Genomic incompatibilities in the diploid and tetraploid offspring of the goldfish × common carp cross

Shaojun Liu1,2, Jing Luo3,4,5, Jing Chai6,7,8, Li Ren3,4, Yi Zhou3,4, Feng Huang9,10, Xiaochuang Liu11,12, Yubao Chen13, Chun Zhang14, Min Tao15, Bin Lu9,10, Wei Zhou16, Guoliang Lin8,9, Chao Mai17, Shuo Yuan18, Jun Wang19, Tao Li20, Qinbo Qin21, Zhao Ya22, Kaikun Luo23, Jun Xia24, Huan Zhong25, Rurong Zhao26, Wei Duan27, Zhenyan Song28, Yanqin Wang29,30, Chun Zhang21, Min Tao21, Wei Zhou21, Qinbo Qin21, Yun Liu31, Axel Meyer32, and Ya-Ping Zhang23,24,25,26

*Key Laboratory of Protein Chemistry and Fish Developmental Biology of Education Ministry of China, College of Life Sciences, Hunan Normal University, Changsha, 410081, Hunan, China; 2State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan University, Kunming, 650091, Yunnan, China; 3Key Laboratory for Animal Genetic Diversity and Evolution of High Education in Yunnan Province, School of Life Sciences, Yunnan University, Kunming, 650091, Yunnan, China; 4State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, 650223, Yunnan, China; 5Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, 650204, China; 6School of Software, Yunnan University, Kunming, 650091, Yunnan, China; 7Core Genomic Facility, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, 100029, China; 8Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, 100101, China; 9Centre for Biodiversity and Conservation Biology, Royal Ontario Museum, Toronto, ON, Canada M5S 2C6; and 10Department of Biology, University of Konstanz, 78457 Konstanz, Germany

Edited by Z. Jeffrey Chen, The University of Texas at Austin, Austin, TX, and accepted by the Editorial Board December 15, 2015 (received for review July 29, 2015)

Polyplody is much rarer in animals than in plants but it is not known why. The outcome of combining two genomes in vertebrates remains unpredictable, especially because polyploidization seldom shows positive effects and more often results in lethal consequences because viable gametes fail to form during meiosis. Fortunately, the goldfish (maternal) × common carp (paternal) hybrids have reproduced successfully up to generation 22, and this hybrid lineage permits an investigation into the genomics of hybridization and polyploidization. The first two generations of these hybrids are diploids, and subsequent generations are tetraploids. Liver transcriptomes from four generations and their progenitors reveal chimeric genes (≥9%) and mutations of orthologous genes. Characterizations of 18 randomly chosen genes from genomic DNA and cDNA confirm the chimerism. Some of the chimeric and differentially expressed genes relate to maternal-gene repair, and cancer-related pathways in 2nF2. Erroneous DNA excision between homologous parental genes may drive the high percentage of chimeric genes, or even more potential consequences because viable gametes fail to form during meiosis. The goldfish (maternal) × common carp (paternal) hybrids have reproduced successfully up to generation 22, and this hybrid lineage permits an investigation into the genomics of hybridization and polyploidization. The first two generations of these hybrids are diploids, and subsequent generations are tetraploids. Liver transcriptomes from four generations and their progenitors reveal chimeric genes (≥9%) and mutations of orthologous genes. Characterizations of 18 randomly chosen genes from genomic DNA and cDNA confirm the chimerism. Some of the chimeric and differentially expressed genes relate to maternal-gene repair, and cancer-related pathways in 2nF2. Erroneous DNA excision between homologous parental genes may drive the high percentage of chimeric genes, or even more potential consequences because viable gametes fail to form during meiosis.

This is a PNAS Direct Submission. Z.J.C. is a guest editor invited by the Editorial Board. Freely available online through the PNAS open access option.

Data deposition: All short-read data have been deposited in the Sequence Read Archive database (accession no. SRX668451–SRX671567). The genome assembly of one gynogenetic goldfish (Carassius auratus) has been deposited in the DNA Data Bank of Japan/European Molecular Biology Laboratory/GenBank (project accession no. PRJNA289059). The sequences of nuclear genes used in estimating heterozygosity and cDNA sequences for chimeric validation have been deposited in the GenBank database (for a list of accession numbers, see Appendix, Table S10).


The authors declare no conflict of interest.

This article is a PNAS Direct Submission. Z.J.C. is a guest editor invited by the Editorial Board.

www.pnas.org/cgi/doi/10.1073/pnas.1512955113
unpredictable effects of hybridization and polyploidization (e.g., transcriptomic shock) occur in many plant systems, such as in allopolyploid Brassica (5–8), cotton (9), and rice (10). So far, genome-level changes in the initial stages of allopolyploidization remain unknown in vertebrates.

Bisexual, diploid (based on karyotype) goldfish (Carassius auratus red var., 2n = 100) × common carp (Cyprinus carpio L., 2n = 100) (11) hybrids allow for investigations into genomic consequences of allotetraploidization. These allopolyploids offer several advantages. For example, their known parentage separates them from natural polyploids (12), and it is easy to trace the fate of progenitor genes. The parental species seem to have originated from the same allopolyploidization event; based on the number of genomic alleles, both species would be tetraploids (13). Alignment of randomly chosen genes from the genomes of goldfish [DNA Data Bank of Japan (DDBJ)/European Molecular Biology Laboratory (EMBL)/GenBank project accession no. PRJNA28905] and common carp (European Nucleotide Archive project accession no. PRJEB7241) reveals that more than 5% of nucleotide positions differ between the two copies in both species, yet <5% variation occurs within copies of both species. Twenty-two generations of hybrids were created ex situ to study the genomic processes of this allopolyploidization event (11). The first two generations after hybridization consisted of diploids. Only 4.33% of 2nF2 offspring survived embryogenesis. From the third generation onward, offspring were allotetraploids (two maternal- and two paternal-origin sets of chromosomes); survival increased to 79.33% in F4 (SI Appendix, Table S1) and remained stable at least until 4nF22. Karyotypic, fluorescence in situ hybridization (FISH), and cellular DNA content studies confirmed tetraploidy from the third generation (4nF3) onward (11, 12, 14).

The interplody crossing of tetraploid fish with diploid Carassius auratus generates sterile triploid fish on a large scale (11). The sterile triploids grow faster than their parental diploids, and, consequently, they are bred commercially in vast aquaculture facilities in the Yangtze River drainage (14). Although the initial research documented that rapid and extensive genomic changes follow tetraploidization (15–18), many questions about allopolyploidization remain unanswered.

Comparative genomics provides insights into dramatic genomic restructuring of allopolyploid hybrid offspring of the goldfish (♀ × common carp) (♂), which differs from that of plants (19, 20). Herein, we use next generation sequencing (NGS), including Roche 454 FLX (GS-FLX) and Illumina (GAI1 and HiSeq2000) technologies for RNA-seq, to investigate changes in the genomes of hybrid fish. By using the genomes of gynogenetically bred goldfish and common carp as references, we identify the rapid changes that occur immediately after allopolyploidization, explore what drives changes in the offspring compared with their parents, and determine whether allotetraploid offspring have recombined genes. Thus, we seek to detail how polyploidization and subsequent changes may contribute to the diversification of vertebrates. We also characterize the differences of gene expression between the offspring and their parents because this change might facilitate environmental adaptations that follow hybridization and allotetraploidization.

Results
Sample Discrimination, Chromosomes and FISH, and Confirmed Ploidy of Liver Cells. Before transcriptomic assessments, metaphase chromosome assays of cultured blood cells confirmed that 2nF1 and 2nF2 hybrids were diploids (2n = 100) and that 4nF18 and 4nF22 hybrids were allotetraploids (4n = 200) (Fig. 1D, E, H and I). All diploid and tetraploid offspring originated from both progenitors (Fig. 1). Flow cytometry did not find a significant difference (P > 0.01) between ploidy levels of liver cells and erythrocytes in diploid goldfish and common carp, diploid 2nF2 hybrids, and tetraploid 4nF18 and 4nF22 hybrids (SI Appendix, Table S2 and Fig. S1). After ploidy confirmation and having ruled out endoreduplication, we sequenced CDNA from only liver, the most metabolically active organ in vertebrates and the tissue with the most abundant gene expression (21, 22).

Sequencing, Transcript Reconstruction, and Annotation. We used GS-FLX, GAI1, and HiSeq2000 platforms to sequence transcriptomes and to compare changes in the genomes of 2nF1, 2nF2, 4nF18, and 4nF22 (three individuals) to their diploid parents (eight samples in total). Obtained data ranged from 9.19 gigabases (Gb) to 13.01 Gb for both parents and their diploid and tetraploid offspring after quality control (Q ≥ 30) (SI Appendix, Table S3). Using TopHat2, we aligned Illumina reads to the genomes of the gynogenetically bred goldfish and common carp. Ultimately, we obtained from 34,026 to 36,353 annotated genes from the eight individuals (SI Appendix, Table S3). A Venn diagram depicted 27,681 annotated genes shared by progenitors and offspring in all individuals (SI Appendix, Fig. S2). The hybrid generations and their parents shared from 29,375 to 30,036 genes (SI Appendix, Fig. S3).

Chimeric Genes and Unique Mutations in Offspring. First, we determined parent-specific and offspring-specific variations. Based on sequences mapped using the BWAligner tool, variations of the eight samples were called by using both SAMTools and GATK. After pairwise comparing the maternal goldfish, the
Table 1. Patterns of genomic variation in the goldfish × common carp hybrid offspring based on two reference genomes

<table>
<thead>
<tr>
<th>Categories</th>
<th>2nF₁</th>
<th>2nF₂</th>
<th>4nF₁8</th>
<th>4nF₂₁-1</th>
<th>4nF₂₂-2</th>
<th>4nF₂₂-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chimeric (goldfish as reference)</td>
<td>1,073</td>
<td>1,229</td>
<td>1,051</td>
<td>1,037</td>
<td>1,086</td>
<td>1,027</td>
</tr>
<tr>
<td>% No. of genes</td>
<td>3.59</td>
<td>4.09</td>
<td>3.57</td>
<td>3.53</td>
<td>3.63</td>
<td>3.46</td>
</tr>
<tr>
<td>Chimeric (common carp as reference)</td>
<td>1,831</td>
<td>1,949</td>
<td>1,879</td>
<td>1,831</td>
<td>1,833</td>
<td>1,722</td>
</tr>
<tr>
<td>% No. of genes</td>
<td>6.13</td>
<td>6.49</td>
<td>6.39</td>
<td>6.23</td>
<td>6.12</td>
<td>5.80</td>
</tr>
<tr>
<td>Chimeric (by both references)</td>
<td>132</td>
<td>145</td>
<td>147</td>
<td>133</td>
<td>134</td>
<td>122</td>
</tr>
<tr>
<td>% No. of genes</td>
<td>0.44</td>
<td>0.48</td>
<td>0.50</td>
<td>0.45</td>
<td>0.45</td>
<td>0.41</td>
</tr>
<tr>
<td>Potentially chimeric</td>
<td>125</td>
<td>126</td>
<td>211</td>
<td>148</td>
<td>142</td>
<td>147</td>
</tr>
<tr>
<td>% No. of genes</td>
<td>0.42</td>
<td>0.42</td>
<td>0.72</td>
<td>0.50</td>
<td>0.47</td>
<td>0.49</td>
</tr>
<tr>
<td>Maternal-origin genes</td>
<td>588</td>
<td>522</td>
<td>635</td>
<td>634</td>
<td>588</td>
<td>644</td>
</tr>
<tr>
<td>% No. of genes</td>
<td>1.97</td>
<td>1.74</td>
<td>2.16</td>
<td>2.16</td>
<td>1.96</td>
<td>2.17</td>
</tr>
<tr>
<td>Paternal-origin genes</td>
<td>1,788</td>
<td>1,691</td>
<td>1,698</td>
<td>1,678</td>
<td>1,729</td>
<td>1,664</td>
</tr>
<tr>
<td>% No. of genes</td>
<td>5.99</td>
<td>5.63</td>
<td>5.77</td>
<td>5.71</td>
<td>5.78</td>
<td>5.60</td>
</tr>
<tr>
<td>Genes with specific mutations</td>
<td>317</td>
<td>341</td>
<td>341</td>
<td>327</td>
<td>328</td>
<td>302</td>
</tr>
<tr>
<td>% No. of genes</td>
<td>1.06</td>
<td>1.14</td>
<td>1.16</td>
<td>1.11</td>
<td>1.10</td>
<td>1.02</td>
</tr>
<tr>
<td>Total no. of shared genes</td>
<td>29,852</td>
<td>30,036</td>
<td>29,427</td>
<td>29,375</td>
<td>29,928</td>
<td>29,713</td>
</tr>
</tbody>
</table>

Fig. 2. Schematic diagrams of gene patterns for offspring from hybridizing the goldfish (R) and the common carp (C). Orange bars marked R variation denote offspring fragments with the goldfish-specific variants; blue bars marked C variation show common carp-specific variants; and red bars marked F variation show offspring-specific variants. Genes were classified as three categories. The first category includes patterns 1–8 (A–H, respectively) in which chimeric genes have single or multiple fragments consisting of continuous, alternating variations from parent-specific variants. The second category includes patterns 9–13 (I–M, respectively), which are not chimeras and in which the genes are derived exclusively from one parent. Category three includes patterns 14–18 (N–R, respectively) where genes are derived from either or both progenitors but with offspring-specific mutations.

paternal common carp, and the offspring, offspring variations were classified as R-variations (goldfish-specific variations), C-variations (common carp-specific variations), and F-variations (offspring-specific variations) (SI Appendix, Table S4). Chimeric patterns were identified in gene regions by the distributions of variations. We classified 18 patterns within three categories (Fig. 2, Table 1, Dataset S1, and SI Appendix, Table S5). The first category included nine patterns of likely chimeric genes. Patterns 1–6 contained genes with a single chimeric fragment, either with or without offspring-specific mutations. Patterns 7 and 8 contained chimeric genes consisting of multiple exons, either with or without offspring-specific mutations. Chimeric genes from patterns 1–8 comprised from 9.67% to 11.06% of genes in overlapping mapped regions of all offspring. Among these chimeric genes, only 122–147 genes (0.41–0.50%) were detected as chimeras based on both reference genomes, and the remaining chimeric genes (9.26–10.58%) were identified from one reference genome only. Pattern 9 comprised from 0.42% to 0.72% of genes that included possible chimeric genes, depending on their splicing pattern. These genes had multiple exon sequences that consisted of alternating progenitor fragments. Sanger sequencing validated 18 of the tested 23 apparent chimeric genes (>75% correct bioinformatic identification of chimeric genes) (SI Appendix, Figs. S4–S21). The second category included either paternal-origin or maternal-origin genes. Maternal-origin genes (patterns 10 and 11) were less common than paternal-origin genes (patterns 12 and 13) (Fig. 2, Table 1, and SI Appendix, Table S5). The third category included genes with mutations unique to offspring. These genes grouped into patterns 14–18, which consisted of genes derived from both progenitors but with offspring-specific mutations. They comprised from 1.02% to 1.16% of genes in overlapping mapped regions in the six offspring (Fig. 2, Table 1, and SI Appendix, Table S5). Genes with concordant variation assessments of being chimeras were retained for further analyses (Table 1 and SI Appendix, Table S5).

Chimeric genes (9.67–11.06%) and mutation events (1.02–1.16%) were revealed in different generations of nascent allopolyploids. Genes with multiple recombinations that involved both parents were enriched significantly in more than 1,000 functional terms (P < 0.05) (SI Appendix and Dataset S2). There were 617 of these terms shared by all offspring, and the terms of “mutagenesis site” and “disease mutation” had high gene counts (P < 3.6E–22). In all offspring, chimeric genes were involved directly in spindle assembly [e.g., casein kinase (CSNK)], spliceosome (e.g., TRA2, PRPF8), RNA polymerase (e.g., RPB, RPC), or chromatin modification (e.g., SMYD, JHDMID E_F) (23–28). Chimeric genes also participated in the activities of mitogen-activated protein kinase (MAPK) [e.g.,...
MAPKAPK2, nemo-like kinase (NLK), TAO, Ser/Thr protein kinase [e.g., CSNK, cell division control protein (CDC)], and the related MAPK signaling pathway. These potentially interacting kinases were shown to be crucial in the regulation of cell fate (29–31). Other chimeric genes were also directly involved in the regulation of cell cycle [e.g., CDC, DNA repair and recombination protein (RAD), VCP] and DNA damage response and repair (via recombination) [e.g., ubiquitin-conjugating enzyme (UBE), single-strand DNA-binding protein (sdb)]. Many chimeric genes in 2nF1 were specifically enriched in cancer-related pathways, including the Wnt (e.g., NLK, DAAM), mTOR (HIF1A), VEGF (e.g., VEGF, TGFβ3), and PPAR (NR2B1/RXRα) signaling pathways (Datasets S1 and S2 and SI Appendix, Fig. S22).

Expression Patterns in Hybrids. Analyses revealed pairwise alterations of gene expression. Diploid offspring clustered with their paternal progenitor, yet tetraploid offspring clustered with the maternal one (Fig. 3). Expression analyses (SI Appendix, Figs. S23 and S24) yielded varying results in group comparisons among both parents and one offspring. In 2nF1, 2nF2, and 4nF22-1, comparing with both parents, significantly up-regulated genes (3.20–2.20) yielded varying results in group comparisons between the two patterns. In addition, the expression of maternal- or paternal-biased genes did not associate with ploidy (between the two patterns). In 5.87% and down-regulation in 11.49% of genes, maternal one (Fig. 3). Expression analyses (SI Appendix, Figs. S24 and Table S6). The differentially expressed up- and down-regulated genes in each offspring showed enrichment from 10 to 80 functional terms (SI Appendix and Dataset S2). Most differentially expressed genes of offspring were important components of liver tissues, or they played crucial roles in essential liver processes (SI Appendix and Datasets S2 and S3). Notably, some genes were specifically enriched in mutagenesis site (P = 1.40E−03 in 2nF1 and in the regulation of cell death and apoptosis (P < 0.05) in 2nF1 and 2nF2 (SI Appendix and Datasets S2 and S3). Upon checking, very few chimeric genes exhibited up- or down-regulation (Dataset S3). The expression of most chimeric genes (>98%) did not differ significantly from either parent [P > 0.05, false discovery rate (FDR) > 0.05].

Analyses Using a de Novo Assembly Strategy Obtained Results Compatible with the Mapping Results (SI Appendix, Table S7). Based on extracted orthologous sequences, orthologs of all six offspring also classified into three categories, in which the chimeric genes comprised from 24.76% to 27.24% of all genes or unknown ORFs. Further, the six offspring had from 0.38% to 1.32% maternal-origin or paternal-origin genes or ORFs. Intriguingly, from 2.99% to 4.01% of the offspring had genes or ORFs that differed from both parents at more than 3.5% of base pair positions (SI Appendix, Table S8). Expression analyses showed that up-regulated genes (19.98–23.47%) were more common than down-regulated genes (1.08–2.03%). Both 2nF1 and 2nF2 significantly expressed fewer maternal-biased (P < 0.05) but more paternal-biased genes (P < 0.05) than all allotetraploid offspring. For the other genes, diploid and allotetraploid offspring did not differ significantly in the expressions of their genes. Within all patterns, the three 4nF22 individuals did not differ significantly from one another (SI Appendix, Table S9). Quantitative real-time PCR (qRT-PCR) validated expressional changes for 21 of 25 chosen genes detected by de novo assembly without distinguishing alleles (SI Appendix, Fig. S25 and Table S10).

Discussion Analyses based on genome mapping seem to provide more reliable results than analyses of the de novo assembly alone due to alleles with short reads in next generation sequencing resulting in assembly errors (SI Appendix, Tables S11 and S12). Further, genome sizes of liver cells and erythrocytes do not differ significantly within diploid and allotetraploid individuals. Thus, our analyses on chimeras and changes in gene expression relate to hybridization and tetraploidization rather than effects of endoreduplication, as was reported in human liver cells (32–34).

The analyses of liver transcriptomes provide results previously unidentified into allopolyploidization of these fishes and beyond. Like these fishes, polyploidization in plants involves genomic reorganization and massive gene loss (35–42). Some polyploid plants, such as Brassica (41, 43), Tragopogon (42), and wheat (36, 40), exhibit relatively high levels of genomic rearrangements. However, in other plants, such as Arabidopsis (44) and cotton (45), changes in gene expression predominate.

The potential synthetic effects caused by both genomic structural change and alteration of expression might severely constrain the survival of vertebrate offspring. A high level of genomic restructuring occurs in the offspring, and these changes include genetic recombination, offspring-specific mutation, and significant alteration of gene expression. In contrast, different factors seem to determine the survival of polyploid plant progeny. All allopolyploid fish offspring exhibit a high rate of chimeric change and mutation (Fig. 2 and Table 1). Overall, most chimeras of maternal/paternal origin do not overlap, and this phenomenon indicates nonreciprocal structural change. Chimeric genes might have formed after the two genomes merged and before whole genome duplication that leads to tetraploidization (46). Both chimeric genes and nonsynonymous mutations might produce structural changes that reduce enzyme activity or fidelity by affecting normal transcriptional processing (47). The high frequency of chimeras and mutation also might result from large-scale DNA repair via recombination or nonhomologous end-joining, or even transposon activity (48–51). The common occurrence of chimeric genes that persists throughout the initial 22 generations of hybrid fishes might result from different processes that relate to chimeras: replication slippage or the imprecise cutting of an unpaired duplication during large-loop mismatch repair (46). Abnormalities of DNA or RNA repair, such as dysfunction of RAD (52, 53) or other genes and pathways (e.g., UBE2N/UBC13,
ssb in Datasets S1 and S2 and SI Appendix, Fig. S22) (54–57), might drive, or contribute to, the high rate of crossing-over in the offspring. Recent reports of higher mutation rates in heterozygotes support this possibility (58). Chaos caused by polyploidization also might result in multiple failures of chromosomal pairing (53). These changes may trigger complicated interactions between the repair of DNA damage and the regulation of cell cycles. Another aspect of genome shock may also include parent-biased patterns of gene expression and the abnormal up- and down-regulation of offspring genes and abnormality in the very nascent hybrid genomes: e.g., 2nF1 and 2nF2 (59, 60). We hypothesize that the alteration of expression of these genes drives polyploidization. In addition, polyploidization might cause extensive transposon activity, which can result in extensive chimeric regions (49–51). Our analyses cannot discern the causative mechanism(s).

Long-term genome shock may be responsible for the rarity of polyploidization in vertebrates. Plants achieve amelioration within four or five generations (21, 61–63). In contrast, vertebrates require more than 22 generations for achieving it. This discovery also indicates that the initial stage of allopolyploidization involves a struggle for survival, probably in terms of alternative selection acting on developmental processes. Severe synthetic effects include changes in genomic structure and alteration of expression, accompanying genome shock. These happenings explain the rarity of allopolyploid speciation in vertebrates, and this result may apply to other animals as well. In addition, allopolyploidization is rare in vertebrates possibly due to the greatly reduced viability of 2nF2 hybrid offspring caused by the severe and synthetic effect of genome shock (2–4). Ultimately, much work remains on exactly how polyploid plants and animals survive genome shock, which happens more frequently in allopolyploid plants than in animals (5, 6). Further theoretical work on survival and the viability of allopolyploids might benefit from taking into consideration these results. Additional functional analyses at the genomic (genetic/epigenetic) level are also necessary (3, 5, 64–67).

Methods

SI Appendix has additional information relating to the methodologies described below.

FISH Detection, Ploidy Confirmation of Liver Cells, and RNA Isolation. All experiments were approved by Animal Care Committee of Hunan Normal University and followed guidelines of the Administration of Affairs Concerning Experimental Animals of China. We collected eight individuals, including three 2-y-old 4nF22 hybrids [the goldfish (♂) × common carp (♀) descendants (F2−1, male; F2−2, female; and F2−3, male)], one 2-y-old hybrid male 4nF2, one 2-y-old female hybrid 2nF1, one 2-y-old female hybrid 2nF2, one 2-y-old female goldfish, and one 2-y-old male common carp. Ploidy of the eight samples was confirmed by metaphase chromosome assay of cultured blood cells (Fig. 1). A flow cytometer was used to measure the DNA content of liver cells and erythrocytes in 2-y-old samples of three diploid goldfish, two blood cells (Fig. 1). A flow cytometer was used to measure the DNA content of liver tissue of all eight samples. RNA was extracted from liver tissue of all eight samples.

Transcriptome Reconstruction and Differential Expression Analyses. After isolation of mRNA from liver tissues of all the samples, we constructed three Illumina libraries for all samples using an mRNA-Seq Sample Prep Kit (Illumina Inc.) and two 454 libraries for the goldfish, the common carp, and 4nF22, and performed sequencing steps on Illumina GAIIx, HiSeq2000, and Roche 454 FLX (GS-FLX). After performing a series of strict filtering steps to remove adapter contamination and low-quality reads, we performed alignment of reads by using TopHat2 v2.0.4 (68) and transcript reconstruction by Cufflinks v2.2.1 (69). Cuffdiff and downstream tools (tools from DESeq package, detailed in SI Appendix, Part 1) performed the differentially expressed analyses.

Variation and Detection of Chimeric Patterns. To detect variants among the transcriptomes of progenitors and offspring, and their distributions, high-quality reads were remapped to the goldfish and common carp reference genomes by using the Burrows-Wheeler Aligner (BWA) v0.7.10 (70). After obtaining BAM files, we recorded the mapped region of each sample on the reference genomes. Variations from regions that overlapped in both progenitors and one offspring were extracted from the alignments using both mpileup in the SAMtools package v0.1.19 (71) and the GATK v3.4.0 (72–74) pipeline for RNA-seq. Candidate variations were filtered based on a variation-quality score of ≥20, and depth of >3× reads. VCFtools v0.1.12 (75) was used to compare variations from both progenitors and one offspring as a group that were found by both methods. Offspring loci were compared with those of both parents with the same coordinates. Mutation patterns were defined as R-variation, C-variation, and F-variation. The distribution patterns of these variations were analyzed, and the distributions of chimeric loci were retained for downstream analysis.

Gene Annotation and Shared Relationships. Annotations via the Cufflinks pipeline were based on information from both reference genomes. Accession numbers of annotated genes were obtained using BLASTX along with their GO terms and accession numbers. The Database for Annotation, Visualization and Integrated Discovery (DAVID) tool v6.7 (76, 77) was used for GO enrichment analyses. Venn diagrams reflecting the shared relationships of genes among different individuals were generated by using the in-house software VennPainter (see URLs). Predicted pathways of all orthologous sequences were analyzed by the KEGG Automatic Annotation Server (KAAS) (see URLs).

PCR Validation for Chimeric Genes and Quantitative Real-Time PCR Validation for 25 Differentially Expressed Genes. A set of primers used in PCR reactions and clone numbers for chimeric genes are displayed in SI Appendix, Table S13. Primers for expression validation were designed according to transcriptional sequences (SI Appendix, Table S14).

ACKNOWLEDGMENTS. We thank many researchers who helped with this study: Lin Chen, Fan Yu, Cuiping You, Weiguo He, Zhen Liu, Jie Hu, Duo Wang, Li Zou, Can Song, Tanguo Li, Rong Yang, Xingguo Yi, Gang Liu, Honglin Liu, Xuewei Kang, Kang Xu, Xiang Xie, Ning Li, Wei Liu, Zhiouling He, Lihua Xie, Lihai Ye, Yi Zhang, Zhuhuai Zhang, Ding Dai, Zehou Zhong, Ming Zeng, Zhanzhao Yao, Jinxiu He, Hui Zhao, Linliu Liu, Siqi Wang, and Jinxiu Li. This research was supported by National Natural Science Foundation of China Grants 30930071, 91331105, 31360514, 31430088, and 31210103918, the Cooperative Innovation Center of Engineering and New Products for Cotton, in Polyploid and Hybrid Genomics, eds Chen ZJ, eds Chen ZJ, Birchler JA (Wiley, Oxford), pp 171–186.

Liu et al. PNAS Early Edition | 5 of 6


