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

Shaojun Liu[a,1,2], Jing Luo[b,c,1], Jing Chai[b,c,d,e,1], Li Ren[a,1], Yi Zhou[a,1], Feng Huang[b,c,1], Xiaochuan Liu[b,c,1], Yubao Chen[a], Chun Zhang[a], Min Tao[a], Bin Lu[b,c], Wei Zhou[f], Guoliang Lin[b,c], Chao Mai[f], Shuo Yuan[f], Jun Wang[a], Tao Li[a], Qinbo Qin[a], Hao Feng[a], Kaikun Luo[a], Jun Xiao[a], Huan Zhong[a], Rurong Zhao[a], Wei Duan[a], Zhenyan Song[a], Yanqin Wang[b,c], Jing Wang[b,c], Li Zhong[b,c], Lu Wang[b,c], Zhaoli Ding[d], Zhenglin Du[g], Xuemei Lu[h], Yun Gao[d], Robert W. Murphy[d,i], Yun Liu[a], Axel Meyer[j,2], and Ya-Ping Zhang[b,c,d,2]

[a]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; [b]State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan University, Kunming, 650091, Yunnan, China; [c]Key Laboratory for Animal Genetic Diversity and Evolution of High Education in Yunnan Province, School of Life Sciences, Yunnan University, Kunming, 650091, Yunnan, China; [d]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, 650223, Yunnan, China; [e]Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, 650204, China; [f]School of Software, Yunnan University, Kunming, 650091, Yunnan, China; [g]Core Genomic Facility, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, 100029, China; [h]Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, 100101, China; [i]Centre for Biodiversity and Conservation Biology, Royal Ontario Museum, Toronto, ON, Canada M5S 2C6; and [j]Department 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)

Polyploidy 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 tetraploidization. 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 chimera. Some of the chimeric and differentially expressed genes relate to mutagenesis, repair, and cancer-related pathways in $2nF_1$. Erroneous DNA excision between homologous parental genes may drive the high percentage of chimeric genes, or even more potential mechanisms may result in this phenomenon. Meanwhile, diploid offspring show paternal-biased expression, yet tetraploids show maternal-biased expression. These discoveries reveal that fast and unstable changes are mainly deleterious at the level of transcriptomes although some offspring still survive their genomic abnormalities. In addition, the synthetic effect of genome shock might have resulted in greatly reduced viability of $2nF_2$ hybrid offspring. The goldfish × common carp hybrids constitute an ideal system for unveiling the consequences of intergenomic interactions in hybrid vertebrate genomes and their fertility.

allopolyploidization | chimeric genes | transcriptomes | sequence validation | vertebrate

Polyploidization is much rarer in vertebrates than in plants, and the reasons for this difference remain a mystery (1–3). Traditional explanations include barriers to sex determination, physiological and developmental constraints (especially nuclear–cytoplasmic interactions and related factors) (2, 3), and genome shock or dramatic genomic restructuring (2–4). One type of polyploidization, allopolyploidization, involves the genomes of two species. Hybridization, accompanied by polyploidization, triggers vast genetic and genomic imbalances, including abnormal quadrivalent chromosomal groups, dosage imbalances, a high rate of DNA mutations and combinations, and other non-Mendelian phenomena (5–7). The effects of these imbalances are usually deleterious and are rarely advantageous. Imbalances in many plant crops determine the fate of the allopolyploid offspring. Genomic changes immediately follow allopolyploidization. Various and

Significance

Why is polyploidization rarer in animals than in plants? This question remains unanswered due to the absence of a suitable system in animals for studying instantaneous polyploidization and the crucial changes that immediately follow hybridization. RNA-seq analyses discover extensive chimeric genes and immediate mutations of orthologs in both diploid and tetraploid offspring of the goldfish (♀) × common carp (♂) hybrids. Overall, diploid offspring show paternal-biased expression, yet tetraploids show maternal-biased expression. Some chimeric and differentially expressed genes relate to crucial functions of normal cell cycle activities, and cancer-related pathways in $2nF_1$. The discovery of fast changes at the levels of chromosomes, genomic DNA, and transcriptomes suggests that allopolyploidization hinders genomic functions in vertebrates, and this conclusion may extend to all animals.

Author contributions: S.L., A.M., and Y.P.Z. designed research; J.C., L.R., Y.Z., F.H., X. Liu, C.Z., M.T., B.L., W.Z., G.L., C.M., S.Y., Jun Wang, T.L., Q.Q., K.L., J.X., H.Z., R.Z., W.D., Z.S., Y.W., Jing Wang, L.Z., L.W., Z. Ding, Z. Du, X. Lu, and Y.G. analyzed data; J.C. and B.L. conducted cloning and data validation; B.L., L.Z., L.W., and Z. Ding conducted sequencing; W.Z. led the software design; G.L., C.M., and S.Y. did the program writing; T.L., Q.Q., K.L., and J.X. conducted sampling; R.Z., W.D., and Z.S. conducted species identification; L.Z., L.W., and Z. Ding performed library construction; L.W. and Z. Ding conducted raw data quality control; Z. Du and X. Lu conducted assembly; and S.L., J.L., J.C., Y.Z., Y.C., H.F., R.W.M., Y.L., and A.M. wrote the paper.

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.

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 nos. SRA049856, SRA049856.1, SRA056516, SRX668435, SRX668436, SRX668451, SRX668453, SRX668467, SRX669310, SRX671568–SRX671571, SRX175396, SRX175397, SRX176547, and SRX177691). The FISH probe-cloned DNA fragment has been deposited in the GenBank database (accession no. JQ086761). 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 validations have been deposited in the GenBank database (for a list of accession numbers, see *SI Appendix*, Table S15).

[1]S.L., J.L., J.C., L.R., Y.Z., F.H., and X. Liu contributed equally to this work.

[2]To whom correspondence may be addressed. Email: zhangyp@mail.kiz.ac.cn, lsj@hunnu.edu.cn, or axel.meyer@uni-konstanz.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1512955113/-/DCSupplemental.

GENETICS

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 $2nF_2$ offspring survived embryogenesis. From the third generation onward, offspring were allotetraploids (two maternal-origin and two paternal-origin sets of chromosomes); survival increased to 79.33% in F_4 (*SI Appendix*, Table S1) and remained stable at least until $4nF_{22}$. Karyotypic, fluorescence in situ hybridization (FISH), and cellular DNA content studies confirmed tetraploidy from the third generation ($4nF_3$) onward (11, 12, 14). The interploidy crossing of tetraploid fish with diploid *Carassius cuvieri* 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 (GAII and Hiseq2000) technologies for RNA-seq, to investigate changes in the genomes of hybrid fish. By using the genomes of gynogenetic 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 $2nF_1$ and $2nF_2$ hybrids were diploids (2n = 100) and that $4nF_{18}$ and $4nF_{22}$ hybrids were allotetraploids (4n = 200) (Fig. 1 *D*, *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 $2nF_2$ hybrids, and tetraploid $4nF_{18}$ and $4nF_{22}$ 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).

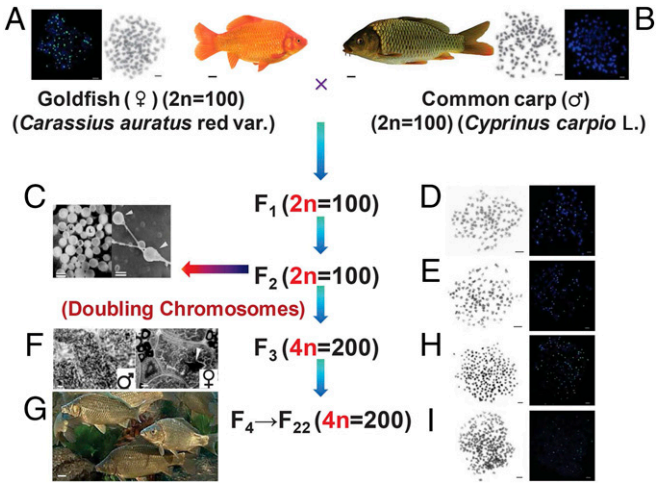

Fig. 1. The chromosomal trait, gonadal development, and appearance of goldfish (2n = 100), common carp (2n = 100), and their $2nF_1$ (2n = 100), $2nF_2$ (2n = 100), $4nF_{18}$ (4n = 200), and $4nF_{22}$ (4n = 200) hybrid offspring. (*A*) The goldfish (*Right*) has 100 chromosomes (*Middle*) and 100 signals (*Left*) after the chromosomes are stained with DNA probe (probe A) (9,468-bp fragment of 36 copies of a repetitive 263-bp fragment). (Scale bar: 1 cm.) (*B*) The common carp (*Left*) has 100 chromosomes (*Middle*) and no signal (*Right*) after the chromosomes are stained with probe A. (Scale bar: 1 cm.) (*C*) Light microscopy of eggs produced by $2nF_2$ female (*Left*) showing large eggs (L), the midsized eggs (M), and small eggs (S) (11). (Scale bar: 0.2 cm.) Scanning electron micrograph of spermatozoa in semen stripped out from a 1-y-old $2nF_2$ male (*Right*) showing diploid (lower arrowhead) and haploid spermatozoon (upper arrowhead) (11). (Scale bar: 1.9 μm.) (*D*) The $2nF_1$ has 100 chromosomes (*Left*) and 50 signals (*Right*) after the chromosomes are stained with probe A. (Scale bar: 3.0 μm.) (*E*) The $2nF_2$ has 100 chromosomes (*Left*) and 50 signals (*Right*) after the chromosomes are stained with probe A. (Scale bar: 3.0 μm.) (*F*) Histology of normal mature testis (*Left*) and ovary (*Right*; white arrow represents the cell nuclei) in $4nF_{18}$. (Scale bars: *Left*, 10 μm; *Right*, 0.02 cm.) (*G*) Image of $4nF_{18}$. (Scale bar: 1.2 cm.) (*H*) The $4nF_{18}$ has 200 chromosomes (*Left*) and 100 signals (*Right*) after the chromosomes are stained with probe A. (Scale bar: 3.0 μm.) (*I*) The $4nF_{22}$ has 200 chromosomes (*Left*) and 100 signals (*Right*) after the chromosomes are stained with probe A. (Scale bar: 3.0 μm.)

Sequencing, Transcript Reconstruction, and Annotation. We used GS-FLX, GAII, and Hiseq2000 platforms to sequence transcriptomes and to compare changes in the genomes of $2nF_1$, $2nF_2$, $4nF_{18}$, and $4nF_{22}$ (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 ($Q \geq 30$) (*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 BWA aligner 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

	$2nF_1$		$2nF_2$		$4nF_{18}$		$4nF_{22}$-1		$4nF_{22}$-2		$4nF_{22}$-3	
Categories	No. of genes	%	No. of genes	%	No. of genes	%	No. of genes	%	No. of genes	%	No. of genes	%
Chimeric (goldfish as reference)	1,073	3.59	1,229	4.09	1,051	3.57	1,037	3.53	1,086	3.63	1,027	3.46
Chimeric (common carp as reference)	1,831	6.13	1,949	6.49	1,879	6.39	1,831	6.23	1,833	6.12	1,722	5.80
Chimeric (by both references)	132	0.44	145	0.48	147	0.50	133	0.45	134	0.45	122	0.41
Potentially chimeric	125	0.42	126	0.42	211	0.72	148	0.50	142	0.47	147	0.49
Maternal-origin genes	588	1.97	522	1.74	635	2.16	634	2.16	588	1.96	644	2.17
Paternal-origin genes	1,788	5.99	1,691	5.63	1,698	5.77	1,678	5.71	1,729	5.78	1,664	5.60
Genes with specific mutations	317	1.06	341	1.14	341	1.16	327	1.11	328	1.10	302	1.02
Total no. of shared genes	29,852		30,036		29,427		29,375		29,928		29,713	

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*, *JHDM1D_E_F*) (23–28). Chimeric genes also participated in the activities of mitogen-activated protein kinase (MAPK) [e.g.,

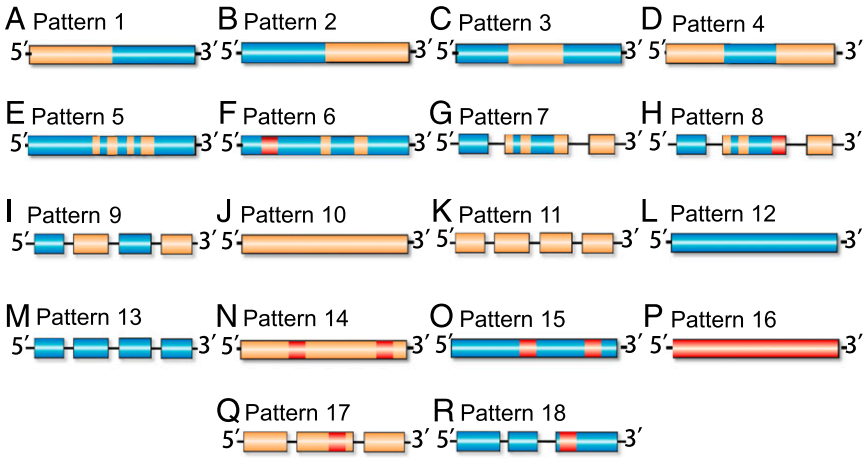

Fig. 2. Schematic diagrams of gene patterns for offspring from hybridizing the goldfish (R) and the common carp (C). Orange bars marked R variation denote offspring fragments with the goldfish-specific variants; blue bars marked C variation show common carp-specific variants; and red bars marked F variation 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, and with or without offspring-specific variations, and pattern 9 (*I*), in which potentially chimeric genes have multiple fragments or exons consisting of alternating progenitor-derived fragments. The second category includes patterns 10–13 (*J*–*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.

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 (*ssb*)]. Many chimeric genes in $2nF_1$ were specifically enriched in cancer-related pathways, including the Wnt (e.g., *NLK*, *DAAM*), mTOR (*HIF1A*), VEGF (e.g., *VEGF*, *TGFB3*), and PPAR (*NR2B1*/*RXRA*) 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 $2nF_1$, $2nF_2$, and $4nF_{22}$-1, comparing with both parents, significantly up-regulated genes (8.55–13.26%) were more common than down-regulated genes (2.20–8.30%). Differing, the $4nF_{22}$-2 individual exhibited up-regulation in 5.87% and down-regulation in 11.49% of genes, and yet $4nF_{18}$ and $4nF_{22}$-3 showed no significant difference between the two patterns. In addition, the expression of maternal- or paternal-biased genes did not associate with ploidy (*SI Appendix*, Fig. 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 $2nF_1$ and in the regulation of cell death and apoptosis ($P < 0.05$) in $2nF_1$ and $2nF_2$ (*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].

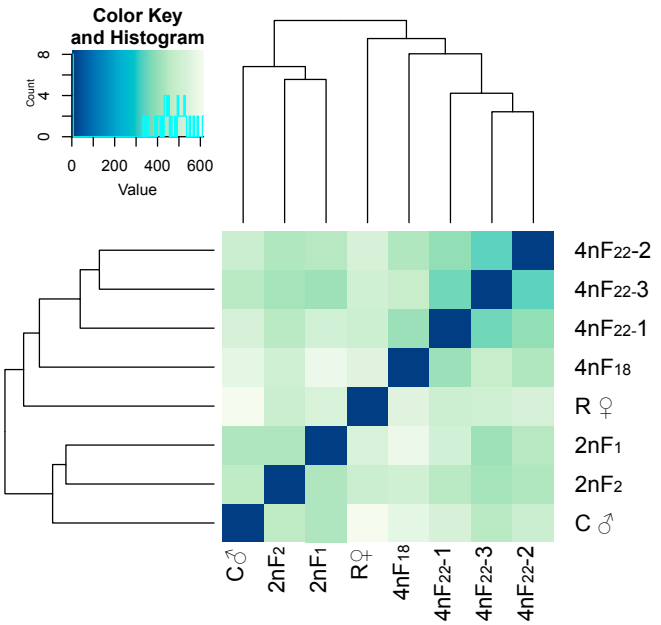

Fig. 3. Overall cluster in eight samples of hybrids and their progenitors from hybridizing goldfish (R) and common carp (C) using all normalized count data calculated by Cuffdiff. The heatmap, drawn from all gene count data by both genome references based on Euclidean distances, depicts the relationships of all transcriptomes.

Analyses Using a de Novo Assembly Strategy Obtained Results Compatible with the Mapping Results (*SI Appendix*, Table S7). Based on extracted orthologous sequences, the genes 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 $2nF_1$ and $2nF_2$ 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 $4nF_{22}$ 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 allopolyploidization 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 maybe also include parentage-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., $2nF_1$ and $2nF_2$ (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 $2nF_2$ 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 $4nF_{22}$ hybrids [the goldfish (♀) × common carp (♂) descendants (F_{22}-1, male; F_{22}-2, female; and F_{22}-3, male)], one 2-y-old hybrid male $4nF_{18}$, one 2-y-old female hybrid $2nF_1$, one 2-y-old female hybrid $2nF_2$, 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 common carp, two $2nF_2$, four $4nF_{18}$, and four $4nF_{22}$ hybrids (*SI Appendix*, Table S2 and Fig. S1). All liver tissue samples were excised carefully to avoid floral, fungal, bacterial, and faunal contamination from the gut. Samples were stored in RNAlater (Ambion) at −80 °C. A bacterial artificial chromosome (BAC) library of the goldfish was constructed, and a 9,468 bp-sized DNA fragment, with 36 copies of a repetitive 263 bp-sized DNA fragment, was sequenced and used as the FISH probe to characterize the chromosomes. The genome constitution of all eight samples was confirmed. 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 $4nF_{18}$, and performed sequencing steps on Illumina GAII, 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 transcriptome sequences (*SI Appendix*, Table S14).

URLs. The following URLs are used in this article: www.uniprot.org/ (Uniprot); https://github.com/linguoliang/VennPainter/ (VennPainter); www.genome.jp/tools/kaas/ (KAAS); and www.ebi.ac.uk/Tools/es/cgi-bin/clustalw2 (Clustalw2).

ACKNOWLEDGMENTS. We thank many researchers who helped with this study: Lin Chen, Fan Yu, Cuiping You, Weiguo He, Zhen Liu, Jie Hu, Dao Wang, Li Zou, Can Song, Tangnuo Li, Rong Yang, Xingguo Yi, Gang Liu, Honglin Liu, Xuewei Kang, Kang Xu, Xiang Xie, Ning Li, Wei Liu, Zhouling He, Lihua Xie, Lihai Ye, Yi Zhang, Zhuohui Zhang, Jing Dai, Zezhou Zhong, Ming Zeng, Zhanzhou Yao, Jinhui Liu, Hui Zhao, Linlin 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 Developmental Biology of Hunan Province (20134486), the Construction Project of Key Discipline of Hunan Province and China, the State Key Laboratory of Genetics, Resources and Evolution, Kunming Institute of Zoology, CAS, Strategic Priority Research Program of the Chinese Academy of Sciences (Grant XDB13040300), and the Program for Innovative Research Team (in Science and Technology) in the University of Yunnan Province.

1. Muller HJ (1925) Why polyploidy is rare in animals than in plants. *Am Nat* 59:346–353.
2. Mable BK (2004) 'Why polyploidy is rarer in animals than in plants': Myths and mechanisms. *Biol J Linn Soc Lond* 82(4):453–466.
3. Mable BK (2013) Polyploids and hybrids in changing environments: Winners or losers in the struggle for adaptation? *Heredity (Edinb)* 110(2):95–96.
4. Wertheim B, Beukeboom LW, van de Zande L (2013) Polyploidy in animals: Effects of gene expression on sex determination, evolution and ecology. *Cytogenet Genome Res* 140(2-4):256–269.
5. Chen ZJ (2007) Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu Rev Plant Biol* 58:377–406.
6. Otto SP, Whitton J (2000) Polyploid incidence and evolution. *Annu Rev Genet* 34: 401–437.
7. Liu B, Wendel JF (2002) Non-mendelian phenomena in allopolyploid genome evolution. *Curr Genomics* 3(6):489–505.
8. Jenczewski E, Chèvre AM, Alix K (2013) Chromosomal and gene expression changes in *Brassica* allopolyploids. *Polyploid and Hybrid Genomics*, eds Chen ZJ, Birchler JA (Wiley, Oxford), pp 171–186.
9. Adams KL, Wendel JF (2013) Dynamics of duplicated gene expression in polyploid cotton, in polyploid and hybrid genomics. *Polyploid and Hybrid Genomics*, eds Chen ZJ, Birchler JA (Wiley, Oxford), pp 187–194.

10. Xu C, et al. (2014) Genome-wide disruption of gene expression in allopolyploids but not hybrids of rice subspecies. *Mol Biol Evol* 31(5):1066–1076.
11. Liu S, et al. (2001) The formation of tetraploid stocks of red crucian carp x common carp hybrids as an effect of interspecific hybridization. *Aquaculture* 192(2-4):171–186.
12. He W, et al. (2013) The formation of diploid and triploid hybrids of female grass carp × male blunt snout bream and their 5S rDNA analysis. *BMC Genet* 14:110.
13. Ma W, et al. (2014) Allopolyploidization is not so simple: Evidence from the origin of the tribe Cyprinini (Teleostei: Cypriniformes). *Curr Mol Med* 14(10):1331–1338.
14. Liu S (2010) Distant hybridization leads to different ploidy fishes. *Sci China Life Sci* 53(4):416–425.
15. Bessert M, Sitzman C, Ortí G (2007) Avoiding paralogy: Diploid loci for allotetraploid blue sucker fish (*Cycleptus elongatus*, Catostomidae). *Conserv Genet* 8(4):995–998.
16. Ohno S, Muramoto J, Christian L, Atkin NB (1967) Diploid-tetraploid relationship among old-world members of the fish family Cyprinidae. *Chromosoma* 23(1):1–9.
17. Mank JE, Avise JC (2006) Phylogenetic conservation of chromosome numbers in Actinopterygiian fishes. *Genetica* 127(1-3):321–327.
18. Sun Y, et al. (2003) The chromosome number and gonadal structure of F_9~F_{11} allotetraploid crucian-carp. *Acta Genetica Sin* 30(5):414–418.
19. Coate JE, Doyle JJ (2013) Genomics and transcriptomics of photosynthesis in polyploids. *Polyploid and Hybrid Genomics*, eds Chen ZJ, Birchler JA (Wiley, Oxford), pp 153–169.
20. Hegarty MJ, Hiscock SJ (2008) Genomic clues to the evolutionary success of polyploid plants. *Curr Biol* 18(10):R435–R444.
21. Maekawa S, Matsumoto A, Takenaka Y, Matsuda H (2007) Tissue-specific functions based on information content of gene ontology using cap analysis gene expression. *Med Biol Eng Comput* 45(11):1029–1036.
22. Bilyk KT, Cheng CHC (2013) Model of gene expression in extreme cold: Reference transcriptome for the high-Antarctic cryopelagic notothenioid fish *Pagothenia borchgrevinki*. *BMC Genomics* 14:634.
23. Kusuda J, Hidari N, Hirai M, Hashimoto K (1996) Sequence analysis of the cDNA for the human casein kinase I delta (*CSNK1D*) gene and its chromosomal localization. *Genomics* 32(1):140–143.
24. Sciabica KS, Hertel KJ (2006) The splicing regulators *Tra* and *Tra2* are unusually potent activators of pre-mRNA splicing. *Nucleic Acids Res* 34(22):6612–6620.
25. Keightley MC, et al. (2013) In vivo mutation of pre-mRNA processing factor 8 (*Prpf8*) affects transcript splicing, cell survival and myeloid differentiation. *FEBS Lett* 587(14): 2150–2157.
26. Acker J, Murroni O, Mattei MG, Kedinger C, Vigneron M (1996) The gene (*POLR2L*) encoding the hRPB7.6 subunit of human RNA polymerase. *Genomics* 32(1):86–90.
27. Abu-Farha M, et al. (2011) Proteomic analyses of the *SMYD* family interactomes identify *HSP90* as a novel target for *SMYD2*. *J Mol Cell Biol* 3(5):301–308.
28. Osawa T, et al. (2011) Increased expression of histone demethylase *JHDM1D* under nutrient starvation suppresses tumor growth via down-regulating angiogenesis. *Proc Natl Acad Sci USA* 108(51):20725–20729.
29. Wada T, et al. (2008) Antagonistic control of cell fates by *JNK* and p38-MAPK signaling. *Cell Death Differ* 15(1):89–93.
30. Tomida T (2015) Visualization of the spatial and temporal dynamics of MAPK signaling using fluorescence imaging techniques. *J Physiol Sci* 65(1):37–49.
31. Avruch J (1998) Insulin signal transduction through protein kinase cascades. *Mol Cell Biochem* 182(1-2):31–48.
32. Ullah Z, Lee CY, Depamphilis ML (2009) Cip/Kip cyclin-dependent protein kinase inhibitors and the road to polyploidy. *Cell Div* 4:10.
33. Edgar BA, Orr-Weaver TL (2001) Endoreplication cell cycles: More for less. *Cell* 105(3): 297–306.
34. Duncan AW, et al. (2010) The ploidy conveyor of mature hepatocytes as a source of genetic variation. *Nature* 467(7316):707–710.
35. Zhang T, et al. (2015) Sequencing of allotetraploid cotton (*Gossypium hirsutum L. acc.* TM-1) provides a resource for fiber improvement. *Nat Biotechnol* 33(5):531–537.
36. Brenchley R, et al. (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* 491(7426):705–710.
37. Otto SP (2007) The evolutionary consequences of polyploidy. *Cell* 131(3):452–462.
38. Soltis PS, Soltis DE (2009) The role of hybridization in plant speciation. *Annu Rev Plant Biol* 60:561–588.
39. Woodhouse MR, et al. (2010) Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. *PLoS Biol* 8(6):e1000409.
40. Feldman M, et al. (1997) Rapid elimination of low-copy DNA sequences in polyploid wheat: A possible mechanism for differentiation of homoeologous chromosomes. *Genetics* 147(3):1381–1387.
41. Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC (2007) Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. *Plant Cell* 19(11):3403–3417.
42. Buggs RJ, et al. (2012) Rapid, repeated, and clustered loss of duplicate genes in allopolyploid plant populations of independent origin. *Curr Biol* 22(3):248–252.
43. Liu S, et al. (2014) The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nat Commun* 22(3):248–252.
44. Wang J, et al. (2006) Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. *Genetics* 172(1):507–517.
45. Liu B, Brubaker CL, Mergeai G, Cronn RC, Wendel JF (2001) Polyploid formation in cotton is not accompanied by rapid genomic changes. *Genome* 44(3):321–330.
46. Rogers RL, Bedford T, Hartl DL (2009) Formation and longevity of chimeric and duplicate genes in *Drosophila melanogaster*. *Genetics* 181(1):313–322.
47. Koyama H, Ito T, Nakanishi T, Sekimizu K (2007) Stimulation of RNA polymerase II transcript cleavage activity contributes to maintain transcriptional fidelity in yeast. *Genes Cells* 12(5):547–559.
48. Wang HC, Chou WC, Shieh SY, Shen CY (2006) Ataxia telangiectasia mutated and checkpoint kinase 2 regulate BRCA1 to promote the fidelity of DNA end-joining. *Cancer Res* 66(3):1391–1400.
49. Fedoroff NV (2012) Presidential address. Transposable elements, epigenetics, and genome evolution. *Science* 338(6108):758–767.
50. Bao J, Yan W (2012) Male germline control of transposable elements. *Biol Reprod* 86(5):162, 1-14.
51. Levin HL, Moran JV (2011) Dynamic interactions between transposable elements and their hosts. *Nat Rev Genet* 12(9):615–627.
52. Huang J, et al. (2009) RAD18 transmits DNA damage signalling to elicit homologous recombination repair. *Nat Cell Biol* 11(5):592–603.
53. Delmas S, Shunburne L, Ngo HP, Allers T (2009) Mre11-Rad50 promotes rapid repair of DNA damage in the polyploid archaeon *Haloferax volcanii* by restraining homologous recombination. *PLoS Genet* 5(7):e1000552.
54. Kobayashi J, et al. (2008) Current topics in DNA double-strand break repair. *J Radiat Res (Tokyo)* 49(2):93–103.
55. Zhao GY, et al. (2007) A critical role for the ubiquitin-conjugating enzyme *Ubc13* in initiating homologous recombination. *Mol Cell* 25(5):663–675.
56. van Gent DC, Hoeijmakers JH, Kanaar R (2001) Chromosomal stability and the DNA double-stranded break connection. *Nat Rev Genet* 2(3):196–206.
57. Cui P, Lin Q, Ding F, Hu S, Yu J (2012) The transcript-centric mutations in human genomes. *Genomics Proteomics Bioinformatics* 10(1):11–22.
58. Yang S, et al. (2015) Parent-progeny sequencing indicates higher mutation rates in heterozygotes. *Nature* 523(7561):463–467.
59. Fu S, et al. (2013) Alterations and abnormal mitosis of wheat chromosomes induced by wheat-rye monosomic addition lines. *PLoS One* 8(7):e70483.
60. Wu J, et al. (2014) Comparative cytological and transcriptomic analysis of pollen development in autotetraploid and diploid rice. *Plant Reprod* 27(4):181–196.
61. Jiang J, et al. (2013) Use of digital gene expression to discriminate gene expression differences in early generations of resynthesized *Brassica napus* and its diploid progenitors. *BMC Genomics* 14:72.
62. Xu Y, Zhao Q, Mei S, Wang J (2012) Genomic and transcriptomic alterations following hybridisation and genome doubling in trigenomic allohexaploid Brassica carinata × Brassica rapa. *Plant Biol (Stuttg)* 14(5):734–744.
63. Hegarty MJ, et al. (2006) Transcriptome shock after interspecific hybridization in senecio is ameliorated by genome duplication. *Curr Biol* 16(16):1652–1659.
64. Van de Peer Y, Maere S, Meyer A (2009) The evolutionary significance of ancient genome duplications. *Nat Rev Genet* 10(10):725–732.
65. Gerstein AC, Otto SP (2009) Ploidy and the causes of genomic evolution. *J Hered* 100(5):571–581.
66. Cuypers TD, Hogeweg P (2014) A synergism between adaptive effects and evolvability drives whole genome duplication to fixation. *PLOS Comput Biol* 10(4):e1003547.
67. De Smet R, Van de Peer Y (2012) Redundancy and rewiring of genetic networks following genome-wide duplication events. *Curr Opin Plant Biol* 15(2):168–176.
68. Kim D, et al. (2013) TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14(4):R36.
69. Trapnell C, et al. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 7(3):562–578.
70. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.
71. Li H, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
72. McKenna A, et al. (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20(9):1297–1303.
73. DePristo MA, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43(5):491–498.
74. Van der Auwera GA, et al. (2013) From FastQ data to high-confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinfomatics* 11(1110):11.10.1–11.10.33.
75. Danecek P, et al. (2011) The variant call format and VCFtools. *Bioinformatics* 27(15): 2156–2158.
76. Huang W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4(1):44–57.
77. Huang W, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37(1):1–13.