table S8 in Supporting Information). Additionally,

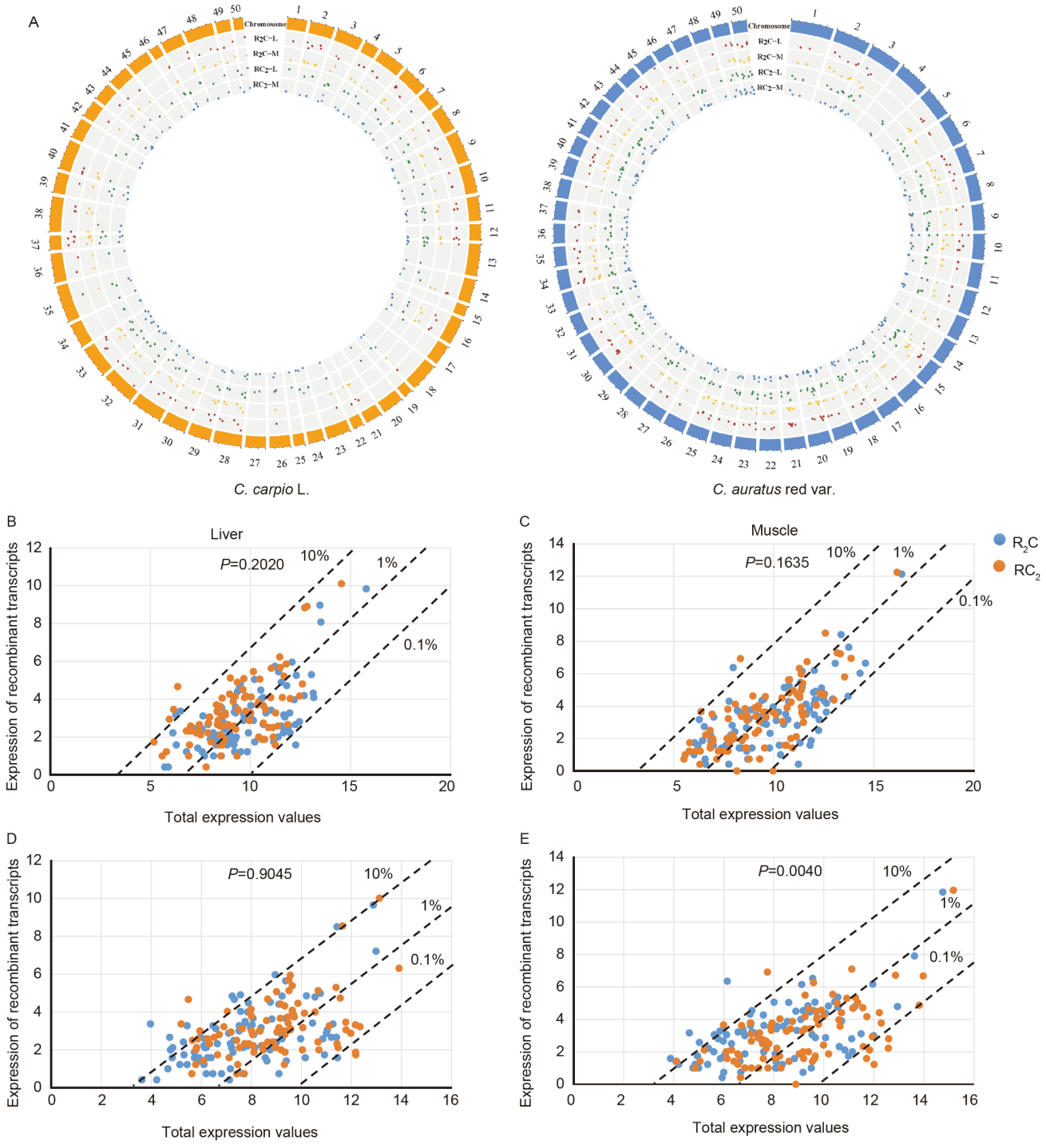

Figure 4 Expression ratios and the distribution of homoeologous recombinant transcripts. A, ratio of shared recombinant transcripts in the *C. auratus* red var. (RCC) and *C. carpio* L. (CC) subgenomes. Outer ring, chromosomes; middle four rings, ratios of recombinant transcripts in liver (L) and muscle (M) of the two triploids. Recombinant transcripts were determined based on a threshold of read coverage>4 in each gene. B, Ratios of liver detected in RCC subgenomes. C, Ratios of muscle detected in RCC subgenomes. D, Ratios of liver detected in CC subgenomes. E, The ratios of muscle detected in CC subgenomes. Differential analysis was analyzed using Student's *t* test.

1,743 recombination events in RC_2 were found in coding sequences of RCC and CC homoeologous genes, whereas 126 recombination events were distributed in non-coding sequences. (Table S8 in Supporting Information). Then, an examination of the 13,008 orthologous pairs of *C. auratus* red var. and *C. carpio* L. found that only 577 (33.33%) recombination events in R_2C coding sequences were included in 414 HR genes, whereas 553 (31.73%) in RC_2 were classified in 344 HR genes (Figure S6 in Supporting Information).

Distribution of HRs relative to stress resistance and immune response

Fifty-four and sixty-five HR genes shared between Illumina and PacBio data were predicted in R_2C and RC_2, respectively (Figure 5). Among these, 12 HR genes were shared between the two triploids. The prediction of the few shared HRs of the two methods came about as a result of the low sequencing depth and large amount of alternative splicing in PacBio data. Functional analysis revealed that six recombinant genes (*abca1a*, *ada10*, *tns2*, *acoc*, *hg2a*, and *sulf2*) shared between Illumina and PacBio data were classified as "response to stimulus" (level 2 in GO classification). Interestingly, a novel recombinant pattern with two additional exons was observed in *abca1a* transcripts of RC_2 based on PacBio results (Figure 6). In eight exons of the recombinant transcripts, exons 2 and 3 originated from *C. auratus* red var., while the other exons were from *C. carpio* L. (Figure 6). In addition, the two recombination events of *tns2* transcript in RC_2 were detected in sequence regions of 4,054–4,091 bp and 4,157–4,248 bp based on the five species-specific loci across the region of 4,092–4,146 bp (Figure S7 in Supporting Information). Most *sulf2* transcript regions in RC_2 were similar to CC homoeologs, while the sequence region of 344–370 bp was similar to those of RCC homoeologs (Figure S8 in Supporting Information).

DISCUSSION

After whole-genome duplication, the genetic variability, and heritable phenotypic diversity were described in the early generations of 4NAT, which have now been bred over 26 generations (Liu et al., 2001; Liu et al., 2016). The remodeling of the structure of genomic DNA and expression profiles, including HR (Liu et al., 2016) and expression dominance at the transcript level (Ren et al., 2016), were observed. As the most pervasive and immediate genetic consequence of nascent polyploidy, HR was thought to disrupt normal meiosis in allopolyploids (Gaeta and Chris Pires, 2010). For handling the paring of multiple sets of homologous and homoeologous chromosomes, the novel meiotic mechanism, differing from that in diploids, must be formatted (Pecinka et al., 2011; Liu et al., 2016). As a result of homoeologous chromosome exchange, the emergence of many recombinant genes gradually increases in nascent polyploidy accompanied with genotype changes. Under this situation, the conflict between subgenomes that originate from different species will gradually decrease. However, analyses of HR in allotetraploids were hindered by the complexity of multiple sets of homologous and homoeologous chromosomes. We provide a novel way to detect HRs in allotetraploids by backcrossing them to their diploid inbred parents. The two triploid fish (R_2C and RC_2) were obtained from backcrossing of 4NAT to their two diploid inbred parents (Shen et al., 2006; Chen et al., 2009), and only included one set of homoeologous chromosomes of 4NAT. Analysis of their HRs could help elucidate HR occurrence in 4NAT.

Using fluorescence in situ hybridization (FISH, Ren et al., 2019) two sets of RCC homoeologous chromosomes and one set of CC homoeologous chromosome were detected in R_2C, while two sets of CC homoeologous chromosomes and one set of RCC homoeologous chromosome were detected in RC_2. The subgenome expression dominance caused by imbalances of the two subgenomes reshaped the gene expression profiles, contributing to the phenotype changes in the triploids (Ren et al., 2019). Similar phenomena were described in rice (Huang et al., 2015), wheat (Li et al., 2014), and cyprinid fish (Ren et al., 2016). Additionally, dosage effects may also shape the gene expression and lead to heterosis (Yao et al., 2013). Under these circumstances, the emergence of HR in the two triploids might change the imbalanced ratios of the two subgenomes and further alter the expression dominance of homoeologs. The shared HRGs between the two triploids (175 in liver and 247 in muscle) reflected the stable HRs from their paternal allotetraploids (Figures 3 and 5, Figure S1 in Supporting Information). Although the short reads of Illumina data are not ideal for detecting recombination in high repeat content (van Nieuwerburgh et al., 2011), the prediction of shared HRs in the two triploid offspring could effectively reflect the HRs in the paternal allotetraploid population. In addition, the comparative analyses of HRs between genome and transcriptome could also help to obtain more reliable data from the two triploids. In comparison with Illumina data, the long length transcripts in PacBio sequencing help to further check HR events because sequenced transcripts require no assembling. On the other hand, the unique HRGs of the two triploids provide insight into the diverse distribution of HRGs of the paternal allotetraploid population (Figures 4 and 5, Figure S1 in Supporting Information).

Expressed HRs in triploids may contribute to various phenotypes, including heterosis in growth and disease resistance (Shen et al., 2006; Chen et al., 2009). In our study, the ratios of HRT *vs*. total transcription of corresponding genes were mainly observed from 0.1% to 10% in all samples, whereas various levels of HR-expressed transcripts were observed (Figure 4, Figures S4 and S5 in Supporting Information). The effects of these various molecular structures can have wide ranging impacts on biological function (Huertas and Aguilera, 2003). In addition, there were some significant differences in the expression level of recombinant transcripts ($P<0.05$) between the two tissues; this provides insight into some potential mechanisms of regulation in various tissues (Figure S5 in Supporting Information), such

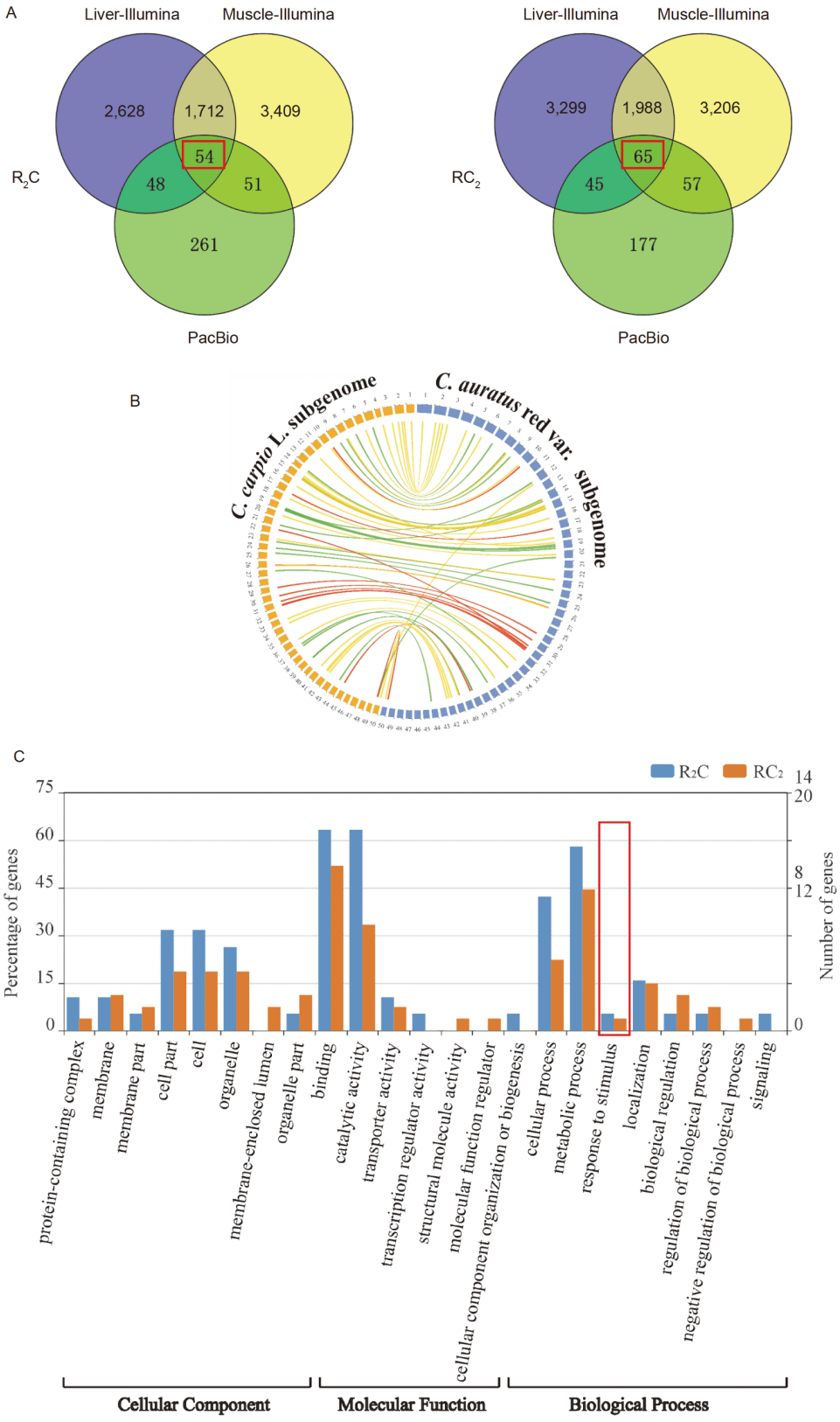

Figure 5 Homoeologous recombination predicted based on Illumina and PacBio data. A, Distribution of recombinant genes in R_2C and RC_2 based on the two sequencing methods. B, Circos plot that depicts shared homoeologous recombinant transcripts between two subgenomes. Outer ring, chromosomes; inner ring, lines that denote the recombinant transcripts of two homoeologs: the red line represents recombinant genes shared between the two triploids; the green line represents recombinant genes unique to R_2C; the yellow line represents recombinant genes unique to RC_2. C, Shared homoeologous recombinant genes. The genes classified as "response to stimulus" (GO: 0050896) based on gene ontology represent potential changes caused by recombination events.

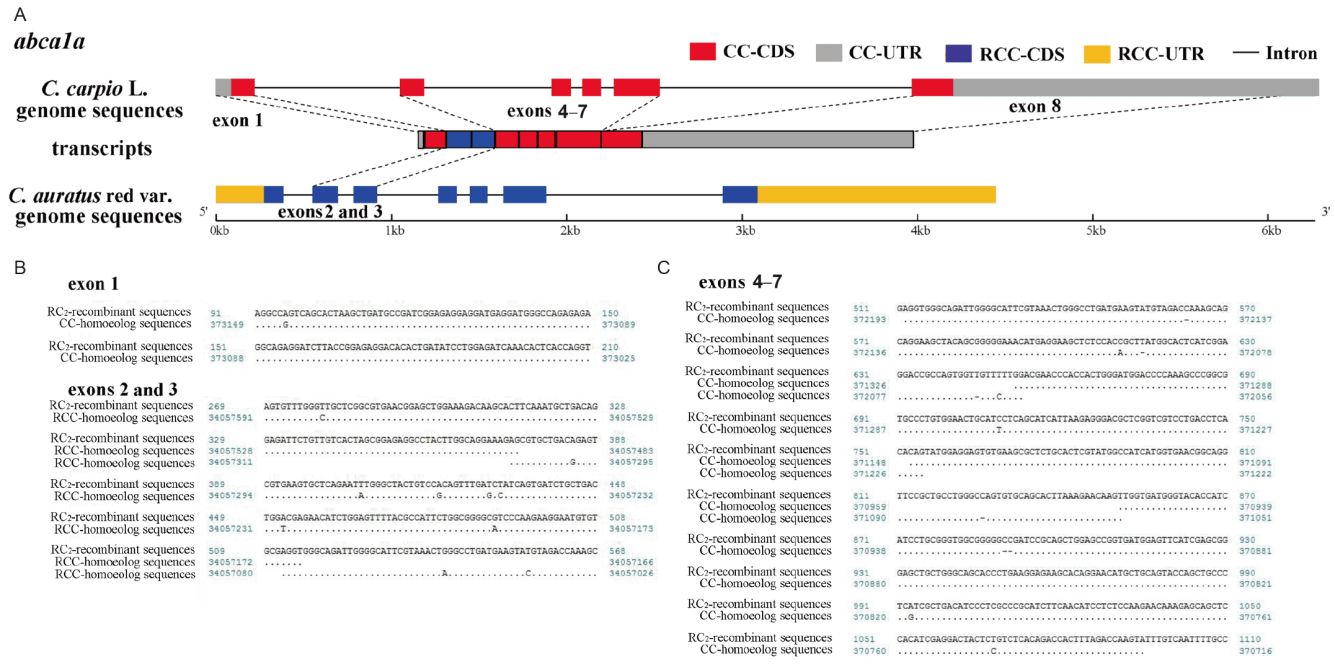

Figure 6 Gene structure of *abca1a* and homoeologous recombination (HR) detection in RC_2. HRs were observed by comparing *C. auratus* red var. (RCC) and *C. carpio* L. (CC) homoeologs. A, The transcript of HR was observed in *abca1a* of RC_2. B, Sequence alignments of exons 1–3 were shown between progeny and their inbred parents. C, Sequence alignments of exons 4–7 were shown between progeny and their inbred parents.

as tissue-specific expression and various *trans*-splicing events (Mitchell, 2000; Mayer and Floeter-Winter, 2005; Yin et al., 2010). Interestingly, a HRT of RC_2 detected in the *abca1a* gene, which plays a role in cholesterol metabolism, fat digestion, and fat absorption (Goodman, 2010) (Figure 6). Furthermore, the two HR events were detected in transcripts of *tns2* and *sulf2* genes, respectively (Figures S7 and S8 in Supporting Information). These rapid genetic changes reflect the reshaping of the phenotype due to a "response to stimulus." These results revealed the potential adaptation related to innate immunity in the triploid fish and their parental allotetraploid population. Many reports have focused on the molecular basis of heterosis using transcriptomic, proteomic, metabolomic, epigenomic, and systems biology approaches (Chen, 2013). Our study provides a novel way of investigating HRs in an allotetraploid population based on the cross prediction at genomic and transcriptional levels using both Illumina and PacBio results. The detection of HR events related to innate immunity provides insight into the potential adaptation of triploids and allotetraploids associated with self-crossing and backcrossing.

MATERIALS AND METHODS

Animal materials

The triploid R_2C fish was bred using hybridization between diploid *C. auratus* red var. (maternal) and an allotetraploid *C. auratus* red var.×*C. carpio* L.(paternal) (4NAT), whereas the triploid RC_2 fish was produced via hybridization between diploid *C. carpio* L. (maternal) and 4NAT (paternal). These triploid fish were bred in a pool with suitable illumination, water temperature, dissolved oxygen content, and adequate forage in the Engineering Center of Polyploidy Fish Breeding of the National Education Ministry located at Hunan Normal University, China. All experiments were approved by the Animal Care Committee of Hunan Normal University. Three biological replicates of two-year-old male individuals of each genotype were collected. Then, the ploidy level of these fishes was determined based on the DNA content of erythrocytes, established using flow cytometry (Liu et al., 2001; Liu et al., 2016; Ren et al., 2018). In addition, the genotypes of these fishes were detected using FISH (Liu et al., 2016). Fish were deeply anesthetized with 100 mg L^{-1} tricaine (MS-222, Sigma-Aldrich, USA) before dissection. All samples of liver, muscle, and gonad tissues were carefully excised to avoid contamination from the gut, and used for DNA and RNA extraction.

Identification of homoeologs in the two subgenomes

Alignment of the homoeologs between *C. auratus* red var. and *C. carpio* L. were obtained by all-against-all reciprocal BLASTP (v 2.2.26) comparisons using default parameters based on the protein sequences. The genome of *C. auratus* red var. (1.82 Gb) and of *C. carpio* L. (1.42 Gb) were ob-

tained from the NCBI database (accession Nos.: PRJNA487739 and PRJNA510861, respectively). The alignments were examined using the MCScanX toolkit to determine syntenic blocks (Wang et al., 2012), which were then displayed as a schematic diagram created with Circos (v 0.69-6) (Krzywinski et al., 2009). Transcripts that lacked annotated coding regions and those <100 bp were discarded. The GO and KEGG databases were used to assess the distribution of gene functions and signaling pathways of recombinant genes.

Prediction of genomic HRs using Illumina sequencing

To investigate the HRs between the two subgenomes at the genomic level, total genomic DNA of the two triploids was extracted from liver tissues of three individuals using DNeasy Blood & Tissue Kit (QIAGEN, Germany) and mixed in equal proportions. After DNA quality assessment, the RC_2 and R_2C libraries (300–500 bp) were constructed, and raw data were produced following the standard operating procedures of Illumina sequencing. Then, whole-genome sequencing was performed using HiSeq X Ten (paired-end, 150 bp) based on Illumina protocols (van Dijk et al., 2014). Following this, adapters were removed using Cutadapt (v 1.2.1) (Martin, 2011) and quality was checked using Trim Galore (v 0.4.0) (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with default parameters. The high-quality reads were mapped to combined genome of *C. auratus* red var. (https://www.ncbi.nlm.nih.gov/genome/?term=goldfish) and *C. carpio*. Next, the duplication, translocation, deletion, and insertion events were detected based on the bam file of the two triploid fish using Manta (v 1.5.0) (Chen et al., 2015). Default parameters were used in this analysis. Furthermore, HRs were obtained from inter-chromosomal translocations in the homoeologs between the *C. auratus* red var. and *C. carpio* L. subgenomes. Circos was used as a visualization tool for the recombination events in chromosomes (Krzywinski et al., 2009).

Expressed HRs detected by Illumina sequencing

The short-read RNA-seq data of four species, including *C. auratus* red var., *C. carpio* L., and two triploids, were downloaded from the NCBI Short Read Archive database (accession Nos.: SRS4475349, SRS4475350, SRS4475351, SRS4475352, SRS4475353, and SRS4475354). The samples from which the *C. auratus* red var. and *C. carpio* L. data were obtained were fed under the same conditions as the two triploids. These transcriptomes were based on three biological replicates and were paired-end (2×150 bp) sequenced using the Illumina HiSeq X Ten system. Then, we quality checked the data with FastQC (Andrews, 2010) and adapter trimmed with Trimmomatic (Bolger et al., 2014).

Recombinant transcript was detected by Illumina sequencing using a few methods (Piferrer et al., 2009; Sabio and Davis, 2010; Liu et al., 2016). Because of a lack of reference genomes of the two triploids, the high-quality reads of hybrids were mapped to the combined genome of *C. auratus* red var. and *C. carpio* L. using Star (v 2.4.0) with the parameters "–outFilterMismatchNoverReadLmax 0.01/0.02/0.03 –chimSegmentMin 50/70/90/105" (Bennett et al., 2001). Based on the mapping files, pairwise alignments of homoeologs were used to assess the differences between each hybrid. The parameter "–outFilterMismatchNoverReadLmax 0.02" was selected based on low mutation rates and sequencing errors. The parameter "–chimSegmentMin 70" was selected to detect the recombinant reads in hybrids. Then, only the reads distributed in homoeologs of both *C. auratus* red var. and *C. carpio* L. were considered recombinant reads. The gene distributions of recombinant reads were depicted at the chromosome level with shinyCircos (Yu et al., 2017). Additionally, the transcription of recombinant genes was calculated based on the average of three biological replicates using custom perl scripts.

Library construction and PacBio sequencing

To obtain the full-length transcripts of the two triploids, we extracted the total RNA from three tissues, including liver, muscle, and gonad; these were treated with DNase I (Invitrogen, USA) to remove any contaminating genomic DNA. RNA concentration was measured using Nanodrop technology, and the purified RNA was quantified using a 2100 Bioanalyzer system (Agilent, USA). After mixing the three tissue samples in equal amounts, the mixture was then reverse-transcribed using the SMARTer PCR cDNA synthesis kit and PCR-amplified using KAPA HiFi PCR kits (KAPA Biosystems, USA). The PCR product (size=0.5–6 kb and >6 kb) was selected using agarose gel electrophoresis. Then, the libraries were constructed from these cDNA products using a SMRTbell template prep kit 1.0 (Pacific Biosciences, USA). After library preparation, the library template and enzyme mixture were used to perform sequencing with the PacBio Sequel™ system.

PacBio-detected recombinant transcripts

After PacBio sequencing, the adapter sequences and low-quality sequences (subread length <50 bp, accuracy rate<0.75) were removed. Then, the sequences were processed through PacBio's SMRT-Portal analysis suite to generate circular consensus sequences (CCSs). The high-quality sequences were obtained based on the threshold of the number of CCS cycles>1 and accuracy>0.8, and full-length sequences were obtained based on threshold lengths>300 bp, poly (A) tails, 5′ primers and 3′ primers

using the SMRT Iso-Seq analysis pipeline (http://www.pacb.com/products-and-services/analytical-software/smrt-analysis/).

To distinguish RCC and CC homoeolog transcripts and their recombinant transcripts in the two triploids, the above high-quality sequences were aligned to the combined RCC and CC genome using GMAP with the following parameters: identity>90%, coverage>90% (Wu and Watanabe, 2005). All mapped sequences were classified into three patterns: uniquely mapped to the RCC subgenome, uniquely mapped to the CC subgenome, or mapped to both subgenomes. Then, the RCC and CC species-specific SNPs were predicted from the transcriptome data of *C. auratus* red var. and *C. carpio* L. using Genome Analysis Toolkit (GATK) with default parameters (McKenna et al., 2010). These RCC and CC species-specific SNPs were used to check the sequences that mapped to the genomes of both inbred parents using the PacBio data. Using custom python scripts, only the sequences mapped to both genomes (>95% sequence similarity ratio) were selected in our analyses. Recombinant transcripts were considered sequences that were mapped to each subgenome and that were >500 bp. In addition, some transcripts uniquely mapped to the RCC and CC subgenomes were considered recombinant transcripts if the five continuous RCC or CC species-specific SNPs coexisted in one transcript.Declarations

Availability of data and materials

The raw genome data from Illumina sequencing, raw transcriptome data from Illumina sequencing, and raw transcriptome data from PacBio sequencing were deposited in the NCBI Short Read Archive (accession Nos.: SRR9983189 and SRR9983190; SRR8735277–SRR8735279, SRR8712975–SRR8712976, and SRR8712978–SRR8712991; and SRR9203584, SRR9185089, and SRR9185090, respectively).

Compliance and ethics *The author(s) declare that they have no conflict of interest.*

Acknowledgements *This work was supported by the National Natural Science Foundation of China (31702334, 31730098, U19A2040 and 31430088), the Key Research and Development Program of Hunan Province (2018NK2072 and 2017NK1031), the Earmarked Fund for China Agriculture Research System (CARS-45), 111 Project (D20007), High-Level Talent Agglomeration Program of Hunan, China (2019RS1044), the Cooperative Innovation Center of Engineering and New Products for Developmental Biology of Hunan Province (20134486). We thank Mallory Eckstut, PhD, from Liwen Bianji, Edanz Editing China (www.liwenbianji.cn/ac), for editing the English text of a draught of this manuscript.*

References

Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data.

Bayne, B.L., Hedgecock, D., McGoldrick, D., and Rees, R. (1999). Feeding behaviour and metabolic efficiency contribute to growth heterosis in Pacific oysters [*Crassostrea gigas* (Thunberg)]. J Exp Mar Biol Ecol 233, 115–130.

Bennett, B.L., Sasaki, D.T., Murray, B.W., O'Leary, E.C., Sakata, S.T., Xu, W., Leisten, J.C., Motiwala, A., Pierce, S., Satoh, Y., et al. (2001). SP600125, an anthrapyrazolone inhibitor of Jun N-terminal kinase. Proc Natl Acad Sci USA 98, 13681–13686.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

Chen, S., Wang, J., Liu, S.J., Qin, Q.B., Xiao, J., Duan, W., Luo, K.K., Liu, J.H., and Liu, Y. (2009). Biological characteristics of an improved triploid crucian carp. Sci China Ser C-Life Sci 52, 733–738.

Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., Cox, A.J., Kruglyak, S., and Saunders, C.T. (2015). Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics 32, 1220–1222.

Chen, Z.J. (2013). Genomic and epigenetic insights into the molecular bases of heterosis. Nat Rev Genet 14, 471–482.

Christiansen, D.G., and Reyer, H.U. (2009). From clonal to sexual hybrids: genetic recombination via triploids in all-hybrid populations of water frogs. Evolution 63, 1754–1768.

Dickerson, G.E. (1973). Inbreeding and heterosis in animals. J Anim Sci 1973(Symposium), 54–77.

Eichler, E.E. (2001). Recent duplication, domain accretion and the dynamic mutation of the human genome. Trends Genet 17, 661–669.

Gaeta, R.T., and Chris Pires, J. (2010). Homoeologous recombination in allopolyploids: the polyploid ratchet. New Phytol 186, 18–28.

Goodman, B.E. (2010). Insights into digestion and absorption of major nutrients in humans. Adv Physiol Educ 34, 44–53.

Guo, X., Liu, S., and Liu, Y. (2006). Evidence for recombination of mitochondrial DNA in triploid crucian carp. Genetics 172, 1745–1749.

Huang, X., Yang, S., Gong, J., Zhao, Y., Feng, Q., Gong, H., Li, W., Zhan, Q., Cheng, B., Xia, J., et al. (2015). Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. Nat Commun 6, 6258.

Huertas, P., and Aguilera, A. (2003). Cotranscriptionally formed DNA: RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. Mol Cell 12, 711–721.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res 19, 1639–1645.

Li, A., Liu, D., Wu, J., Zhao, X., Hao, M., Geng, S., Yan, J., Jiang, X., Zhang, L., 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, 1878–1900.

Liu, S., Liu, Y., Zhou, G., Zhang, X., Luo, C., Feng, H., He, X., Zhu, G., and Yang, H. (2001). The formation of tetraploid stocks of red crucian carp×common carp hybrids as an effect of interspecific hybridization. Aquaculture 192, 171–186.

Liu, S., Luo, J., Chai, J., Ren, L., Zhou, Y., Huang, F., Liu, X., Chen, Y., Zhang, C., Tao, M., et al. (2016). Genomic incompatibilities in the diploid and tetraploid offspring of the goldfish×common carp cross. Proc Natl Acad Sci USA 113, 1327–1332.

Mallet, J. (2007). Hybrid speciation. Nature 446, 279–283.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17, 10.

Masson, J.Y., and West, S.C. (2001). The Rad51 and Dmc1 recombinases: a non-identical twin relationship. Trends Biochem Sci 26, 131–136.

Mayer, M.G., and Floeter-Winter, L.M. (2005). Pre-mRNA *trans*-splicing: from kinetoplastids to mammals, an easy language for life diversity. Mem Inst Oswaldo Cruz 100, 501–513.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20, 1297–1303.

Mitchell L.G. (2000). Chimeric RNA molecules generated by *trans*-splicing. Google Patents.
Ni, Z., Kim, E.D., Ha, M., Lackey, E., Liu, J., Zhang, Y., Sun, Q., and Chen, Z.J. (2009). Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. Nature 457, 327–331.
Pecinka, A., Fang, W., Rehmsmeier, M., Levy, A.A., and Mittelsten Scheid, O. (2011). Polyploidization increases meiotic recombination frequency in *Arabidopsis*. BMC Biol 9, 24.
Piferrer, F., Beaumont, A., Falguière, J.C., Flajšhans, M., Haffray, P., and Colombo, L. (2009). Polyploid fish and shellfish: Production, biology and applications to aquaculture for performance improvement and genetic containment. Aquaculture 293, 125–156.
Qi, L., Friebe, B., Zhang, P., and Gill, B.S. (2007). Homoeologous recombination, chromosome engineering and crop improvement. Chromosome Res 15, 3–19.
Ren, L., Gao, X., Yang, C., Tan, H., Cui, J., Wang, S., Li, W., Zhang, C., Tao, M., Qin, Q., et al. (2018). Comparison of diploid and triploid *Carassius auratus* provides insights into adaptation to environmental change. Sci China Life Sci 61, 1407–1419.
Ren, L., Li, W., Tao, M., Qin, Q., Luo, J., Chai, J., Tang, C., Xiao, J., Tang, X., Lin, G., 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.
Ren, L., Yan, X., Cao, L., Li, J., Zhang, X., Gao, X., Liu, J., Cui, J., and Liu, S. (2019). Combined effects of dosage compensation and incomplete dominance on gene expression in triploid cyprinids. DNA Res 26, 485–494.
Rieseberg, L.H., van Fossen, C., and Desrochers, A.M. (1995). Hybrid speciation accompanied by genomic reorganization in wild sunflowers. Nature 375, 313–316.
Sabio, G., and Davis, R.J. (2010). cJun NH2-terminal kinase 1 (JNK1): roles in metabolic regulation of insulin resistance. Trends Biochem Sci 35, 490–496.
Schumer, M., Xu, C., Powell, D.L., Durvasula, A., Skov, L., Holland, C., Blazier, J.C., Sankararaman, S., Andolfatto, P., Rosenthal, G.G., et al. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. Science 360, 656–660.
Shen, J.M., Liu, S.J., Sun, Y.D., Zhang, C., Luo, K.K., Tao, M., Zeng, C., and Liu, Y (2006). A new type of triploid crucian crap-red crucian crap (♀)×allotetraploid (♂). Prog Nat Sc 16, 1348–1352.
Shinohara, M., Gasior, S.L., Bishop, D.K., and Shinohara, A. (2000). Tid1/Rdh54 promotes colocalization of Rad51 and Dmc1 during meiotic recombination. Proc Natl Acad Sci USA 97, 10814–10819.
Soltis, D.E., Albert, V.A., Leebens-Mack, J., Bell, C.D., Paterson, A.H., Zheng, C., Sankoff, D., de Pamphilis, C.W., Wall, P.K., and Soltis, P.S. (2009). Polyploidy and angiosperm diversification. Am J Bot 96, 336–348.
Soltis, D.E., and Soltis, P.S. (1999). Polyploidy: recurrent formation and genome evolution. Trends Ecol Evol 14, 348–352.
Starcevic, A., Diminic, J., Zucko, J., Elbekali, M., Schlosser, T., Lisfi, M., Vukelic, A., Long, P.F., Hranueli, D., and Cullum, J. (2011). A novel docking domain interface model predicting recombination between homoeologous modular biosynthetic gene clusters. J Ind Microbiol Biotechnol 38, 1295–1304.
van Dijk, E.L., Auger, H., Jaszczyszyn, Y., and Thermes, C. (2014). Ten years of next-generation sequencing technology. Trends Genets 30, 418–426.
van Nieuwerburgh, F., Thompson, R.C., Ledesma, J., Deforce, D., Gaasterland, T., Ordoukhanian, P., and Head, S.R. (2011). Illumina mate-paired DNA sequencing-library preparation using Cre-Lox recombination. Nucleic Acids Res 40, e24.
Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J., Wang, X., Lee, T., Jin, H., Marler, B., Guo, H., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res 40, e49.
Wu, T.D., and Watanabe, C.K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21, 1859–1875.
Xiao, J., Fu, Y., Zhou, W., Peng, L., Xiao, J., Liu, S., and Feng, H. (2018). Establishment of fin cell lines and their use to study the immune gene expression in cyprinid fishes with different ploidy in rhabdovirus infection. Dev Comp Immunol 88, 55–64.
Yao, H., Dogra Gray, A., Auger, D.L., and Birchler, J.A. (2013). Genomic dosage effects on heterosis in triploid maize. Proc Natl Acad Sci USA 110, 2665–2669.
Yin, J., Yu, L., and Savage-Dunn, C. (2010). Alternative *trans*-splicing of *Caenorhabditis elegans sma-9/schnurri* generates a short transcript that provides tissue-specific function in BMP signaling. BMC Mol Biol 11, 46.
Yu, Y., Ouyang, Y., and Yao, W. (2017). shinyCircos: an R/Shiny application for interactive creation of Circos plot. Bioinformatics 34, 1229–1231.
Zhang, Q.Y., and Gui, J.F. (2018). Diversity, evolutionary contribution and ecological roles of aquatic viruses. Sci China Life Sci 61, 1486–1502.
Zhou, L., and Gui, J.F. (2017). Natural and artificial polyploids in aquaculture. Aquac Fish 2, 103–111.

SUPPORTING INFORMATION

The supporting information is available online at https://doi.org/10.1007/s11427-020-1749-9. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.