Developmental and Comparative Immunology 135 (2022) 104476

Contents lists available at ScienceDirect

Developmental and Comparative Immunology

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

Immunoglobulin heavy-chain loci in ancient allotetraploid goldfish

Linmei Han [1], Jihong Li [1], Wen Wang [1], Kaikun Luo, Mingli Chai, Caixia Xiang, Ziye Luo, Li Ren, Qianhong Gu, Min Tao, Chun Zhang, Jing Wang [*], Shaojun Liu [**]

State Key Laboratory of Developmental Biology of Freshwater Fish, College of Life Sciences, Engineering Research Center of Polyploid Fish Reproduction and Breeding of the State Education Ministry, Hunan Normal University, Changsha, 410081, Hunan, People's Republic of China

ARTICLE INFO

Keywords:
Immunoglobulin heavy-chain (*IGH*) genes
IGH loci
Allotetraploid
Chimeric gene
Carassius auratus

ABSTRACT

As an ancient allotetraploid species, goldfish (*Carassius auratus*) have two sets of subgenomes. In this study, immunoglobulin heavy-chain (*IGH*) genes were cloned from the red crucian carp (*Carassius auratus* red var.), and the corresponding loci were identified in the gynogenetic diploid red crucian carp (GRCC) genome as well as the genomes of three other goldfish strains (Wakin, G-12, and CaTCV-1). Examination showed that each goldfish strain possessed two sets of parallel *IGH* loci: a complete *IGHA* locus and a degenerated *IGHB* locus that was nearly 40 × smaller. In the *IGHA* locus, multiple τ chain loci were arranged in tandem between the μ&δ chain locus and the variable genes, but no τ-like genes were found in the *IGHB* locus.

1. Introduction

Immunoglobulin heavy-chain (*IGH*) germ-line genes include multiple variable (V), diversity (D), joining (J), and constant (C) genes (Lucas et al., 2015); the immunoglobulin (Ig) is classified based on the constant domain (CH) of the immunoglobulin heavy-chain. Six immunoglobulin isotypes have been identified in mammals (IgM, IgG, IgA, IgE, IgD, and IgO), but only three have been identified in teleosts (IgM, IgT/IgZ, and IgD) (Bilal et al., 2021; Sun et al., 2020).

The whole genome duplication (WGD) event (Glasauer and Neuhauss, 2014) in the evolutionary history of teleost fish greatly enriched the diversity and complexity of the teleost *IGH* genes (Bengtén and Wilson, 2015; Fillatreau et al., 2013). Whole *IGH* genes are duplicated in a tandem array on one chromosome in the three-spined stickleback (*Gasterosteus aculeatus*) (Bao et al., 2010), channel catfish (*Ictalurus punctatus*) (Bengtén et al., 2006), and medaka (*Oryzias latipes*) (Magadán-Mompó et al., 2011), while whole *IGH* loci are duplicated on different chromosomes in the autotetraploid Atlantic salmon (*Salmo salar*) and rainbow trout (*Oncorhynchus mykiss*); these duplications consist of two *IGH* loci (locus A and locus B; Magadan et al., 2019; Yasuike et al., 2010). *IGH* locus duplication was not detected in zebrafish (*Danio rerio*) (Danilova et al., 2005), grass carp (*Ctenopharyngodon idella*) (Xiao et al., 2010), or loach (*Misgurnus anguillicaudatus*) (Xu et al., 2018).

Extreme genomic volatility characterizes the evolution of the τ-like genes (Bradshaw and Valenzano, 2020; Mirete-Bachiller et al., 2021). The τ-like genes from *Trematomus bernacchii* (Giacomelli et al., 2015), fugu (*Fugu rubripes*) (Savan et al., 2005b), torafugu (*Takifugu rubripes*) (Fu et al., 2015), and the three-spined stickleback (Gambón-Deza et al., 2010) lack one or two CH exons; the chimeric τ-like genes from common carp (*Cyprinus carpio*) (Ryo et al., 2010; Zhang et al., 2021) and grass carp (Xiao et al., 2010) contain C regions composed of Cμ-like and Cτ-like exons; and species such as the channel catfish (Bengtén et al., 2006) and the medaka (Magadán-Mompó et al., 2011) lack τ-like genes entirely.

It is generally believed that goldfish (*Carassius auratus*) and common carp are ancient allotetraploid species (Luo et al., 2014; Ma et al., 2014), with about twice as many chromosomes as the diploid cyprinids (Yu and Zhou, 1989). Previous genomic studies of goldfish (Chen et al., 2019; Li et al., 2021; Luo et al., 2020) and common carp (Xu et al., 2014, 2019) have introduced allotetraploids derived from the fusion of two subgenomes. The evolutionary history of goldfish is somewhat complicated: Wild crucian carp evolved into red/yellow crucian carp firstly. Through long-term natural selection or artificial breeding, up to 300 different varieties developed subsequently, including red crucian carp (*Carassius auratus* red var.), grass goldfish (with a long tail), dragon-eye goldfish (with bulging eyes), oval goldfish (lacking a dorsal fin), wen goldfish

Abbreviations: VH, V genes in the immunoglobulin heavy-chain loci.

* Corresponding author.
** Corresponding author.
E-mail addresses: hwangjing0826@163.com (J. Wang), lsj@hunnu.edu.cn (S. Liu).

[1] Authors contributed equally to this work.

https://doi.org/10.1016/j.dci.2022.104476
Received 2 March 2022; Received in revised form 13 June 2022; Accepted 13 June 2022
Available online 16 June 2022

Abbreviations

IGH	immunoglobulin heavy-chain
RCC	Red crucian carp
C	constant
CH	constant domain of the immunoglobulin heavy-chain
TM	transmembrane
HK	head kidney
GU	gut
Ig	immunoglobulin
V	variable
D	diversity
J	joining
BLAST	basic local alignment search tool
RT-qPCR	quantitative reverse transcription PCR
RACE	rapid amplification of cDNA ends
Cys	cysteine residues
Trp	tryptophan residues
FR	framework region
UTR	untranslated region
WGD	whole genome duplication

(with a forked tail), and so on (Chen, 1956). Therefore, goldfish are an ideal model for studying *IGH* gene evolution in ancient allotetraploid fish.

Most studies of the *IGH* gene in ancient allotetraploid fish to date have focused on the common carp and comprehensively address the introduction of isotypes of μ-like genes (Nakao et al., 1998) and chimeric τ-like genes (Ryo et al., 2010; Savan et al., 2005a; Zhang et al., 2021). However, no comprehensive description of all *IGH* isotypes and the arrangement patterns of *IGH* loci in ancient allotetraploid carp are available. In addition, the published carp genome is not as complete as the published goldfish genome. Therefore, we employed four goldfish strains to explore the arrangement pattern of ancient allotetraploids *IGH* loci. The four goldfish strains used in this study were Wakin (*Carassius auratus*), G-12 (*Carassius auratus*), CaTCV-1 (*Carassius auratus*), and GRCC (*Carassius auratus* red var.). Strain Wakin is a common goldfish in Japan; strains G-12, CaTCV-1, and GRCC originate from the Chinese provinces Fujian, Hubei, and Hunan, respectively. In addition, strain GRCC is a gynogenetic diploid red crucian carp. Red crucian carp from our laboratory were used as experimental samples for molecular verification.

2. Materials and methods

2.1. Animals

Red crucian carp (*Carassius auratus* red var. RCC; body weight 400 ± 50 g) were obtained from an aquaculture base in Wang Cheng district, Changsha, Hunan province, China. Fish were anesthetized in 100 mg/L MS-222 (Sigma-Aldrich, USA) prior to dissection or blood extraction.

2.2. RNA isolation and cDNA synthesis

The head kidney (HK) and gut (GU) from three anesthetized RCC were isolated and immediately frozen in liquid nitrogen. The phenol-chloroform method was used to isolate total RNA with Trizol Reagent (Invitrogen, USA). The integrity of the isolated RNA was determined using gel electrophoresis with 1% RNase-free agarose, and RNA concentration was measured using a Synergy 2 Multi-Mode Microplate Reader (BioTek2, USA). Random hexamer and Oligo $(dT)_{18}$ primers were used to synthesize cDNA for quantitative reverse transcription PCR (RT-qPCR) using the RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific, USA) and following the manufacturer's instructions. The synthesis of cDNA for cloning was carried out using Oligo $(dT)_{18}$ primers only.

2.3. Cloning and the expression of IGH genes

Degenerate forward primers [annealed to framework region 3 (FR3)] and specific reverse primers were designed based on the cDNA sequences of Trinity-assembled transcriptome data (BioSample accession no. SAMN17515892) to amplify the cDNA sequences of the RCC *IGH* genes. Because the Trinity-assembled transcriptome sequence is not a full-length sequence, the secretory and membrane primers of *IGHMA* were annealed to the UTR region to obtain the clones; multiple pairs of primers were used for the *IGHD* gene due to its long length, and 3′ rapid amplification of cDNA ends (RACE) was used to amplify the complete 3′ UTRs of RCC *IGHD*, *IGHT2D*, *IGHT4D,* and *IGHMB*. PCR products were amplified using LA Taq with a volume of 20 μl and 30 cycles (TaKaRa, JP) and cleaned with a SanPrep Spin Column & Collection Tube (Sangon, CN); then, 0.5 μl of the gel-cleaned PCR product cloned into the pMD18-T vector (TaKaRa, JP), and eight monoclones were randomly selected and sequenced. All primers used in this study are listed in Supplementary Table 1. RT-qPCRs were performed using PowerUp SYBR Green Master Mix (ThermoFisher, USA) with a volume of 20 μl and 40 cycles. qPCR results were analyzed using QuantStudio 5 (ThermoFisher, USA), and the relative expression levels of the *IGH* genes were calculated using the $2^{-\triangle\triangle Ct}$ method with β-actin gene as the internal reference (Livak and Schmittgen, 2001).

2.4. Identifying IGH loci in goldfish genomes

First, we roughly located the *IGH* loci in the genomes of four goldfish strains (strains GRCC, G-12, Wakin, and CaTCV-1) by using the V and the C genes of the immunoglobulin heavy-chain loci of zebrafish and grass carp to determine the total span range of the goldfish *IGH* loci. Then, the boundaries of V, D, J, and C were determined as follows: the CH region was manually checked by the cDNA sequences of the RCC *IGH* genes using BLAST (Altschul et al., 1990); the V genes were detected using IgBLAST (https://www.ncbi.nlm.nih.gov/igblast) (Ye et al., 2013) and FGENESH (Salamov and Solovyev, 2000), and then refined using homology-based prediction with FGENESH+ (Solovyev, 2007) and splice site prediction with NNSPLICE (http://www.fruitfly.org/seq_tools/splice.html) (Reese et al., 1997); D and J genes were identified using a pattern search with EMBOSS fuzznuc software (v.6.6.0.0) (Rice et al., 2000), allowing a maximum of four mismatches. All complete V genes begin with the octamer sequence ATG(C/T)AAA(G/T) followed by two exons (first leader peptide exon and second V exon) and terminate with the recombination signal sequence (RSS); D genes have a 3′-flanking RSS with the pattern CACAGTG-$N_{11\text{-}13}$-ACAAAAACC no more than 40 nucleotides from the 5′-flanking RSS (we used CACAGTG-$N_{11\text{-}13}$-ACAAAAACC and GGTTTTTGT-$N_{11\text{-}13}$-CACAGTG as the D-gene search string), while J genes not only possess a conserved FDY motif and an ambiguous VTVSS termination motif, but also a 5′-flanking RSS (we used CACAGTG-$N_{22\text{-}24}$-ACAAAAACC as the J-gene search string). Using the above sequence criteria for V, D, and J genes, which were summarized based on the sequences of these genes in other species (Lucas et al., 2015), the goldfish V, D, and J genes were manually checked and annotated. V genes were named following standard IMGT nomenclature (Lefranc, 2011), requiring complete second V exon. Gene structure was visualized using IBS software (Liu et al., 2015).

2.5. Southern blotting

Whole blood was extracted from the coccygeal veins of four RCCs (RCC1–4), and a 1/10 volume of anticoagulant ACD solution was added. The whole blood was washed once with 0.7% NaCl solution and three times with D-Hanks solution to separate and purify the erythrocytes.

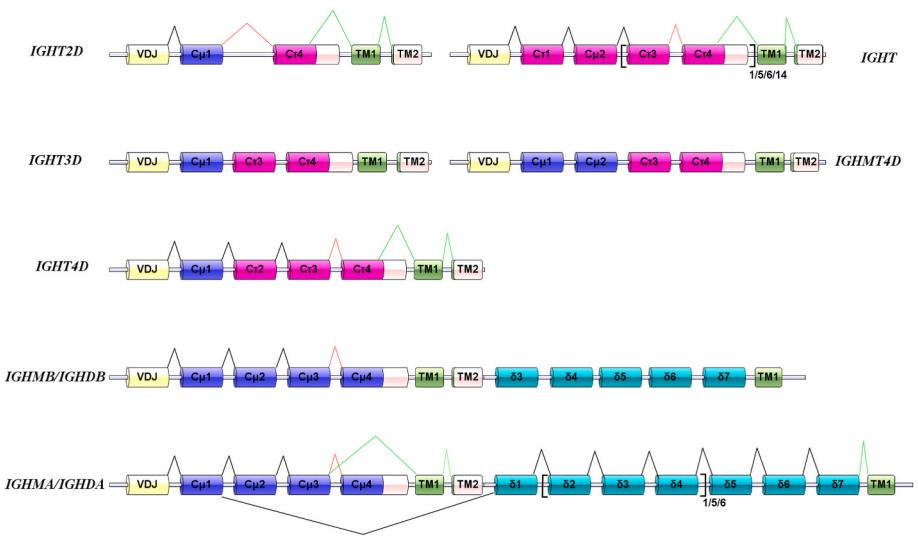

Fig. 1. A schematic representation of the immunoglobulin heavy-chain genes recovered in four goldfish genomes (*Carassius auratus* and *Carassius auratus* red var.). Bent lines above and below the gene schematic represent alternative splicing of the Ig-like exons (black), transmembrane isoforms (green), and secreted isoforms (red). The cDNA sequences of *IGHT3D*, *IGHMT4D,* and *IGHDB* were not successfully cloned in the red crucian carp.

Genomic DNA was extracted from the erythrocytes using a Rapid Animal Genomic DNA Isolation Kit (Sangon, CN). We digested 30 μg of genomic DNA from each of the four fish (RCC1–4) to completion with 100 units of PstI and 100 units of EcoRI at 37 °C for 6 h. The digests were electrophoresed on a 0.8% agarose gel and transferred to a Hybond-N^+ membrane (GE, USA). We fixed the nucleic acid to the membrane via cross-linking using UV-light (Mamiatis et al., 1985). The membranes were hybridized with probes at 42 °C using the DIG High Prime DNA Labeling and Detection Starter Kit II (Roche, CH), following the manufacturer's instructions.

The genomic DNA sequences of *IGHMA* and *IGHMB* (the sequences between the Cμ2 and Cμ4 exons) from RCC1 and RCC3 (in Southern blotting results, the band patten of RCC2 was similar to that of RCC1, and RCC4 was similar to that of RCC3, and thus only RCC1 and RCC3 were selected for molecular cloning) were cloned into the pMD18-T vector. Eight clones were then randomly selected for sequencing.

To determine the copy numbers of the RCC *IGHM* genes, DIG-labeled DNA probes specific to the *IGHMA* (probe MA) and *IGHMB* (probe MB) were hybridized with PstI/EcoRI digests of erythrocyte DNA from the four individual RCC (no DIG-labeled MA or MB were selected as positive controls). The MA probe was synthesized from the amplified products of the *RCC1_IGHMA1*-containing plasmid, and the MB probe was synthesized from the amplified products of the *RCC1_IGHMB1*-containing plasmid. Neither probe sequence contained a PstI or an EcoRI site. However, the sequence of *RCC1_IGHMA2* contained a PstI site.

2.6. Sequence analysis and phylogenetic analysis

The ExPASy translate tool (https://web.expasy.org/translate/) (Duvaud et al., 2021) was used to deduce the immunoglobulins (Igs) amino acid sequences of RCC and zebrafish. Multiple sequence alignments were made with the clustalW algorithm (Ma et al., 2007). The amino acid sequences alignment and DNA sequences alignment were performed by BioEdit software. For genomic DNA sequences alignment of RCC *IGHM* genes, *GRCC_IGHMA* (GenBank accession no. the reverse complement of LGRG01000905.1: 338577–339617) and *GRCC_IGHMB* (GenBank accession no. the reverse complement of LGRG01000274.1: 509764–510793) were selected as references. For amino acid sequence alignment of RCC Igs, zebrafish Igs were selected as references (the GenBank accession numbers of secretory and membrane isoforms of IgM and IgZ are AY643753, AY643751, AY643752, and AY643750). The secondary structure of the Igs amino acid sequence was predicted using Prosite (https://prosite.expasy.org/prosite.html) (Duvaud et al., 2021), and N-Glycosylation sites were predicted using the etNGlyc 1.0 Server (http://www.cbs.dtu.dk/services/NetNGlyc/) (Gupta and Brunak, 2002).

Table 1
The *IGH* cDNA sequences amplified from a single RCC individual.

IGH gene	Amplified region	Length of secretory type (bp)	Length of membrane type (bp)	GenBank accession no
IGHMA	FR3 to 3′UTR	1433	1279	MZ558192; MZ558193
IGHMB	FR3 to end of 3′UTR	1601	/	MZ558191
IGHT	Cτ1 to 3′UTR	1313	/	MZ558198
IGHT4D	FR3 to end of 3′UTR	1715	1738	MZ558196; MZ558197
IGHT2D	FR3 to end of 3′UTR	1128	1188	MZ558194; MZ558195
IGHD	FR3 to end of 3′UTR	/	2977	OM104098

The total numbers of V genes in the immunoglobulin heavy-chain loci (VH) of G-12 and CaTCV-1 were almost completely identified. The second exon sequences of the VH genes of the G-12 and CaTCV-1 were used to construct separate datasets. ModelFinder (Kalyaanamoorthy et al., 2017) predicted that the best model was GTR; two phylogenetic trees were constructed using MrBayes (Huelsenbeck and Ronquist, 2001) as implemented in PhyloSuite (Zhang et al., 2020). Two MrBayes runs were performed, each for 20 million generations with four chains in parallel. The first 25% of the trees were discarded as a burnin; the sampling frequency was 100, and the analysis was terminated after the average standard deviation of the split frequencies fell below 0.01. For the VH subgroup, a Bayesian phylogeny of VH genes was constructed by requiring at least 75% nucleotide identity among VH genes within the same subgroup. The phylogenetic tree was displayed using FIGTREE (V1.4.4) software.

3. Results

3.1. Identification of IGH genes in goldfish genomes

We identified nine different *IGH* genes in the previously published

Table 2
Difference in *IGH* genes among the four goldfish strains.

Goldfish strain	*IGH* locus	The arrangement of the conserved region of the *IGH* genes
G-12	CτA1: *IGHT2D*	Cμ1-Cτ4-TM1-TM2
	CτA2: *IGHT2D*	Cμ1-Cτ4-TM1-TM2
	CτA3: *IGHMT4D*	Cμ1-Cμ2-Cτ3-Cτ4-TM1-TM2
	CμA&CδA	Cμ1-Cμ2-Cμ3-Cμ4-TM1-TM2-Cδ1-(Cδ2-Cδ3-Cδ4)5-Cδ5-Cδ6-Cδ7-TM
	CμB&CδB	Cμ1-Cμ2-Cμ3-Cμ4-TM1-TM2-Cδ3-Cδ4-Cδ5-Cδ6-Cδ7-TM
GRCC	CτA1: *IGHT*	Cτ1-Cτ2-(Cτ3-Cτ4)5↑-TM1-TM2
	CτA2: *IGHT2D*	Cμ1-Cτ4-gap
	CτA3: *IGHT2D*	Cμ1-Cτ4-TM1-TM2
	CτA4: *IGHT3D*	Cμ1-Cτ3-Cτ4-TM1-TM2
	CτA5: *IGHMT4D*	gap-Cμ2-Cτ3-Cτ4-TM1-TM2
	CτA6: *IGHT4D*	Cμ1-Cτ2-Cτ3-Cτ4-TM1-TM2
	CμA&CδA	Cμ1-Cμ2-Cμ3-Cμ4-TM1-TM2-Cδ1-Cδ2-Cδ3-Cδ4-Cδ5-Cδ6-Cδ7-TM
	CμB&CδB	Cμ1-Cμ2-Cμ3-Cμ4-TM1-TM2-Cδ3-Cδ4-Cδ5-Cδ6-Cδ7-TM
Wakin	CτA1	gap-(Cτ3-Cτ4)11-TM1-TM2
	CτA2: *IGHT*	Cτ1-Cτ2-(Cτ3-Cτ4)5-TM1-TM2
	CτA3: *IGHT4D*	Cμ1-Cτ2-Cτ3-Cτ4-TM1-TM2
	CτA4: *IGHT4D*	Cμ1-Cτ2-Cτ3-Cτ4-TM1-TM2
	CτA5: *IGHT4D*	Cμ1-Cτ2-Cτ3-Cτ4-TM1-TM2
	CτA6	gap-Cτ2-Cτ3-Cτ4-TM1-TM2
	CμA1&CδA1	Cμ1-Cμ2-Cμ3-Cμ4-TM1-TM2-Cδ1-(Cδ2-Cδ3-Cδ4)6-Cδ5-Cδ6-Cδ7-TM
	CμA2&CδA2	Cμ1-Cμ2-Cμ3-Cμ4-TM1-TM2-Cδ1-(Cδ2-Cδ3-Cδ4)6-Cδ5-Cδ6-Cδ7-TM
	CμB&CδB	Cμ1-Cμ2-Cμ3-Cμ4-TM1-TM2-Cδ3-Cδ4-Cδ5-Cδ6-Cδ7-TM
CaTCV-1	CτA1: *IGHT3D*	Cμ1-Cτ3-Cτ4-TM1-TM2
	CτA2: *IGHT4D*	Cμ1-Cτ2-Cτ3-Cτ4-TM1-TM2
	CτA3: *IGHT*	Cτ1-Cτ2-Cτ3-Cτ4
	CτA4: *IGHT*	Cτ1-Cτ2-(Cτ3-Cτ4)14-TM1-TM2
	CτA5: *IGHT*	Cτ1-Cτ2-Cτ3-Cτ4-TM1-TM2
	CτA6	Cτ2- TM1-TM2
	CτA7: *IGHT*	Cτ1-Cτ2-(Cτ3-Cτ4)6-TM1-TM2
	CμA1&CδA1	Cμ1-Cμ2-Cμ3-Cμ4-TM1-TM2-Cδ1-(Cδ2-Cδ3-Cδ4)6-Cδ5-Cδ6-Cδ7-TM
	CμA2&CδA2	Cμ1-Cμ2-Cμ3-Cμ4-TM1-TM2-Cδ1-(Cδ2-Cδ3-Cδ4)6-Cδ5-Cδ6-Cδ7-TM
	CμB&CδB	Cμ1-Cμ2-Cμ3-Cμ4-TM1-TM2-Cδ3-Cδ4-Cδ5-Cδ6-Cδ7-TM

Note: ↑ indicates that the Cτ3-Cτ4 blocks were repeated more than 5 times; The exact number of repetitions could not be determined due to gaps.

genomes of four goldfish strains: two δ-like genes (*IGHDA* and *IGHDB*; *IGHDB* was truncated compared to *IGHDA*), two μ-like genes (*IGHMA* and *IGHMB*), five τ-like genes (classified based on their C region; *IGHT*, *IGHT2D*, *IGHT3D*, *IGHT4D*, and *IGHMT4D*) (Fig. 1).

The cDNA sequences of the *IGH* genes were successfully cloned from the RCC, except for *IGHDB*, *IGHT3D,* and *IGHMT4D* (Table 1). Only the membrane isoform of *IGHD* and the secretory isoforms of *IGHMB* and *IGHT* were present in RCC. Both the membrane and the secretory isoforms of *IGHMA*, *IGHT2D,* and *IGHT4D* were successfully cloned in RCC. Cloning of *IGHDB*, *IGHT3D,* and *IGHMT4D* may have failed due to low levels of expression; alternatively, these could be pseudogenes. In addition, the τ-like genes share high identity at the nucleotide level, and thus cloning may have failed due to insufficient primer specificity.

The overall structure of the μ-like genes in the *IGHA* and *IGHB* locus was conserved across the four goldfish strains, while δ and τ-like genes were structurally variable (Table 2). The μ-like genes consisted of four constant Ig-like exons and two transmembrane exons (Cμ1-Cμ2-Cμ3-Cμ4-TM1-TM2); Cμ3 was directly spliced to TM1 to form the membrane isoform. The δ-like genes in the *IGHA* locus consisted of multiple constant Ig-like exons and one transmembrane exon (Cδ1-(Cδ2-Cδ3-Cδ4)$_{1/5/6}$-Cδ5-Cδ6-Cδ7-TM1); Cμ1 was spliced to Cδ1 to form the membrane *IGHD*. The number of Cδ2-Cδ3-Cδ4 repeats depends on the strains and the gene locus; G-12 *IGHDA* was 5; GRCC *IGHDA* was 1; Wakin *IGHDA1*, *IGHDA2*, CaTCV-1 *IGHAD1*, and *IGHAD2* were 6. However, *IGHDB* was truncated among the four goldfish strains (Cδ3-Cδ4-Cδ5-Cδ6-Cδ7-TM1). Different isotypes of τ-like genes have different numbers of constant Ig-like exons plus two transmembrane exons, and the Cτ4-like exon is always followed by two transmembrane exons. In combination with the RCC cDNA sequences of *IGHT2D* and *IGHT4D*, this suggests that the one isotype of τ-like genes consisted of the same number of constant Ig domains in both the membrane and the secreted isoforms, and all of the τ-like genes may share similar alternative splice modes: a cryptic splice site near the 3′ end of the Cτ4 exon connects the transmembrane exon (*IGHT2D*: Cμ1-Cτ4-TM1-TM2; *IGHT3D*: Cμ1-Cτ3-Cτ4-TM1-TM2; *IGHT4D*: Cμ1-Cτ2-Cτ3-Cτ4-TM1-TM2; *IGHMT4D*: Cμ1-Cμ2-Cτ3-Cτ4-TM1-TM2; *IGHT*: Cτ1-Cτ2-(Cτ3-Cτ4)$_{1/5/6/14}$-TM1-TM2) (Fig. 1). The number of Cτ3-Cτ4 repeats also differed among three goldfish strains at the *IGHT* locus: GRCC *IGHT* had more than five (the number could not be exactly determined due to gap sequences), Wakin *IGHT* had five, CaTCV-1 *IGHT* of CτA3 and CτA5 had one, CaTCV-1 *IGHT* of CτA7 had six, and CaTCV-1 *IGHT* of CτA4 had 14 (Table 2).

Multiple sequence alignment of the amino acid sequences of IgMA, IgMB, IgT, IgT2D, and IgT4D of RCC with IgM and IgZ from zebrafish showed that all CH domains had two conserved cysteine residues (Cys) that formed the intradomain disulfide bridge (Supplementary Fig. 1). This bridge is a key feature of Ig-like domains (Kabat, 1991). In the Cτ3 region, the space between the two cysteine residues was much shorter than that in the Cμ3 region. Three "extra" cysteine residues were identified in IgMA and IgMB: the first was located at the beginning of Cμ1, forming an interdomain disulfide bridge to an immunoglobulin light chain (IgL); the second was located at the beginning of Cμ3, and the third was located in the secretory tail, forming a bridge between the heavy chains. IgT2D also had three extra cysteine residues, but the last two of these residues were located in the secretory tail. However, only two "extra" cysteine residues were identified in IgT and IgT4D. Both IgT and IgT4D had an "extra" cysteine, which formed interchain disulfide bonds with the light chains (H2L2), as well as an "extra" cysteine in the secretory tail. This implied that H2L2 monomers are held together by covalent bonds, and IgT or IgT4D tetramers, when present, may associate by non-covalent interactions (Yu et al., 2019). The number of N-linked glycosylation sites differed among Igs: seven in IgMA, six in IgMB and IgT, four in IgT4D, and three in IgT2D. The secretory tail of IgM was longer than that of IgT, IgT2D, and IgT4D, and no N-linked glycosylation sites were found in the Cτ tail (Supplementary Fig. 1).

3.2. The arrangement of IGH genes in the four goldfish genomes

In the goldfish genomes, the *IGH* loci were identified on different chromosomes (Chr) in two parallel subgenomes (Supplementary Table 2). The *IGHA* locus had multiple τ chain loci before the μ&δ chain locus; at the *IGHB* locus the τ-like gene was lost and the δ chain locus was truncated, suggesting that only the μ-like gene may be functional (Fig. 2). The CτA types and copy numbers in the *IGHA* loci varied greatly among the four goldfish strains (Table 2).

Using the high-quality genome assembly of strain G-12, *IGHA* (1.12 Mb) was located on Chr3, and *IGHB* (32.66 Kb) was located on Chr28. Three τ chain loci were arranged in tandem between clusters of V genes and the μ&δ chain locus in the *IGHA* locus. CτA1 and CτA2 had only two Ig-like exons encoding *IGHT2D*. CτA3 had four Ig-like exons encoding *IGHMT4D* (Table 2). All τ chain loci and μ&δ chain loci had independent D and J genes, with up to 166 V genes in the *IGHA* locus. However, only the μ&δ chain locus was identified in the *IGHB* locus; the δ1 and δ2 exons had been lost. In addition, far fewer V, D, and J genes were identified in the *IGHB* locus (5, 1, and 3, respectively) as compared to the *IGHA* locus (166, 18, and 13, respectively; Fig. 2A).

In the genome assembly of strain GRCC, none of the scaffolds were anchored to chromosomes. Two μ&δ chain loci were identified, one in the *IGHA* locus and another in the *IGHB* locus. Up to six τ chain loci

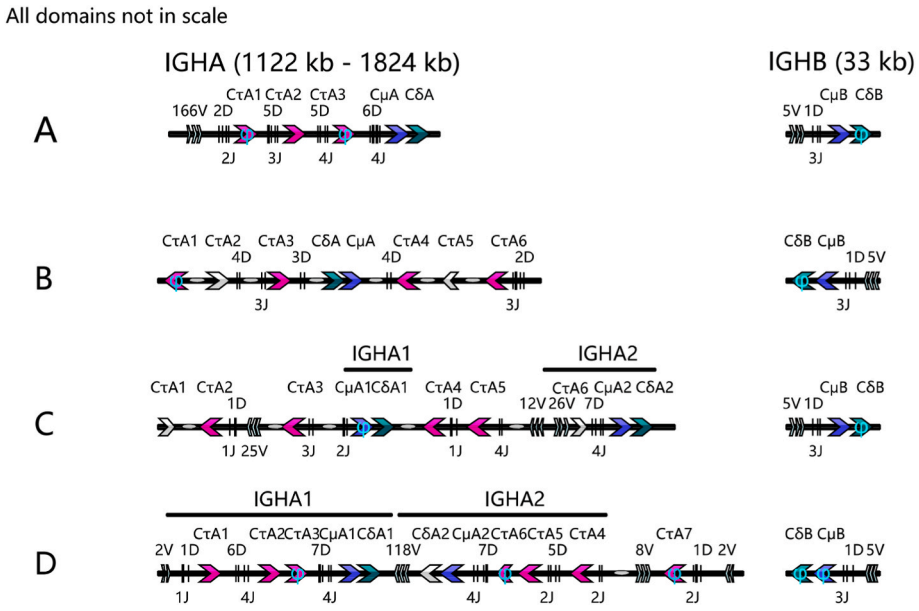

Fig. 2. Configuration of the immunoglobulin heavy-chain loci in the genomes of four goldfish strains. (A) Strain G-12 (GenBank accession no. GCA_014332655.1). (B) Strain GRCC (GenBank accession no. GCA_009069565.1). (C) Strain Wakin (GenBank accession no. GCA_003368295.1). (D) Strain CaTCV-1 (GenBank accession no. GCA_019720715.1). Grayed out areas indicate that the sequence in the corresponding region could not be correctly annotated due to sequencing or assembly problems, and the relative orientation of the *IGH* genes in the region can thus not be determined. φ indicates that the corresponding gene is a pseudogene.

encoded 5 *IGH* genes (CτA1: *IGHT*; CτA2 and CτA3: IGT2D; CτA4: *IGHT3D*; CτA5: *IGHMT4D*; CτA6: *IGHT4D*) (Fig. 2B and Table 2). The GRCC CτA5 locus had a gap sequence ahead of the Cμ2 exon (gap-Cμ2-Cτ3-Cτ4-TM1-TM2), the same as G-12 *IGHMT4D* (Cμ1-Cμ2-Cτ3-Cτ4-TM1-TM2). Therefore, we speculated that the gene type of GRCC CτA5 was *IGHMT4D*.

In the genome assembly of strain Wakin, three μ&δ chain loci and six τ chain loci were identified (CτA3, CτA4 and CτA5: *IGHT4D*; CτA2: *IGT*; Fig. 2C and Table 2). The genes encoded by CτA1 and CτA6 in the *IGHA* locus could not be determined due to gap sequences. In addition, the *IGHA* locus in the Wakin genome assembly was incomplete compared to the *IGHA* locus in the G-12 genome assembly: the Wakin assembly included fewer VH genes and the incomplete τ chain loci. The configuration of the *IGHB* locus in the Wakin assembly was identical to that in the G-12 assembly.

In the genome assembly of strain CaTCV-1, the transcription orientation of *IGHA2* was the reverse of that of *IGHA1*; the length of the *IGHA* locus was about 1.82 Mb (Fig. 2D). In the *IGHA* locus, most V genes were located between *IGHA1* and *IGHA2*. These V genes had the same transcription orientation as *IGHA2*. In the *IGHA1* locus, only two V genes were located before the τ chain loci; these two V genes had the same transcription orientation as *IGHA1*. There were three τ chain loci before the μ&δ chain locus in *IGHA1* and *IGHA2*. Of the seven τ-like genes, two were chimeric; one was truncated, and the remainder were *IGHT* genes (CτA1: *IGHT3D*; CτA2: *IGHT4D*; CτA6: truncated; CτA3, CτA4, CτA5 and CτA7: *IGHT*; Table 2). It was unclear whether CτA7 was linked to the *IGHA* locus. The configuration of the *IGHB* locus in the CaTCV-1 assembly was identical to that of the *IGHB* locus in the G-12 assembly, but in the CaTCV-1 assembly, *IGHMB* in the CμB chain locus was a pseudogene.

It remains to be determined whether the tandem duplication of *IGHA1* and *IGHA2* in the strain CaTCV-1 and Wakin *IGHA* loci was the result of intervarietal differences or incomplete genome assembly.

3.3. The V, D, and J genes in the goldfish IGH loci

In strain G-12, there were 166 V genes in the *IGHA* locus and 5 V genes in the *IGHB* locus; 27 of the V genes were truncated (*IGHA* was 26 and *IGHB* was 1), but 11 (*IGHA* was 10 and *IGHB* was 1) of these truncated V genes had a complete second V exon (Supplementary Table 2). We therefore identified and named a total of 155 V sequences across both loci (5 from *IGHB* and 150 from *IGHA*). We constructed a

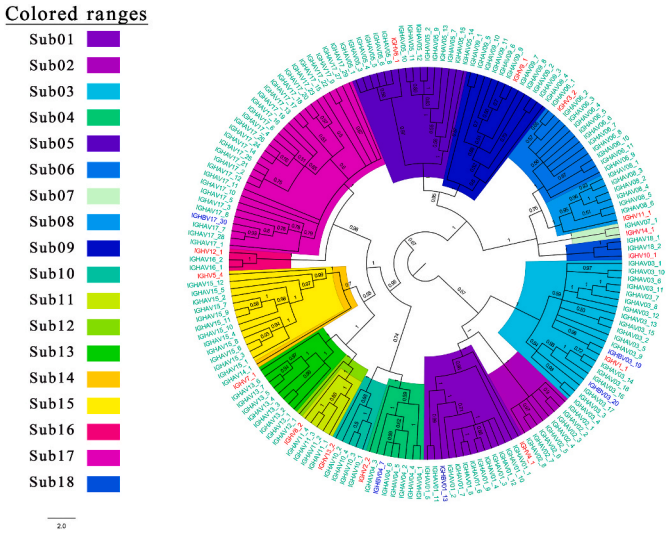

Fig. 3. Bayesian phylogeny of VH genes from zebrafish and goldfish strain G-12. The tree was constructed using the second exon of the G-12 VH genes and 14 representative zebrafish VH genes (without the RSS sequence; IMGT/LIGM accession number BX649502). The representative zebrafish sequences are labeled in red, the G-12 *IGHAV* sequences are labeled in light blue, and the G-12 *IGHBV* sequences are labeled in purple. Clades are shaded by VH subgroup; subgroups 02, 12, 14, and 17 are not found in zebrafish.

phylogenetic tree based on representative sequences from the 14 VH subgroups in zebrafish and the VH sequences from strain G-12, and we classified the 155 VH sequences into 18 different VH subgroups; four of these VH subgroups (02, 12, 14, and 17) were not present in zebrafish (Fig. 3). In the G-12 genome, we identified only one copy of VH subgroups 12 and 14, but 8 copies of subgroup 02 and 30 copies of subgroup 17. Subgroup 17 had the greatest number of members across all VH subgroups in the G-12 genome. The conserved tryptophan (Trp) residues in FR2 were replaced by phenylalanine (Phe) residues in the most members of subgroup 15, similar to the replacement of Trp residues by Phe residues in zebrafish VH subgroup 5 (Danilova et al., 2005).

Strain CaTCV-1 had fewer VH genes than G-12: we identified a total of 135 VH genes (*IGHA* had 130 and *IGHB* had 5), but only 103 VH sequences (99 from *IGHA* and 4 from *IGHB*) were nominated. To

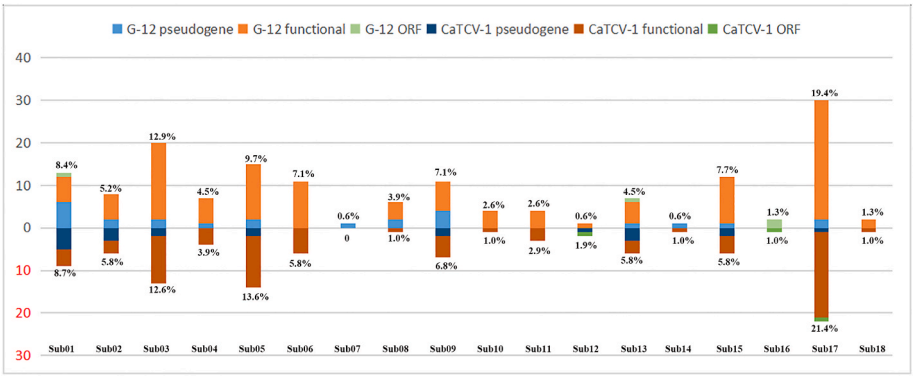

Fig. 4. Absolute numbers and relative percentages of VH sequences from goldfish strains G-12 and CaTCV-1 classified into each VH subgroup. Only named VH genes are included.

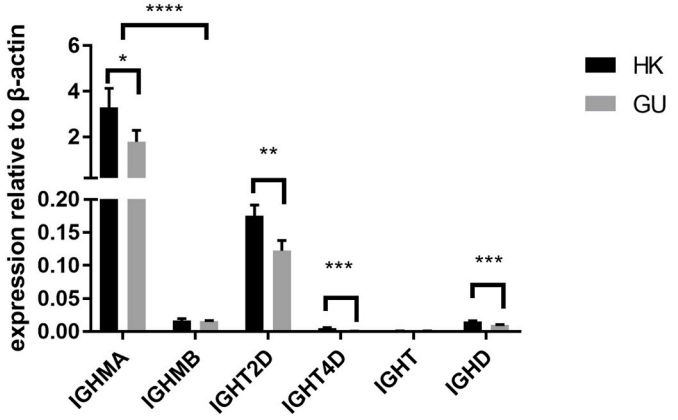

Fig. 5. Relative expression of *IGH* genes in the head kidney (HK) and gut (GU) of RCC. Relative mRNA expression levels were normalized against the expression of β-actin. Each value represents the mean of three replicates; data are expressed as mean ± SD of three fish, and statistical analysis between HK and GU was performed by a *t*-test. *P < 0.05 to ****P < 0.0001.

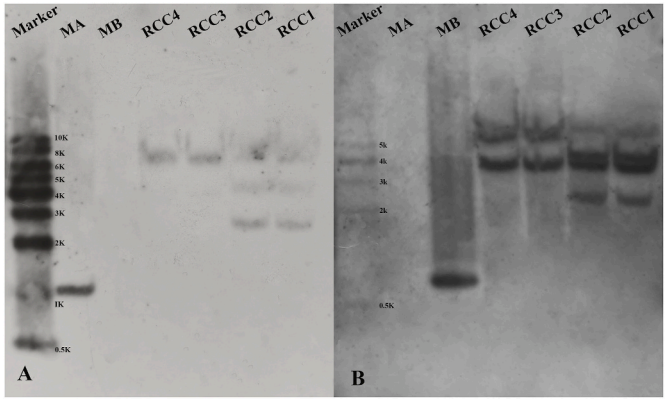

Fig. 6. Southern blots showing the hybridization patterns of RCC erythrocyte genomic DNA digests using CH DNA probes specific for MA and MB. 1 Kb marker was at the first panel with 1 ng. Lanes RCC1–4 corresponding to RCC individuals. (**A**) Southern blot showing hybridization with probe MA. No DIG-labeled MA and MB (each 100 pg) were used as positive controls. (**B**) Southern blot showing hybridization with probe MB. No DIG-labeled MA and MB (each 300 pg) were used as positive controls.

evaluate the relationship among VH genes of the goldfish strains of G-12 and CaTCV-1, we constructed a phylogenetic tree of 18 representative G-12 VH sequences and 103 CaTCV-1 VH genes. These 103 VH sequences were classified into 17 different VH subgroups; subgroup 07 was not identified in strain CaTCV-1 (Supplementary Fig. 2). Other than subgroup 07, the proportion of CaTCV-1 sequences classified into each subgroup was roughly similar to proportions in strain G-12 (Fig. 4). However, 41.48% of the CaTCV-1 VH sequences encoded pseudogenes (56/135), which was much higher than the proportion of pseudogenes in strain G-12 (26.90%; 46/171).

The D and J genes in goldfish strains G-12 and CaTCV-1 are presented in Supplementary Table 3. The *IGHA* locus of strain G-12 had 18 D genes and 13 J genes, while the *IGHA* locus of strain CaTCV-1 had 27 D genes and 19 J genes; the *IGHB* loci of both strains had 5 V, 1 D, and 3 J genes. It is noteworthy that some D or J sequences were identical among different *IGH* loci.

3.4. Expression of IGH genes

Comparison of *IGH* gene expression between the head kidney and gut of RCC showed that *IGHMA*, *IGHT2D*, *IGHT4D,* and *IGHD* were more strongly expressed in the head kidney, while there were no significant differences in the expression levels of other genes (Fig. 5). In the head kidney, *IGHMA* was about 20-fold more strongly expressed than *IGHT2D* and about 100-fold more strongly expressed than either *IGHMB* or *IGHD* (Fig. 5). Among these three τ-like genes, the expression of *IGHT2D* was dominant in both tissues, while expression levels of *IGHT* and *IGHT4D*

Table 3
Pairwise sequence identity between *IGHMA* and *IGHMB* (genomic DNA sequence between the Cμ2 and Cμ4 exons) in RCC1 and RCC3.

	Length (bp)	Identity (%)	1	2	3	4	5	6	7	8
RCC1_IGHMA1	1042	1	/	97.88	99.90	99.04	78.31	79.35	79.67	79.46
RCC1_IGHMA2	1054	2		/	97.79	98.37	78.36	79.39	79.49	79.28
RCC3_IGHMA1	1042	3			/	98.94	79.19	79.23	79.56	79.35
RCC3_IGHMA2	1043	4				/	78.22	79.25	79.58	79.37
RCC1_IGHMB1	875	5					/	96.43	94.27	95.40
RCC1_IGHMB2	870	6						/	96.55	98.52
RCC3_IGHMB1	874	7							/	97.81
RCC3_IGHMB2	870	8								/

were extremely low in both tissues (Fig. 5).

3.5. The copy numbers of IGHM genes

The DNA sequences of *IGHMA* and *IGHMB* were cloned in RCC1 and RCC3 (Table 3). In RCC1, *IGHMA1* was 1042 bp, *IGHMA2* was 1054 bp, *IGHMB1* was 875 bp, and *IGHMB2* was 870 bp; in RCC3, *IGHMA1* was 1042 bp, *IGHMA2* was 1043 bp, *IGHMB1* was 874 bp, and *IGHMB2* was 870 bp. Nucleotide sequence identity among the *IGHMA* and *IGHMB* sequences was less than 80%, but sequence identity among the *IGHMA1* and *IGHMA2* sequences, as well as among the *IGHMB1* and *IGHMB2* sequences, was more than 97%. The alignment of *IGHMA* and *IGHMB* sequences from RCC1, RCC3, and GRCC also showed the low identity between *IGHMA* and *IGHMB* (Supplementary Fig. 3).

Southern blots of each of the *IGHM* probes revealed two different band patterns across the four individual RCCs: in both the MA and MB results, the band patten of RCC1 was similar to that of RCC2, while that of RCC3 was similar to that of RCC4 (Fig. 6). However, the band patterns exhibited by RCC1 and RCC2 differed from those exhibited by RCC3 and RCC4. Cross-hybridization between the MA and MB probe was not observed, indicating the specificity of the MA and MB probes. The MA results showed that RCC1 and RCC2 had three bands, while RCC3 and RCC4 had one band (nearly 8 kb) each, similar to the longest band in RCC1 and RCC2 (Fig. 6A). The MB results showed that RCC1 and RCC2 had four bands each, while RCC3 and RCC4 had two bands (more than 5 kb and nearly 4 kb) each; these two bands were the same as the corresponding bands from RCC3 and RCC4 (Fig. 6B). In addition, there was no obvious correspondence in band patterns between the probes for any individual (Fig. 6). Therefore, *IGHMA1* and *IGHMA2* or *IGHMB1* and *IGHMB2* might be alleles on homologous chromosomes or copies of different *IGHM* loci. However, *IGHMA* and *IGHMB* represent distinct *IGHM* genes. These results were consistent our analysis of *IGH* loci in the four goldfish genomes, where we observed two parallel *IGH* loci.

4. Discussion

We compared *IGH* loci among four goldfish genomes. We found that all four goldfish had two parallel *IGH* loci that may be inherited in a disomic manner from two subgenomes, and we verified this hypothesis via Southern blots of the *IGHMA* and *IGHMB* results. The findings were consistent with what has been reported in common carp (Nakao et al., 1998). The allotetraploid will revert to the diploid level by gene loss in the evolutionary process, and the remaining functionally similar duplicates have a dosage compensation effect (Birchler and Veitia, 2007). The same conclusion was found for *IGH* loci of ancient allotetraploid goldfish; significant bias of the *IGHA* locus was observed in the number of genes (Fig. 2) and in the expression of homologous genes (Fig. 6) compared to the *IGHB* locus.

The gene numbers and configuration of the *IGHB* locus were conserved among the four goldfish strains, while types and numbers of the τ-like genes in the *IGHA* loci were strain-specific. The *IGH* locus of five τ-like genes (*IGHT*, *IGHT2D*, *IGHT3D*, *IGHT4D*, and *IGHMT4D*) were all found in GRCC *IGHA* loci; however, only two τ-like genes *IGHT2D* and *IGHMT4D* were found in G-12; two τ-like genes *IGHT4D* and *IGHT* in Wakin; and three τ-like genes *IGHT3D*, *IGHT4D* and *IGHT* in CaTCV-1. Several studies have demonstrated that the subclasses of τ-like genes have different immune responses to pathogen infection (Ryo et al., 2010; Ji et al., 2021), suggesting that goldfish may selectively inherit different types of τ-like genes to adapt to the local environment (Bradshaw and Valenzano, 2020). Chimeric Igs have been widely reported in teleosts; chimeric τ-like Igs have independent corresponding genomic DNA sequences, unlike IgD that forms chimeras with the Cμ1 exon by alternative splicing to Cδ exons (Ohta and Flajnik, 2006). IgD/IgT chimeras in the European sea bass (*Dicentrarchus labrax* L.) have also been identified (Buonocore et al., 2020). To date, the reason for the formation of chimeric Igs in teleost fish has not been well explained, but it must be related to WGD events and the high plasticity of the *IGH* gene structure (large structural variants and copy number variants) (Watson et al., 2013), and the mechanism of chimeric gene formation needs further study. *IGHT* and *IGHD* displayed remarkable structural plasticity among the four goldfish strains; the arrangement configuration of *IGHT* in the *IGHA* locus possessed repeated blocks of exons encoding Cτ3-Cτ4 among three goldfish stains (strain G-12 lost the *IGHT* gene); this repeated block Cδ2-Cδ3-Cδ4 has previously been reported only in the *IGHD* genes (Sun et al., 2011; Edholm et al., 2011) in teleost fish, and the repeat numbers of Cτ3-Cτ4 in the *IGHT* locus or Cδ2-Cδ3-Cδ4 in the *IGHDA* locus also varied among the four goldfish strains.

CaTCV-1 had two *IGHA* loci. The numbers of D, J, and C genes were nearly twice that of G-12, which had only one *IGHA* locus, but the number of VH genes was lower. The G-12 VH sequences included 26.90% pseudogenes (46/171), similar to the proportion of pseudogenes in the zebrafish VH sequences (23.40%; 11/47) (Danilova et al., 2005), but much lower than the proportion of VH pseudogenes identified in Atlantic salmon (>68%; Yasuike et al., 2010). The total number of CaTCV-1 VH sequences was less than in G-12, but the proportion of pseudogenes (41.48%; 56/135) was much higher than in G-12. The percentage of VH sequences in each subgroup did not differ significantly between G-12 and CaTCV-1 (Fig. 4), but a small number of gene subgroups were permanently lost from a line. For example, subgroup 07 was lost in the CaTCV-1 strain. Therefore, it can be speculated that differences in the number of genes in *IGH* loci in related species will affect the stress resistance of the species, which is information important for carrying out breeding and selection.

The N-linked glycosylation sites of IgMA, IgMB, IgT2D, IgT4D, and IgT in RCC were also predicted. The results showed that there were some differences in the number and location between IgMA and IgMB or among τ-like Igs in N-linked glycosylation sites. Previous studies suggested that N-linked glycosylation affects antibody stability (Su et al., 2018; Zhou and Qiu, 2019); the differences in presence and numbers of N-linked glycosylation sites in subclasses of Igs may facilitate different effector functions. This possibility requires further investigation.

5. Conclusions

This study provides a comprehensive introduction to goldfish *IGH* gene isotypes. The configuration of the parallel *IGHA* and *IGHB* loci in GRCC and three other ancient allotetraploid goldfish are presented. Our research results provide frameworks for studies of the evolution of *IGH* genes in ancient allotetraploid fish, examinations of the immune response modes of different *IGH* genes, and comparisons of the composition and function of subclasses of *IGH* genes between related species.

Author contributions

Shaojun Liu, Jing Wang, and Linmei Han designed the study. Linmei Han, Jihong Li and Wen Wang provided the preliminary data that supported this study and wrote the manuscript. Linmei Han and Jihong Li participated in Southern blot. Li Ren and Linmei Han performed the bioinformatics analysis. Wen Wang and Mingli Chai participated in *IGH* genes cloning and sequence alignment. Caixia Xiang and Ziye Luo participated in RT-qPCR. Qianhong Gu participated in phylogenetic analysis and reconstructed the phylogenetic tree. Kaikun Luo, Min Tao, Chun Zhang provided expert comments. All authors have read and approved the final manuscript.

Compliance and ethics

Fish treatments were carried out according to the regulations for protected wildlife and the Administration of Affairs Concerning Animal Experimentation and were approved by the Science and Technology Bureau of China and the Department of Wildlife Administration.

Funding

This research was funded by the National Natural Science Foundation of China (Grant Nos. 31872549 and 31730098), the National Key R&D Program of China (2018YFD0900200), Laboratory of Lingnan Modern Agriculture Project (Grant No. NT2021008), China Agriculture Research System of MOF and MARA (Grant No. CARS-45), 111 Project (D20007) and the Key Research and Development Program of Hunan Province (Grant No. 2020NK2016).

Acknowledgements

We are grateful to Dr. Kaikun Luo for providing us with the experimental fish. We thank LetPub (www.letpub.com) for linguistic assistance and pre-submission expert review.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dci.2022.104476.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215 (3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Bao, Y., Wang, T., Guo, Y., Zhao, Z., Li, N., Zhao, Y., 2010. The immunoglobulin gene loci in the teleost Gasterosteus aculeatus. Fish Shellfish Immunol. 28 (1), 40–48. https://doi.org/10.1016/j.fsi.2009.09.014.

Bengtén, E., Quiniou, S., Hikima, J., Waldbieser, G., Warr, G.W., Miller, N.W., Wilson, M., 2006. Structure of the catfish IGH locus: analysis of the region including the single functional IGHM gene. Immunogenetics 58 (10), 831–844. https://doi.org/10.1007/s00251-006-0139-9.

Bengtén, E., Wilson, M., 2015. Antibody repertoires in fish. Results. Probl. Cell. Differ. 57, 193–234. https://doi.org/10.1007/978-3-319-20819-0_9.

Bilal, S., Etayo, A., Hordvik, I., 2021. Immunoglobulins in teleosts. Immunogenetics 73 (1), 65–77. https://doi.org/10.1007/s00251-020-01195-1.

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

Bradshaw, W.J., Valenzano, D.R., 2020. Extreme genomic volatility characterizes the evolution of the immunoglobulin heavy chain locus in cyprinodontiform fishes. Proc. Biol. Sci. 287, 20200489 https://doi.org/10.1098/rspb.2020.0489, 1927.

Buonocore, F., Scapigliati, G., Pallavicini, A., Gerdol, M., 2020. Identification of an IgD/IgT chimera in the European sea bass (Dicentrarchus labrax L.). Fish Shellfish Immunol. 105, 224–232. https://doi.org/10.1016/j.fsi.2020.07.041.

Chen, S.C., 1956. A history of the domestication and the factors of the varietal formation of the common goldfish, Carassius auratus. Sci. Sin. 5 (2), 287–321.

Chen, Z., Omori, Y., Koren, S., Shirokiya, T., Kuroda, T., Miyamoto, A., Wada, H., Fujiyama, A., Toyoda, A., Zhang, S., Wolfsberg, T.G., Kawakami, K., Phillippy, A.M., , NISC Comparative Sequencing Program, Mullikin, J.C., Burgess, S.M., 2019. De novo assembly of the goldfish (Carassius auratus) genome and the evolution of genes after whole-genome duplication. Sci. Adv. 5 (6), eaav0547 https://doi.org/10.1126/sciadv.aav0547.

Danilova, N., Bussmann, J., Jekosch, K., Steiner, L.A., 2005. The immunoglobulin heavy-chain locus in zebrafish: identification and expression of a previously unknown isotype, immunoglobulin Z. Nat. Immunol. 6 (3), 295–302. https://doi.org/10.1038/ni1166.

Duvaud, S., Gabella, C., Lisacek, F., Stockinger, H., Ioannidis, V., Durinx, C., 2021. Expasy, the Swiss bioinformatics resource portal, as designed by its users. Nucleic Acids Res. 49 (W1), W216–W227. https://doi.org/10.1093/nar/gkab225.

Edholm, E.S., Bengtén, E., Wilson, M., 2011. Insights into the function of IgD. Dev. Comp. Immunol. 35 (12), 1309–1316. https://doi.org/10.1016/j.dci.2011.03.002.

Fillatreau, S., Six, A., Magadan, S., Castro, R., Sunyer, J.O., Boudinot, P., 2013. The astonishing diversity of Ig classes and B cell repertoires in teleost fish. Front. Immunol. 4 (28), 28. https://doi.org/10.3389/fimmu.2013.00028.

Fu, X., Zhang, H., Tan, E., Watabe, S., Asakawa, S., 2015. Characterization of the torafugu (Takifugu rubripes) immunoglobulin heavy chain gene locus. Immunogenetics 67 (3), 179–193. https://doi.org/10.1007/s00251-014-0824-z.

Gambón-Deza, Sánchez-Espinel, C., Magadán-Mompó, S., 2010. Presence of an unique IgT on the IGH locus in three-spined stickleback fish (Gasterosteus aculeatus) and the very recent generation of a repertoire of VH genes. Dev. Comp. Immunol. 34 (2), 114–122. https://doi.org/10.1016/j.dci.2009.08.011.

Giacomelli, S., Buonocore, F., Albanese, F., Scapigliati, G., Gerdol, M., Oreste, U., Coscia, M.R., 2015. New insights into evolution of IgT genes coming from Antarctic teleosts. Mar. Genomics. 24 (1), 55–68. https://doi.org/10.1016/j.margen.2015.06.009.

Glasauer, S.M., Neuhauss, S.C., 2014. Whole-genome duplication in teleost fishes and its evolutionary consequences. Mol. Genet. Genom. 289 (6), 1045–1060. https://doi.org/10.1007/s00438-014-0889-2.

Gupta, R., Brunak, S., 2002. Prediction of glycosylation across the human proteome and the correlation to protein function. Pac. Symp. Biocomput. 7, 310–322.

Huelsenbeck, J.P., Ronquist, F., 2001. Mrbayes: bayesian inference of phylogenetic trees. Bioinformatics 17 (8), 754–755. https://doi.org/10.1093/bioinformatics/17.8.754.

Ji, J.F., Hu, C.B., Shao, T., Fan, D.D., Zhang, N., Lin, A.F., Xiang, L.X., Shao, J.Z., 2021. Differential immune responses of immunoglobulin Z subclass members in antibacterial immunity in a zebrafish model. Immunology 162 (1), 105–120. https://doi.org/10.1111/imm.13269.

Kabat, E.A., 1991. Sequences of Proteins of Immunological Interest, fifth ed. U.S. Dept. of Health and Human Services, Public Health Service, National Institutes of Health, Bethesda, MD.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., Haeseler, A.V., Jermiin, L.S., 2017. Modelfinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14 (6), 587–589. https://doi.org/10.1038/nmeth.4285.

Lefranc, M.P., 2011. From IMGT-ONTOLOGY CLASSIFICATION Axiom to IMGT standardized gene and allele nomenclature: for immunoglobulins (IG) and T cell receptors (TR). Cold Spring Harb. Protoc. 2011 (6), 627–632. https://doi.org/10.1101/pdb.ip84.

Li, J.T., Wang, Q., Huang, Y.M.D., Li, Q.S., Cui, M.S., Dong, Z.J., Wang, H.W., Yu, J.H., Zhao, Y.J., Yang, C.R., Wang, Y.X., Sun, X.Q., Zhang, Y., Zhao, R., Jia, Z.Y., Wang, X.Y., 2021. Parallel subgenome structure and divergent expression evolution of allotetraploid common carp and goldfish. Nat. Genet. 53 (10), 1493–1503. https://doi.org/10.1038/s41588-021-00933-9.

Liu, W., Xie, Y., Ma, J., Luo, X., Nie, P., Zuo, Z., Lahrmann, U., Zhao, Q., Zheng, Y., Zhao, Y., Xue, Y., Ren, J., 2015. IBS: an illustrator for the presentation and visualization of biological sequences. Bioinformatics 31 (20), 3359–3361. https://doi.org/10.1093/bioinformatics/btv362.

Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the$2^{-\triangle\triangle Ct}$ method. Methods 25 (4), 402–408. https://doi.org/10.1006/meth.2001.1262.

Lucas, J.S., Murre, C., Feeney, A.J., Riblet, R., 2015. Chapter 1 - the structure and regulation of the immunoglobulin loci. In: Alt, F.W., Honjo, T., Radbruch, A., Reth, M. (Eds.), Molecular Biology of B Cells, second ed. Academic Press., London, pp. 1–11.

Luo, J., Chai, J., Wen, Y., Tao, M., Lin, G., Liu, X., Ren, L., Chen, Z., Wu, S., Li, S., Wang, Y., Qin, Q., Wang, S., Gao, Y., Huang, F., Wang, L., Ai, C., Wang, X., Li, L., Ye, C., Yang, H., Luo, M., Chen, J., Hu, H., Yuan, L., Zhong, L., Wang, J., Xu, J., Du, Z., Ma, Z.S., Murphy, R.W., Meyer, A., Gui, J., Xu, P., Ruan, J., Chen, Z.J., Liu, S., Lu, X., Zhang, Y.P., 2020. From asymmetrical to balanced genomic diversification during rediploidization: subgenomic evolution in allotetraploid fish. Sci. Adv. 6 (22), eaaz7677.

Luo, J., Gao, Y., Ma, W., Bi, X.Y., Wang, S.Y., Wang, J., Wang, Y.Q., Chai, J., Du, R., Wu, S.F., Meyer, A., Zan, R.G., Xiao, H., Murphy, R.W., Zhang, Y.P., 2014. Tempo and mode of recurrent polyploidization in the Carassius auratus species complex (Cypriniformes, Cyprinidae). Heredity 112 (4), 415–427. https://doi.org/10.1038/hdy.2013.121.

Ma, L., Blackshields, G., Np, B., Chenna, R., Pa, M.G., Mcwilliam, H., Valentin, F., Im, W., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. ClustalW and ClustalX version 2.0. Bioinformatics 23 (21), 2947–2948.

Ma, W., Zhu, Z.H., Bi, X.Y., Murphy, R.W., Wang, S.Y., Gao, Y., Xiao, H., Zhang, Y.P., Luo, J., 2014. Allopolyploidization is not so simple: evidence from the origin of the tribe cyprinini (teleostei: cypriniformes). Curr. Mol. Med. 14 (14), 1331–1338. https://doi.org/10.2174/1566524014666141203101543.

Magadán-Mompó, S., Sánchez-Espinel, C., Gambón-Deza, F., 2011. Immunoglobulin heavy chains in medaka (Oryzias latipes). BMC Evol. Biol. 11 (1), 165. https://doi.org/10.1186/1471-2148-11-165, 165.

Magadan, S., Krasnov, A., Hadi-Saljoqi, S., Afanasyev, S., Mondot, S., Lallias, D., Castro, R., Salinas, I., Sunyer, O., Hansen, J., Koop, B.F., Lefranc, M.P., Boudinot, P., 2019. Standardized IMGT nomenclature of salmonidae IGH genes, the paradigm of atlantic salmon and rainbow trout: from genomics to repertoires. Front. Immunol. 10, 2541. https://doi.org/10.3389/fimmu.2019.02541.

Mamiatis, T., Fritsch, E.F., Sambrook, J., Engel, J., 1985. Molecular Cloning–A Laboratory Manual, first ed. Cold Spring Harbor Laboratory, New York.

Mirete-Bachiller, S., Olivieri, D.N., Gambon-Deza, F., 2021. Immunoglobulin T genes in actinopterygii. Fish Shellfish Immunol. 108, 86–93. https://doi.org/10.1016/j.fsi.2020.11.027.

Nakao, M., Moritomo, T., Tomana, M., Fujiki, K., Yano, T., 1998. Isolation of cDNA encoding the constant region of the immuoglobulin heavy-chain from common carp (Cyprinus carpioL.). Fish Shellfish Immunol. 8 (6), 425–434.

Ohta, Y., Flajnik, M., 2006. IgD, like IgM, is a primordial immunoglobulin class perpetuated in most jawed vertebrates. Proc. Natl. Acad. Sci. U.S.A. 103 (28) https://doi.org/10.1073/pnas.0601407103.

Reese, M.G., Eeckman, F.H., Kulp, D., Haussler, D., 1997. Improved splice site detection in genie. J. Comp. Biol. 4 (3), 311–323. https://doi.org/10.1089/cmb.1997.4.311.

Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: the European molecular biology open software suite. Trends Genet. 16 (6), 276–277. https://doi.org/10.1016/s0168-9525(00)02024-2.

Ryo, S., Wijdeven, R.H.M., Tyagi, A., Hermsen, T., Kono, T., Karunasagar, I., Rombout, J.H.W.M., Sakai, M., Verburg-van Kemenade, B.M.L., Savan, R., 2010. Common carp have two subclasses of bonyfish specific antibody IgZ showing differential expression in response to infection. Dev. Comp. Immunol. 34 (11), 1183–1190. https://doi.org/10.1016/j.dci.2010.06.012.

Salamov, A.A., Solovyev, V.V., 2000. Ab initio gene finding in Drosophila genomic DNA. Genome Res. 10 (4), 516–522. https://doi.org/10.1101/gr.10.4.516.

Savan, R., Aman, A., Nakao, M., Watanuki, H., Sakai, M., 2005a. Discovery of a novel immunoglobulin heavy chain gene chimera from common carp (Cyprinus carpio L.). Immunogenetics 57 (6), 458–463. https://doi.org/10.1007/s00251-005-0015-z.

Savan, R., Aman, A., Sato, K., Yamaguchi, R., Sakai, M., 2005b. Discovery of a new class of immunoglobulin heavy chain from fugu. Eur. J. Immunol. 35 (11), 3320–3331. https://doi.org/10.1002/eji.200535248.

Solovyev, V., 2007. Statistical approaches in Eukaryotic gene prediction. In: Balding, D. J., Bishop, M., Cannings, C. (Eds.), Handbook of Statistical Genetics. Wiley Online Library, pp. 97–159. https://doi.org/10.1002/9780470061619.ch4.

Su, Y.L., Wang, B., Hu, M.D., Cui, Z.W., Wan, J., Bai, H., Yang, Q., Cui, Y.F., Wan, C.H., Xiong, L., Zhang, Y.A., Geng, H., 2018. Site-specific N-glycan characterization of grass carp serum IgM. Front. Immunol. 9, 2645. https://doi.org/10.3389/fimmu.2018.02645.

Sun, Y., Huang, T., Hammarström, L., Zhao, Y., 2020. The immunoglobulins: new insights, implications, and applications. Annu. Rev. Anim. Biosci. 8 (1), 145–169. https://doi.org/10.1146/annurev-animal-021419-083720.

Sun, Y., Wei, Z., Hammarstrom, L., Zhao, Y., 2011. The immunoglobulin δ gene in jawed vertebrates: a comparative overview. Dev. Comp. Immunol. 35 (9), 975–981. https://doi.org/10.1016/j.dci.2010.12.010.

Watson, C.T., Steinberg, K.M., Huddleston, J., Warren, R.L., Malig, M., Schein, J., Willsey, A.J., Joy, J.B., Scott, J.K., Graves, T.A., Wilson, R.K., Holt, R.A., Eichler, E. E., Breden, F., 2013. Complete haplotype sequence of the human immunoglobulin heavy-chain variable, diversity,and joining genes and characterization of allelicand copy-number variation. Am. J. Hum. Genet. 92 (4), 530–546. https://doi.org/10.1016/j.ajhg.2013.03.004.

Xiao, F.S., Wang, Y.P., Yan, W., Chang, M.X., Yao, W.J., Xu, Q.Q., Wang, X.X., Gao, Q., Nie, P., 2010. Ig heavy chain genes and their locus in grass carp Ctenopharyngodon idella. Fish Shellfish Immunol. 29 (4), 594–599.

Xu, P., Xu, J., Liu, G., Chen, L., Zhou, Z., Peng, W., Jiang, Y., Zhao, Z., Jia, Z., Sun, Y., Wu, Y., Chen, B., Pu, F., Feng, J., Luo, J., Chai, J., Zhang, H., Wang, H., Dong, C., Jiang, W., Sun, X., 2019. The allotetraploid origin and asymmetrical genome evolution of the common carp Cyprinus carpio. Nat. Commun. 10 (1), 4265. https://doi.org/10.1038/s41467-019-12644-1.

Xu, P., Zhang, X., Wang, X., Li, J., Liu, G., Kuang, Y., Xu, J., Zheng, X., Ren, L., Wang, G., Zhang, Y., Huo, L., Zhao, Z., Cao, D., Lu, C., Li, C., Zhou, Y., Liu, Z., Fan, Z., Shan, G., Li, X., Wu, S., Song, L., Hou, G., Jiang, Y., Jeney, Z., Yu, D., Wang, L., Shao, C., Song, L., Sun, J., Ji, P., Wang, J., Li, Q., Xu, L., Sun, F., Feng, J., Wang, C., Wang, S., Wang, B., Li, Y., Zhu, Y., Xue, W., Zhao, L., Wang, J., Gu, Y., Lv, W., Wu, K., Xiao, J., Wu, J., Zhang, Z., Yu, J., Sun, X., 2014. Genome sequence and genetic diversity of the common carp, Cyprinus carpio. Nat. Genet. 46 (11), 1212–1219. https://doi.org/10.1038/ng.3098.

Xu, J., Yu, Y.Y., Huang, Z.Y., Dong, S., Luo, Y.Z., Yu, W., Yin, Y.X., Li, H.L., Liu, Y.A., Zhou, X.Y., Xu, Z., 2018. Immunoglobulin (Ig) heavy chain gene locus and immune responses upon parasitic, bacterial and fungal infection in loach, Misgurnus anguillicaudatus. Fish Shellfish Immunol. 86, 1139–1150. https://doi.org/10.1016/j.fsi.2018.12.064.

Yasuike, M., de Boer, J., von Schalburg, K.R., Cooper, G.A., McKinnel, L., Messmer, A., So, S., Davidson, W.S., Koop, B.F., 2010. Evolution of duplicated IgH loci in Atlantic salmon, Salmo salar. BMC Genom. 11, 486. https://doi.org/10.1186/1471-2164-11-486.

Ye, J., Ma, N., Madden, T.L., Ostell, J.M., 2013. IgBLAST: an immunoglobulin variable domain sequence analysis tool. Nucleic Acids Res. 41, W34–W40. https://doi.org/10.1093/nar/gkt382.

Yu, X.J., Zhou, T., 1989. Chromosomes of Chinese Fresh-Water Fishes. Science Press.

Yu, Y.Y., Kong, W.G., Xu, H.Y., Huang, Z.Y., Zhang, X.T., Ding, L.G., Dong, S., Yin, G.M., Dong, F., Yu, W., Cao, J.F., Meng, K.F., Liu, X., Fu, Y., Zhang, X.Z., Zhang, Y.A., Sunyer, J.O., Xu, Z., 2019. Convergent evolution of mucosal immune responses at the buccal cavity of teleost fish. iScience 19, 821–835. https://doi.org/10.1016/j.isci.2019.08.034.

Zhang, D., Gao, F., Jakovlic, I., Zou, H., Zhang, J., Li, W.X., Wang, G.T., 2020. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol. Ecol. Resour 20 (1), 348–355. https://doi.org/10.1111/1755-0998.13096.

Zhang, F., Li, M., Lv, C., Wei, G., Wang, C., Wang, Y., An, L., Yang, G., 2021. Molecular characterization of A new IgZ3 subclass from common carp (Cyprinus carpio) and comparative expression analysis of IgH transcripts during ontogeny of larvae. BMC Vet. Res. 17 (1), 159. https://doi.org/10.1186/s12917-021-02844-7.

Zhou, Q., Qiu, H., 2019. The mechanistic impact of N-glycosylation on stability, pharmacokinetics, and immunogenicity of therapeutic proteins. J. Pharmacol. Sci. 108 (4), 1366–1377. https://doi.org/10.1016/j.xphs.2018.11.029.