

Article

The Transcriptomes of the Crucian Carp Complex (*Carassius auratus*) Provide Insights into the Distinction between Unisexual Triploids and Sexual Diploids

Chun-Yan Li ^{1,2,3,4}, Jiong-Tang Li ^{3,*}, You-Yi Kuang ¹, Ru Xu ^{3,5}, Zi-Xia Zhao ³,
Guang-Yuan Hou ³, Hong-Wei Liang ⁶ and Xiao-Wen Sun ^{1,2,3,*}

¹ Heilongjiang River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Harbin 150001, China; E-Mails: xiaoyz0511@163.com (C.-Y.L.); youyikuang@hotmail.com (Y.-Y.K.)

² College of Fisheries and Life Science, Shanghai Ocean University, Shanghai 201306, China

³ The Centre for Applied Aquatic Genomics, Chinese Academy of Fishery Sciences, Beijing 100141, China; E-Mails: sdxuru@163.com (R.X.); zhaozx@cafs.ac.cn (Z.-X.Z.); hougy@cafs.ac.cn (G.-Y.H.)

⁴ Tianjin Fisheries Research Institute, Tianjin 300221, China

⁵ College of Life Science and Technology, Dalian Ocean University, Dalian 116023, China

⁶ Yangtze River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Wuhan 430223, China; E-Mail: lianghw@yfi.ac.cn

* Authors to whom correspondence should be addressed;

E-Mails: lij83@gmail.com (J.-T.L.); sunxw2002@163.com (X.-W.S.);

Tel.: +86-10-6867-3905 (J.-T.L. & X.-W.S.);

Fax: +86-10-6869-7522 (J.-T.L. & X.-W.S.).

Received: 29 January 2014; in revised form: 15 May 2014 / Accepted: 16 May 2014 /

Published: 27 May 2014

Abstract: Both sexual reproduction and unisexual reproduction are adaptive strategies for species survival and evolution. Unisexual animals have originated largely by hybridization, which tends to elevate their heterozygosity. However, the extent of genetic diversity resulting from hybridization and the genomic differences that determine the type of reproduction are poorly understood. In *Carassius auratus*, sexual diploids and unisexual triploids coexist. These two forms are similar morphologically but differ markedly in their modes of reproduction. Investigation of their genomic differences will be useful to study genome diversity and the development of reproductive mode. We generated transcriptomes for the unisexual and sexual populations. Genes were identified using homology searches and an *ab initio* method. Estimation of the synonymous substitution rate in the orthologous pairs indicated that the hybridization of gibel carp occurred 2.2 million years ago.

Microsatellite genotyping in each individual from the gibel carp population indicated that most gibel carp genes were not tri-allelic. Molecular function and pathway comparisons suggested few gene expansions between them, except for the progesterone-mediated oocyte maturation pathway, which is enriched in gibel carp. Differential expression analysis identified highly expressed genes in gibel carp. The transcriptomes provide information on genetic diversity and genomic differences, which should assist future studies in functional genomics.

Keywords: crucian carp; gibel carp; RNA-seq; genetic diversity; unisexual reproduction

1. Introduction

Among vertebrates, unisexual individuals have been reported in fish, amphibians, and reptiles [1]. Three modes of unisexual reproduction have been identified, gynogenesis, hybridogenesis, and parthenogenesis [1], but the cellular regulatory mechanisms that maintain unisexual reproduction are still poorly understood. The known unisexual teleost fishes are believed to have arisen by inter-species hybridization of sexual species [2]. Hybridization elevates ploidy and increases genetic diversity of the hybrid at its inception but subsequent genetic drift or diploidization could decrease genomic diversity. Therefore, it is important to explore the genomic diversity of hybrid vertebrates. Comparative genomic analysis between unisexual animals and their closely related sexual species will provide clues to the mechanisms of regulation of unisexual reproduction and allows estimation of the genomic diversity in the hybrid species.

Carassius auratus complex is characterized by the coexistence of sexual diploids and unisexual triploids [3]. The diploids and the triploids are quite similar morphologically but differ markedly in their modes of reproduction. The diploid form has 100 chromosomes and reproduces sexually. In the sexual reproduction, the sperm nuclei are able to transform into male pronuclei and fuse with the eggs. According to current taxonomy, the triploid individuals belong to the subspecies *Carassius auratus gibelio*, also named gibel carp, silver crucian carp, or prussian carp [4]. Gibel carp is believed to have originated by the ancient hybridization of a diploid female crucian carp gamete and a male common carp genome gamete [5]. It possesses 156–162 chromosomes [6,7] and has dual reproduction modes of unisexual gynogenesis and sexual reproduction [8,9]. Gynogenesis by heterogeneous spermatozoa activation is the dominant mode of reproduction and produces all female triploid offspring. In gynogenesis, gibel carp egg development is activated by sperm of other fish, but the incorporated sperm nucleus is kept in condensation and fails to form a male pronucleus [10]. Therefore, the heterologous sperm does not contribute genetically to the offspring. Sexual reproduction, the minor mode, generates sexual triploid progenies. It is estimated that a male ratio of about 20% in gibel carp population are produced from the sexual reproduction [11]. The difference of sexual reproduction and unisexual gynogenesis is likely related to some unknown regulatory mechanisms. The co-existence of the sexual form and the unisexual form makes this complex a promising model in which to study the mechanisms underlying their distinct reproductive modes and genetic diversity following hybridization.

C. auratus complex exhibits an additional round of genome duplication compared with other teleosts and this feature has been used to study the consequences of genome duplication [12]. In addition, fish in the *C. auratus* complex are hypoxia tolerant but the mechanisms are not fully understood [13]. These special genetic and phenotypic characteristics suggest that the *C. auratus* complex may be a suitable model to study genome duplication and physiological adaptation. Considering the promising applications of *C. auratus* in reproductive biology, genome duplication, and adaptive evolution, construction of the genomic resources of the complex will facilitate applications of this system in a number of fields of study.

In this research, we performed RNA sequencing for gibel carps and for diploid crucian carps to determine functional differences between them, and differentially expressed genes. We also investigated tri-allelic polymorphism in gibel carp. These studies provide hints on the genetic diversity within unisexual fish and the regulatory mechanisms concerned with unisexual reproduction. The transcriptomes of the crucian carp complex provide a representative resource for further functional and comparative analyses.

2. Results and Discussion

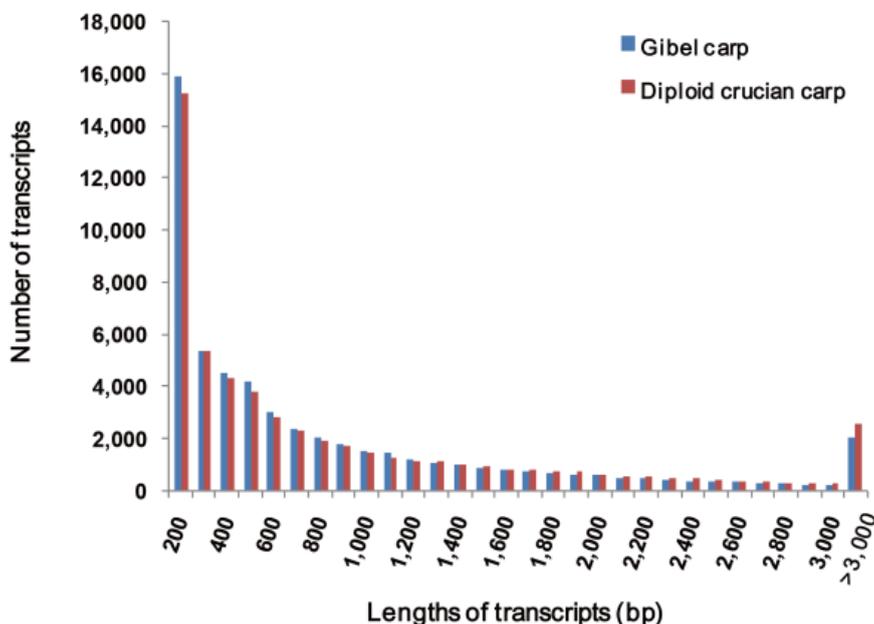
2.1. Transcriptome Sequencing and Assembly

The diploid and triploid individuals are quite similar morphologically and are difficult to differentiate by their appearance. The ploidy of the sequenced populations was determined by flow cytometry (Figure S1). The mean DNA content of gibel carps is 480, which is 1.5 times that of the diploid crucian carps (320). The ratio is consistent with their different ploidy types. Transcriptome sequencing yielded 11,669,953 and 12,135,538 pairs of reads for gibel carps and diploid crucian carps, respectively. Raw RNA-sequencing reads have been deposited at the NCBI Sequence Read Archive (SRA) under accession numbers SRR922167 and SRR924100. After filtering out the low-quality bases and *de novo* assembly, 65,476 and 67,297 transcripts were obtained in gibel carp and crucian carp, respectively. Based on sequence similarity, transcripts were clustered into genes and the longest transcript in each gene was selected as representative. Finally, we obtained 54,459 non-redundant transcripts (an N50 length of 1463 bp) in gibel carp and 53,839 non-redundant transcripts (an N50 length of 1672 bp) in diploid crucian carp. The length distributions of all sequences in the two fish are shown in Figure 1.

One goal of this study was to construct a representative transcriptome resource for the crucian carp complex. We applied saturation analysis to ascertain whether sequencing coverage was sufficient to draw a comprehensive picture of the transcriptome for the crucian carp complex. For each fish, rarefied libraries were constructed by randomly sampling from 10% to 100% of the transcriptome data. Then we produced new assemblies at each of the defined levels to illustrate possible differences in gene discovery rates. The curve for each species was already saturated (Figure S2), indicating that a large part of the genes were detected and that our study provided a comprehensive transcriptome resource for the crucian carp complex. Recently, Liao *et al.*, generated transcriptomes for four tissues of diploid crucian carp and identified differentially expressed genes among four tissues [14]. However, the transcriptome was collected from only four tissues, leading to short transcripts (N50: 547 bp) and

a large number of transcripts (127,711 unigenes). Compared with their result, our data provides a more representative collection of diploid crucian carp genes. In addition, we provide the most comprehensive transcriptome resource for gibel carp to date.

Figure 1. The length distributions of gibel carp transcripts and diploid crucian carp transcripts.



2.2. Sequence Annotation and Validation

To identify the protein-coding genes, we first carried out homolog searches. Homolog searches against fish Ensembl proteins, fish Ensembl transcripts, the NCBI non-redundant protein (nr) database, and UTRdb revealed 46,630 putative protein-coding genes in gibel carp and 44,402 in diploid crucian carp (Table 1). The *ab initio* prediction identified an additional 433 and 1260 coding genes in these two fish, respectively. Because these transcripts had no hits to known proteins, it is possible that they are species-specific protein-coding genes. The proportions of protein-coding transcripts among all transcripts here were 85.6% in gibel carp and 82.5% in diploid crucian carp. These values are higher than those recorded in other transcriptome studies in non-model organisms [15–17].

Table 1. Summary of the annotation of transcripts of gibel carp and diploid crucian carp.

	Database	Gibel Carp	Diploid Crucian Carp
Homolog search	Fish Ensembl proteins	32,030	31,620
	Fish Ensembl transcripts	13,513	10,956
	NCBI “nr” protein database	235	56
	UTRdb	852	1770
	NONCODE database	336	491
	NCBI “nt” transcript database	4004	1386
<i>ab initio</i> prediction	BESTORF	433	1260
Unknown		3056	6300
Total		54,459	53,839

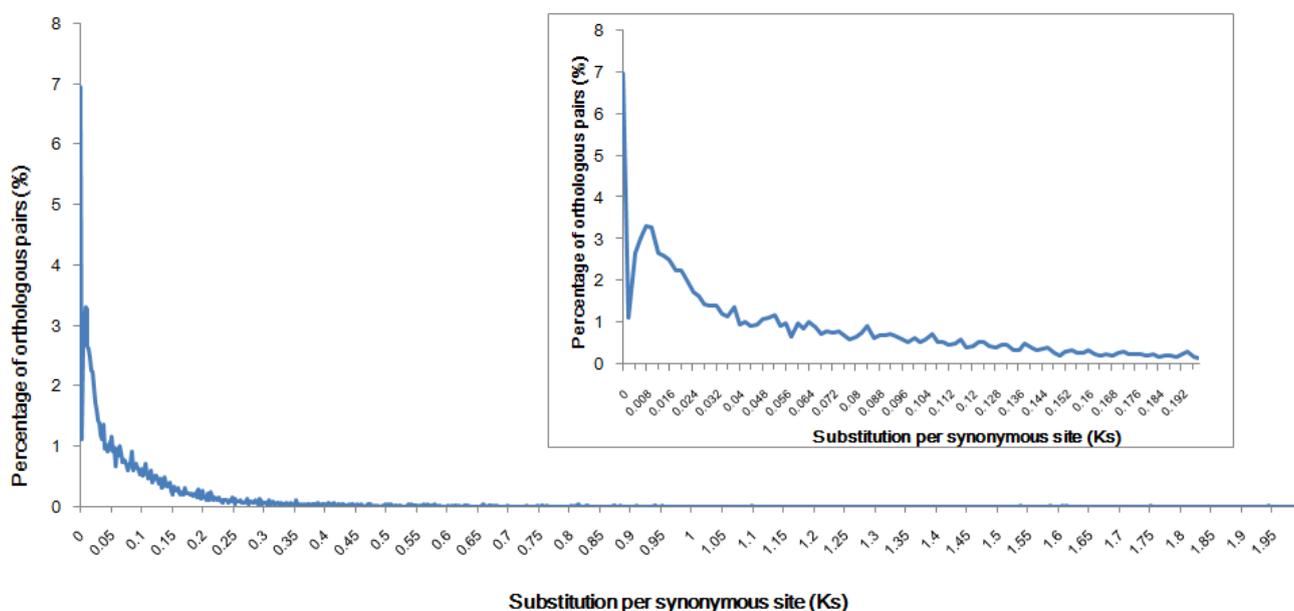
In gibel carp, we also found 4340 transcripts either homologous to known non-coding genes in the NONCODE database or aligned to NCBI “nt” transcripts, indicating that they may be non-coding RNAs (ncRNAs). A total of 1877 transcripts in diploid crucian carp were putative ncRNAs. The remaining unknown 3056 gibel carp transcripts and 6300 diploid crucian carp transcripts had neither protein-coding potential nor known homologs, indicating that they were probably transcribed from intergenic regions of the gibel carp and diploid crucian carp genomes.

To evaluate the accuracy of our assemblies, 12 pairs of paralogs in gibel carp and diploid crucian carp were randomly selected for RT-PCR and specific primers were designed for the selected transcripts. All of the selected transcripts could be amplified (Figure S3), indicating that they were actually expressed and correctly assembled.

2.3. Genome Speciation Event Deduced from Orthologous Pairs

The speciation time between diploid crucian carp and gibel carp has not been reported previously. A secondary peak in the distribution of orthologous K_s values indicates a speciation event [18]. We estimated the genome speciation time based on the K_s distribution of orthologous pairs between the two species. We identified 18,974 orthologous pairs between diploid crucian carp and gibel carp using the reciprocal best blast hit approach. The K_s distribution of these orthologous pairs showed a distinct secondary peak at 0.008 (with a mode at $K_s = 0.006$ to 0.01) (Figure 2).

Figure 2. K_s distribution for identification of the speciation event. Data were grouped into bins of 0.002 K_s units for graphing. The K_s distribution of the orthologous pairs is plotted as the blue line and shows a distinct secondary peak, indicating the time of divergence of these two species.



For analysis of speciation time, we used the rate of 3.51×10^{-9} substitutions/synonymous site per year obtained for fourfold degenerate sites of 47 mammalian genes [19]. The rate was estimated using the method of Pamilo and Bianchi [20], assuming that the divergence time between the human and rodent lineages was 80 million years ago. With the constant clock-like rate, the speciation between

diploid crucian carp and gibel carp was estimated to have occurred 2.2 million years ago (with a range of 1.7–2.8 million years). It is believed that the triploid gynogen originated from an ancient hybridization event with crucian carp the maternal ancestor and common carp (*Cyprinus carpio*) the paternal ancestor [5]. Therefore, this is possibly one of the most recent inter-species hybridizations among vertebrates.

2.4. Low Tri-Allelic Polymorphism and Heterozygosity in Gibel Carp

Hybridization could elevate the genetic diversity of the unisexual species at its inception [1]. However, subsequent ploidy restoration could lead to greatly reduced or completely absent heterozygosity [21,22]. In addition, it is well known that sexual reproduction increases the genetic diversity of offspring while gynogenesis does not. Therefore, it is interesting to compare the genomic diversity between gibel carp and diploid crucian carp. We identified 628,711 polymorphic sites in 21,615 diploid crucian carp transcripts and 509,699 sites in 22,463 gibel carp transcripts. The average diversity level of gibel carp was 22.7 polymorphic sites per transcript, lower than the one of diploid crucian carp (29.1 sites per transcript).

Given the tri-ploidy in gibel carp, theoretically tri-allelic polymorphism might be observed in many genomic regions. However, the frequency of tri-allelic polymorphism in gibel carp is unknown. To investigate the level of tri-allelic polymorphism, we plotted the allele number per polymorphic site. Among 509,699 polymorphic sites in gibel carp, only 7681 belonged to tri-allelic polymorphism while the remaining loci were di-allelic.

The di-allelic polymorphism in the triploid population could come from the following genotypes in individuals, including AAB, BBA, BBB, or AAA. Because di-allelic polymorphism was prevalent in the genome, mono-allelic genotypes and di-allelic genotypes widely existed in most polymorphic loci while the tri-allelic genotypes (ABC) would distribute in only few loci. To validate our hypothesis, we further investigated the proportion of tri-allelic genotypes by microsatellite genotyping in each individual. A total of 3953 microsatellites were identified in 3258 gibel carp transcripts and 4153 microsatellites in 3382 diploid crucian carp transcripts. In each of the two species, the top microsatellites were di-nucleotide motifs (54.2% and 53.7%, Table S1). Eighteen conserved microsatellite loci were genotyped in the gibel carp population. This analysis showed that most loci were still homozygous or di-allelic heterozygous while the loci of tri-allelic heterozygosity accounted for only 22% (Table S2).

A previous study using microsatellite genotyping in 94 gibel carp individuals also found that most loci were homozygous or diploid heterozygous [23]. One explanation of the low tri-allelic polymorphism is that the two inter-breeding species were genetically very similar. The extant triploid represents the fusion of a diploid female crucian carp gamete (AB) and a male common carp genome gamete (C). The various genotypes of triploids can be classified on the basis of their genome constitution, as ABC (if the three alleles are heterozygous to one another), AAC (if A and B are identical), and AAA (if all three alleles are identical). Crucian carp and common carp both belong to the family Cyprinidae and because of their great similarity both AAC and AAA types would be prevalent in the triploid genome. An alternative explanation is that genetic drift following the hybridization event pushed most loci to fixation. In the ancient triploid, each locus might initially have

had three alleles. If a reasonable length of time elapses before the species becomes diploidized, genetic drift could cause some loci to become fixed for alleles that originated from one parent, whereas other loci might retain three alleles from the two parents. A third possible explanation is that diploidization followed inter-species hybridization. Diploidization is assumed to occur through the accumulation of DNA sequence mutations and/or deletions between sister chromosomes [24], leading to the loss of alleles. The diploidized loci then exhibit homozygosity or di-allelic heterozygosity. Summarizing, whatever the explanation for the low frequency of tri-allelic polymorphisms, investigations into tri-allelic polymorphisms will open the door to understanding the whole-genome ploidy level of other polyploids that originated by hybridization.

Microsatellites and SNPs identified within these two species will be useful for future molecular selection and breeding. Because most transcripts in the two fish were protein-coding, the identified SNPs were mainly located in protein-coding genes. The SNPs in the coding regions could change protein sequences and functions. In addition, SNPs in the UTRs might affect the regulation of miRNAs to target genes [25]. Therefore, the SNPs mined here may help to advance functional studies of SNPs and the identification of phenotype-associated SNPs.

2.5. Few Gene Expansions in Gibel Carp

As described above, only a few genomic loci in gibel carp showed tri-allelic heterozygosity, reflecting the evolution of ploidy after hybridization. To investigate gene expansions in gibel carp resulting from inter-species hybridization, we compared molecular functions and pathways between the two forms. Such comparisons may also indicate possible mechanisms underlying the phenotypic differences. Using homologous assignment, we assigned 4041 GO terms to 50.1% of the gibel carp transcripts (27,310 of 54,459). A similar percentage was assigned in diploid crucian carp (50%, 4066 terms to 26,957 transcripts). These two fish shared 3891 GO terms, indicating that most of their molecular functions were common to the two fish. We then used WEGO to find significantly enriched GO terms in gibel carp using diploid crucian carp transcripts as the background. In the molecular function category, only four GO terms, including binding and transferase activity, were significantly overrepresented while four terms, mainly transporter activities, were underrepresented in gibel carp (Table 2).

KEGG pathway analysis mapped 7132 gibel carp transcripts to 162 zebrafish KEGG pathways. A total of 6966 diploid crucian carp transcripts were mapped to 162 zebrafish pathways and used as a background to compare pathway differences. The statistically enriched pathways are shown in Table 3. Only three pathways were enriched in gibel carp and four pathways in diploid crucian carp, consistent with the observation of few GO differences between them. The consistency of these two independent analyses indicates that few genes are expanded in gibel carp.

Table 2. The enriched molecular functions in gibel carp and diploid crucian carp (p value < 0.05).

Species	GO Term	GO ID	Percentage of Transcripts (%) *	p Value
Gibel carp	nucleic acid binding	GO:0003676	13.9:13.2	0.009
	transferase activity	GO:0016740	10.4:9.8	0.012
	transferase activity, transferring phosphorus-containing groups	GO:0016772	6.6:6.2	0.028
	DNA binding	GO:0003677	7.6:7.1	0.033
Diploid crucian carp	transporter activity	GO:0005215	4.1:4.7	0.000
	transmembrane transporter activity	GO:0022857	3.2:3.9	0.000
	passive transmembrane transporter activity	GO:0022803	1.0:1.4	0.000
	substrate-specific transmembrane transporter activity	GO:0022891	2.7:3.3	0.000
	substrate-specific transporter activity	GO:0022892	4.1:4.7	0.001

* Percentage of transcripts: the first number is the proportion of gibel carp transcripts in this molecular function and the second number is the percentage of diploid crucian carp transcripts.

Table 3. The enriched pathways in gibel carp and diploid crucian carp identified by KOBAS (p value < 0.05).

Species	KEGG Pathway	Percentage of Gibel Carp Transcripts (%)	Percentage of Diploid Crucian Carp Transcripts (%)	p Value
Gibel carp	Ubiquitin mediated proteolysis	1.70	1.38	0.013
	Progesterone-mediated oocyte maturation	1.13	0.89	0.018
	Fanconi anemia pathway	0.61	0.46	0.038
Diploid crucian carp	Neuroactive ligand-receptor interaction	1.07	1.53	0.000
	Cell adhesion molecules	0.91	1.18	0.010
	Ribosome	0.55	0.74	0.024
	Tight junction	1.51	1.76	0.045

Gene expansion or loss may lead to some enriched functions or pathways in one species. As expected, most molecular functions and pathways were common to these two fish and only a small number were enriched in each species. These differences in molecular functions and pathways might be responsible for the unisexual gynogenesis of gibel carp. Interestingly, the gibel carp transcripts were enriched in the progesterone-mediated oocyte maturation pathway. This pathway is reported to increase the maturation promoting factor (MPF) [26], which might lead to the tripolar spindle formation and the later modification of microtubule dynamics [27]. The enriched GO terms in this fish, concerned with transferase activity and the transferring phosphorus-containing groups, are much related involved in the enriched progesterone-mediated oocyte maturation pathway [28]. The ubiquitin-mediated proteolysis pathway and the Fanconi Anemia pathway were also significantly enriched in gibel carp. The MPF consists of *cdc2* and *cyclin B* [29]. The degradation of cyclins is essential to regulate the transition from mitosis to the next cell cycle and is regulated by the ubiquitin-mediated proteolytic pathway [30]. The Fanconi Anemia pathway participates in DNA repair

and restores chromosomal integrity [31]. These two pathways might also be genetic mechanisms responsible for unisexual gynogenesis in gibel carp.

2.6. Differentially Expressed Orthologous Genes in Pooled Tissues and Gonads

Gene expansion analysis can provide hints on the possible mechanisms leading to phenotypic differences. Differential gene expression analysis may also help us investigate these mechanisms. We identified 314 differentially expressed genes (DEGs) in the pooled tissues between diploid crucian carp and gibel carp, which accounted for only 1.7% of orthologous pairs. That is, most of the orthologous genes had similar expression levels in these two forms. Hierarchical clustering on the basis of expression patterns indicated that the DEGs were classified into two major groups (Figure 3). In the first group, the expression of orthologous genes was high in both fish but was higher in gibel carp than in diploid crucian carp. There were seven genes in this group. Of particular interest was the fish-egg lectin-like precursor, which is reported to have a function in embryonic development [32]. How this precursor functions in the specialized mode of reproduction of gibel carp needs further investigation.

Figure 3. Expression profiling of differentially expressed orthologous genes in the two species. Hierarchical clustering shows that these genes can be classified into two groups (Group I and Group II). The group II could be further divided into two types.



To validate our RNA-Seq results, 20 genes with statistically significant differential expression were selected for quantitative reverse transcription-polymerase chain reaction (RT-qPCR) analysis in the two populations. These genes participated in functions associated with the immune-response, cell junctions, transferases, transporter activities, molecular binding, and signaling. Overall, except for two genes without significant differences, the expression patterns of 18 genes by RT-qPCR were significantly different and similar to those indicated by the RNA-Seq analysis (Figure 4a). Thus, the similar patterns suggested by RNA-Seq analysis and RT-qPCR validated the genome-wide expression profiles in pooled tissues of the two subspecies.

The differential expression analysis with the RNA-Seq and RT-qPCR suggests that these molecular functions might be involved in the phenotypic differences. The most important phenotypic differences relate to their modes of reproduction. Therefore, we further investigated the expression patterns of these 20 genes in the gonads with RT-qPCR. Of 17 genes with significant difference, the patterns of 12 (71%) in the gonads were consistent with those in the pooled tissues (Figure 4b). Interestingly, some immune-response genes, including *c5* and *vtnb*, were up-regulated in gibel carp gonads compared with diploid crucian carp. It was reported that the complement 1q (*C1q*) family of proteins function as both immunological mediators and in vertebrate oogenesis and oocyte maturation [33]. Hence, these immune-response proteins in gibel carp may play roles in the special mode of unisexual reproduction. These data also demonstrated up-regulation of other genes in gibel carp gonads. In particular, the fish-egg lectin-like precursor was markedly up-regulated more than 4.5 times in gibel carp gonads. This up-regulation is consistent with our observation in the pooled tissues. Fish-egg lectin was reported to be expressed differently in gibel carp and diploid crucian carp and up-regulated in gibel carp [34]. Our comparative data are consistent with this study.

Transcriptome comparison provided another layer of possible mechanisms for phenotypic differences resulting from differences in gene expression. Although the comparison demonstrated that most orthologs had similar expression levels, our analysis identified some differentially expressed genes. Taken together, these data provide important information on the distinction between unisexual triploids and sexual diploids.

3. Experimental Section

3.1. Sample Collection and Transcriptome Sequencing

We collected two populations of mature female gibel carp and female diploid crucian carp (each population, $N = 10$, 100–150 g, non-infectious) from Yangtze River Fisheries Research Institute, Wuhan, China. Use of the samples for all experiments was approved by the Ethics Committee of the institution. The sex of each fish was determined by examination of the gonads. The ploidy types of these fish were confirmed based on their DNA contents using Cell Lab Quanta SC flow cytometer (Beckman Coulter, Brea, CA, USA). Red blood cells were collected from the caudal vein of each individual with syringes containing sodium heparin. The blood samples were resuspended in Nuclear Isolation Media (NIM)-DAPI staining solution (NPE Systems, Pembroke Pines, FL, USA) for 10 min. The DNA content of each individual was compared with that of chicken. The DNA of each sample was isolated using the QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany). DNA concentrations were

measured with NanoVue Plus Spectrophotometer (GE Healthcare, Little Chalfont, UK) and the integrity was confirmed by analysis on a 1% agarose gel.

In each fish, tissue samples were excised from the brain, muscle, liver, skin, kidney, gill, intestine, gonad, spleen, and heart. Total RNA was extracted from each of the ten tissues using the Trizol Kit (Invitrogen, Carlsbad, CA, USA) and followed by DNase I (Invitrogen, Carlsbad, CA, USA) treatment according to the manufacturer's protocol. For each individual, equal quantities (1 µg) of the RNA from the ten tissues were pooled. Finally, the pooled RNA from ten individuals of the same population was checked with Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA) for the integrity. The RNA integrity number (RIN) values for both fish were 7.8, indicating that the quality satisfied the following commercial Illumina sequencing and RT-qPCR [35].

The library construction from purifying mRNA to enriching DNA fragments was performed with the Illumina TruSeq RNA Sample Prep Kit (Illumina, San Diego, CA, USA). Briefly, for each fish, polyA mRNA was extracted from 10 µg RNA using poly-T oligo attached magnetic beads. During the elution of the polyA RNA, the RNA was fragmented and primed with random hexamers using the fragment and prime mix (Illumina, San Diego, CA, USA). Following the Prep Kit, the purified RNA was fragmented under the conditions: 94 °C for 8 min followed by a final hold at 4 °C. The fragmentation resulted in libraries with sizes ranging from 300 to 400 bp. The cleaved RNA fragments primed with random hexamers were reversely transcribed into first strand cDNA with Superscript II reverse transcriptase (Invitrogen, Carlsbad, CA, USA). The process was performed under the following conditions: incubation at 25 °C for 10 min followed by 15 min at 42 °C; then 70 °C for 15 min; final hold at 4 °C. After the second strand cDNA was synthesized and a single "A" nucleotide was added to the 3' end of the blunt ds cDNA fragment, the adapters with a single "T" nucleotide on the 3' end were ligated to the fragments. Finally, those DNA fragments with adapters on both ends were enriched using PCR. The amplification was performed under the conditions: initial denaturation at 98 °C for 30 s; then 15 cycles at 98 °C for 10 s, at 60 °C for 30 s, and then at 72 °C for 30 s; final extension at 72 °C for 5 min. The libraries were then sequenced on an Illumina HiSeq 2000 platform (Illumina, Inc., San Diego, CA, USA) with read lengths of 2 × 100 nt.

3.2. Transcriptome Assembly, Annotation, and Validation

For each individual, the raw transcriptome reads were processed using SolexaQA [36] to filter low-quality reads. The high-quality reads were assembled using Trinity [37]. The Trinity contigs were further assembled using CAP3 [38]. Because the similarity between two crucian carp paralogs was as high as 96% [39], to avoid misassembly of paralogs into one gene, we set the overlap percent identity cutoff as 97. That is, if two transcripts had an overlapping region with an identity over 97%, they were considered from the same gene and should be assembled further. The output contigs were subjected to SSPACE, a stand-alone scaffolder of pre-assembled contigs using paired-read data [40], for scaffolding. To avoid the identification of redundant genes as a result of alternative splicing, we followed the strategy of Wang *et al.* [17], where all-against-all BLASTN (National Center for Biotechnology Information, Bethesda, MD, USA) searches were performed using SSPACE transcripts. If the alignment of two transcripts had 100% identity over 100 bp, then these two transcripts were considered as spliced variants and the longest transcript was selected to represent this gene.

To ascertain whether sequencing depth was sufficient to draw a comprehensive picture of the transcriptome for the crucian carp complex, we constructed different libraries by randomly sampling from 10% to 100% of the transcriptome data. The reads from different libraries were assembled using Trinity and CAP3. Then, the assembled transcript numbers were plotted for different read numbers. A curve was drawn to indicate if sequencing effort was deep enough.

To identify protein-coding genes, we used a combination of homolog searches and an *ab initio* prediction method. First, we searched homologous fish proteins (zebrafish, fugu, tetraodon, medaka, and stickleback) for our assembled transcripts using BLASTX (National Center for Biotechnology Information, Bethesda, MD, USA) with an e-value of 10^{-5} . The e-value cutoff was used in many homology detection strategies [41,42]. The fish proteins were obtained from the Ensembl database [43]. The transcripts without BLASTX hits were then searched against fish transcripts (zebrafish, fugu, tetraodon, medaka, and stickleback) using BLASTN. To identify as many homologous genes as possible, we increased the e-value cutoff to 10^{-2} . The transcripts were downloaded from the Ensembl database [43] and the UCSC genome database [44]. Transcripts that had no match to either of these databases were aligned against the NCBI “nr” protein database using BLASTX. The remaining unmatched transcripts were further aligned against UTRdb [45] using BLASTN. The hit transcripts were considered to be the untranslated regions (UTRs) of protein-coding genes. For the transcripts without hits to UTRdb, we carried out a BLASTN search against the NONCODE database [46] and the NCBI “nt” nucleotide database. Second, for the transcripts without homology to the above databases, we predicted their putative open reading frames (ORFs) using BESTORF (<http://linux1.softberry.com/berry.phtml?topic=bestorf&group=programs&subgroup=gfind>). Because species-specific protein-coding genes exist, these genes could not be identified using homolog searches. Because short putative ORFs could be predicted by chance within noncoding RNAs (ncRNAs), a minimum ORF cutoff is usually applied to reduce the likelihood of falsely categorizing ncRNAs as mRNAs [47]. To distinguish mRNAs from transcriptomes, we adopted a cutoff of 150 nt (50 amino acids) [48]. That is, if BESTORF predicted an intact ORF over 50 amino acids, this transcript was considered as a protein-coding gene.

To annotate the functions of the genes, we assigned Gene Ontology information [49] of homologous fish proteins to gibel carp genes and crucian carp genes. To study the pathways that the genes may participate in, we ran KOBAS software [50] to map transcripts to zebrafish pathways in the Kyoto Encyclopedia of Genes and Genomes (KEGG) [51].

To validate the assembled transcripts, the pooled RNAs were used for cDNA synthesis. We randomly selected 12 pairs of paralogs and designed specific primers for them (Table S3). PCR reactions were conducted in a 15 μ L volume containing 300 ng cDNA, 10 μ M primers, 10 \times universal PCR buffer, and 0.5 U of *Taq* polymerase (Fermentas, Burlington, ON, Canada). PCR was performed under the following conditions: initial denaturation at 95 $^{\circ}$ C for 5 min; then 35 cycles at 94 $^{\circ}$ C for 30 s, at a primer-specific annealing temperature for 30 s, and then at 72 $^{\circ}$ C for 45 s; final extension at 72 $^{\circ}$ C for 10 min. PCR products were then checked on a 1.5% agarose gel.

3.3. Genome Speciation Event Deduced from the *Ks* Distribution of Orthologous Pairs

Since there are no available annotated protein sequences or complete genome sequences for these two fish, the proteome-based ortholog detection strategies, such as Inparanoid [52] and OrthoMCL [41], are not suitable for our analysis. To identify reliable orthologs between gibel carp and diploid crucian carp, we adopted the BLAST-based Reciprocal Best Hit (RBH) method by our previous strategy [17] and Blanc *et al.* [18]. The sequences from the two species were aligned using the reciprocal BLAST (BLASTN) hit method. If each of two aligned sequences was the best hit of the other, and if they were aligned over 300 bp, they were defined as orthologs. Chen *et al.* concluded that the BLAST-based RBH method had a low false positive rate of 8% [53]. The RBH method has been widely applied into identifying orthologs in those species of which proteins were not annotated [18,53].

The approach used to estimate the *Ks* of orthologous pairs was also adapted from previous strategy [18,54]. Briefly, the gibel carp transcripts were aligned with their orthologous diploid crucian carp transcripts with TBLASTX. The longest alignment was selected for analysis. In each pair, the corresponding aligned sequence was extracted and translated with the *getorf* program in the EMBOSS package [55]. Then, the translated amino acid sequences were aligned using Clustalw [56]. The corresponding codon alignments were generated using PAL2NAL [57]. Finally, we calculated the *Ks* of each orthologous pair using a maximum likelihood method in the CODEML program (runmode-2) of the PAML package [58].

3.4. The Analysis of Tri-Allelic Polymorphism and Heterozygosity in Gibel Carp

To compare the diversity level between two populations and study the proportion of tri-allelic polymorphism in gibel carp, we estimated the SNP allele number per polymorphic locus. For each species, we aligned the sequencing reads to the representative transcripts with the CLC Genomics Workbench (<http://www.clcbio.com/products/clc-genomics-workbench/>). The general alignment parameters were set to the default values except that non-specific matches were ignored. We adjusted the minimum read coverage to 5 [15].

The prevalent di-allelic polymorphism in the population indicated that the tri-allelic genotypes (ABC) would exist in only few loci in genome. Thus, we examined the proportion of the tri-allelic heterozygosity by genotyping microsatellite loci in each individual of the sequencing population. Microsatellites were identified using Msatfinder (<http://www.genomics.ceh.ac.uk/msatfinder/>). The repeat thresholds for di-, tri-, tetra-, penta-, and hexa-nucleotide motifs were set as 8, 5, 5, 5, and 5 respectively. Only microsatellite sequences with flanking sequences longer than 50 bp on both sides were collected. Eighteen microsatellites (Table S4, Supplementary Information) were genotyped in individuals following the pipeline of Zhang *et al.* [59].

3.5. Gene Expansions in Special Molecular Functions and Pathways

As a consequence of inter-species hybridization, the proportion of genes in a protein family might be higher in gibel carp than in diploid crucian carp, leading to species-specific gene expansion of particular protein-families, and enriched molecular functions. This preliminary comparative study of gene expansion in molecular functions could provide clues concerning the functions that might

underlie the phenotypic differences between these two fish, particularly in their reproductive modes. To study the distinct molecular functions and pathways within the different phenotypes, we identified significantly enriched GO terms in gibel carp transcripts using diploid crucian carp genes as the background using WEGO [60]. Terms with p value <0.05 were considered to be enriched in gibel carp. To study the enriched pathways in gibel carp, we compared the proportion of gibel carp transcripts in each KEGG pathway against the proportion of diploid crucian carp genes in the same pathway using KOBAS, assuming a hypergeometric distribution. If the proportion of gibel carp transcripts in one KEGG pathway was significantly higher than that of diploid crucian carp transcripts (p value < 0.05), this KEGG pathway was considered statistically enriched in gibel carp.

3.6. Analysis and Validation of Differentially Expressed Orthologous Genes

Differential gene expression might also explain the phenotypic differences. Sequencing reads were mapped to the assembled reference transcriptomes of the two species using Bowtie [61]. RSEM (RNA-Seq by Expectation Maximization), an accurate method of transcript quantification from RNA-Seq data with or without a reference genome [62], was used to estimate the gene abundance. The expressions of orthologous genes were joined and then normalized with edgeR (Empirical analysis of digital gene expression data in R) [63]. After pair-wise comparison, the differentially expressed genes (DEGs) were obtained with stringent cutoffs: FDR (false discovery rate)-corrected p value cutoff of 0.001 and minimum fold change of 4. The FDR correction is designed to control the expected proportion of incorrectly rejected null hypotheses and used in multiple-hypothesis testing to reduce Type-1 errors [64].

To categorize the DEGs according to their expression patterns, a heat map chart was constructed by transforming the normalized data to a log₂ scale for visualization. Hierarchical clustering on the basis of expression was performed using the “gplots” package of the R program. The DEGs were further clustered into different groups based on their functional annotations.

We selected twenty DEGs and used RT-qPCR to validate their expression levels in the pooled tissues and in the gonads of triploid gibel carp and diploid crucian carp. The pooled RNA samples were used for RT-qPCR. We followed the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) Guidelines [65] for quality assessment. Purity of the four RNA samples was checked by measuring the A_{260}/A_{280} ratio with a NanoVue Plus Spectrophotometer (GE Healthcare, Little Chalfont, UK). These samples all demonstrated high purities with ratios over 2.0 (Table S5). RNA integrities were assessed with Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). For each sample, clearly visible 28S and 18S rRNA peaks demonstrated the high integrity (Figure S4). The RIN values for these samples were 7.8 or higher (Table S6), also indicating the high integrity [66,67]. Then, cDNA was synthesized using about 3 μ g of total RNA with the RevertAid™ H Minus First Strand cDNA Synthesis Kit (Fermentas, Burlington, ON, Canada). The beta-actin gene was used as the reference gene because of its validated stable expression in different tissues and conditions of crucian carp [68,69]. Since β -actin has been widely applied in gene expression as an internal control gene in crucian carp [70,71], we consider that the expression of β -actin is stable and could be used as an efficient and single reference gene in our study [72,73]. To determine whether inhibition was occurring during PCR, β -actin was amplified in different dilution series of cDNA. The

amplification plots in these samples were showed in Figure S5. The significant linear correlation between quantification cycle values (C_q) and DNA concentrations demonstrated the absence of inhibition during PCR. The primers of the analyzed genes were shown in Table S7. RT-qPCR was performed on an ABI PRISM 7500 Real-time Detection System (Applied Biosystems, Foster City, CA, USA). The amplification was performed in a total volume of 15 μ L, containing 7.5 μ L $2\times$ SYBR Green Realtime PCR Master Mix (Toyobo, Osaka, Japan), 1 μ L cDNA (100 ng/ μ L), and 0.3 μ L of 10 μ M of each gene-specific primer. The PCR cycle was 50 °C for 2 min, 95 °C for 10 min, 40 cycles of 95 °C for 15 s, and 60 °C for 1 min. All reactions were set up in triplicate. The amplification efficiencies of the twenty DEGs and the reference gene were 1, calculated with 7500 Real-Time PCR System version 2.0 software (Applied Biosystems, Foster City, CA, USA). The expression of each DEG was normalized to that of beta-actin. The comparative C_q method ($2^{-\Delta\Delta C_q}$ method) was used to analyze the expression of the target genes. After normalization, the levels of the genes of gibel carp are stated relative to the orthologous genes of diploid crucian carp. Data are expressed as means \pm SE. Statistical differences between the two groups were determined by the Student's t -test.

4. Conclusions

In this study, we performed transcriptome sequencing for two forms of *C. auratus*, one reproducing sexually and the other unisexually. The assembled transcriptomes and their annotations permitted functional characterization and comparative analysis. The analysis enabled us to investigate: (1) the extent of tri-allelic polymorphism in gibel carp; and (2) the putative regulatory mechanisms for their different modes of reproduction. The transcriptome sequences, annotations, microsatellites, and SNPs obtained in this study will be valuable resources for basic research, including investigations of reproductive biology, genome duplication, and physiological adaptation.

Acknowledgments

This study was supported by the grants from China Ministry of Science and Technology 863 Hi-Tech Research and Development Program (2011AA100401 and 2011AA100402), Special Scientific Research Funds for Central Non-profit Institutes, Chinese Academy of Fishery Sciences (2014A11JC07 and 2014C010).

Author Contributions

Jiong-Tang Li and Xiao-Wen Sun conceived and designed the experiments. Chun-Yan Li and Jiong-Tang Li conducted the bioinformatic analysis and were involved in writing the manuscript. You-Yi Kuang, Ru Xu, Zi-Xia Zhao, Guang-Yuan Hou, and Hong-Wei Liang were involved in one or more processes of samples collection, data analysis and manuscript preparation.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Neaves, W.B.; Baumann, P. Unisexual reproduction among vertebrates. *Trends Genet.* **2011**, *27*, 81–88.
2. Lamatsch, D.K.; Stöck, M. Sperm-dependent parthenogenesis and hybridogenesis in teleost fishes. In *Lost Sex*; Springer: Dordrecht, The Netherlands, 2009; pp. 399–432.
3. Xiao, J.; Zou, T.; Chen, Y.; Chen, L.; Liu, S.; Tao, M.; Zhang, C.; Zhao, R.; Zhou, Y.; Long, Y.; *et al.* Coexistence of diploid, triploid and tetraploid crucian carp (*Carassius auratus*) in natural waters. *BMC Genet.* **2011**, *12*, 20.
4. Jiang, F.F.; Wang, Z.W.; Zhou, L.; Jiang, L.; Zhang, X.J.; Apalikova, O.V.; Brykov, V.A.; Gui, J.F. High male incidence and evolutionary implications of triploid form in northeast Asia *Carassius auratus* complex. *Mol. Phylogenet. Evol.* **2013**, *66*, 350–359.
5. Chun, F.L.; Yang, S.T.; Gui, J.F. Differential screening and characterization analysis of the egg envelope glycoprotein ZP3 cDNAs between gynogenetic and gonochoristic crucian carp. *Cell Res.* **2001**, *11*, 17–27.
6. Yi, M.S.; Li, Y.Q.; Liu, J.D.; Zhou, L.; Yu, Q.X.; Gui, J.F. Molecular cytogenetic detection of paternal chromosome fragments in allogynogenetic gibel carp, *Carassius auratus gibelio* Bloch. *Chromosome Res.* **2003**, *11*, 665–671.
7. Zhou, L.; Gui, J.F. Karyotypic diversity in polyploid gibel carp, *Carassius auratus gibelio* Bloch. *Genetica* **2002**, *115*, 223–232.
8. Zhu, H.P.; Ma, D.M.; Gui, J.F. Triploid origin of the gibel carp as revealed by 5S rDNA localization and chromosome painting. *Chromosome Res.* **2006**, *14*, 767–776.
9. Gui, J.F.; Zhu, Z.Y. Molecular basis and genetic improvement of economically important traits in aquaculture animals. *Chin. Sci. Bull.* **2012**, *57*, 1751–1760.
10. LI, C.J.; GUI, J.F. Comparative studies on *in vitro* sperm decondensation and pronucleus formation in egg extracts between gynogenetic and bisexual fish. *Cell Res.* **2003**, *13*, 159–169.
11. Wang, D.; Mao, H.L.; Peng, J.X.; Li, X.Y.; Zhou, L.; Gui, J.F. Discovery of a male-biased mutant family and identification of a male-specific SCAR marker in gynogenetic gibel carp *Carassius auratus gibelio*. *Prog. Nat. Sci.* **2009**, *19*, 1537–1544.
12. Yuan, J.A.; He, Z.Z.; Yuan, X.N.; Jiang, X.Y.; Sun, X.W.; Zou, S.M. Speciation of polyploid cyprinidae fish of common carp, crucian carp, and silver crucian carp derived from duplicated hox genes. *J. Exp. Zool. Part B Mol. Dev. Evol.* **2010**, *314B*, 445–456.
13. Bickler, P.E.; Buck, L.T. Hypoxia tolerance in reptiles, amphibians, and fishes: Life with variable oxygen availability. *Annu. Rev. Phys.* **2007**, *69*, 145–170.
14. Liao, X.; Cheng, L.; Xu, P.; Lu, G.; Wachholtz, M.; Sun, X.; Chen, S. Transcriptome analysis of Crucian Carp (*Carassius auratus*), an important aquaculture and hypoxia-tolerant species. *PLoS One* **2013**, *8*, e62308.
15. Zhou, Z.C.; Dong, Y.; Sun, H.J.; Yang, A.F.; Chen, Z.; Gao, S.; Jiang, J.W.; Guan, X.Y.; Jiang, B.; Wang, B. Transcriptome sequencing of sea cucumber (*Apostichopus japonicus*) and the identification of gene-associated markers. *Mol. Ecol. Resour.* **2014**, *14*, 127–138.

16. Robertson, L.S.; Cornman, R.S. Transcriptome resources for the frogs *Lithobates clamitans* and *Pseudacris regilla*, emphasizing antimicrobial peptides and conserved loci for phylogenetics. *Mol. Ecol. Resour.* **2014**, *14*, 178–183.
17. Wang, J.T.; Li, J.T.; Zhang, X.F.; Sun, X.W. Transcriptome analysis reveals the time of the fourth round of genome duplication in common carp (*Cyprinus carpio*). *BMC Genomics* **2012**, *13*, 96.
18. Blanc, G.; Wolfe, K.H. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* **2004**, *16*, 1667–1678.
19. David, L.; Blum, S.; Feldman, M.W.; Lavi, U.; Hillel, J. Recent duplication of the common carp (*Cyprinus carpio* L.) genome as revealed by analyses of microsatellite loci. *Mol. Biol. Evol.* **2003**, *20*, 1425–1434.
20. Pamilo, P.; Bianchi, N. Evolution of the *Zfx* and *Zfy* genes: Rates and interdependence between the genes. *Mol. Biol. Evol.* **1993**, *10*, 271–281.
21. Sinclair, E.A.; Pramuk, J.B.; Bezy, R.L.; Crandall, K.A.; Sites, J.W., Jr. DNA evidence for nonhybrid origins of parthenogenesis in natural populations of vertebrates. *Evolution* **2010**, *64*, 1346–1357.
22. Watts, P.C.; Buley, K.R.; Sanderson, S.; Boardman, W.; Ciofi, C.; Gibson, R. Parthenogenesis in Komodo dragons. *Nature* **2006**, *444*, 1021–1022.
23. Bai, Z.Y.; Liu, F.; Li, J.L.; Yue, G.H. Identification of Triploid Individuals and Clonal Lines in *Carassius auratus* Complex Using Microsatellites. *Int. J. Biol. Sci.* **2011**, *7*, 279–285.
24. Wolfe, K.H. Yesterday's polyploids and the mystery of diploidization. *Nat. Rev. Genet.* **2001**, *2*, 333–341.
25. Zhu, Y.P.; Xue, W.; Wang, J.T.; Wan, Y.M.; Wang, S.L.; Xu, P.; Zhang, Y.; Li, J.T.; Sun, X.W. Identification of common carp (*Cyprinus carpio*) microRNAs and microRNA-related SNPs. *BMC Genomics* **2012**, *13*, 413.
26. Voronina, E.; Marzluff, W.F.; Wessel, G.M. Cyclin B synthesis is required for sea urchin oocyte maturation. *Dev. Biol.* **2003**, *256*, 258–275.
27. Yang, Z.A.; Li, Q.H.; Wang, Y.F.; Gui, J.F. Comparative investigation on spindle behavior and MPF activity changes during oocyte maturation between gynogenetic and amphimictic crucian carp. *Cell Res.* **1999**, *9*, 145–154.
28. Morrill, G.A.; Dowd, T.L.; Kostellow, A.B.; Gupta, R.K. Progesterone-induced changes in the phosphoryl potential during the meiotic divisions in amphibian oocytes: Role of Na/K-ATPase. *BMC Dev. Biol.* **2011**, *11*, 67.
29. Pace, M.C.; Thomas, P. Steroid-induced oocyte maturation in Atlantic croaker (*Micropogonias undulatus*) is dependent on activation of the phosphatidylinositol 3-kinase/Akt signal transduction pathway. *Biol. Reprod.* **2005**, *73*, 988–996.
30. Glotzer, M.; Murray, A.W.; Kirschner, M.W. Cyclin is degraded by the ubiquitin pathway. *Nature* **1991**, *349*, 132–138.
31. Moldovan, G.L.; D'Andrea, A.D. How the fanconi anemia pathway guards the genome. *Annu. Rev. Genet.* **2009**, *43*, 223–249.
32. Galliano, M.; Minchiotti, L.; Campagnoli, M.; Sala, A.; Visai, L.; Amoresano, A.; Pucci, P.; Casbarra, A.; Cauci, M.; Perduca, M.; *et al.* Structural and biochemical characterization of a new type of lectin isolated from carp eggs. *Biochem. J.* **2003**, *376*, 433–440.

33. Chen, B.; Gui, J. Identification of a novel C1q family member in color crucian carp (*Carassius auratus*) ovary. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **2004**, *138*, 285–293.
34. Dong, C.H.; Yang, S.T.; Yang, Z.A.; Zhang, L.; Gui, J.F. A C-type lectin associated and translocated with cortical granules during oocyte maturation and egg fertilization in fish. *Dev. Biol.* **2004**, *265*, 341–354.
35. Eizirik, D.L.; Sammeth, M.; Bouckenooghe, T.; Bottu, G.; Sisino, G.; Igoillo-Esteve, M.; Ortis, F.; Santin, I.; Colli, M.L.; Barthson, J. The human pancreatic islet transcriptome: Expression of candidate genes for type 1 diabetes and the impact of pro-inflammatory cytokines. *PLoS Genet.* **2012**, *8*, e1002552.
36. Cox, M.P.; Peterson, D.A.; Biggs, P.J. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinform.* **2010**, *11*, 485.
37. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.D.; *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644–652.
38. Huang, X.; Madan, A. CAP3: A DNA sequence assembly program. *Genome Res.* **1999**, *9*, 868–877.
39. Sandvik, G.K.; Tomter, A.B.; Bergan, J.; Zoppellaro, G.; Barra, A.L.; Rohr, A.K.; Kolberg, M.; Ellefsen, S.; Andersson, K.K.; Nilsson, G.E. Studies of ribonucleotide reductase in crucian carp—an oxygen dependent enzyme in an anoxia tolerant vertebrate. *PLoS One* **2012**, *7*, e42784.
40. Boetzer, M.; Henkel, C.V.; Jansen, H.J.; Butler, D.; Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **2011**, *27*, 578–579.
41. Li, L.; Stoeckert, C.J.; Roos, D.S. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.* **2003**, *13*, 2178–2189.
42. Enright, A.J.; van Dongen, S.; Ouzounis, C.A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **2002**, *30*, 1575–1584.
43. Flicek, P.; Ahmed, I.; Amode, M.R.; Barrell, D.; Beal, K.; Brent, S.; Carvalho-Silva, D.; Clapham, P.; Coates, G.; Fairley, S.; *et al.* Ensembl 2013. *Nucleic Acids Res.* **2013**, *41*, D48–D55.
44. Meyer, L.R.; Zweig, A.S.; Hinrichs, A.S.; Karolchik, D.; Kuhn, R.M.; Wong, M.; Sloan, C.A.; Rosenbloom, K.R.; Roe, G.; Rhead, B.; *et al.* The UCSC Genome Browser database: Extensions and updates 2013. *Nucleic Acids Res.* **2013**, *41*, D64–D69.
45. Pesole, G.; Liuni, S.; Grillo, G.; Ippedico, M.; Larizza, A.; Makalowski, W.; Saccone, C. UTRdb: A specialized database of 5' and 3' untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res.* **1999**, *27*, 188–191.
46. Bu, D.; Yu, K.; Sun, S.; Xie, C.; Skogerbo, G.; Miao, R.; Xiao, H.; Liao, Q.; Luo, H.; Zhao, G.; *et al.* NONCODE v3.0: Integrative annotation of long noncoding RNAs. *Nucleic Acids Res.* **2012**, *40*, D210–D215.
47. Dinger, M.E.; Pang, K.C.; Mercer, T.R.; Mattick, J.S. Differentiating Protein-Coding and Noncoding RNA: Challenges and Ambiguities. *PLoS Comput. Biol.* **2008**, *4*, e1000176.
48. Zhang, G.F.; Fang, X.D.; Guo, X.M.; Li, L.; Luo, R.B.; Xu, F.; Yang, P.C.; Zhang, L.L.; Wang, X.T.; Qi, H.G.; *et al.* The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **2012**, *490*, 49–54.

49. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; *et al.* Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **2000**, *25*, 25–29.
50. Xie, C.; Mao, X.Z.; Huang, J.J.; Ding, Y.; Wu, J.M.; Dong, S.; Kong, L.; Gao, G.; Li, C.Y.; Wei, L.P. KOBAS 2.0: A web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* **2011**, *39*, W316–W322.
51. Kanehisa, M.; Goto, S.; Hattori, M.; Aoki-Kinoshita, K.F.; Itoh, M.; Kawashima, S.; Katayama, T.; Araki, M.; Hirakawa, M. From genomics to chemical genomics: New developments in KEGG. *Nucleic Acids Res.* **2006**, *34*, D354–D357.
52. Östlund, G.; Schmitt, T.; Forslund, K.; Köstler, T.; Messina, D.N.; Roopra, S.; Frings, O.; Sonnhammer, E.L. InParanoid 7: New algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Res.* **2010**, *38*, D196–D203.
53. Chen, F.; Mackey, A.J.; Vermunt, J.K.; Roos, D.S. Assessing performance of orthology detection strategies applied to eukaryotic genomes. *PLoS One* **2007**, *2*, e383.
54. Srivastava, A.; Rogers, W.L.; Breton, C.M.; Cai, L.M.; Malmberg, R.L. Transcriptome analysis of sarracenia, an insectivorous plant. *DNA Res.* **2011**, *18*, 253–261.
55. Rice, P.; Longden, I.; Bleasby, A. EMBOSS: The european molecular biology open software suite. *Trends Genet.* **2000**, *16*, 276–277.
56. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; *et al.* Clustal W and clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948.
57. Suyama, M.; Torrents, D.; Bork, P. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **2006**, *34*, W609–W612.
58. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591.
59. Zhang, X.; Zhang, Y.; Zheng, X.; Kuang, Y.; Zhao, Z.; Zhao, L.; Li, C.; Jiang, L.; Cao, D.; Lu, C.; *et al.* A consensus linkage map provides insights on genome character and evolution in common carp (*Cyprinus carpio* L.). *Mar. Biotechnol.* **2013**, *15*, 275–312.
60. Ye, J.; Fang, L.; Zheng, H.K.; Zhang, Y.; Chen, J.; Zhang, Z.J.; Wang, J.; Li, S.T.; Li, R.Q.; Bolund, L.; *et al.* WEGO: A web tool for plotting GO annotations. *Nucleic Acids Res.* **2006**, *34*, W293–W297.
61. Langmead, B.; Trapnell, C.; Pop, M.; Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **2009**, *10*, R25.
62. Li, B.; Dewey, C.N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **2011**, *12*, 323.
63. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010**, *26*, 139–140.
64. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **1995**, *57*, 289–300.
65. Bustin, S.A.; Benes, V.; Garson, J.A.; Hellems, J.; Huggett, J.; Kubista, M.; Mueller, R.; Nolan, T.; Pfaffl, M.W.; Shipley, G.L. The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* **2009**, *55*, 611–622.

66. Schroeder, A.; Mueller, O.; Stocker, S.; Salowsky, R.; Leiber, M.; Gassmann, M.; Lightfoot, S.; Menzel, W.; Granzow, M.; Ragg, T. The RIN: An RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol. Biol.* **2006**, *7*, 3.
67. Fleige, S.; Pfaffl, M.W. RNA integrity and the effect on the real-time qRT-PCR performance. *Mol. Asp. Med.* **2006**, *27*, 126–139.
68. Cheng, H.L.; Wang, X.; Peng, Y.X.; Meng, X.P.; Sun, S.P.; Shi, X.Y. Molecular cloning and tissue distribution of lipoprotein lipase full-length cDNA from Pengze crucian carp (*Carassius auratus* var. Pengze). *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **2009**, *153*, 109–115.
69. Somamoto, T.; Yoshiura, Y.; Sato, A.; Nakao, M.; Nakanishi, T.; Okamoto, N.; Ototake, M. Expression profiles of TCRbeta and CD8 α mRNA correlate with virus-specific cell-mediated cytotoxic activity in ginbuna crucian carp. *Virology* **2006**, *348*, 370–377.
70. An, L.H.; Lei, K.; Zheng, B.H. Use of heat shock protein mRNA expressions as biomarkers in wild crucian carp for monitoring water quality. *Environ. Toxicol. Pharmacol.* **2014**, *37*, 248–255.
71. Shi, J.; Zhang, Y.B.; Liu, T.K.; Sun, F.; Gui, J.F. Subcellular localization and functional characterization of a fish IRF9 from crucian carp *Carassius auratus*. *Fish. Shellfish. Immunol.* **2012**, *33*, 258–266.
72. Bustin, S. Transparency of reporting in molecular diagnostics. *Int. J. Mol. Sci.* **2013**, *14*, 15878–15884.
73. Bustin, S.A.; Benes, V.; Garson, J.; Hellemans, J.; Huggett, J.; Kubista, M.; Mueller, R.; Nolan, T.; Pfaffl, M.W.; Shipley, G.; *et al.* The need for transparency and good practices in the qPCR literature. *Nat. Methods* **2013**, *10*, 1063–1067.

© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).