Symmetric expression of ohnologs encoding conserved antiviral responses in tetraploid common carp suggest absence of subgenome dominance after whole genome duplication

Annemiek Blasweiler, Hendrik Jan Jan Megens, Mark R.G. Goldman, Roni Tadmor-Levi, Jackie Lighten, Martien A.M. Groenen, Ron P.H. Dirks, Hans J. Jansen, Herman P. Spaink, Lior David, et al.

Genomics 115 (2023) 110723

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Symmetric expression of ohnologs encoding conserved antiviral responses in tetraploid common carp suggest absence of subgenome dominance after whole genome duplication

A. Blasweiler [a,b,*], H.-J. Megens [b], M.R.G. Goldman [a], R. Tadmor-Levi [c], J. Lighten [d], M.A.M. Groenen [b], R.P. Dirks [e], H.J. Jansen [e], H.P. Spaink [f], L. David [c], P. Boudinot [g], G. F. Wiegertjes [a]

[a] Aquaculture and Fisheries, Wageningen University, the Netherlands
[b] Animal Breeding and Genomics, Wageningen University, the Netherlands
[c] Dept. of Animal Sciences, RH Smith Faculty of Agriculture Food and Environment, The Hebrew University of Jerusalem, Israel
[d] Biosciences, University of Exeter, United Kingdom
[e] Future Genomics Technologies B.V., the Netherlands
[f] Institute of Biology, Leiden University, Leiden, Netherlands
[g] Université Paris-Saclay, INRAE, UVSQ, VIM, Jouy-en-Josas, 78350, France

ARTICLE INFO

Keywords:
Whole genome duplication
Subgenome dominance
Antiviral response
Common carp
Ohnolog expression

ABSTRACT

Allopolyploids often experience subgenome dominance, with one subgenome showing higher levels of gene expression and greater gene retention. Here, we address the functionality of both subgenomes of allotetraploid common carp (*Cyprinus carpio*) by analysing a functional network of interferon-stimulated genes (ISGs) crucial in anti-viral immune defence. As an indicator of subgenome dominance we investigated retainment of a core set of ohnologous ISGs. To facilitate our functional genomic analysis a high quality genome was assembled (WagV4.0). Transcriptome data from an in vitro experiment mimicking a viral infection was used to infer ISG expression. Transcriptome analysis confirmed induction of 88 ISG ohnologs on both subgenomes. In both control and infected states, average expression of ISG ohnologs was comparable between the two subgenomes. Also, the highest expressing and most inducible gene copies of an ohnolog pair could be derived from either subgenome. We found no strong evidence of subgenome dominance for common carp.

1. Introduction

Whole genome duplications (WGDs) have played a major role in the evolution of plants and animals [1,2,3,4]. Although WGDs have been observed more frequently in plants and invertebrates, the root of the vertebrate evolutionary tree is marked by two WGD events [5,6–8]. In addition, WGDs have occurred in various branches of the vertebrate tree, including in ray finned fishes known as Teleostei. More recent rounds of WGDs occurred in several teleostei groups [9,10,11,12], in most cases recent enough for these species to be considered tetraploids.

Duplicate gene copies originating from a WGD event are referred to as ohnologs [13,14]. Duplicate genes may eventually be lost over time, a process known as rediploidisation ([15,16]; [17]). Gene loss occurs by relieve of evolutionary constraints, due to functional redundancy and accumulation of deleterious mutations. Although generally one gene copy is enough to maintain the ancestral function, some ohnologs attain non-overlapping functions (subfunctionalization or neo-functionalization) and both gene copies are then conserved. The degree to which ohnologs are maintained, therefore, depends on time since duplication and on functional significance [14].

WGD events can be divided into two forms, alloploidy, which refers to interspecific hybridization, and autoploidy, which refers to direct intraspecific genome doubling [18]. A teleost example of an ancient autotetraploidy event are salmonids. The ancestor of salmonids has undergone a direct genome doubling event about 80 million years ago (MYA) [10]. More recent teleost examples of allotetraploids are presented by the common carp (*Cyprinus carpio*) and goldfish (*Carassius auratus*), who share an allotetraploid ancestor resulting from

* Corresponding author at: Aquaculture and Fisheries, Wageningen University, the Netherlands.
E-mail address: blasweiler_a@outlook.com (A. Blasweiler).

https://doi.org/10.1016/j.ygeno.2023.110723
Received 13 June 2023; Received in revised form 22 August 2023; Accepted 25 September 2023
Available online 5 October 2023

interspecific hybridization about 12 MYA [19,20,21,22].

The allotetraploid common carp has two subgenomes, 'A' and 'B' [20,21,23,24], which do not recombine [19]. An important question in such allotetraploid genomes is if in the absence of recombination subgenomes each retain function of all genes, or, if not, how rediploidization occurs, and how fast. In this process it may be especially relevant to understand if one subgenome is dominant as shown by consistently higher gene expression in one subgenome compared to the other. If subgenome dominance exists it may, in part, explain asymmetries in rates of gene loss or subfunctionalization, and therefore is an important feature to quantify in detail.

It has been proposed that in common carp both subgenomes are equally under purifying selection, retaining a high degree of ohnologs in near equal numbers [11,12]. Recently, research on the common carp suggested a certain degree of subgenome dominance; although both subgenomes A and B actively express their genes, subgenome B expression on average is slightly higher than subgenome A expression [22,12]. For a total of 7536 expressed ohnologous gene pairs investigated, [12] reported expression differences >2-fold change in at least one tissue for 5403 homeologs in subgenome B compared to 4719 homeologs in subgenome A.

Such a broad characterization of mean expression level, however, does not address the full functional contributions of each subgenome. It omits important properties such as the degree to which ohnologous genes of either subgenome may be induced when required. Function of well-annotated ohnologs in allotetraploid common carp are best studied addressing genes with known phylogenetic history and functional roles in molecular pathways [25,26]. Here, to study common carp subgenome expression and function of ohnologous genes in a biological context, we created a high quality, chromosome-level assembly. We focused on a broad but conserved set of induced genes united in their general function - but not their molecular mechanisms or phylogenetic origin - to combat expansion of viral infections within the infected host. A high quality assembly at chromosome level is essential for an unambiguous allocation of duplicated genes to the two subgenomes, while the choice to study a conserved set of induced genes considered important to combat viral infections provides a broad analysis across the genome.

During anti-viral immune responses, secreted interferon initiates a cascade of expression of interferon stimulated genes (ISGs), resulting in various enzymes, transcription-factors and secreted molecules to be released within and outside the virus-infected cell. Central to the induced expression of ISGs is activation of the interferon type I pathway, an immune pathway well-studied across vertebrates [27,28,29]. ISGs collectively comprise a broad set of genes, diverse in function and distributed across the genome, well-characterized for their induced expression in response to interferon type-1 following viral infection. As such the ISGs form an interesting case into subgenome constituent and induced expression and function. We studied the evolution of common carp ISGs function based on information from 72 ortholog groups including most of the known signalling components of the IFN system as well as key effectors, previously identified as a core ancestral ISG repertoire present in both zebrafish and humans [30]. We address whether the full core set of ISGs has been retained in duplicate, and whether there is a difference in ohnolog basal and induced gene expression at subgenome level. We discuss to which degree our data influence conclusions on the presence of subgenome dominance for allotetraploid common carp.

2. Results

2.1. A new common carp genome assembly for subgenome identification

The genome of common carp was sequenced, first with Oxford nanopore long-read technology (Supplementary Table 1) and then with Illumina short-read technology (resulting in 483,633,600 paired end reads) to decrease sequencing error rates and assembly polishing. The reads were assembled into 50 major scaffolds, corresponding to the 50 expected chromosomes. These scaffolds showed high continuity (Fig. 1) and high congruence with linkage map data (Supplementary material 2) The HiC contact map shows that there are no major interactions between chromosomes and thus, indicates no major remaining large-scale assembly inconsistencies (Fig. 1). The resulting genome assembly Cypcar_WagV4.0 is therefore assembled at chromosome level (Supplementary Table 1). The assembly, and gene annotation, is available as Cypcar_WagV4.0 (GCA_905221575.1) since Ensembl Release 106.

2.2. The tetraploid structure of the carp genome

We identified 25 pairs of chromosomes formed by the alloploidy event with the help of existing linkage map markers. Linkage map markers from one chromosome secondarily mapped to the ohnologous chromosome of the chromosome pair. The 25 chromosome pairs of the common carp were divided into two subgenomes (A and B) by nucleotide sequence comparisons with the Tapien genome (Fig. 2). Sequence comparisons consistently showed one of the chromosomes in a carp chromosome pair always having a higher mapping percentage to Tapien compared to its pair counterpart, hence sorted as part of subgenome B (Supplementary Table 2). In general, all chromosome pairs have comparable sizes, except for the pair of chromosome 6 and 44. Chromosome 6 has a size of 31,259,323 bp, comparable to other chromosomes, while chromosome 44 is the smallest chromosome in the assembly with only 18,875,570 bp (Supplementary Table 2). Next, the common carp ($N = 50$) assembly was mapped to the Ensembl zebrafish assembly ($N = 25$). The carp chromosome pairs showed sequence similarity to zebrafish chromosomes allowing unambiguous inference of synteny to zebrafish (Supplementary Table 3, Fig. 3).

2.3. Genome completeness and subgenome retention

A BUSCO analysis was performed on Cypcar_WagV4.0 (Table 1) with a vertebrate and Actinopterygii gene set of 2586 and 4584 genes respectively, to assess genome and subgenome completeness. The Cypcar_WagV4.0 assembly showed BUSCO completeness between 98.3% - 97.0% for complete genes, indicating the carp assembly is nearly complete. The analysis also showed the high degree of duplicated genes that are still present in the common carp genome by complete and duplicated BUSCOs ranging from 64.2% - 66.2% (Table 1). For further insight in functional completeness of each of the common carp subgenomes, the BUSCO analyses were repeated for each subgenome separately. The BUSCO analyses of both subgenomes showed high gene completeness; 86.4–92%, and low presence of duplicated genes within subgenome A; 2.6%–3.5%, and subgenome B; 2.9%–3.6%. This demonstrates that the high degree of duplication in common carp is the result of the genome duplication, as expected, with subgenome B showing slightly higher intactness compared to subgenome A. Differences in the number of missing BUSCO genes between the subgenomes were not significant (Chi-square test statistic 0.2986 *P*-value 0.585).

2.4. Interferon stimulated genes in the carp genome

An established ancestral set of zebrafish ISGs ($n = 103$), belonging to 72 ortholog groups that are conserved between humans and zebrafish, was used to annotate ISGs in the new carp assembly. In total, 88 ISGs were identified in the Cypcar_WagV4.0 genome (Fig. 4 & Supplementary Table 4). The protein domains of the 88 ISGs identified in carp were predicted and their orthologs were identified also in zebrafish confirming their conservation in carp. The 88 ISGs included members of all 72 conserved ortholog groups except *cxcl11.3*. Though we did find a candidate region for *cxcl11.3*, we could not identify a predicted gene in either subgenome. The number of 88 ISGs identified in carp is smaller than the original set of 103 ISGs identified in zebrafish for several reasons. From the 103 ISGs, *ifi44a5* and o*gfrl1* belonging to ortholog groups

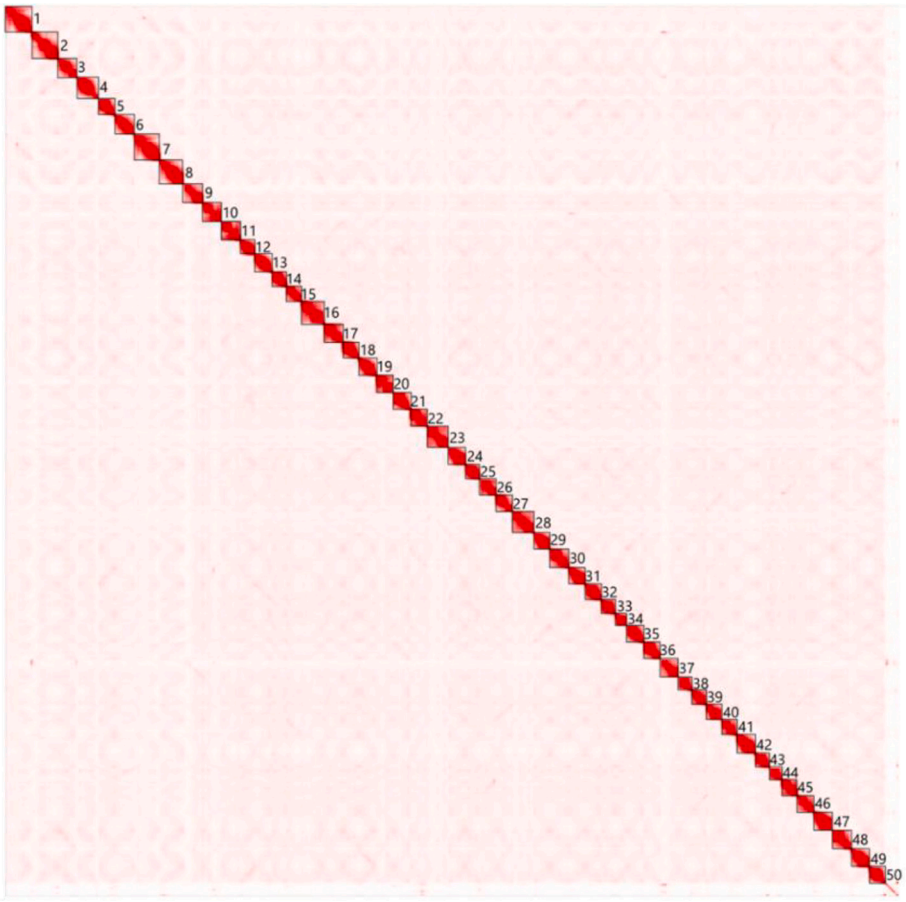

Fig. 1. HiC contact map of carpV4 genome assembly.
Shows chromosome conformation capture interaction matrix. Degree of red colouring indicates intensity of interaction between and within chromosomes. Chromosomes 1–50 are shows by black outlined squires. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

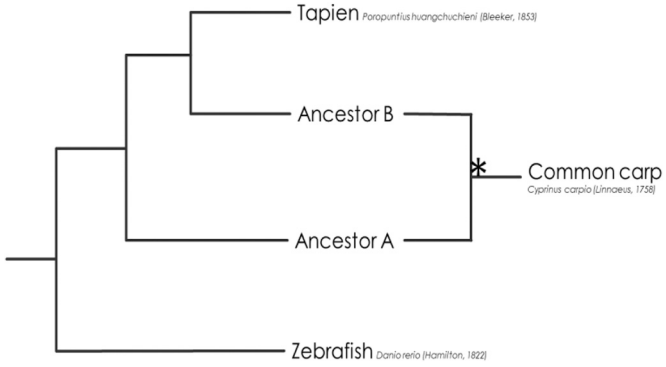

Fig. 2. Common carp Phylogeny.
graphical representation of phylogenetic relationships between zebrafish (*Danio rerio*), Tapien (*Poropuntius huangchuchieni*), common carp (*Cyprinus carpio*). * sign indicates WGD event.

ifi44 and ogfrl respectively, could not be identified (Supplementary Table 4), suggesting they were lost in the common carp. Further, six single copy ISGs of common carp are duplicated several times in zebrafish. (Supplementary Table 4). For example, from two zebrafish ISG genes adjacent to each other, known annotated in zebrafish as myxovirus (influenza virus) resistance C and E (*mxc* and *mxe)*, only one gene was found in the orthologous position of the carp genome. Based on sequence similarity it was not possible to distinguish which zebrafish mx gene the carp gene was most similar to. The carp gene is referred to as *mxc_e* in this study. Of the 88 ISGs present in carp, eight were present in only one subgenome (Fig. 4, Supplementary Table 4 & Supplementary Fig. 2). For subgenome A these were *rigi*, *apol and mxc_e* and for subgenome B, *samd9*, *parp12b*, *tap1*, *b2mlike* and *trex3*. Neither protein or nucleotide sequence BLAST searches, synteny analysis or RNA-seq reads comparison indicated gene presence of these eight genes in their subgenome counterpart or unassembled contigs. We identified three ISGs that have undergone extra duplications in only one subgenome resulting in carp specific paralogs that were duplicated post WGD (Supplementary Table 4 and Supplementary Fig. 3). The ISG *gimap*, *rarres3* and *lgals3bpb* are found on both subgenomes but also have an extra duplicate on subgenome B.

Of the 88 ISGs identified in common carp, 74 (84%) were present as ohnologs, one in each subgenome (column A1 and B1 Supplementary Table 4). One ohnolog of the gene *rnf213a* had to be excluded from further analysis due to the gene annotation not being consistent with the latest Ensembl annotation. The final result was a set of 73 ohnologs to compare with a 1:1 ratio and synteny between the carp subgenomes.

2.5. Gene expression of ohnologous genes

RNA-seq data were mapped to Cypcar_WagV4.0 and showed high unique mapping scores with little evidence of subgenome mapping ambiguity. A principal component analysis of 6 control and 6 Poly I:C treated samples distinctively separated control from treated samples, explaining 94% of the experimental variation (Supplementary Fig. 1) and confirming the poly I:C treatment induced consistent results

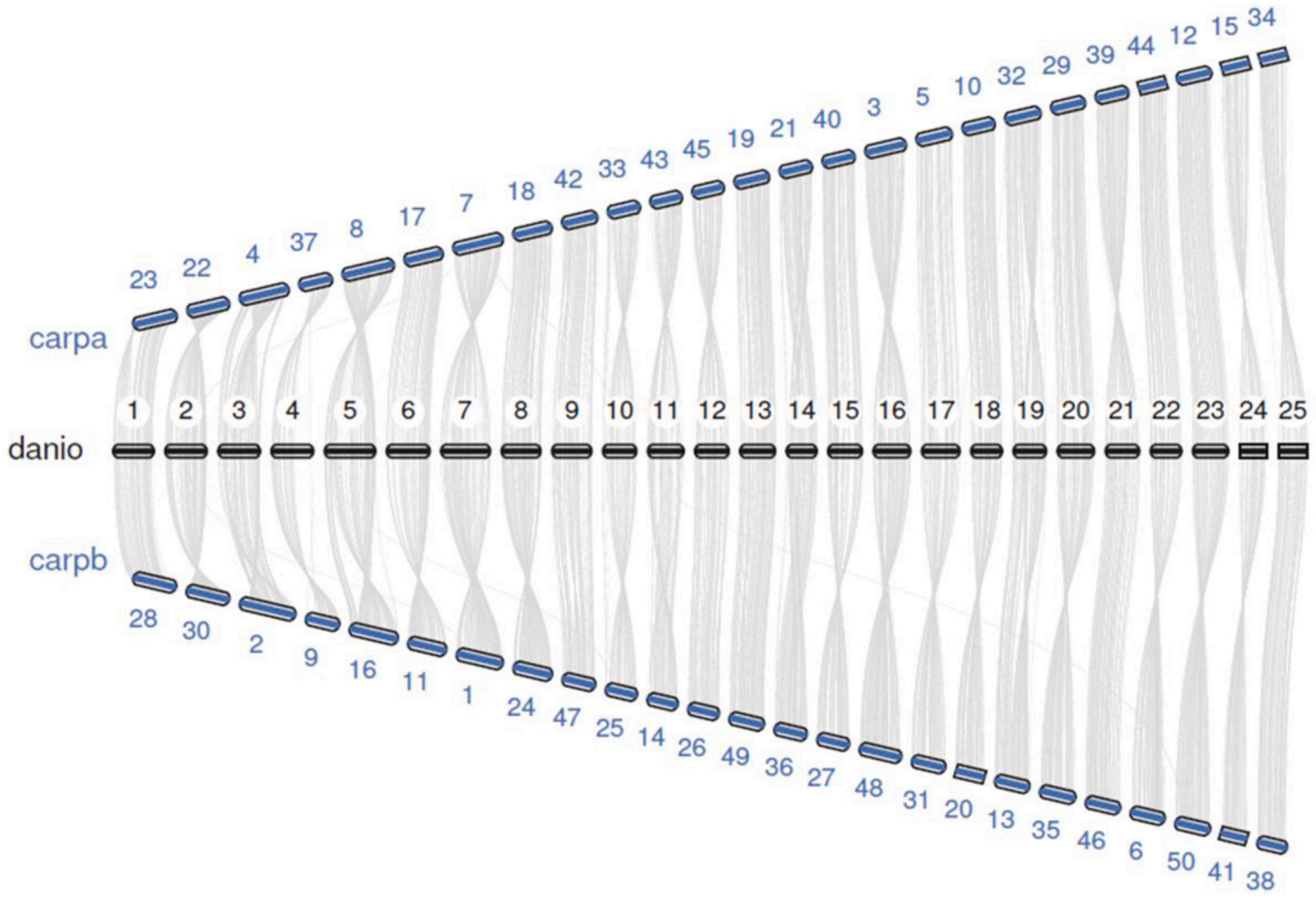

Fig. 3. Graphical representation of synteny between zebrafish (*Danio rerio*) and subgenomes of the common carp (*Cyprinus carpio*).
Zebrafish chromosomes are represented in black with title danio. Zebrafish chromosome numbers correspond with chromosome id of zebrafish genome (Danio_rerio. GRCz11). Common carp subgenome A is represented in blue with header "carpa". Common carp subgenome B is represented in blue with header "carpb". Chromosome numbers of common carp subgenome A and B correspond with linkage map. Grey lines connect 2306 synteny anchors based on protein similarity between common carp and zebrafish. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
BUSCO analysis genome carpV4 & subgenomes A and B.

	Cypcar_WagV4.0	Subgenome A	Subgenome B
Vertebrata (Odb9, $N = 2586$)			
Complete BUSCOs	98.3%	86.4%	91.4%
Complete and single-copy BUSCOs	34.1%	83.8%	88.5%
Complete and duplicated BUSCOs	64.2%	2.6%	2.9%
Fragmented BUSCOs	0.4%	2.9%	2.0%
Missing BUSCOs	1.3%	10.7%	6.6%
Actinopterygii (Odb9, $N = 4584$)			
Complete BUSCOs	97.0%	88.6%	92.0%
Complete and single-copy BUSCOs	30.8%	85.1%	88.4%
Complete and duplicated BUSCOs	66.2%	3.5%	3.6%
Fragmented BUSCOs	0.6%	1.7%	1.4%
Missing BUSCOs	2.4%	9.7%	6.6%

significantly different from the basal condition. Out of the 41,372 genes with nonzero total read count (adj. p-value <0.0.5), 8162 (19,7%) genes were significantly upregulated, 8671 (20,9%) were significantly down regulated. The remaining 53% of genes did not significantly change expression.

To address differences in expression between ohnologous gene pairs, we first composed a reference set of 3183 genes with a 1:1:1 ohnologous relationship between subgenome A, subgenome B and zebrafish. For subgenome A, 2762 (86.7%) of genes reached the minimum read count threshold set by Deseq2 and were used for further analysis. Out of the 2762 subgenome A genes, 600 (22%) and 676 (24%) were significantly upregulated and downregulated, respectively, in response to Poly I:C (adj. p-value <0.05). For subgenome B, 2820 (88.6%) genes had read counts higher than 0 and thus were used for further analysis. Out of the 2820 subgenome B genes, 618 (22%) and 719 (25%) were significantly upregulated and downregulated, respectively (adj. p-value <0.05). Thus, for both subgenomes, a similar proportion of upregulated or down regulated genes was found out of the 3183 genes. When comparing the significantly expressing ohnologs between the subgenomes, only about 50% of the ohnolog pairs showed similar expression levels of both copies. Whereas 353 and 360 pairs were significantly up regulated and down regulated, respectively, in one copy compared to the other (Fig. 6a–b).

We calculated the average log2fold change and counts for all the 3183 genes in the ohnolog set for each subgenome separately. The log2fold changes (treatment vs. control) varied between genes (Supplementary Fig. 5a). For subgenome A log2fold changes differed between genes with a standard deviation of 0.91, with an average log2fold change of −0.019 (Supplementary Fig. 6). Similarly for subgenome B the log2fold changes differed between genes with a standard deviation of 0.78, with an average log2fold change of −0.024 (Supplementary

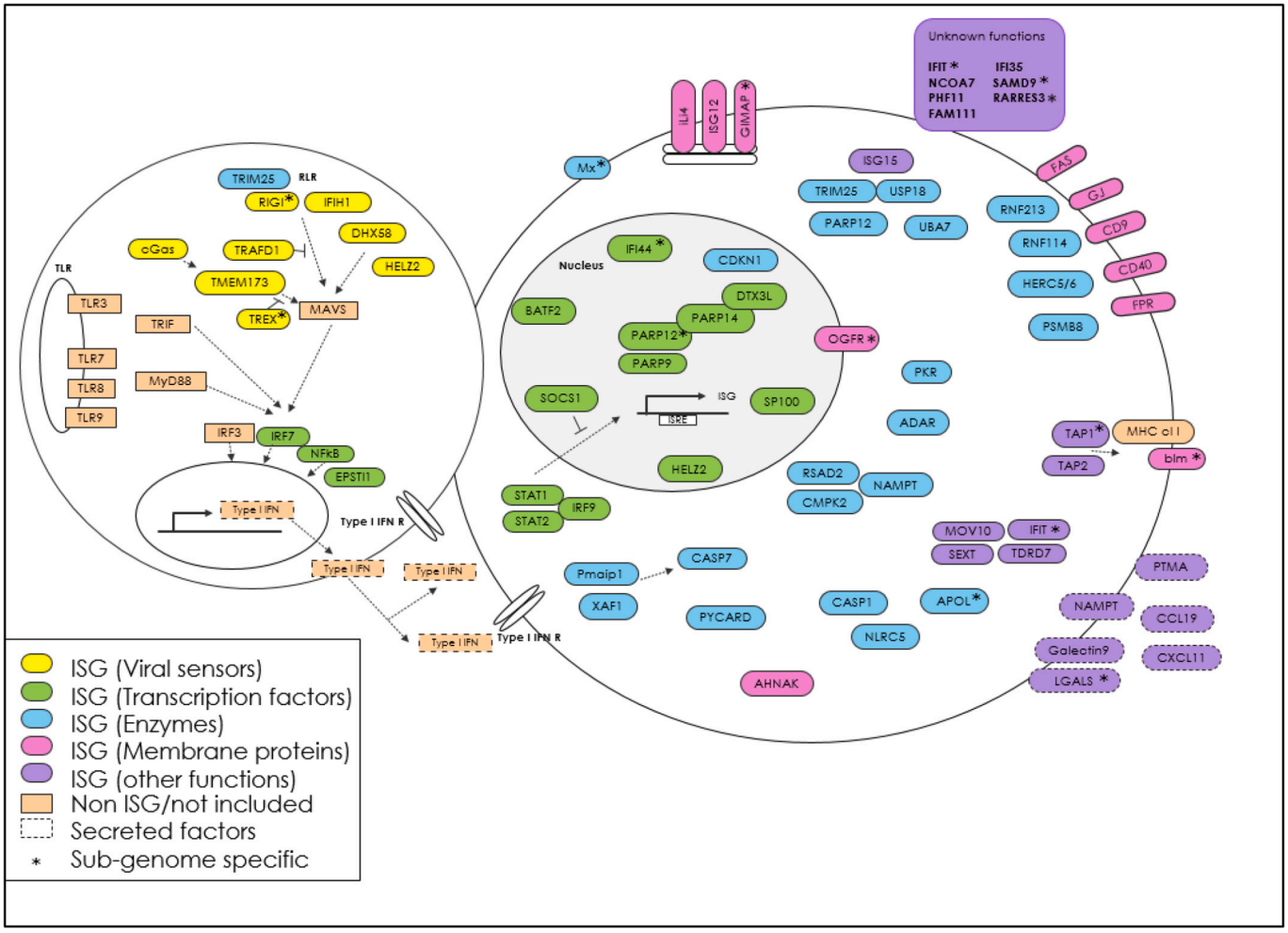

Fig. 4. Graphical Overview of Ancestral ISG ortholog groups in common carp based on Levraud et al. [30].
Graphical representation of all 72 ancestral ISG ortholog groups annotated in carpV4. A ortholog group may contain multiple ISGs. Rounded outlines with colours reflect functional groups of ISGs. Squire orange boxes indicate non ISGs, with exception of *MHC cII*, which group was not included in this study. Dotted arrows indicate positive feedback function, while bold lines with minus end stripe indicate inhibitory function. Dotted shape outlines indicate secreted factors both for ISGs and non ISGs. An asterisk (*) indicates ancestral ISG ortholog groups that had carp subgenome specific deletions and duplications. This figure is based on functional ordering of orthologous zebrafish (*Danio rerio*) ancestral ISGs in Levraud et al. [30].

Fig. 6). A high standard deviation for both subgenomes was observed, indicating high variation of inducibility between ohnologous genes pairs. Average counts and standard deviations were comparable between subgenome A and B, both in control and treatment conditions (Supplementary Fig. 5b and 7). We observed no significant subgenome dominance between subgenomes based on median counts and fold-changes (unpaired-two-samples-wilcoxon-test, p-value <0.05).

From the ISG annotated in the new assembly of the carp genome (88/103 ancestral zebrafish ISGs), 83% (73/88) were present in a 1:1 ratio, defined as ohnologs, and analysed for gene expression from both subgenomes (Supplementary Table 4, Fig. 5). First, 72 and 71 ISGs, for subgenome A and B respectively, were accepted by Deseq2 standard filtering process based on normalized read counts. For subgenome A, 61 out of 72 genes were significantly upregulated (adj.p-value <0.05). Three ISGs were significantly down regulated, and the 8 remaining ISGs did not show significant expression changes due to the treatment. For subgenome B, 62 out of 71 ISGs were significantly upregulated (adj. p-value <0.05). Five ISGs were significantly down regulated and three ISGs that showed no significant expression change due to the treatment. When comparing the expressed ISGs of both subgenomes, 61 out of 73 showed similar expression results in both subgenomes; 56 sig. up regulated and two sig. down regulated, and three ISGs that showed non-significant expression in both subgenomes. (Fig. 6d-e).

As could be expected from a broad network of genes such as ISGs with different functions (e.g. membrane proteins, transcription factors, enzymes; see Fig. 4), inducibility (i.e. fold changes) differed greatly between different ISGs. Genes could be induced >15 fold, or could be negatively regulated at −5 fold, in line with different functions within the interferon network. We did not observe a functional division between the subgenomes based on ISG functions, For example induced expression of transcription factors would not be consistently highest for subgenome B ohnologs. Most importantly, the viral mimic Poly I:C led to the induction of ISG gene expression in both subgenomes, without evidence of subgenome dominance (Fig. 5a). When looking at absolute counts, we did not observe subgenome dominance for basal gene expression of ISGs nor for the counts observed in the Poly I:C condition (Fig. 5b).

For example, the gene for radical SAM domain-containing 2; *rsad2*, also known as viperin, was only significantly expressed in subgenome B, while the subgenome A ohnolog showed significant higher counts in basal conditions (34,949.03 compared to 19,278.42). Also, the genes for signal transducer and activator of transcription *stat1a* and for interferon regulatory factor *irf7*, two transcription factors, had completely different subgenome expression, with *stat1a* having significantly lower basal expression in subgenome A compared to B, while *irf7* showed comparable basal expression in both ohnologs. *Stat1a* was significantly induced

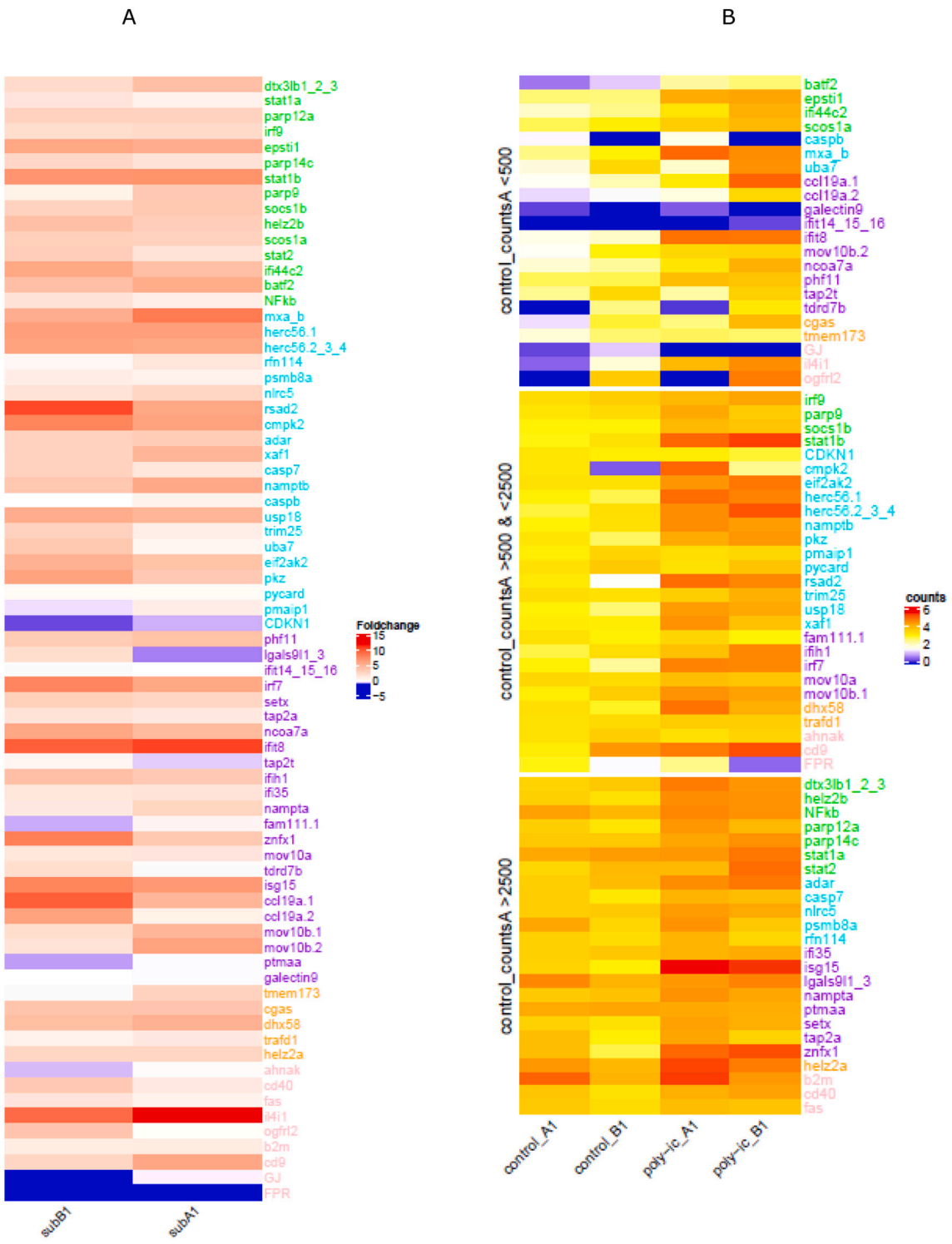

Fig. 5. Fold change & normalized counts of induced ISGs.
5a. Shows the log2foldchange (PolyIC exposed samples compared to control) for subgenome A and B ohnologs that were present in 1:1 on both subgenomes. 5b. Shows the log10 normalized counts of subgenome A and B ohnologs in control and PolyIC conditions. Three groups were made based on counts in control conditions of subgenome A; subgenome A genes <500 read counts, subgenome A genes with read counts between 500 and 2500 and subgenome A genes with > 2500 read counts. Gene name colours correspond to the functional groups of Fig. 4.

and expressed in both subgenomes (adj. *p*-value 0.05), while *irf7* only showed significant upregulation in subgenome A (adj. p-value <0.05). In contrast, the gene for interferon induced protein 44, *ifi44c2*, had significantly higher counts for basal and induced expression in subgenome B but only significantly expressed in subgenome A.

These observations showed no indication of subgenome dominance, but rather showed the gene expression patterns of the carp ISG ohnologs appeared highly variable, and randomly distributed over the two subgenomes. Statistical testing for consistent subgenome differences in counts for carp ISGs were insignificant (unpaired-two-samples-wilcoxon-test, p-value <0.05), in both basal and poly I:C conditions (Supplementary Fig. 4). When taking all foldchanges of ISG ohnologs

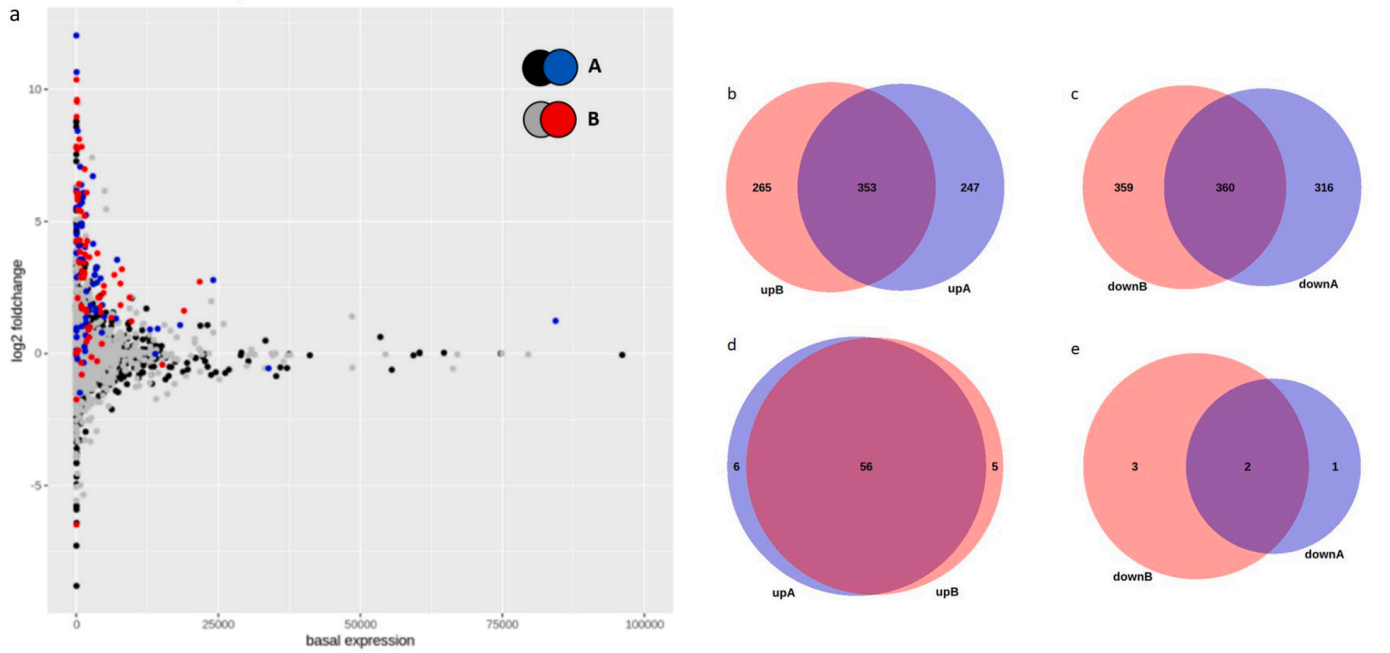

Fig. 6. a. Log2 foldchange compared with basal expression of subgenome 1:1.
Ohnologous genes represented as dots within in the Figure. Ancestral ISGs represented in Blue and Red. The set of ~3000 selected Orthologous genes represented in Black and Grey. All included genes are present in a 1:1 ratio between subgenome A and B. Basal expression was the average of read numbers mapped onto a given gene in control (pbs injected) head kidney leukocytes. Log2 Foldchange is the ratio of reads numbers in PolyIC exposed head kidney leukocytes divided by basal expression.
b–e Significantly expressed ohnolog of viral mimic experiments for both subgenomes. Fig. 6b shows significantly upregulated genes of the 3183 ohnolog gene set (*P*-value <0.05). 6c shows significantly down regulated genes of the 3183 ohnolog gene set. 6d shows significantly upregulated genes of the ISG gene set (P-value <0.05). 6e shows significantly down regulated genes of the ISG set. Overlap shows gene pairs that were significantly expressed in both subgenomes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

together we did not observe significant differences between subgenome A and B (unpaired-two-samples-wilcoxon-test, p-value <0.05).

When comparing the basal response to the log2fold change for both the ISGs and the ohnolog gene set we observed that the ISGs ohnologs are on average higher induced than the ~3000 random non ISG ohnolog gene set within this experiment (Fig. 6). The non-ISG ohnolog set showed a higher subgenome expression variation, with fewer gene pairs showing similar significant upregulated gene expression patterns compared to the ISG ohnolog pairs (Fig. 6 b-e). In the Poly I:C challenged fish the ohnolog gene set showed a lower average count compared to the basal expression observed in the ISGs (Supplementary Fig. 5b and 7). This further demonstrates the consistent function of the ISGs where over 80% of the ohnolog pairs showed similar significant expression, while for a random selected ohnolog set we observed that ~50% of pairs only the copy on one subgenome showing significant expression. Fig. 6 further demonstrates that the two subgenomes overlap in their expression by showing clear overlap of ISGs (red and blue dots) belonging to the two subgenomes, as well as the non ISG ohnolog set genes showing no clear separate clustering of subgenomes (grey and black dots). Though the subgenomes may differ in ohnolog basal expression and inducibility, when taking the average expression we do not observe consistent subgenome dominance in both gene sets.

BUSCO analysis with both vertebrata and Actinopterygii gene datasets as input. BUSCO completeness percentages were used as an indication of genome and subgenome completeness. BUSCO analysis of whole genome included all scaffolds. For subgenome A and B, 25 chromosomes respectively were used as input for BUSCO analysis.

3. Discussion

The allotetraploid common carp is a relatively recent (12 MYA) hybrid between two related cyprinid species. Assuming the absence of recombination between homologous chromosomes of the two donor species, it has two separate subgenomes, named A and B. Thus, the two-thirds of the duplicated genes that have been retained since the duplications [12], are expected to be ohnologs. The overarching question addressed here is whether we can observe subgenome functional dominance: does one subgenome retain a higher number of genes, are genes expressed more in one subgenome, or are genes of one subgenome systematically induced more upon stimulus. This question was addressed by studying a conserved set of ISG genes that are known for their common trait of Interferon-type 1 inducibility during an antiviral response. Due to this consistent shared trait we could directly relate induction of ISGs to subgenome gene expression and function.

To this end, a common carp genome was assembled to chromosome-level with a specific R3xR8 genetic history[31,32]. The Nanopore long read sequencing produced a primary assembly with lengths contained half to sometimes complete chromosomes (Supplementary Table 1). Combining our primary assembly with HiC scaffolding resulted in a chromosome level final assembly (Fig. 1). The carp used for the assembly is genetically similar to the carp used for the immunological stimulus experiment. Having a close genetic relationship between reference individual and RNA-seq data is expected to limit ambiguity in mapping of RNA-seq data to the two subgenomes.

To get a first impression of the completeness of our genome a BUSCO gene analyses was done for the new genome as well as the identified subgenomes separately. The BUSCO analysis of our new genome indicated the presence of 63% duplicates, which mostly reflects the duplication present due to the two subgenomes (Table 1). Similar patterns of gene duplication have been observed in the closely related goldfish, where 58% of BUSCO genes were found in duplicate [33]. Previous studies hinted at a degree of dominance in the allotetraploid carp genome, with one subgenome (A) showing lower gene expression and lower gene completeness [22,12]. We also observed more complete

BUSCOs in subgenome B compared to A, suggesting that, in the process of rediploidization, subgenome A, more than B, would be on a path to lower functionality. However, BUSCOs are genes expected to be highly conserved and thus less likely to experience gene loss. Neither of the two subgenomes had a complete BUSCO gene set, indicating that the common carp would need both subgenomes for complete genome function.

Next we looked for possible subgenome-specific retainment of genes in well studied molecular network (Fig. 4). Similar to the findings for BUSCO genes, none of two subgenomes had a complete ISG core gene set, again indicating that the common carp would need both subgenomes for a complete ISG repertoire. However, subgenome specific gene loss was small, with 86% 1:1 ohnologous gene pairs remaining between the subgenomes. The 86% percentage of conserved gene duplication between the subgenomes could suggest a higher degree of retained ohnologs in specific gene networks when compared to the ~60% duplication observed in the BUSCO gene sets.

In contrast with gene number asymmetry observed in the closely related allotetraploid goldfish (*Carassius auratus*) and the allotetraploid frog (*Xenopus laevis*), where pseudogenes evolved asymmetrically between the subgenomes [11,34], our detailed core ISG annotation provided no compelling evidence for asymmetrical gene loss within a functional network of carp ohnologs. Our finding is in line with [11], who observed symmetric accumulation of pseudo genes of the subgenomes of common carp.

The retention of ISGs in both subgenomes can partly be explained by the WGD event having occurred recently (12MY). It is likely that over time more duplicates will be lost. For example, after WGD events certain genes have a higher chance to be maintained as duplicate due to the evolutionary advantage of maintaining a stoichiometric balance [35,36] or can even be maintained based for the higher expression needs after whole genome duplication alone. However, explanations of retentions of duplicates through a single process are unlikely to be complete, as multiple processes can influence a gene towards altered expression and new functions [37]. This was also observed for the autotetraploid Atlantic salmon, where the gene balance hypothesis alone could not explain the retention of certain gene copies [10].

To study the asymmetry of subgenome gene expression, we first investigated a larger set of ~3000 ohnologous genes that have a 1:1:1 relationship between subgenome A, subgenome B, and zebrafish. Comparable with the ISG ohnologs, this reference set of ohnologs was spread across the whole genome of carp, with genes being present on almost every chromosome. By calculating the average expression of the ~3000 ohnologous genes on both subgenomes, we did not observe significant subgenome dominant expression. Interestingly we did observe expression asymmetry between the subgenomes when directly comparing ohnolog pairs. About 40% of the ohnologs were significantly differentially expressed upon exposure to Poly I:C (Fig. 6b-c). We observed that of these significantly differentially expressed genes only about half of the ohnolog pairs were expressed similarly between both subgenomes (Fig. 6b-c). This could indicate divergence in gene regulation and function between subgenome A and B.

In contrast with our reference set of ohnologs, expression of practically all ISG ohnologs was evident, both under control conditions and under conditions where they were induced by exposure to Poly I:C. Almost all ISG ohnologs were induced by the viral mimic, but with high variation in activation between the different genes (Fig. 5). The high variation can be explained by the different functions these ISGs perform in the gene network (Fig. 4). As expected, the ISGs show higher induction compared to our reference ohnolog set (Fig. 6a), as our experimental stimulus was selected for these ISGs specifically. In contrast to our reference set, >75% of our ISGs ohnolog pairs showed similar significant expression patterns between the pairs (Fig. 6d-e).

The significant induction in the same direction of the majority of ISG pairs demonstrates both subgenomes are responding to the immunological stimulus in a similar manner. Expression patterns between ohnologs differed in basal expression and inducibility, but showed no clear evidence of subgenome dominance, here defined as a systematic expression difference between two ohnologs depending on subgenome. The pattern that ISG ohnolog expression was randomly higher in either subgenome, was mirrored by a comparable pattern for the observed expression of the reference ohnolog set (Supplementary Fig. 5–7). In contrast to the higher expression of subgenome B that has been reported [22,12], our data did not provide evidence for consistent higher gene expression of subgenome B, neither for the reference ohnolog set nor for the ISGs. Therefore we found no evidence for subgenome dominance in common carp.

In addition the induction patterns of ohnologs can be conserved over time. For example, the duplicated genes *stat1a* and *stat1b* are known to differ in inducibility in zebrafish [30]. Both gene copies are present in duplicate in common carp resulting a 2:4 ratio of *stat1* between zebrafish and common carp. Comparable to zebrafish, *stat1b* expression and inducibility is higher than *stat1a* in both subgenomes of carp, conserving expression patterns over evolutionary time.

Though we do not observe significant expression dominance on average, we do observe different expression patterns within ohnologs pairs. Even under situations where genes, e.g. annotated to subgenome A, had low constituent expression, they sometimes proved even more inducible than their ohnolog counterpart. As a result, a gene with low constituent expression can contribute equal or more than its ohnolog post-stimulus. These differences in inducibility between ohnolog pairs could also point to subfunctionalization. In this way both gene copies could maintain an important role in a gene network, by one expressing more in basal conditions while the other showing increased expression after viral infection. That the higher inducible gene copy is not consistently from the same subgenome may indicate that both subgenomes are on a similar rediploidisation path. Copied genes may disappear randomly from each subgenome without substantially increasing the dominance of either.

In summary, by studying common carp ohnologs present in a 1:1 ratio between the subgenomes we did not observe subgenome dominant expression. Both subgenomes were near equal in their average basal expression and average induction. Specifically, with the ISGs we could directly compare subgenome functionality and gene expression in a network of genes with known functions and phylogenetic history. Here, no subgenome dominance was observed in the common carp either. Rather, the highest expressing and inducible gene of an ohnolog pair may be derived from either subgenome. The allotetraploid common carp is therefore an example of a recently duplicated vertebrate genome consisting of balanced subgenomes.

4. Materials & methods

4.1. Primary genome sequencing and assembly

To minimize genetic divergence between the carp (an R3xR8 cross) used for the functional study and the reference used for mapping RNA-seq data, first a new chromosome-level assembly was made of a closely related individual. Genomic DNA was isolated from red blood cells of a female double haploid gynogenetic common carp line as in Henkel et al. [20,21]. The original genetic background of this individual was a cross of two pure lines or Polish (R3) and Hungarian (R8) origin [31,32], as previously described in Henkel et al. [20,21].

To maximize the length of unique sequence reads and thus optimize subgenome assembly, a combination of three next generation sequencing methods was used.

Isolated genomic DNA was used to prepare libraries using the SQK-108 and SQK-109 Ligation Sequencing kits from Oxford Nanopore Technologies. Libraries were sequenced using R9.4.1-type Minion and Promethion flow cells from Oxford Nanopore Technologies Raw reads statistics were summarized with NanoStat (v.1.6.0) [38] (Supplementary Table 1). The two obtained libraries of Nanopore long reads were base called with Guppy (v.3.2.2; Oxford Nanopore Technologies) and

then filtered for reads >10 kb (Supplementary Table 1). Filtered reads were assembled using Flye (v.2.4.2) [39] at default settings. For polishing remaining sequence errors in the primary assembly, DNA extracted from a blood sample from the same homozygous gynogenetic carp line was used to prepare an Illumina library using the Nextera DNA Flex Library Prep Kit according to the manufacturer's instructions (Illumina Inc. San Diego, CA, USA). The genomic paired-end (PE) library was sequenced with a read length of 2 × 150 nt using the Illumina NovaSeq6000 system. The consensus tool from wtdbg2 [40] was applied for polishing with default settings.

4.1.1. Chromosome-level scaffolding and linkage group assignment

HiC-scaffolding was done at Dovetail® and combined with the primary genome assembly. In short, a Dovetail HiC library was prepared from an 8 mL blood sample of a female common carp individual of R3 x R8 origin (each are pure carp lines; R3 and R8, inbred for 12 generations by brother sister mating) [20,21], according to the methods of Lieberman-Aiden et al. [41]. The library was sequenced on a Illumina HiSeqX to produce 609 million 2 × 150 bp paired-end reads, which provided 501.48× physical coverage of the genome (10–10,000 kb). Prior to final assembly of the common carp genome, HiC data was used to identify contigs erroneously fused in the primary Nanopore-based assembly. The generated HiC contact maps were used to check for large assembly inconsistencies. After breaking up contig errors of the primary assembly, the corrected primary assembly was subjected to a final round of HiC scaffolding using Dovetail® Omni-C® scaffolding technology and HiRise® pipeline. Scaffolds <1 kb were removed from the common carp genome assembly and the 50 largest scaffolds were named according to the linkage groups described in Tadmor-Levi et al. [42]. Chromosome pairs were identified by linkage group marker positioning and confirmed by MCscan synteny mapping [43] between common carp chromosomes and the zebrafish (*Danio rerio*) reference genome (Danio_rerio.GRCz11) (Supplementary Table 2).

4.2. Subgenome identification

The 50 chromosomes were mapped against the genome of the diploid species Tapien (*Poropuntius huangchuchieni*) [12] in sister pairs, using Mummer [44]. Tapien is a species in the Barbinae subfamily of fish included in the family Cyprinidae. Tapien is assumed to have a common ancestor to one of diploid ancestors involved in the inbreeding event that resulted in the carp specific WGD [12]. Two subgenomes were identified by comparing the Mummer mapping percentage obtained from mapping the 50 chromosomes to Tapien (Supplementary Table 3). Sister chromosomes that had a higher mapping percentage with Tapien were placed into subgenome B; the other sister chromosomes were placed into subgenome A. The two identified subgenomes were subsequently compared to Li et al. [22], who used another diploid species of the family Barbinae (*P. tetrazona*) to divide the subgenomes of common carp. The comparison was made to assess if our subgenome division was consistent with other published subgenome divisions.

4.3. Genome annotation

To assess genome completeness, two BUSCO analyses were performed, applying the vertebrate ortholog set (vertebrata_odb9) and the Actinopterygii ortholog set (Actinopterygii_odb9), using zebrafish (*Danio rerio*) as Augustus species [45]. The BUSCO analysis was done for both, the complete assembly (Cypcar_WagV4.0), as well as for each subgenome separately.

Cypcar_WagV4.0 was submitted to NCBI [46] and annotated by Ensembl with the Ensembl pipeline [47]. The ortholog sets and the zebrafish genome were used to visualize macro synteny blocks between the subgenomes of Cypcar_WagV4.0 and zebrafish. By use of the MCscan package [48], putative homologous chromosomal regions were detected between zebrafish and the two subgenomes of common carp using vertebrate ortholog genes as anchors.

4.4. ISG annotation

An established core gene set was selected for ISG annotation in Cypcar_WagV4.0 based on two criteria. Firstly, all ISGs in our core set are Interferon type I inducible. In this way we could use gene inducibility as a direct measure of gene function [30]. Lastly, this gene set contained only ISGs with an established orthologous relationship between zebrafish and human(*Homo sapiens*) [30]. Due to this established orthology we could infer the phylogenetic history of all our included ISGs. With these two conditions met, we used 103 ancestral ISGs found in zebrafish belonging to 72 ortholog groups as the starting point. A comprehensive gene list can be found in Supplementary Table 4. Zebrafish ISG protein sequences were obtained from Ensembl and aligned to Cypcar_WagV4.0 annotated gene protein sequences using Diamond [49]. Diamond results were manually checked for best matches with carp ISG candidates by comparing identity scores. Identity scores 75% or higher were considered to be ISG candidate genes.

In addition, chromosome locations of carp ISG candidates were compared to zebrafish ISG chromosome locations for conserved synteny. Sets of 10 neighbouring genes upstream and downstream of the zebrafish ISGs were selected as micro synteny anchors using the Ensembl genome browser. The protein sequences of neighbouring genes were blasted against the protein sequences of Cypcar_WagV4.0 using Diamond. Best common carp gene matches, based on the protein blasting, were manually checked for their genome location. If a genome location of a neighbour gene matched with the zebrafish equivalent, it was marked as a synteny anchor. The established synteny anchors were used as confirmation of micro synteny blocks between zebrafish and common carp. Micro synteny blocks were successfully identified for all ISG carp candidates. Carp ISG candidates where further manually validated from RNA-seq visualization (Jbrowse [50]).

Next, manually selected carp ISG candidates were aligned back against all Ensembl protein sequences of zebrafish using Diamond. Carp ISG candidates were selected only if their alignment against the protein zebrafish database provided a clear hit with the orthologous ISGs of zebrafish. Zebrafish ISGs without a candidate in carp based on protein-to-protein alignments were analysed further by aligning to the common carp genome Cypcar_WagV4.0 using Diamond. These protein-to-nucleotide alignment results pointed towards areas of interest in the Cypcar_WagV4.0 genome and were manually checked for missing genes by using the micro synteny approaches as earlier described. By this method we could identify possible ISG pseudogenes in common carp.

4.5. Protein domain analysis

For all Cypcar_WagV4.0 ISG candidates, protein domains were predicted using InterProscan (v.5.50–84.0)[51]. Predicted protein domains were compared manually with the protein domains of the original ISG zebrafish gene set. This comparison was used to identify if the annotated carp ISGs had the same protein domains as their zebrafish counterparts.

4.6. Final annotation

Our manual annotation was compared with the current Ensembl annotation release (version 108) of our carp genome Cypcar_WagV4.0. Over 80% of our manually annotated ISGs matched Ensembl predictions completely. However, automated annotations may be error-prone. Annotation errors could only be solved by manually correcting Ensembl annotation files, resulting in some ISGs to not have an Ensemble id (Supplementary Table 4).

4.7. Additional ohnolog set

To compare ISGs 1:1 ohnologous expression to other common carp

ohnologs, an additional gene set was included in the RNA-seq expression analysis. This additional set of genes was used as a neutral reference set to compare basal gene expression patterns observed in the ISG set. The genes of the additional set were selected based on an ohnolog prediction between zebrafish and common carp using ProteinOrtho6 [52] with default settings. Subgenome A and B were compared separately to the zebrafish genome for ohnolog prediction at the same time with ProteinOrtho. The predicted ohnologs were filtered down to a set containing only genes with a 1:1:1 relationship between subgenome A, subgenome B and zebrafish. The gene set was further filtered to contain only ohnologs that were present on the same corresponding sister chromosome pairs for subgenome A and B. Finally, 3183 ohnologous neighbouring genes were selected with a 1:1:1 relationship between subgenome A and subgenome B and zebrafish.

4.8. Transcriptomic analysis of ISGs and additional ohnolog gene set

4.8.1. Animals

European common carp (*Cyprinus carpio carpio*) of the R3xR8 strain were used, a cross between the Polish R3 strain and Hungarian R8 strain [31,32]. Carp were bred and raised in the aquatic research facility of Wageningen university at 21–23 °C in recirculating water. Fish were fed twice daily with pelleted dry food (Skretting, Nutreco). All experiments were performed with the approval of the animal experiments committee of Wageningen University DEC: 2019.W-0048.009.

4.8.2. Head kidney leukocyte isolation

Carp (12-month-old) were starved for 16–18 h and euthanized with 0.3 g/l tricaine methane sulfonate (TMS) (Crescent Research Chemicals, Phoenix, USA) in aquarium water buffered with 0.6 g/l sodium bicarbonate and bled via the caudal vein. The head kidney was isolated and total head kidney leukocytes were separated on a 51% Percoll density gradient (GE Healthcare, Thermo Fisher Scientific) as previously described[53].

4.8.3. In vitro stimulation of head kidney leukocytes

Head kidney leukocytes (2×10^6 cells) were seeded in a 6-wells plate (Corning) in DMEM/l15 supplemented with penicillin G (100 U/ml), streptomycin sulfate (100 μg/ml, Gibco) and FBS (2%, Gibco). The head kidney leukocytes were stimulated with 50 μg/ml Poly I:C (Sigma), or with culture medium (as negative control) for 24 h at 27 °C in the presence of 5% CO^2.

4.8.4. RNA isolation

After 24 h of stimulation, five replicate wells were pooled and 1×10^7 head kidney leukocytes were lysed in 700 μL RLT buffer (Qiagen) and stored at −80 °C until RNA isolation. Total RNA was isolated using the RNeasy Mini kit, including on-column DNase treatment (according to manufacturer's protocol). Isolated RNA was send for sequencing to Novogene Co (Beijing, China).

4.8.5. RNA sequencing and expression analysis of ISGs

RNA-seq reads obtained by Novogene Co (Beijing, China) were mapped to Cypcar_WagV4.0 with hisat2 [54]. RNA-seq transcripts were quantified with StringTie (v.2.1.5)[55] into a read count matrix. We analysed the assembled RNA transcripts with Bioconductor (v.3.10.1), R (v.3.6.3) and package Deseq2 (v.1.26.0) [56]. Deseq2 calculated foldchanges were shrunken towards zero by use of the apeglm method to correct for high variation of counts for each gene separately [57]. RNA expression was visualized by use of package Complex Heatmap (v 2.2.0) [58] and ggplot2 (v 3.3.5)[59]. Annotated ISGs that occurred as singletons on only one subgenome or occurred as multiple paralogs (Supplementary Table 4), were not included in the final analysis because of the expression bias they would create when comparing subgenomes 1:1. Details on expression data can be found in Supplementary Figs. 3–4. Instead, only ISGs that occurred in both subgenomes were used for further expression analysis. To compare overall expression levels between the subgenomes A and B, the difference between the average normalized read count of these 1:1 ohnologs was tested for significance using an unpaired-two-samples-wilcoxon-test, p-value <0.05. The same approach was used to test for a significant difference in the average foldchange of these ISG ohnologs between subgenome A and B. The RNA expression of the additional gene set of 3183 ohnologous genes was analysed and visualized using the same methods described above (Supplementary Figs. 5–7).

CRediT authorship contribution statement

A. Blasweiler: Writing – original draft, Conceptualization, Formal analysis, Methodology, Visualization, Project administration. **H.-J. Megens:** Supervision, Writing – review & editing, Conceptualization, Formal analysis, Methodology, Project administration. **M.R.G. Goldman:** Investigation, Writing – review & editing. **R. Tadmor-Levi:** Investigation. **J. Lighten:** Writing – review & editing, Funding acquisition. **M.A.M. Groenen:** Supervision, Writing – review & editing, Funding acquisition. **R.P. Dirks:** Investigation, Writing – review & editing. **H.J. Jansen:** Investigation, Writing – review & editing. **H.P. Spaink:** Investigation, Writing – review & editing. **L. David:** Writing-review & editing, Methodology. **P. Boudinot:** Writing – review & editing, Conceptualization, Methodology. **G.F. Wiegertjes:** Supervision, Writing – review & editing, Conceptualization, Resources, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

raw data of our experiments has been made available trough the FAANG data portal: https://data.faang.org/. Our genome assembly is publicly available under ID: Cypcar_WagV4.0 (GCA_905221575.1)

Acknowledgements

The authors would like to thank Kaylee van Dijk, Maria Forlenza, Cornelius Gunter and Jules Petit for their technical support of the in vitro experiment. Martijn Derks, Henri van Kruistum and Langqing Liu provided advice on genome assembly and synteny analysis. This research has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. (817923) (AquaFAANG), and from a Future Leader Fellowship awarded to Dr. J Lighten by the Biotechnology and Biological Sciences Research Council (BB/R010870/1). Dr. R. Tadmor-Levi was funded by the Israel Science Foundation Grant No. (981/22).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2023.110723.

References

[1] Braasch Ingo, A.R. Gehrke, J.J. Smith, K. Kawasaki, T. Manousaki, J. Pasquier, A. Amores, T. Desvignes, P. Batzel, J. Catchen, et al., The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons, Nat. Genet. 48 (4) (2016) 427–437, https://doi.org/10.1038/ng.3526.

[2] A. Christoffels, E.G.L. Koh, J.M. Chia, S. Brenner, S. Aparicio, B. Venkatesh, Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes, Mol. Biol. Evol. 21 (6) (2004) 1146–1151, https://doi.org/10.1093/molbev/msh114.

[3] P.S. Soltis, Ancient and recent polyploidy in angiosperms, New Phytol. (2005) 5–8.

[4] S.R. Voss, D.K. Kump, S. Putta, N. Pauly, A. Reynolds, R.J. Henry, S. Basa, J. A. Walker, J.J. Smith, Origin of amphibian and avian chromosomes by fission, fusion, and retention of ancestral chromosomes, Genome Res. 21 (8) (2011) 1306–1312, https://doi.org/10.1101/gr.116491.110.

[5] J.-M. Aury, O. Jaillon, L. Duret, B. Noel, C. Jubin, B.M. Porcel, B. Ségurens, P. Wincker, Global trends of whole-genome duplications revealed by the ciliate Paramecium tetraurelia, Nature. 444 (2006) 171–178.

[6] J. Pasquier, C. Cabau, T. Nguyen, E. Jouanno, D. Severac, I. Braasch, L. Journot, P. Pontarotti, C. Klopp, J.H. Postlethwait, et al., Gene evolution and gene expression after whole genome duplication in fish: the PhyloFish database, BMC Genomics 17 (1) (2016) 1–10, https://doi.org/10.1186/s12864-016-2709-z.

[7] Y. Van de Peer, S. Maere, A. Meyer, The evolutionary significance of ancient genome duplications, Nat. Rev. Genet. 10 (10) (2009) 725–732, https://doi.org/10.1038/nrg2600. http://www.ncbi.nlm.nih.gov/pubmed/19652647.

[8] Y. Van De Peer, E. Mizrachi, K. Marchal, The evolutionary significance of polyploidy, Nat. Rev. Genet. 18 (7) (2017) 411–424, https://doi.org/10.1038/nrg.2017.26.

[9] C. Berthelot, F. Brunet, D. Chalopin, A. Juanchich, M. Bernard, B. Noël, P. Bento, C. Da Silva, K. Labadie, A. Alberti, et al., The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates, Nat. Commun. 5 (2014), https://doi.org/10.1038/ncomms4657.

[10] S. Lien, B.F. Koop, S.R. Sandve, J.R. Miller, M.P. Kent, T. Nome, T.R. Hvidsten, J. S. Leong, D.R. Minkley, A. Zimin, et al., The Atlantic salmon genome provides insights into rediploidization, Nature. 533 (7602) (2016) 200–205, https://doi.org/10.1038/nature17164.

[11] J. Luo, J. Chai, J. Chai, Y. Wen, Y. Wen, Y. Wen, M. Tao, G. Lin, X. Liu, L. Ren, et al., From asymmetrical to balanced genomic diversification during rediploidization: subgenomic evolution in allotetraploid fish, Sci. Adv. 6 (22) (2020), https://doi.org/10.1126/sciadv.aaz7677.

[12] P. Xu, J. Xu, G. Liu, L. Chen, Z. Zhou, W. Peng, Y. Jiang, Z. Zhao, Z. Jia, Y. Sun, et al., The allotetraploid origin and asymmetrical genome evolution of the common carp Cyprinus carpio, Nat. Commun. 10 (1) (2019) 1–11, https://doi.org/10.1038/s41467-019-12644-1.

[13] T. Makino, A. McLysaght, Ohnologs in the human genome are dosage balanced and frequently associated with disease, Proc. Natl. Acad. Sci. U. S. A. 107 (20) (2010) 9270–9274, https://doi.org/10.1073/pnas.0914697107.

[14] S. Ohno, Evolution by Gene Duplication. Berlin, New York, 1970.

[15] J. Inoue, Y. Sato, R. Sinclair, K. Tsukamoto, M. Nishida, Rapid genome reshaping by multiple-gene loss after whole-genome duplication in teleost fish suggested by mathematical modeling, Proc. Natl. Acad. Sci. 112 (48) (2015) 14918–14923, https://doi.org/10.1073/pnas.1507669112.

[16] M. Kasahara, K. Naruse, S. Sasaki, Y. Nakatani, W. Qu, B. Ahsan, T. Yamada, Y. Nagayasu, K. Doi, Y. Kasai, et al., The medaka draft genome and insights into vertebrate genome evolution, Nature. 447 (7145) (2007) 714–719, https://doi.org/10.1038/nature05846.

[17] I. Braasch, A.R. Gehrke, J.J. Smith, K. Kazuhiko, J.H. Postlethwait, The spotted gar genome illuminates vertebrate evolution and facilitates human-to-teleost comparisons, Nat. Genet. 48 (4) (2016) 427–437, https://doi.org/10.1038/ng.3526.

[18] O. Simakov, F. Marlétaz, J.X. Yue, B. O'Connell, J. Jenkins, A. Brandt, R. Calef, C. H. Tung, T.K. Huang, J. Schmutz, et al., Deeply conserved synteny resolves early events in vertebrate evolution, Nat. Ecol. Evol. 4 (6) (2020) 820–830, https://doi.org/10.1038/s41559-020-1156-z.

[19] L. David, S. Blum, M.W. Feldman, U. Lavi, J. Hillel, Recent duplication of the common carp (Cyprinus carpio L.) genome as revealed by analyses of microsatellite loci, Mol. Biol. Evol. 20 (9) (2003) 1425–1434, https://doi.org/10.1093/molbev/msg173.

[20] C.V. Henkel, R.P. Dirks, H.J. Jansen, M. Forlenza, G.F. Wiegertjes, K. Howe, G.E.E. J.M. van den Thillart, H.P. Spaink, Comparison of the exomes of common carp (*Cyprinus carpio*) and zebrafish (*Danio rerio*), Zebrafish. 9 (2) (2012) 59–67, https://doi.org/10.1089/zeb.2012.0773.

[21] C.V. Henkel, R.P. Dirks, H.J. Jansen, M. Forlenza, G.F. Wiegertjes, K. Howe, G.E.E. J.M. Van Den Thillart, H.P. Spaink, Comparison of the exomes of common carp (Cyprinus carpio) and zebrafish (Danio rerio), Zebrafish. 9 (2) (2012) 59–67, https://doi.org/10.1089/zeb.2012.0773.

[22] J.T. Li, Q. Wang, Huang Yang M. Di, Q.S. Li, M.S. Cui, Z.J. Dong, H.W. Wang, J. H. Yu, Y.J. Zhao, C.R. Yang, et al., Parallel subgenome structure and divergent expression evolution of Allo-tetraploid common carp and goldfish, Nat. Genet. 53 (10) (2021) 1493–1503, https://doi.org/10.1038/s41588-021-00933-9.

[23] S. Ohno, J. Muramoto, L. Christian, N.B. Atkin, Diploid-tetraploid relationship among old-world members of the fish family Cyprinidae, Chromosoma. 23 (1) (1967) 1–9, https://doi.org/10.1007/BF00293307.

[24] P. Xu, X. Zhang, X. Wang, J. Li, G. Liu, Y. Kuang, J. Xu, X. Zheng, L. Ren, G. Wang, et al., Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*, Nat. Genet. 46 (11) (2014) 1212–1219, https://doi.org/10.1038/ng.3098.

[25] J. Petit, L. David, R. Dirks, G.F. Wiegertjes, Genomic and transcriptomic approaches to study immunology in cyprinids: what is next? Dev. Comp. Immunol. 75 (2017) 48–62, https://doi.org/10.1016/j.dci.2017.02.022.

[26] J. Szczygieł, T. Kamińska-Gibas, J. Petit, P. Jurecka, G. Wiegertjes, I. Irnazarow, Re-evaluation of common carp (Cyprinus carpio L.) housekeeping genes for gene expression studies - considering duplicated genes, Fish Shellfish Immunol. aug (115):58–69 (2021), https://doi.org/10.1016/j.fsi.2021.05.013.

[27] V. Briolat, L. Jouneau, R. Carvalho, N. Palha, C. Langevin, P. Herbomel, O. Schwartz, H.P. Spaink, J.-P. Levraud, P. Boudinot, Contrasted innate responses to two viruses in zebrafish: insights into the ancestral repertoire of vertebrate IFN-stimulated genes, J. Immunol. 192 (9) (2014) 4328–4341, https://doi.org/10.4049/jimmunol.1302611.

[28] C. Langevin, E. Aleksejeva, G. Passoni, N. Palha, J.P. Levraud, P. Boudinot, The antiviral innate immune response in fish: evolution and conservation of the IFN system, J. Mol. Biol. 425 (24) (2013) 4904–4920, https://doi.org/10.1016/j.jmb.2013.09.033.

[29] Zhu L. Yun, L. Nie, G. Zhu, Xiang L. Xin, Shao J. Zhong, Advances in research of fish immune-relevant genes: a comparative overview of innate and adaptive immunity in teleosts, Dev. Comp. Immunol. 39 (1–2) (2013) 39–62, https://doi.org/10.1016/j.dci.2012.04.001.

[30] J. Levraud, L. Jouneau, V. Briolat, IFN-stimulated genes in zebrafish and humans define an ancient arsenal of antiviral immunity, J. Immunol. 203 (12) (2019) 3361–3373, https://doi.org/10.4049/jimmunol.1900804.

[31] I. Irnazarow, Genetic variability of Polish and Hungarian carp lines, Aquaculture. 1 (129) (1995) 215.

[32] V.R. Vasilév, Evolutsionnaya Kariologiya Ryb (Evolutionary Caryology of Fish), Nauka, Moscow, 1985.

[33] Z. Chen, Y. Omori, S. Koren, T. Shirokiya, T. Kuroda, A. Miyamoto, H. Wada, A. Fujiyama, A. Toyoda, S. Zhang, et al., De novo assembly of the goldfish (Carassius auratus) genome and the evolution of genes after whole genome duplication, Sci. Adv. 5 (6) (2019), https://doi.org/10.1101/373431.

[34] A.M. Session, Y. Uno, T. Kwon, J.A. Chapman, A. Toyoda, S. Takahashi, A. Fukui, A. Hikosaka, A. Suzuki, M. Kondo, et al., Genome evolution in the allotetraploid frog Xenopus laevis, Nature. 538 (7625) (2016) 336–343, https://doi.org/10.1038/nature19840.

[35] J.A. Birchler, R.A. Veitia, The gene balance hypothesis: from classical genetics to modern genomics, Plant Cell 19 (2) (2007) 395–402, https://doi.org/10.1105/tpc.106.049338.

[36] J.A. Birchler, R.A. Veitia, The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution, New Phytol. 186 (1) (2010) 54–62, https://doi.org/10.1111/j.1469-8137.2009.03087.x.

[37] G.C. Conant, J.A. Birchler, J.C. Pires, Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time, Curr. Opin. Plant Biol. 19 (June) (2014) 91–98, https://doi.org/10.1016/j.pbi.2014.05.008.

[38] W. De Coster, S. D'Hert, D.T. Schultz, M. Cruts, C. Van Broeckhoven, NanoPack: visualizing and processing long-read sequencing data, Bioinformatics. 34 (15) (2018) 2666–2669, https://doi.org/10.1093/bioinformatics/bty149.

[39] M. Kolmogorov, J. Yuan, Y. Lin, P.A. Pevzner, Assembly of long, error-prone reads using repeat graphs, Nat. Biotechnol. 37 (5) (2019) 540–546, https://doi.org/10.1038/s41587-019-0072-8.

[40] J. Ruan, H. Li, Fast and accurate long-read assembly with wtdbg2, Nat. Methods 17 (2019) 155–158, https://doi.org/10.1038/s41592-019-0669-3.

[41] E. Lieberman-Aiden, N. van Berkum, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. Lajoie, P. Sabo, M. Dorschner, R. Sandstrom, et al., Comprehensive mapping of long-range interactions reveals folding principles of the human genome, Science 326 (2009) 289, https://doi.org/10.1126/science.1181369.

[42] R. Tadmor-Levi, G. Hulata, L. David, Multiple interacting QTLs affect disease challenge survival in common carp (Cyprinus carpio), Heredity 1 (2019), https://doi.org/10.1038/s41437-019-0224-0.

[43] H. Tang, X. Wang, J.E. Bowers, R. Ming, M. Alam, A.H. Paterson, Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps, Genome Res 18 (2008) 1944–1954, https://doi.org/10.1126/science.1153917.

[44] S. Kurtz, A. Phillippy, A.L. Delcher, M. Smoot, M. Shumway, C. Antonescu, S. L. Salzberg, Versatile and open software for comparing large genomes, Genome Biol. 5 (2) (2004), https://doi.org/10.1186/gb-2004-5-2-r12.

[45] M. Seppey, M. Manni, E.M. Zdobnov, BUSCO: Assessing Genome Assembly and Annotation Completeness, in: M. Kollmar (Ed.), Gene Prediction. Methods in Molecular Biology, 1962, Humana, New York, NY, 2019, pp. 227–245, https://doi.org/10.1007/978-1-4939-9173-0_14.

[46] E. Sayers, E. Bolton, J. Brister, K. Canese, J. Chan, D. Comeau, R. Connor, K. Funk, C. Kelly, S. Kim, et al., Database resources of the national center for biotechnology information, Nucleic Acids Res. 7 (50 (D1)) (2022) D20–D25, https://doi.org/10.1093/nar/gkab1112.

[47] K.L. Howe, P. Achuthan, Allen James, Allen Jamie, J. Alvarez-Jarreta, M. Ridwan Amode, I.M. Armean, A.G. Azov, R. Bennett, J. Bhai, et al., Ensembl 2021, Nucleic Acids Res. 49 (D1) (2021) D884–D891, https://doi.org/10.1093/nar/gkaa942.

[48] H. Tang, J.E. Bowers, X. Wang, R. Ming, M. Alam, A.H. Paterson, Synteny and collinearity in plant genomes, Science (80-) 320 (5875) (2008) 486–488, https://doi.org/10.1126/science.1153917.

[49] B. Buchfink, C. Xie, D.H. Huson, Fast and sensitive protein alignment using DIAMOND, Nat. Methods 12 (1) (2015) 59–60.

[50] R. Buels, E. Yao, C.M. Diesh, R.D. Hayes, M. Munoz-Torres, G. Helt, D. M. Goodstein, C.G. Elsik, S.E. Lewis, L. Stein, et al., JBrowse: a dynamic web platform for genome visualization and analysis, Genome Biol. 17 (1) (2016) 1–12, https://doi.org/10.1186/s13059-016-0924-1.

[51] M. Blum, H.Y. Chang, S. Chuguransky, T. Grego, S. Kandasaamy, A. Mitchell, G. Nuka, T. Paysan-Lafosse, M. Qureshi, S. Raj, et al., The InterPro protein families and domains database: 20 years on, Nucleic Acids Res. 49 (D1) (2021) D344–D354, https://doi.org/10.1093/nar/gkaa977.

[52] M. Lechner, S. Findeiß, L. Steiner, Proteinortho: detection of (Co-)orthologs in large-scale analysis, BMC Bioinform 12 (2011) 124, https://doi.org/10.1186/1471-2105-12-124.

[53] M. Joerink, C. Ribeiro, R.J. Stet, T. Hermsen, H.F. Savelkoul, G.F. Wiegertjes, Head kidney-derived macrophages of common carp (Cyprinus carpio L.) show plasticity

and functional polarization upon differential stimulation, J. Immunol. 177 (1) (2006) 61–69, https://doi.org/10.4049/jimmunol.177.1.61.
[54] D. Kim, J.M. Paggi, C. Park, C. Bennett, S.L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype, Nat. Biotechnol. 37 (8) (2019) 907–915, https://doi.org/10.1038/s41587-019-0201-4.
[55] M. Pertea, D. Kim, G.M. Pertea, J.T. Leek, S.L. Salzberg, Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown, Nat. Protoc. 11 (9) (2016) 1650–1667, https://doi.org/10.1038/nprot.2016.095.
[56] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (12) (2014) 1–21, https://doi.org/10.1186/s13059-014-0550-8.
[57] Z. Angi, J. Ibrahim, M. Love, Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences, Bioinformatics. 35 (12) (2019) 2084–2092, https://doi.org/10.1093/bioinformatics/bty895.
[58] Z. Gu, R. Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data, Bioinformatics. 32 (18) (2016) 2847–2849, https://doi.org/10.1093/bioinformatics/btw313.
[59] H. Wickham. ggplot2: Elegant Graphics for Data Analysis, springer, 2016..