

# Regulator of complement activation (RCA) group 2 gene cluster in zebrafish: identification, expression, and evolution

Jie Wu · Hongyan Li · Shicui Zhang

Received: 11 October 2011 / Revised: 18 December 2011 / Accepted: 1 January 2012 / Published online: 18 January 2012  
© Springer-Verlag 2012

**Abstract** The activation of the complement system is tightly regulated by a group of plasma and cell membrane-associated proteins for host cell protection. In humans, these regulatory protein genes are clustered in a region named the regulator of complement activation (RCA) gene locus and can be categorized into two groups. The group 1 gene cluster has been reported in zebrafish, but information regarding the RCA locus remains scarce in fish. Here we identified two closely linked RCA group 2 genes in zebrafish, *ZRC1* and *ZRC2*, which had all the features characteristic of known RCA group 2 genes. Both *ZRC1* and *ZRC2* were closely linked to the *PFKFB1* gene and located 17 Mkb downstream of the *PFKFB2* gene; in contrast, RCA group 2 genes are closely linked to *PFKFB2* in frogs, chickens, and humans. However, both the direction of the RCA group 2 genes relative to *PFKFB2* and the order of the RCA group 2 gene-encoded proteins in zebrafish were comparable to those in frogs, chickens, and humans. *ZRC1* and *ZRC2* shared 71.1% identity to each other, implicating that they might have originated by gene duplication after the split of the fish/mammalian common ancestor. Moreover, *ZRC1* and *ZRC2* encoded a membrane-associated protein and a soluble protein, respectively, and displayed different expression patterns, suggesting that functional divergence has already occurred. This is the first report

showing the presence of the RCA group 2 cluster as well as the membrane-associated complement regulatory protein in zebrafish, providing a better understanding of the origin and evolution of RCA proteins.

**Keywords** Complement · Regulator of complement activation (RCA) · Group 2 · Gene cluster · Evolution · Zebrafish

## Abbreviations

RCA	Regulator of complement activation
PFKFB1	6-Phosphofructo-2-kinase/fructose-2,6-biphosphatase 1
PFKFB2	6-Phosphofructo-2-kinase/fructose-2,6-biphosphatase 2
C4bp	C4b-binding protein
C4bp $\alpha$	$\alpha$ Chain of the C4b-binding protein
C4bp $\beta$	$\beta$ Chain of the C4b-binding protein
FH	Factor H
CR1	Complement receptor type 1
CR2	Complement receptor type 2
DAF	Decay-accelerating factor
MCP	Membrane cofactor protein
SCR	Short consensus repeats

**Electronic supplementary material** The online version of this article (doi:10.1007/s10142-012-0262-7) contains supplementary material, which is available to authorized users.

J. Wu · H. Li (✉) · S. Zhang (✉)  
Laboratory for Evolution & Development,  
Institute of Evolution & Marine Biodiversity,  
Qingdao 266003, China  
e-mail: hongyanli@ouc.edu.cn  
e-mail: sczhang@ouc.edu.cn

J. Wu · H. Li · S. Zhang  
Department of Marine Biology, Ocean University of China,  
Qingdao 266003, China

## Introduction

The complement system is a central component of innate immunity and a link between innate and adaptive immune responses, which consists of both effectors for foreign cell clearance and regulators for host cell protection (Boackle 2003). The complement system has more than 20 plasma proteins and more than 10 cell membrane-associated proteins that are both involved in the activation and regulation of the

complement (Sarma and Ward 2011). There are three pathways by which the complement system can be activated: the classical, alternative, and lectin pathways (Boshra et al. 2006; Gasque 2004). These three pathways merge at a common amplification step involving the formation of unstable protease complexes, named C3 convertases (C3bBb in alternative pathway and C4b2a in classical/lectin pathways), and the cleavage of C3 to generate C3b and proceed through a terminal pathway that initiates the assembly of a membrane attack complex, which leads to complement-mediated lysis (de Córdoba and de Jorge 2008).

To prevent autologous tissue damage and to avoid wasteful consumption of components, the activation of the complement system is tightly controlled by a group of soluble and membrane-associated regulatory proteins, which primarily regulate the formation and stability of C3 convertases (Ciulla et al. 2005; de Córdoba et al. 2004). In humans, the soluble regulators comprise C4b-binding protein (C4bp), factor H (FH), and low molecular weight FH-like proteins, which control C3 activation in the fluid phase or on the surface of activating particles (Krushkal et al. 2000). Among them, C4bp is composed of seven identical  $\alpha$  chains and one  $\beta$  chain, which are encoded by *C4bp $\alpha$*  and *C4bp $\beta$*  genes, respectively. The complement regulatory function of C4bp is executed by the  $\alpha$  chain (C4bp $\alpha$ ) whereas the  $\beta$  chain of C4bp (C4bp $\beta$ ) is involved in the regulation of the blood coagulation process (Blom et al. 2001; Pardo-Manuel et al. 1990). The membrane-associated regulators include complement receptor type 1 (CR1/CD35), complement receptor type 2 (CR2/CD21), decay-accelerating factor (DAF/CD55), and membrane cofactor protein (MCP/CD46), which act to prevent uncontrolled activation of C3 (Krushkal et al. 2000). These soluble and membrane-associated complement regulatory proteins share a common tandem repetitive unit of 60–70 amino acid residues, termed short consensus repeats (SCRs) or complement control protein modules (de Córdoba et al. 2004; Janatova et al. 1989). The SCR framework which is highly conserved typically includes four cysteines forming disulfide bonds, which hold the SCRs into a rigid trip-loop structure, resulting in similar three-dimensional structures (Kimura et al. 2004).

Genes for complement regulatory proteins in humans are clustered on the long arm of chromosome 1 (1q32), designated the regulator of complement activation (RCA) gene locus (Heine-Suñer et al. 1997). These RCA genes can be categorized into two groups: group 1 and group 2. Human *FH* and its related genes form RCA group 1, and *CR1*, *CR2*, *DAF*, *MCP*, and *C4bp $\alpha$*  fall into RCA group 2 (Krushkal et al. 2000; Kimura et al. 2004). The RCA group 2 gene locus is close to the 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 2 (*PFKFB2*) gene in humans (Heine-Suñer et al. 1997). The RCA locus has been identified in mammalian species such as humans and mice and non-mammalian vertebrates such as chickens and frogs. Similar but two split loci of the RCA are

found in the mouse (Kingsmore et al. 1989). The chicken has a single RCA locus encoding the regulatory proteins that correspond to the human (Oshiumi et al. 2005). Furthermore, three genes coding for SCR proteins, mapping downstream of the *PFKFB2*, are identified in the frog RCA locus (Oshiumi et al. 2009). Recently, a gene cluster including *FH* and its related genes have been reported in zebrafish (Sun et al. 2010), but the information regarding the RCA locus in fish remains rather limited. Moreover, the origin and evolution of RCA proteins are obscure. Therefore, this study was undertaken to explore the RCA group 2 genes in zebrafish, a widely used model organism. We identified two closely linked RCA group 2 genes, *ZRC1* and *ZRC2*, which encode a type I membrane protein and a soluble protein, respectively. This is the first report showing the presence of the RCA group 2 cluster as well as the membrane-associated complement regulatory protein in zebrafish, pushing the origin of the membrane-associated complement regulatory protein to the basal jawed vertebrate fish.

## Materials and methods

### Fish

Wild-type zebrafish (*Danio rerio*) were obtained from a local fish dealer and maintained in containers with well-aerated tap water at  $27 \pm 1^\circ\text{C}$ . The fishes were fed on live bloodworms and fish flakes (Tetramin, Germany) twice a day.

### RNA extraction and reverse transcriptase polymerase chain reaction

Different tissues including the fin, heart, liver, eye, muscle, gill, and brain were dissected out of zebrafish, ground in RNAiso plus (TaKaRa), and kept at  $-70^\circ\text{C}$  until use. Total RNAs were isolated from the frozen sample according to the manufacturer's instructions. After digestion with recombinant DNase I (RNase free) (TaKaRa) to eliminate the genomic contamination, cDNAs were synthesized with reverse transcription kit (TaKaRa) with oligo (dT) primer. The reaction was carried out at  $42^\circ\text{C}$  for 1 h and inactivated at  $75^\circ\text{C}$  for 15 min. The cDNA was stored at  $-20^\circ\text{C}$  till used.

### Cloning and sequencing of zebrafish RCA group 2 genes (*ZRCs*)

Using the human RCA group 2 amino acid sequences as the queries, we performed a Basic Local Alignment Search Tool (BLAST) search in both National Center for Biotechnology Information (NCBI) database (<http://www.ncbi.nlm.nih.gov/>) and Ensembl database (<http://www.ensembl.org>) and obtained two putative *ZRC* genes (ENSDART00000133372

and XM\_688820) in the zebrafish genome. Based on the sequences of the putative *ZRC* genes in the database, two pairs of specific primers (Table 1) were designed using Primer Premier program (version 5.0), and polymerase chain reaction (PCR) was performed to amplify the *ZRC* cDNAs using the RNAs isolated from adult *D. rerio* liver. The PCR protocol for *ZRC1* gene was as follows: an initial denaturation at 94°C for 5 min, followed by 35 cycles of 94°C for 30 s, 50°C for 30 s, and 72°C for 2 min and a final extension at 72°C for 7 min, and that for the *ZRC2* gene was an initial denaturation at 94°C for 5 min, followed by 35 cycles of 94°C for 30 s, 51°C for 30 s, and 72°C for 2 min and a final extension at 72°C for 7 min. The amplification products were cloned into the pGEM-T vector (TIANGEN) following the manufacturer's instruction and transformed into *Trans5α* bacteria (TRANSGEN). The DNA inserts were then sequenced to verify for authenticity.

#### Sequence, phylogenetic and genomic organization analyses

The cDNAs were analyzed for coding probability with the DNASTAR software package (version 5.0), and the protein domains were analyzed using the SMART program (<http://smart.embl-heidelberg.de/>). Signal predictions were conducted using SignalP 3.0 Server (<http://www.cbs.dtu.dk/services/SignalP/>), and the molecular masses (MW) and isoelectric points (pI) of the mature peptides were determined using ProtParam (<http://www.expasy.ch/tools/protparam.html>).

Homology searches in the GenBank database were carried out by BLAST server (<http://www.ncbi.nlm.nih.gov/BLAST/>) (Altschul et al. 1990), and multiple alignments of the protein sequences and homologies among SCR domains were generated using the Clustal W program (Chenna et al. 2003) within the MegAlign of the DNASTAR software package (version 5.0). Phylogenetic trees were constructed by MEGA (version 4.0) using Poisson Correction distance based on the neighbor-joining method (Kumar et al. 2004). The reliability of each node was estimated by bootstrapping with 1,000 replications (Felsenstein 1985).

The information of exon–intron organization of zebrafish RCA group 2 genes was obtained from NCBI database (<http://www.ncbi.nlm.nih.gov/>) and Ensembl database (<http://www.ensembl.org>).

Chromosomal locations of RCA group 2 loci of humans (*Homo sapiens*), chickens (*Gallus gallus*), frogs (*Xenopus tropicalis*), and zebrafish (*D. rerio*) were obtained from the Sequence Viewer (<http://www.ncbi.nlm.nih.gov/projects/sviewer>) and Ensembl Genome Browser (<http://www.ensembl.org>).

#### Quantitative real-time PCR

To detect the tissue expression patterns of *ZRC* genes in zebrafish, quantitative real-time PCR (qRT-PCR) was performed using the first-strand cDNAs as templates, which were reverse transcribed from total RNAs extracted from the different tissues (fin, heart, liver, eye, muscle, gill, and brain). The two pairs of the primers specific for *ZRC1* and *ZRC2* (Table 1) were designed using Primer Premier program (version 5.0). The reaction mixtures (20 μl final volume) consisted of 10 μl of SYBR® Premix Ex Taq™ (Tli RnaseH Plus), 0.4 μl ROX Reference Dye II, 0.5 μl of template, and 200 nM of each sense and antisense primer. The β-actin gene was chosen as the reference for internal standardization. All qRT-PCR experiments were conducted in triplicate. The amplification was performed on ABI 7500 real-time PCR system (Applied Biosystems) with the following conditions: at 95°C for 15 s, followed by 40 cycles of 95°C for 5 s, 60°C for 15 s, and 72°C for 35 s. The expression levels of *ZRC* genes relative to that of β-actin were calculated by the comparative Ct method ( $2^{-\Delta\Delta C_t}$ ) (Livak and Schmittgen 2001).

## Results

#### General characteristics of *ZRC1* and *ZRC2*

By searching the NCBI and Ensembl databases using human RCA group 2 protein sequences as the queries, two zebrafish cDNAs (ENSDART00000133372 and XM\_688820) coding for putative complement regulatory proteins were identified. Next, we performed RT-PCR to amplify their cDNAs. As expected, the complete open reading frame (ORF) sequences of the two genes were obtained (Fig. 1), suggesting that the genes, designated *ZRC1* and *ZRC2* respectively, are expressed

**Table 1** Primers used for cDNA cloning and qRT-PCR

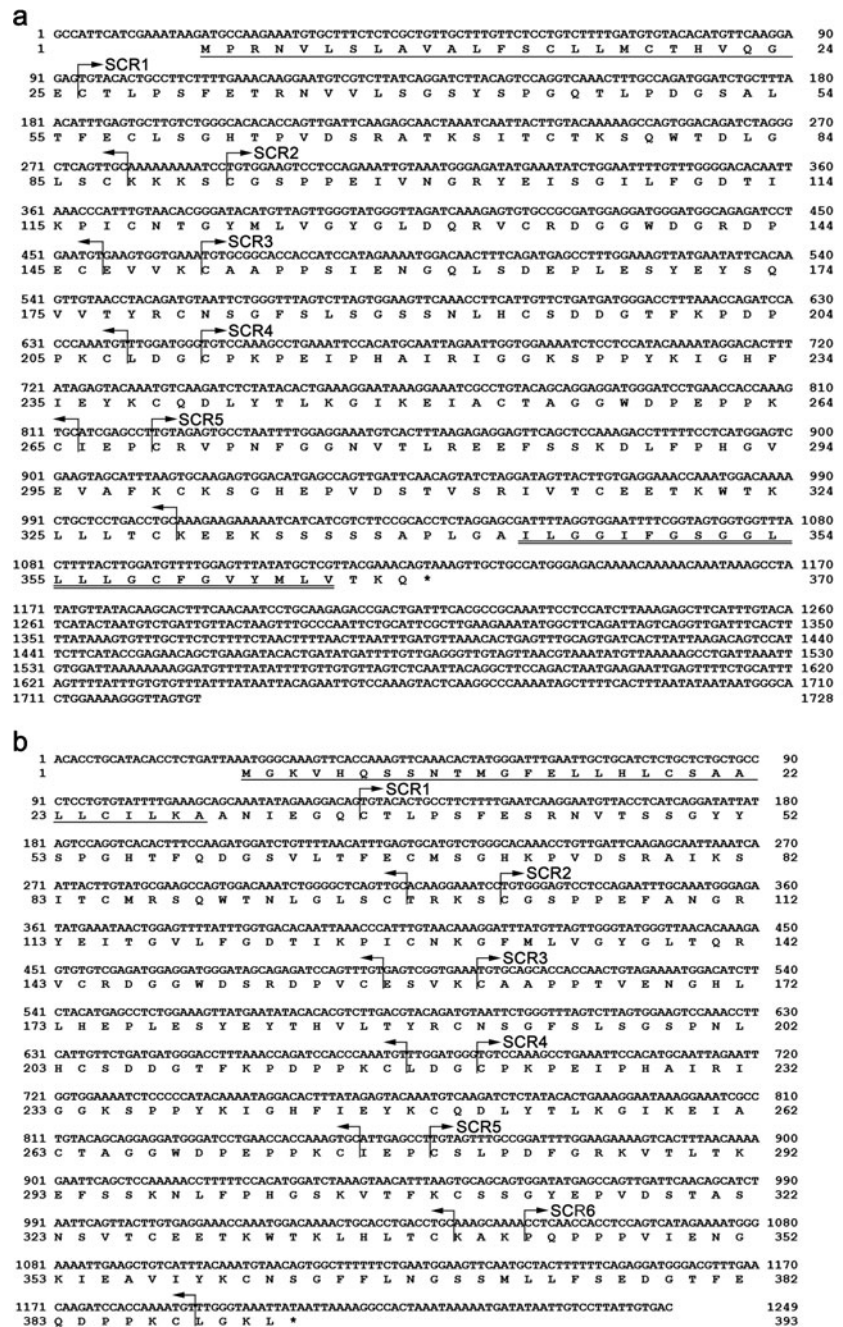
Gene	Accession number		cDNA cloning primers (5'–3')	qRT-PCR primers (5'–3')
<i>ZRC1</i>	ENSDART00000133372	Sense	TGCCATTCATCGAAATAAGATGCCA	CATTGTTCTGATGATGGGACCT
		Antisense	ACACTAACCCTTTTCCAGTGCCCAT	GGCGATTCCTTTATTCTTTTC
<i>ZRC2</i>	XM_688820	Sense	TACACCTGCATACACCTCTGATT	CATTGAGCCTGTAGTTTGCCG
		Antisense	AAGTCACAATAAGGACAATTATATCA	TGGAGGTGGTTGAGGTTTTGC
<i>β-Actin</i>	BC063950	Sense	–	GATGCGGAAACTGGAAAGGG
		Antisense	–	AGGAGGGCAAAGTGGTAAACC



in vivo. The two cDNA sequences have been deposited in Ensembl (ENSDART00000133372) and GenBank (XM\_688820), and each of them contained an initiation codon ATG at the 5' end and a stop codon at its 3' end. As shown in Fig. 1, the ORF of *ZRC1* consisted of 1,110 bp, encoding a protein of 369 amino acids with a calculated MW of ~39.79 kDa and a isoelectric point (pI) of 5.87. The *ZRC1* protein contained a hydrophobic signal peptide, five SCR domains, and a transmembrane region. The ORF of *ZRC2* was 1,179 bp in length, encoding a protein of 392 amino acids,

with a MW of ~42.84 kDa and a pI of 7.42. Like *ZRC1*, the *ZRC2* protein had a signal peptide and six SCR domains, but it had no transmembrane region (Fig. 2). Overall, *ZRC1* shared 71.1% identity to *ZRC2*, suggesting that *ZRC1* and *ZRC2* were highly similar to each other. Alignment analysis revealed that SCRs in *ZRC1* and *ZRC2* were each composed of about 60 aa including highly conserved four cysteines, which is typical of canonical SCRs. These analyses of the primary structures of the predicted proteins suggested that *ZRC1* was a type I membrane protein, while *ZRC2* was a soluble protein.

**Fig. 1** Full-length ORF sequences and deduced amino acid sequences of *ZRC1* (a) and *ZRC2* (b). Nucleotides are numbered in the 5' to 3' direction. The derived amino acid sequences are shown underneath the nucleotide sequences using single-letter codes. The stop codons are indicated by an *asterisk*. The putative signal peptides are *underlined*, and the transmembrane region is *double underlined*. The borders of SCR domains are shown by *arrows*. The nucleotide sequences have been deposited in the Ensembl and GenBank database with the accession number ENS-DART00000133372 (*ZRC1*) and XM\_688820 (*ZRC2*)



### Genomic structures of *ZRC1* and *ZRC2*

Analysis of exon–intron organization showed that zebrafish *ZRC1* consisted of 12 exons interspaced by 11 introns, and *ZRC2* consisted of nine exons plus eight introns (Fig. 2). The deduced proteins *ZRC1* and *ZRC2* both contained a split SCR, SCR2, which was encoded by two exons, and the exon–intron splice sites of the two split SCRs occurred after the second nucleotide (phase II intron) within the conserved Gly codon (Long and Deutsch 1999). Except for the SCR2, all the other SCRs of *ZRC1* and *ZRC2* were encoded by a single exon, and the exon–intron splice sites occurred after the first nucleotide of a codon (phase I intron). Furthermore, genomic structures of *ZRC1* and *ZRC2* both displayed the ag-gt consensus sequence for splicing (Table 2). Interestingly, the splitting feature of SCR2 and the ag-gt consensus sequence for splicing in *ZRC1* and *ZRC2* were also observed in frog, chicken, and human RCA group 2 genes (Oshiumi et al. 2005, 2009; Krushkal et al. 2000). These data indicated that *ZRC1* and *ZRC2* possessed all the features characteristic of frog, chicken, and human RCA group 2 proteins.

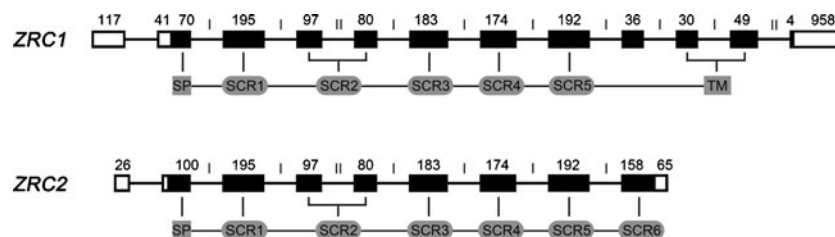
### Linkage and phylogeny of *ZRC1* and *ZRC2*

Genes in the RCA group 2 locus are closely linked to those of *PFKFB2* in humans, mice, chickens, and frogs (Kingsmore et al. 1989; Krushkal et al. 2000; Oshiumi et al. 2005, 2009). Analysis of the genes flanking *ZRCs* revealed that both *ZRC1* and *ZRC2* were syntenic and mapped onto zebrafish chromosome 23, which was closely linked to the *PFKFB1* gene, 17 Mkb downstream of the *PFKFB2*, a marker gene for the RCA loci in frogs, chickens, and humans (Fig. 3). This genomic organization was slightly different from the RCA group 2 loci in frogs, chickens, and humans, where the RCA group 2 loci were closely linked to the *PFKFB2* gene. However, both the direction of the RCA group 2 genes relative to *PFKFB2* and

the order of the RCA group 2 gene-encoded proteins in zebrafish were comparable to those in frogs, chickens, and humans. Both chicken and human RCA group 2 gene-encoded proteins were arranged in an order of PFKFB2-soluble protein (CRES in chickens/C4bp $\alpha$  in humans)-type II membrane protein (DAF in chickens and humans)-type I membrane protein (CREG, CR2, and CREM in chickens/CR2, CR1, and MCP in humans). In contrast, no gene encoding GPI-anchored (type II) membrane protein was identified in frog RCA group 2 locus, and thus, it had an order of PFKFB2-soluble protein (ARC1)-type I membrane protein (ARC2 and ARC3). Similarly, zebrafish RCA group 2 gene-encoded proteins were also arranged in the order of PFKFB2-soluble protein (*ZRC2*)-type I membrane protein (*ZRC1*). These data suggested that the RCA group 2 locus was closely linked to the *PFKFB1* gene (6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 1), located 17 Mkb downstream of the *PFKFB2* gene, in zebrafish.

As shown in Fig. 4, the phylogenetic tree constructed using *ZRC1* and *ZRC2*, zebrafish CFH, and its related proteins as well as all the human RCA proteins revealed that zebrafish CFH and its related proteins clubbed together with human CFH and CFHRs, forming the RCA group 1, consistent with the reports by Sun et al. (2010). In contrast, *ZRC1* and *ZRC2* were clustered with human RCA group 2 proteins, forming an independent group which was branched from the RCA group 1 proteins. These further supported that *ZRC1* and *ZRC2* were members of the RCA group 2 locus.

To determine the evolutionary position of the zebrafish RCA group 2 locus, we constructed a phylogenetic tree using the sequences of *ZRC1* and *ZRC2* as well as the RCA group 2 proteins available (Fig. 5). The phylogenetic tree showed that *ZRC1* and *ZRC2* clubbed together and formed an independent clade, branching from other RCA group 2 proteins. These suggested that *ZRC1* and *ZRC2* might have originated from a gene duplication event after teleosts split from the common ancestor of vertebrates.



**Fig. 2** Genomic structures of *ZRCs* and the primary structure of deduced proteins. Open rectangles and black rectangles represent non-coding exons and coding exons, respectively. The horizontal lines between two rectangles represent introns. The numbers above the rectangles represent the nucleotide number of each exon. The Roman numbers above the horizontal lines were the intron types. Gray figures indicate deduced protein structures (signal peptide region in gray rectangle with the character SP, transmembrane region in gray

rectangle with the character TM, SCR domain in gray oval). The corresponding relationship between exon and coding region was denoted by horizontal lines. The accession numbers of the analyzed sequences are ENSDART00000133372 (*ZRC1*) and XM\_688820 (*ZRC2*). The information of the exon–intron organization was obtained from NCBI database (<http://www.ncbi.nlm.nih.gov/>) and Ensembl database (<http://www.ensembl.org>) and modified based on the sequences cloned in this study

**Table 2** Information on exons and introns of *ZRC1* and *ZRC2*

Gene	Exon no. <sup>a</sup>	5' splice donor	3' splice donor	Intron no. <sup>b</sup>	Intron phase <sup>c</sup>	Amino acid interrupted
<i>ZRC1</i>	0	CTAAAGgtaacgtt	aattgcagAAATCA	0		
	1	TTCAAGgtgagtag	ctccacagGAGAGT	1	I	G24
	2	GCAAAAgtatgtac	gattacagAAAAAT	2	I	K89
	3	CACAGGgtgagtca	tcttcagATACAT	3	II	G121
	4	GTGAAGgtgagtta	ttttcagTGGTGA	4	I	V148
	5	GTTTGgtaagtgtg	tttgcagATGGGT	5	I	D209
	6	GCATTGgtaagaat	tttacagAGCCTT	6	I	E267
	7	GCAAAGgtaagtac	ttccacagAAGAAA	7	I	E331
	8	TAGGAGgtgagtca	tttgcagCGATTT	8	I	A343
	9	GTGGTGgtaagtct	tgtccagGTTTAC	9	I	G353
	10	GAAACAgtagtgtg	aattatagGTAAAG	10	II	Q369
<i>ZRC2</i>	0	CTCTGAgtaagtat	ctctgcagTTAAAT	0		
	1	TAGAAggtgagitt	ctccacagGACAGT	1	I	G34
	2	GCACAAgtatgtac	aactacagGGAAAT	2	I	R99
	3	CAAAGGgtgagtca	acttcagATTTAT	3	II	G131
	4	GTGAATgtgagtta	ctattcagCGGTGA	4	I	S158
	5	GTTTGgtaagtag	tttgcagATGGGT	5	I	D219
	6	GCATTGgtaagaat	ctttatagAGCCTT	6	I	E277
	7	GCAAAGgtaagttc	tgcttagCAAAAC	7	I	A342

<sup>a</sup>Numbered from 5'-terminus.

No. 0 represents non-coding exons, and other numbers represent coding exons

<sup>b</sup>Numbered from 5'-terminus.

No. 0 represents the introns among non-coding region, and other numbers represent the introns among the coding region

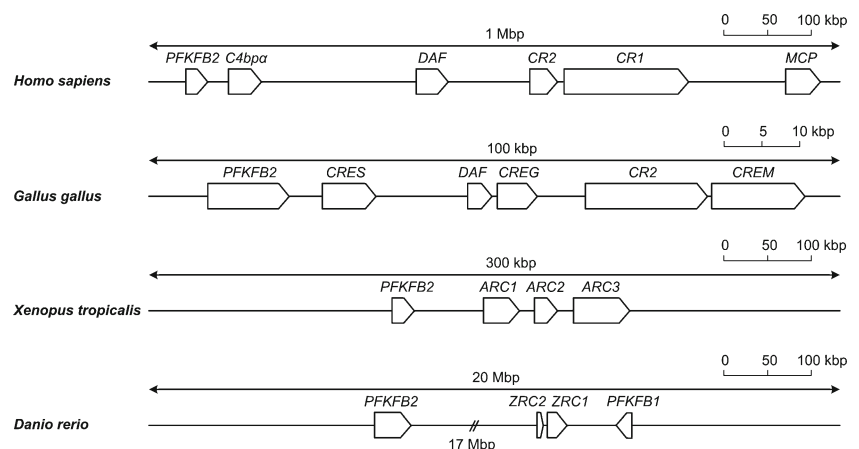
<sup>c</sup>Intron phase: I represents the exon–intron splice site located after the first nucleotide of a codon. II represents the exon–intron splice site located after the second nucleotide of a codon, and 0 represents the exon–intron splice site located between two codons

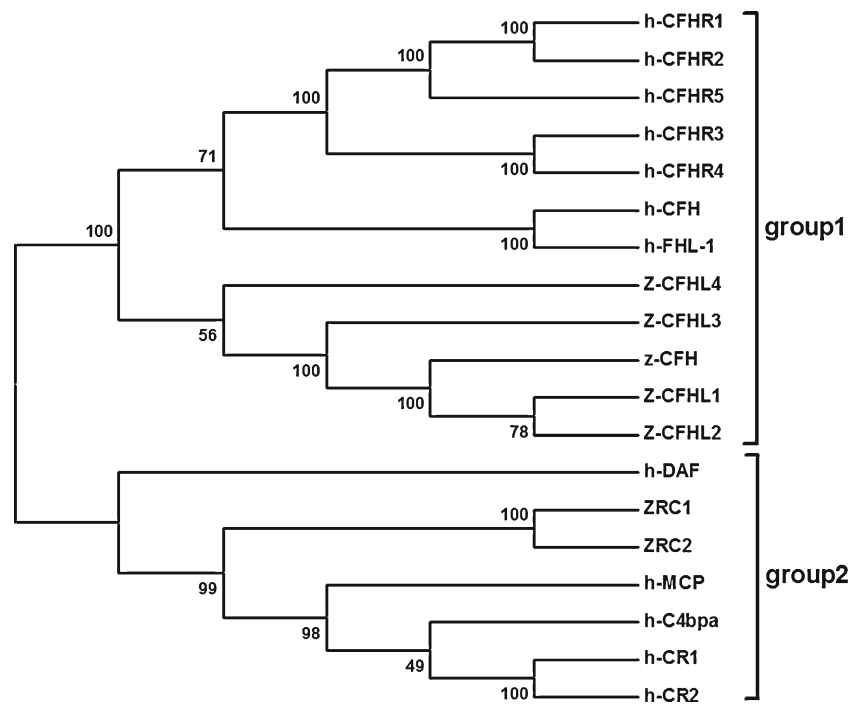
### Relationship among SCRs of *ZRC1* and *ZRC2*

To clarify the homology relationship of *ZRC1* and *ZRC2*, we compared the sequences of the individual SCRs of *ZRC1* and *ZRC2* with each other and then constructed a phylogenetic tree using the neighbor-joining method. It was found that SCR1-5 of *ZRC1* shared a high identity (>72%) to the corresponding SCRs in *ZRC2* (Fig. 6c), consistent with the notion that *ZRC1* and *ZRC2* were generated from an ancestral gene duplication. The phylogenetic tree showed that the SCRs of *ZRC1* and *ZRC2* were clustered into four distinct groups: SCR-A, B, C,

and D (Fig. 6a), resembling those of frog RCA group 2 proteins (Oshiumi et al. 2009). The phylogenetic analyses further demonstrated that zebrafish SCR-A, B, C, and D were clubbed with the corresponding types of SCRs in frog RCA group 2 proteins (Fig. 6b). As shown in Fig. 6c, the SCR domain combination in *ZRC1* was A-B-C-D-A, while that in *ZRC2* was A-B-C-D-A-C. Obviously, the SCR domain combination A-B-C-D existed in both *ZRC1* and *ZRC2*. Because the SCR domain combination A-B-C-D also existed in frog RCA group 2 proteins (Oshiumi et al. 2009), the common ancestral protein in fish and frog RCA group 2 loci might be

**Fig. 3** Comparison of zebrafish RCA group 2 locus with frog, chicken, and human RCA group 2 loci. Boxes represent genes. The information of chromosomal locations of RCA group 2 loci of humans (*H. sapiens*), chickens (*G. gallus*), frogs (*X. tropicalis*), and zebrafish (*D. rerio*) was obtained from the Sequence Viewer (<http://www.ncbi.nlm.nih.gov/projects/sviewer>) and Ensembl Genome Browser (<http://www.ensembl.org>)





**Fig. 4** Phylogenetic tree analysis of zebrafish ZRCs in comparison with zebrafish CFH and its related proteins as well as all the human RCA proteins. The phylogenetic tree was drawn by MEGA (version 4.0) using Poisson Correction distance based on the neighbor-joining method. The reliability of each node was estimated by bootstrapping with 1,000 replications. The *numbers* shown at each node indicate the bootstrap values (percent). All the sequences were clustered into two distinct clades and named group 1 and group 2. Accession numbers of each sequence are as follows: h-CFH (*H. sapiens*, NP\_000177), h-FHL-1 (*H. sapiens*,

NP\_001014975), h-CFHR1 (*H. sapiens*, NP\_002104), h-CFHR2 (*H. sapiens*, NP\_005657), h-CFHR3 (*H. sapiens*, NP\_066303), h-CFHR4 (*H. sapiens*, CAA66980), h-CFHR5 (*H. sapiens*, NP\_110414), h-CR1 (*H. sapiens*, NP\_000564), h-CR2 (*H. sapiens*, NP\_001868), h-MCP (*H. sapiens*, NP\_002380), h-DAF (*H. sapiens*, NP\_000565), h-C4bpa (*H. sapiens*, NP\_000706), z-CFH (*D. rerio*, ADK32771), z-CFHL1 (*D. rerio*, ADK32772), z-CFHL2 (*D. rerio*, ADK32773), z-CFHL3 (*D. rerio*, ADK32774), z-CFHL4 (*D. rerio*, ADK32775)

composed of SCR domain A-B-C-D. Figure 6c also showed that SCR1/SCR5 in both ZRC1 and ZRC2 belonged to the same type domain A, and SCR3/SCR6 in ZRC2 to the same type domain C, suggesting that the duplication of SCR domains might have occurred in ZRC1 and ZRC2.

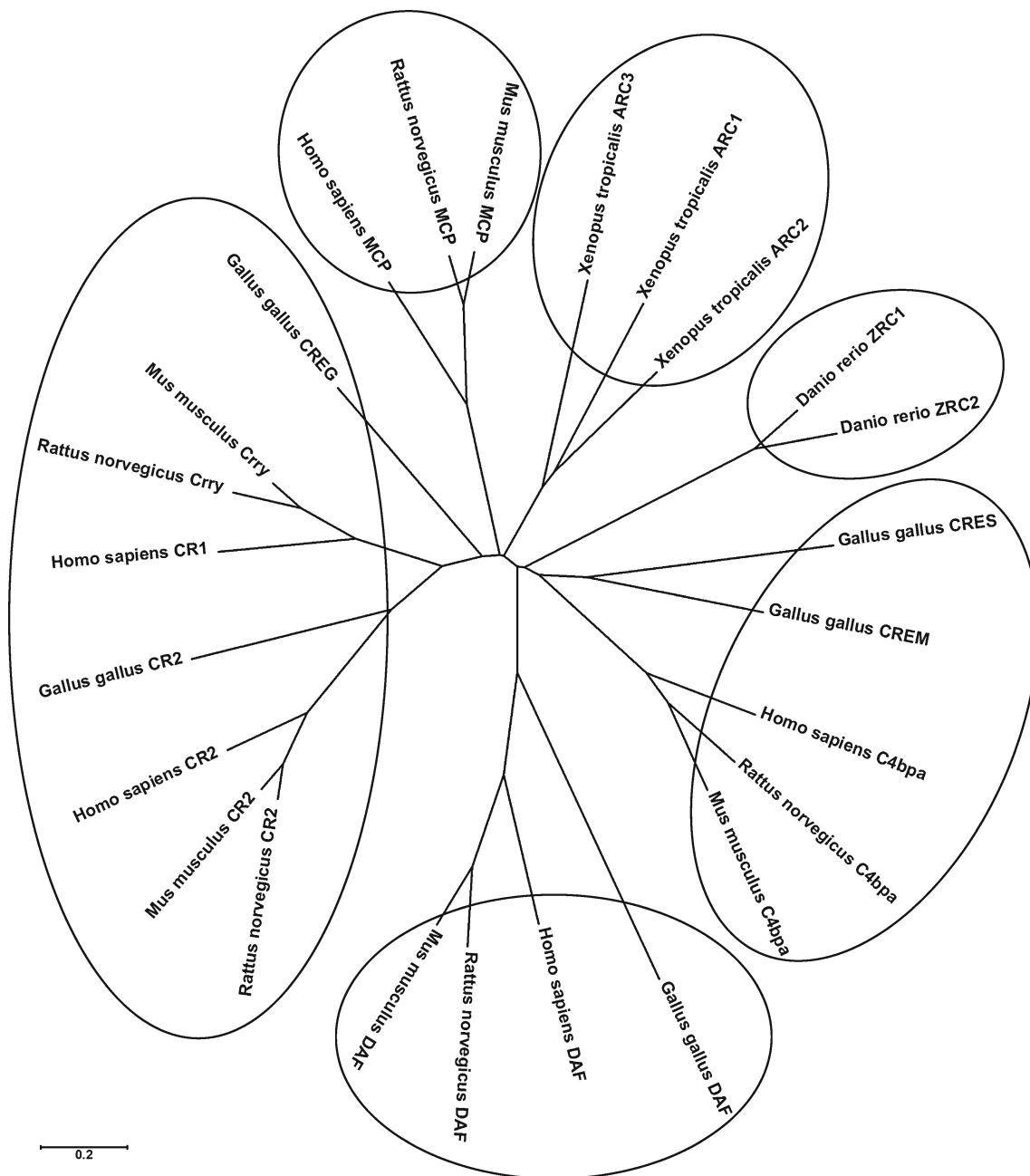
#### Expression profiles of *ZRC1* and *ZRC2*

qRT-PCR was used to examine the expression profiles of *ZRC1* and *ZRC2* genes in the different tissues including the fin, heart, liver, eye, muscle, gill, and brain. The dissociation curve of amplified products in all cases showed a single peak, indicating that the amplifications were specific (data not shown). In humans and frogs, *MCP*, *DAF*, and *ARC3*, which all encoded membrane-associated proteins, were ubiquitously expressed (Oshiumi et al. 2009). Similarly, *ZRC1* encoding a putative membrane protein was constitutively expressed in all the zebrafish tissues examined (Fig. 7a). In contrast, *ZRC2* encoding a soluble protein was primarily expressed in the liver and muscle (Fig. 7b), consistent with the tissue-specific expression pattern of *C4bpa* in humans (Moffat et al. 1992). These indicated that *ZRC1* and *ZRC2* displayed different expression patterns.

#### Discussion

A gene cluster including *FH* and its related genes has recently been identified in zebrafish (Sun et al. 2010), but information regarding the RCA locus remains scarce in fish. We demonstrate in this study that the zebrafish *D. rerio* possesses two SCR domain-containing genes, *ZRC1* and *ZRC2*, which have all the features characteristic of frog, chicken, and human RCA group 2 genes. Genes in the RCA group 2 locus are closely linked to the *PFKFB2* gene in humans, mice, chickens, and frogs (Kingsmore et al. 1989; Krushkal et al. 2000; Oshiumi et al. 2005, 2009). The linkage of *ZRC1* and *ZRC2* is slightly different from this conventional genomic organization. Analysis of the genes flanking *ZRCs* revealed that *ZRC1* and *ZRC2* were tightly linked to the *PFKFB1* gene, 17 Mkb downstream of the *PFKFB2*. However, the direction of the RCA group 2 genes relative to *PFKFB2* in zebrafish is comparable to that in frogs, chickens, and humans, and moreover, the RCA group 2 gene-encoded proteins are arranged in the order of PFKFB2-soluble protein (*ZRC2*)-type I membrane protein (*ZRC1*), which is essentially similar to that in frogs, chickens, and humans. Furthermore, the two genes *ZRC1* and *ZRC2* resemble each other, consistent with the high similarity of the three *ARC*





**Fig. 5** Phylogenetic tree based on the amino acid sequence comparisons of RCA group 2 proteins from various species. The tree was constructed by MEGA (version 4.0) using Poisson Correction distance based upon the neighbor-joining method with 1,000 bootstrap replicates. The bar represents the distance. The protein IDs are as follows: *H. sapiens* CR1(NP\_000564), *H. sapiens* CR2 (NP\_001868), *H. sapiens* MCP (NP\_002380), *H. sapiens* DAF (NP\_000565), *H. sapiens* C4bpa(NP\_000706), *Rattus norvegicus* Crry (NP\_001005330), *R. norvegicus* CR2 (NP\_001099459), *R. norvegicus* MCP (NP\_062063), *R.*

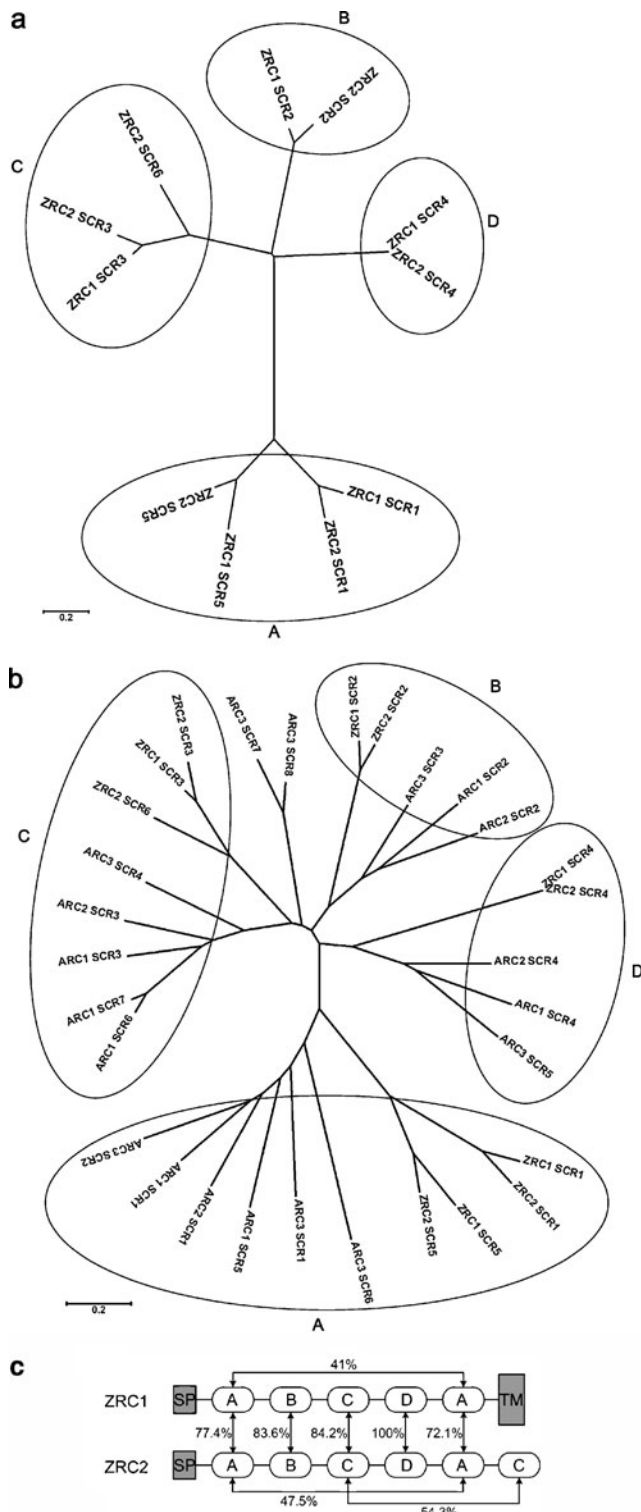
*norvegicus* DAF (NP\_071605), *R. norvegicus* C4bpa(NP\_036648), *Mus musculus* Crry (NP\_038527), *M. musculus* CR2 (NP\_031784), *M. musculus* MCP (NP\_034908), *M. musculus* DAF (NP\_034146), *M. musculus* C4bpa(BAC75871), *G. gallus* CREG (NP\_001028815), *G. gallus* CR2 (XP\_001233602), *G. gallus* CRES (NP\_001028814), *G. gallus* CREM (NP\_989995), *G. gallus* DAF (XP\_417981), *X. tropicalis* ARC1 (BAH58778), *X. tropicalis* ARC2 (NP\_001156329), *X. tropicalis* ARC3 (BAH58780)

genes in frogs (Oshiumi et al. 2009). As with frog *ARCs*, *ZRC1* and *ZRC2* might have originated from an ancestral gene duplication after teleosts split from the common ancestor of vertebrates. Possibly, the linkage of the RCA group 2 locus to *PFKFB2* occurred after the evolutionary split of the zebrafish

lineage, or an insertion may have occurred within these genes during the evolution of zebrafish.

Previous studies have shown the presence of membrane-associated forms of RCA proteins in mammals and chickens (Morgan and Harris 1999; Oshiumi et al. 2005). Recently, a

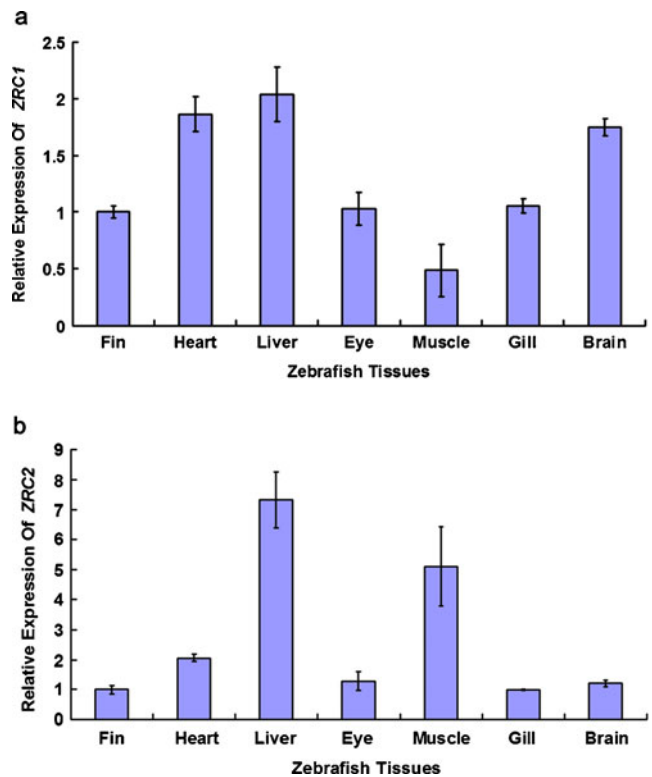




**Fig. 6** Homology relationship analysis for ZRCs based on individual SCR comparison. The phylogenetic tree was constructed by MEGA (version 4.0) using Poisson Correction distance based upon the neighbor-joining method with 1,000 bootstrap replicates. The *bar* represents the distance. **a** Phylogenetic tree analysis for the SCRs of ZRCs. Four clades were formed and termed A, B, C, and D. **b** The phylogenetic tree was constructed based on the alignment among the SCRs of ARCs (frog RCA group 2 proteins) and ZRCs. **c** The amino acid sequence identity among the SCRs of ZRCs. *Gray rectangles* with the characters *SP* and *TM* represent signal peptide regions and transmembrane regions, respectively. The *open ovals* indicate the SCR domains, and the character A, B, C, or D in *open ovals* represent the SCR clade type generated from the phylogenetic tree analysis for the SCRs of ZRCs and ARCs (**a**, **b**). The amino acid sequence identity was calculated using the DNASTAR software package with Clustal W method

protein gene identified in fish, suggesting that the membrane form of RCA proteins already emerged in the basal jawed vertebrate fish. However, this is still hypothetical at this point and needs to be tested by functional experiments.

SCR2 in both ZRC1 and ZRC2 were encoded by two exons (“split” SCR), similar to that of frog, chicken, and human RCA group 2 proteins (Oshiumi et al. 2005, 2009; Krushkal et al. 2000). In human RCA proteins, SCR2 plays a central role in C3b inactivation (Casasnovas et al. 1999; Liszewski et al. 2000). It has been known that all the functional RCA proteins such as human C4bp $\alpha$ , DAF, and CR1



**Fig. 7** Tissue expression profile analysis of zebrafish *ZRC1* (**a**) and *ZRC2* (**b**) gene by qRT-PCR. The tissues include fin, heart, liver, eye, muscle, gill, and brain. The  $\beta$ -actin gene was chosen as internal control for normalization. Relative expression data were calculated by the method of  $2^{-\Delta\Delta Ct}$

membrane-associated form of RCA proteins (ARC2 and ARC3) has also been identified in amphibians (Oshiumi et al. 2009). However, no SCR-containing membrane proteins have been reported so far in fish. Structural analysis clearly showed that *ZRC1* coded for a protein containing a transmembrane region. This is the first membrane-associated RCA

have one or more “split” SCRs (Krushkal et al. 2000). The split SCR2 in ZRCs shares a high degree of identity (up to 37%) to the corresponding SCRs of human C4bp $\alpha$ , DAF, and CR1 (Online resource 1). It is therefore highly likely that ZRCs may be functional, capable of binding to C3b. These indicate that the primary composition of SCR2 is conserved across humans and fishes.

In humans, RCA group 2 proteins, MCP and DAF, are ubiquitously expressed membrane proteins capable of protecting host cells from complement-mediated cell damage (Atkinson 1996), while RCA group 2 proteins, CR1 and CR2, confer cell type-specific expression and specified functions (Ahearn and Fearon 1989). Functional divergence has also been suggested for RCA group 2 genes, *ARC1* and *ARC3*, with ubiquitous expression pattern and RCA group 2 gene, *ARC2*, with tissue-specific expression pattern (Oshiumi et al. 2009). Here we show that *ZRC1* was ubiquitously expressed, while *ZRC2* was primarily expressed in the liver and muscle. The different expression patterns and different expression levels of *ZRC1* and *ZRC2* possibly implicate the occurrence of functional divergence in fish RCA genes, which deserves further study.

The origin and evolution of RCA proteins are still obscure. Our study adds more information to the existing knowledge of the origin of the RCA gene cluster. Oshiumi et al. (2009) proposed that the prototype of the RCA locus contains SCR-domain combination A-B-C-D, as exemplified by *X. tropicalis*. We demonstrated that the same combination A-B-C-D was also present in *D. rerio*, pushing this prototype to fish. Of note, the order of PFKFB2-soluble protein-type II membrane protein-type I membrane protein is essentially the same in humans and chickens. No gene encoding GPI-anchored (type II) membrane protein was identified in frog and zebrafish RCA group 2 locus. Besides, both frog and zebrafish had the same protein order of PFKFB2-soluble protein-type I membrane protein. These hint at the clue that the gene encoding GPI-anchored membrane protein may be a novel addition to the RCA group 2 locus in terrestrial vertebrates.

In summary, this study identifies two SCR domain-containing genes, *ZRC1* and *ZRC2*, which are closely linked to the *PFKFB1* gene and located 17 Mkb downstream of the *PFKFB2* gene, in zebrafish. They might have originated by gene duplication after the split of the fish/mammalian common ancestor but have already displayed functional divergence. It also highlights that the RCA group 2 locus in terrestrial vertebrates might have evolved from the RCA locus prototype of SCR-domain combination A-B-C-D, with an insertion of the gene encoding GPI-anchored membrane protein.

**Acknowledgments** This work was supported by a grant (2012CB114404) of the Ministry of Science and Technology (MOST) of China to SCZ. The authors also thank Dr. Yuan Wang for technical aids.

**Ethical standards** The experiments were carried out in accordance with the guidelines of the Laboratory Animal Administration Law of China, with permit number SD2007695 approved by the Ethics Committee of the Laboratory Animal Administration of Shandong Province.

**Conflict of interest** The authors declare no conflict of interest.

## References

- Ahearn JM, Fearon DT (1989) Structure and function of the complement receptors, CR1 (CD35) and CR2 (CD21). *Adv Immunol* 46:183–219. doi:10.1016/S0065-2776(08)60654-9
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Atkinson JP (1996) Impact of the discovery of membrane inhibitors of complement. *Res Immunol* 147:95–100. doi:10.1016/0923-2494(96)87180-X
- Blom AM, Kask L, Dahlbäck B (2001) Structural requirements for the complement regulatory activities of C4BP. *J Biol Chem* 276:27136–27144. doi:10.1074/jbc.M102445200
- Boackle SA (2003) Complement and autoimmunity. *Biomed Pharmacother* 57:269–273. doi:10.1016/S0753-3322(03)00084-2
- Boshra H, Li J, Sunyer JO (2006) Recent advances on the complement system of teleost fish. *Fish Shellfish Immunol* 20:239–262. doi:10.1016/j.fsi.2005.04.004
- Casasnovas JM, Larvie M, Stehle T (1999) Crystal structure of two CD46 domains reveals an extended measles virus-binding surface. *EMBO J* 18:2911–2922. doi:10.1093/emboj/18.11.2911
- Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 31:3497–3500. doi:10.1093/nar/gkg500
- Ciulla E, Emery A, Konz D, Krushkal J (2005) Evolutionary history of orthopoxvirus proteins similar to human complement regulators. *Gene* 355:40–47. doi:10.1016/j.gene.2005.05.008
- de Córdoba SR, de Jorge EG (2008) Translational mini-review series on complement factor H: genetics and disease associations of human complement factor H. *Clin Exp Immunol* 151:1–13. doi:10.1111/j.1365-2249.2007.03552.x
- de Córdoba SR, Esparza-Gordillo J, de Jorge EG, Lopez-Trascasa M, Sánchez-Corral P (2004) The human complement factor H: functional roles, genetic variations and disease associations. *Mol Immunol* 41:355–367. doi:10.1016/j.molimm.2004.02.005
- Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–791
- Gasque P (2004) Complement: a unique innate immune sensor for danger signals. *Mol Immunol* 41:1089–1098. doi:10.1016/j.molimm.2004.06.011
- Heine-Suñer D, Díaz-Guillén MA, de Villena FP, Robledo M, Benítez J, de Córdoba SR (1997) A high-resolution map of the regulator of the complement activation gene cluster on 1q32 that integrates new genes and markers. *Immunogenetics* 45:422–427. doi:10.1007/s002510050224
- Janatova J, Reid KB, Willis AC (1989) Disulfide bonds are localized within the short consensus repeat units of complement regulatory proteins: C4b-binding protein. *Biochemistry* 28:4754–4761. doi:10.1021/bi00437a036
- Kimura Y, Inoue N, Fukui A, Oshiumi H, Matsumoto M, Nonaka M, Kuratani S, Fujita T, Nonaka M, Seya T (2004) A short consensus repeat-containing complement regulatory protein of lamprey that participates in cleavage of lamprey complement 3. *J Immunol* 173:1118–1128

- Kingsmore SF, Vik DP, Kurtz CB, Leroy P, Tack BF, Weis JH, Seldin MF (1989) Genetic organization of complement receptor-related genes in the mouse. *J Exp Med* 169:1479–84
- Krushkal J, Bat O, Gigli I (2000) Evolutionary relationships among proteins encoded by the regulator of complement activation gene cluster. *Mol Biol Evol* 17:1718–1730
- Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* 5:150–163. doi:10.1093/bib/5.2.150
- Liszewski MK, Leung M, Cui W, Subramanian VB, Parkinson J, Barlow PN, Manchester M, Atkinson JP (2000) Dissecting sites important for complement regulatory activity in membrane cofactor protein (MCP; CD46). *J Biol Chem* 275:37692–37701. doi:10.1074/jbc.M004650200
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta Ct}$  method. *Methods* 25:402–408. doi:10.1006/meth.2001.1262
- Long M, Deutsch M (1999) Association of intron phases with conservation at splice site sequences and evolution of spliceosomal introns. *Mol Biol Evol* 16:1528–1534
- Moffat GJ, Vik DP, Noack D, Tack BF (1992) Complete structure of the murine C4b-binding protein gene and regulation of its expression by dexamethasone. *J Biol Chem* 267:20400–20406
- Morgan BP, Harris CL (1999) Complement regulatory proteins. Academic, London
- Oshiumi H, Shida K, Goitsuka R, Kimura Y, Katoh J, Ohba S, Tamaki Y, Hattori T, Yamada N, Inoue N, Matsumoto M, Mizuno S, Seya T (2005) Regulator of complement activation (RCA) locus in chicken: identification of chicken RCA gene cluster and functional RCA proteins. *J Immunol* 175:1724–1734
- Oshiumi H, Suzuki Y, Matsumoto M, Seya T (2009) Regulator of complement activation (RCA) gene cluster in *Xenopus tropicalis*. *Immunogenetics* 61:371–384. doi:10.1007/s00251-009-0368-9
- Sarma JV, Ward PA (2011) The complement system. *Cell Tissue Res* 343:227–235. doi:10.1007/s00441-010-1034-0
- Pardo-Manuel F, Rey-Campos J, Hillarp A, Dahlbäck B, de Córdoba SR (1990) Human genes for the alpha and beta chains of complement C4b-binding protein are closely linked in a head-to-tail arrangement. *Proc Natl Acad Sci USA* 87:4529–4532
- Sun G, Li HY, Wang Y, Zhang BH, Zhang SC (2010) Zebrafish complement factor H and its related genes: identification, evolution, and expression. *Funct Integr Genomics* 10:577–587. doi:10.1007/s10142-010-0182-3