Identification of cDNA Sequences Encoding the Complement Components of Zebrafish (Danio rerio)

Vo, Kha Tam
Laboratory of Marine Biochemistry, Division of Marine Biological Chemistry, Department of Bioscience and Biotechnology, Graduate School of Bioresource and Bioenvironmental Science, Kyushu University

Tsujikura, Masakazu
Laboratory of Marine Biochemistry, Division of Marine Biological Chemistry, Department of Bioscience and Biotechnology, Graduate School of Bioresource and Bioenvironmental Science, Kyushu University

Somamoto, Tomonori
Laboratory of Marine Biochemistry, Division of Marine Biological Chemistry, Department of Bioscience and Biotechnology, Faculty of Agriculture, Kyushu University

Nakano, Miki
Laboratory of Marine Biochemistry, Division of Marine Biological Chemistry, Department of Bioscience and Biotechnology, Faculty of Agriculture, Kyushu University

https://doi.org/10.5109/16119

出版情報：九州大学大学院農学研究院紀要. 54 (2), pp.373-387, 2009-10-29. Faculty of Agriculture, Kyushu University
バージョン：
権利関係：

J. Fac. Agr., Kyushu Univ., **54** (2), 373–387 (2009)

Identification of cDNA Sequences Encoding the Complement Components of Zebrafish (Danio rerio)

VO Kha Tam[1], Masakazu TSUJIKURA[1], Tomonori SOMAMOTO and Miki NAKAO*

Laboratory of Marine Biochemistry, Division of Marine Biological Chemistry,
Department of Bioscience and Biotechnology, Faculty of Agriculture,
Kyushu University, Hakozaki, Fukuoka 812–8581, Japan
(*Received June 25, 2009 and accepted July 13, 2009*)

Zebrafish can be utilized for immunological research as a fish model that allows gene targeting, but lacks enough molecular and sequence information on the complement system, a humoral innate immune factor. The present study aimed at identification of all the major complement components at the molecular level with phylogenetic evidence. Complementary DNA sequences encoding C1q, C1r/s, C2 (B/C2–B), C3, C4, C5, C6, C7, C8, C9, factor B (B/C2–A), the mannose–binding lectin (MBL), two MBL–associated serine proteases, properdin, and factor I were identified by cloning from the hepatopancreas and by database mining. As reported for carp and trout, the presence of two distinct isotypes were confirmed for C3, C4, and C7 also in zebrafish, indicating the ancient evolutionary origin of their diversifications in the vertebrate lineage.

INTRODUCTION

Zebrafish has been established as an extremely useful model animal for forward genetic study for a wide range of phenotypes. Furthermore, an efficient method of for reverse genetic methodology, termed target–selected inactivation of a gene, has recently been developed and proven useful to obtain a gene–knockout zebrafish, for which any embryonic stem cells are not available (Wienholds *et al*., 2002). Thus, zebrafish is considered to be useful fish species for not only developmental biology and genetics but also comparative immunology. Especially, a recently established rag1 mutant may serve as a unique fish model that lacks totally the adaptive immune response for comprehensive studies on roles of innate immunity in a fish (Swaim *et al*., 2006). Complement is a major player of innate system, comprised of more than 30 proteins in plasma and on cell surfaces (Walport, 2001). To date, however, only a few complement component sequences have been reported in the literature, in spite of the presence of highly organized sequence databases that cover genomic and transcript sequences as well as their chromosomal localization. Although there are a number of complement component–like sequence entries in the zebrafish sequence database, they are annotated based on automatic BLAST homology searches, and therefore, the annotations remain ambiguous or sometimes misleading or confusing.

In the present study, candidate gene sequences for the complement components were obtained by in silico database mining and actual cDNA cloning, and the obtained sequences were carefully evaluated for their authenticity using phylogenetic analyses, resulting in identification of twenty–three zebrafish genes orthologous to complement components/factors known in mammals and other teleost species as follows: C1qA, C1qB, C1qC, C1r/s, C2 (orthologue of carp B/C2–B, Nakao *et al*., 1998), C3 isotypes, C4 isotypes, C5, C6, C7 isotypes, C8, C9, factor B (Bf, orthologue of carp B/C2–A, Nakao *et al*., 1998) isotypes, factor I (fI), MASP2, MASP3, CD11, CD18 and properdin (P).

MATERIALS AND METHODS

Animals

Adult zebrafish (*Danio rerio*), weighing about 1g, were purchased from a local pet shop and kept in a circulating aquarium at 27 ˚C with daily feeding.

Reagents

Oligonucleotide primers shown in Table 1 were synthesized and supplied by GeneNet Co. (Fukuoka, Japan). SMART RACE kit was from Clontec Japan (Tokyo, Japan). KOD Plus DNA polymerase was obtained from Toyobo (Tokyo, Japan). A plasmid vector, pGEM–T, was from Promega (Tokyo, Japan), and QIAEX–II DNA purification kit was a product of Qiagen (Tokyo, Japan).

Database search

Basic Local Alignment Search Tool (BLAST, http://www.ncbi.nlm.nih.bov/blast) (Altschul *et al*., 1990) and SMART (http://smart.embl–heidelberg.de/) were used for homology search, identity/similarity assessment and protein domain prediction. Amino acid sequences of complement components of human, mouse, carp and trout were used as queries of the homology search.

Preparation of RNA and first strand cDNA

Total RNA was extracted from the hepatopancreas of adult zebrafish by the acid–guanidium–phenol–chloroform method (Chomczynski, 1993), using ISOGEN reagent (Nippon Gene, Japan), and 1 µg aliquot of the

[1] Laboratory of Marine Biochemistry, Division of Marine Biological Chemistry, Department of Bioscience and Biotechnology, Graduate School of Bioresource and Bioenvironmental Science, Kyushu University
* Corresponding author (E–mail: mikimnakao@kyudai.jp)

RNA was reverse–transcribed to first strand cDNA using oligo–dT_{16} primer and Moloney murine leukemia virus reverse transcriptase (Invitrogen, Japan), according to the manufacturer's instructions.

RT–PCR amplification of factorB/C2 isotypes, C6 and C7–1 isotypes

Based on respective nucleotide sequence obtained from zebrafish DNA database, oligonucleotide primers were designed to amplify zebrafish cDNA orthologous to B/C2–A1, B/C2–A3, and B/C2–B of carp, and C6 and C7–1 of rainbow trout. RT–PCR amplification using KOD Plus Polymerase Kit (Toyobo, Japan) was performed using zebrafish hepatopancreas cDNA as template. Thirty cycles of amplification were carried under the following condition: 94 °C for 20 sec, 55–60 °C for 30 sec, 68 °C for 3 min.

RACE to amplify full–length cDNA of C3 isotypes

Zebrafish hepatopancreas cDNA was synthesized from the total RNA using the Universal Primer Mix (UPM), the SMART oligonucleotide, and Powerscript reverse transcriptase.

A cDNA fragment encoding zebrafish C3 was amplified using a degenerated sense primer encoding the thioester site (GCGEQNM) and the zebra–2–C3H1 primer, under the following conditions: 30 cycles of 94 °C for 15 sec, 55 °C for 15 sec, 68°C for 2 min. Then, 5'–RACE was performed to amplify cDNA encoding N–terminal two third of zebrafish C3, using Smart RACE cDNA Amplification Kit, employing UPM and C3–H1–N2. Reaction was done with 25 cycles in the following conditions: 94 ° for 5 sec, 60 ° for 10 sec, 68 ° for 5 min. The residual C–terminal one third–encoding region was obtained by 3'–RACE, using C3H1–N4 primer and UPM, under the same condition as above. Finally, the entire coding region was amplified by 5'–RACE using UPM and C3H2–N1 primer.

RACE and Nested–PCR for cDNA sequence of C7–2

Based on a putative partial cDNA sequence of zebrafish C7–2 found in the database, gene–specific primers to perform 3'–RACE, 5'–RACE and Nested–PCR was designed. The 3'–RACE was done with a pair of primers: 3'GSP (sense primer) and UPM (antisense primer) (Table 1). The RACE product was subjected to a nested–PCR with the sense primer, 3'NGSP, and antisense primer, NUP, under the following conditions: 30 cycles of 94 °C for 30 sec, 64 °C for 30 sec, and 68 °C for 3 min.

With the same procedure, 5'–end cDNA sequence was amplified using antisense primer, 5'GSP, and sense primer, UPM, then the amplicon served as a template for the nested–PCR with 5'NGSP primer and NUP, with the following thermal cycling conditions: 30 cycles of 94 °C for 30 sec, 60 °C for 30 sec, 68 °C for 3 min.

Subcloning

All the PCR products were gel–purified using 1–2% agarose gels and QIAEX II purification kit, and ligated with pGEM–T vector to transform *E. coli* DH5 α strain. The positive clones with expected insert DNA were picked up, and recombinant plasmids were purified by alkaline mini–prep method (Sambrook *et al.*, 1989).

Nucleotide sequencing

Nucleotide sequences were determined using dide-

Table 1. Primer sequence of PCR reaction

Gene–primer name	Sequence (5'–3')	Reaction
C1r/s–P1 (sense)	GTCTGCCTAACAGTTATGGATGG	RT–PCR
C1r/s–P2 (antisense)	ACCCCAAGACACAATGCCTC	RT–PCR
fB/A1–P1 (sense)	CCCCTGACTGTTTGCTTCTG	RT–PCR
fB/A1–P2 (antisense)	TCAAACTGCTCTCACACACTTTCAG	RT–PCR
fB/A3–P1 (sense)	TGCTGTGTGAACCTGACAGG	RT–PCR
fB/A3–P2 (antisense)	TCCACAATGCTTTCAATGAGGCACAATACA	RT–PCR
fB/C2B–P1 (sense)	ATCTGAGCGTCCGGGTTACA	RT–PCR
fB/C2B–P2 (antisense)	CCTTGGTGTTTCTCACTGTGAGGT	RT–PCR
C6–P1 (sense)	GGCCTTCATGCTGAAGTCCA	RT–PCR
C6–P2 (antisense)	GACCGTTGTCATGGAGATCATAATGG	RT–PCR
C7–1–P1 (sense)	AAGATGTGCCCTCAGTCCTC	RT–PCR
C7–1–P2 (antisense)	GTGCATATGGGCATTCTTCAT	RT–PCR
C7–2–3'GSP (antisense)	TGAACTGGTGAAAGAAGTGCCGTGT	3'–RACE
C7–2–5'NGSP (sense)	TGTAAACCGGGAACTAGTGGCCAAG	3'–RACE
C7–2–5'GSP (antisense)	GCCCACCGGATACACATCTTTAGGA	5'–RACE
C7–2–5'NGSP (sense)	GCGTTTTACAGCTTTCCCTGGGTTT	5'–RACE
NUP	AAGCAGTGGTATCAACGCAGAGT	Nested–PCR
UPM (long)	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT	RACE
UPM (short)	AAGCAGTGGTATCAACGCAGAGT	RACE
C3–GCGEQNM (sense)	GNTGYGGNGARCARAAYATG	RT–PCR
C3–zebra–2–C3H1 (antisense)	GGTGTAATGCAGTTATGTCAAACATCCTCCTG	RT–PCR
C3–zebra–C3H1–N2 (antisense)	TTCATGCAGACTACCAACTGTTCCCCCACA	5'–RACE
C3–C3H1–N4 (sense)	AGTTATGCACAAACAACTACCAGATGGCTC	3'–RACE
C3–C3H2–N1 (antisense)	ACACACAATTTATTTGGAAATAGGAAAGATGGTGTTG	5'–RACE

Signal peptide

```
Zebrafish C3-1   MEVKLLFLSAVLLSSALLTLCDPLYVLMAPNLLRVGSSENVFVEAQDYSGEPLRVKISVK
Zebrafish C3-2   MEVKLLFLSAVLLSSALLTLCDPLYVLSAPNLLRVGSSENVFVEAQDYSGGPLNVRISVK
Carp C3-H1       MEVKLLFLTVVLLSSPLLTLCNPLYVLSAPNLLRVGSSENVFVEAQDYSGAAFDVKIIVK
Carp C3-S        MGVKLLFLTVVLLSSPLLTLCDPLFVLSAPNLLRVGSSENVFVEAQDYSGAAIEVKIIVK

Zebrafish C3-1   NFPAKNVEMMQKTVTLTG-EKYQILTDIEIPDDKNFFSDD-EKKQYVYLQAHFPSVTLEK
Zebrafish C3-2   NFPAKNVEMMQKTVTLTG-AKYQSLTDIEIPGDRNYFSDD-KIKQYVYLQAHFPSVTLEK
Carp C3-H1       NHPKKDKEILSQSVSLTAANNFQILKDIKIPDDQNYFSDDPLEKQYVYLQAHFPSVILEK
Carp C3-S        NHPKKDREILSQSVRLTADNNFQIIKDIKIPDDQNYFSDNPLEKQYVYLQAQFPSVTLEK

Zebrafish C3-1   VVMVSFQSGYLFVQTDKPIYTPGSNVQYRIFSMTPNLKPLSQPGITVDIMNPQGITVASD
Zebrafish C3-2   VVMVSFQSGYLFVQTDKPIYTPGSNVHYRIFSMTPNLKPFSKPGITVDITNPQGITVASD
Carp C3-H1       VVLLSFQSGYIFVQTDKPIYTPASTVQYRIFSLTPNLEPLSQSGITVEIMNPQGITVSSE
Carp C3-S        EVMVSFQSGYIFVQTDKPIYTPASTVQYRIFSLTPNLEPLSQSGITVEIMNPQGITVSSE

Zebrafish C3-1   AIKAEKGVKSSIYNIPDVTSLGMWKVVTRYTNTPLKTFTAEFEVKEYVLPTFEVKLTPSS
Zebrafish C3-2   TIFAEKGVKSSIYNIPEVTSLGMWKVVTRYANTPQKTFTAEFEVKEYVLPTFEVKLKTSK
Carp C3-H1       KIFPVKGMKSGKYAIPEMASPGIWKVVTLFSNTPQKTFTADFEVKEYVLPTFEVKLKPSK
Carp C3-S        KIFPVKGMKSGKYPIPEIASPGIWKVVTLFSNTPQKKFTADFEVKEYVLPTFEVKLKPSK

Zebrafish C3-1   FFYYVGDNKQENDEEDSLTVDIEAKYLFGKKMDGNAFVVFGVMDGAAKTSIVNSLQKVPV
Zebrafish C3-2   SFYYVGDSKPKNNEDDSLTVDIEAKYLFGKKMDGNAFVVFGVMDGAMKTSIVNSLQKVQV
Carp C3-H1       SFFYVHD--------ESLTVDIEAKYLFGQKVDGNAFVMFGVMEDEKKTSIPASLQKVQI
Carp C3-S        SFFYVGD--------PSLTVDIEAKYLFGQKVDGNAFVVFGVMNGEKKISIPTSLQKVQI

Zebrafish C3-1   VNGVGTAELTREMITKTFKNIKDLVGKSIYVSVNLLTESGSEMVEAEKRGIQIVTSPYTI
Zebrafish C3-2   FNGVGTAELTREMITKTFKNIKDLVGKSIYVSVNLLTESGSEMVEAEKRGIQIVTSPYTI
Carp C3-H1       IKGEGTAELTNQMITKTFPNINQLVGRSIYVSVSLLTESGSEMVEAERRGIQIVTSPYTI
Carp C3-S        IRGEGTAELTSGMITKTFPNINQLVGQSIYVSVSVLTESGSEMVEAERRGIQIVTSPYSI

Zebrafish C3-1   HYKRTSQFFKPGMPLGVSVYVTNPDETPAKDVEVEVLVDGKGGVSGKTRDNGIAKVKVNT
Zebrafish C3-2   HYKRTSQFFKPGMPLGVSVYVTNPDETPAEDVEVEVLVDGVGGVNSKTKANGIAKVTVNT
Carp C3-H1       HFRKTPQFFKPGMPFDVSVYVTNPDQTPAVNVEVEVN---PGGLRGQTRANGIAKVTVNT
Carp C3-S        HFKKTTQFFKPGMPFDVSVYVTNPDQTPAVNVEVEVVGS-QGTVKGQTKHNGIAKVTVNT

Zebrafish C3-1   PGGASTLKITAKTKNTEYTSEQQ-AVKTMTAQAYKTKHDSKNYLHINIESAEFEIGDQMT
Zebrafish C3-2   KAGVSTLKITATTRDPELANEQQ-ADKTMIAQAYKTNLGSKNYLHINIESAEFEIGDQMT
Carp C3-H1       PGGSPTLAITAKTKDPEIKDERQQAEKRMTAQAYIPKGGSNNYLHIGIDAAELQISDPMK
Carp C3-S        LGGSSTQEITAKTKDPQLRDNQQAVKKMTAHAYIPKDASKKNYLHIGIDAAELQMGDSMK

Zebrafish C3-1   VNLITGQSSGDRDQDYTYMILSKGQIVLAERFKRQGQTLVTLSLTITKEMVPSFRFVAYY
Zebrafish C3-2   VYLNTGESPGVKDQDFTYMILSKGQIVLAERFKRQGQTLVSLSLPVTKDMVPSFRFVAYY
Carp C3-H1       VNLNTGQSPGVKDQDYTYMILSKGQIVKVDRFKRKGQSLVTLPVTVTKDMVPSFRFVAYY
Carp C3-S        VFLNTGQSPGVKDQDYTYMILSKGQIVSVDRFKRRGQSLVSLPVPVTKDMVPSFRFVAYY
Carp C3-Q1       -------------------ILSKGQIVKVDRFKRKGQSLVALSLTVTKDMVPSFRFVAYY
                                    ******** .:****:**:**:*.:.:**:***********

Zebrafish C3-1   HVGKDEVVSDSVWVDVKDTCMGKLNVEVKEKKN-TYETGNDVNLHITGDPGARVGLVVVD
Zebrafish C3-2   HVGKDEVVSDSVWVDVKDTCMGKLKVEVTEKQK-FYETGNEVNLQISGDPGARVGLVVVD
Carp C3-H1       HVG-LEVVSDSVWVDVKDTCMGKLQIKVKNKMN-TYGTGDEVKLQITGDPGAKVGLVVVD
Carp C3-S        HVGSSEVVSDSVWVDVKDTCMGKLQVKVKDKMN-TYDTGDEVKLEITGDPGAKVGLVVVD
Carp C3-Q1       LVGSSEVVSDSVWVDVKDTCMGTLQLKVKEKSGKSYDTGDEVKLQITGDPGAKVGLVVVD
                   **  *****************.*:::*.:*    * **::*:*.*:*****:*******

Zebrafish C3-1   KAVQVLNKNRLTQSKVWDVIEKHDTGCTGGGGKDSMGVFSDAGLIFESNTAGGTDTRTVP
Zebrafish C3-2   KAVQVLNKNRLTQSKIWDVIEKHDTGCTGGGGKDSMGVFSDAGLMFESDPAGGTNTRTEP
Carp C3-H1       KAVQVLNKNRLTQTQIWDVIEKHDTGCTAGGGKDSMGVFTDAGLMFESSTAGGTNTRTMP
Carp C3-S        KAVQVLNKNRLTQTQIWDVIEKHDTGCTAGGGRDRMGVFTDAGLMFQSNTAGGTNIRTMP
Carp C3-Q1       KAVQVLNKNRLTQTKIWDVIEKHDTGCTAGGGRDSMGVFTDAGLMFVSNGAGRTNTRTVP
                 *************:::*************.***:* ****:****:* *. ** *: ** *
```

β–α processing site

```
Zebrafish C3-1   ECPAPLKRKRRSESLQTITTTLAGQYTEKIRPCCYDGMRNNRIGYTCERRASYVIDGEEC
Zebrafish C3-2   TCPAPPKRKRRSESLQTITTTLAGQYTEKIRPCCYDGMRNNRIGYTCERRASYVIDGEEC
Carp C3-H1       ECPKTSKRKRRAESLLQITSTLAGKYTGELKQCCVDGMRDNKLGYTCERRATYIVDGEAC
Carp C3-S        ECPKPSKRRRRAESLLKITSTLAEKYSGELKQCCVDGMRENKLGYTCDRRSRYIADGKEC
Carp C3-Q1       ECTIPAKRKRRSESLLQITSTLAGKYPGELKQCCVDGMRNNKFGYTCERRATYIIDGEGC
                  *. . **:**:***  **:*** :*. ::: ** ****:*::****:**: *: **: *
```

/ C3-convertase cleavage site

```
Zebrafish C3-1   VKAFLHCCKEVKNHKEKETEEEELILARSDEDEEGDYKDYDDITSRTQFPESWLWEEFDL
Zebrafish C3-2   VKAFLHCCKEVKNHKETEEEELILARKDDEEIYDEDYEDYDDITSRTQFPESWLWEEFDL
Carp C3-H1       AKAFVDCCNKIKDRKNTETEEEEMLLARSDDD-DDYYTESEEIVSRTQFPESWLWEEIDL
Carp C3-S        VDAFLHCCNQMKTHKDVKDEVEEMVLARSDDD-DDYYTESEEIVSRTQFPESWLWEEIDL
Carp C3-Q1       VKVFLECCNLMKTHKNMKTEEEELILARSDDD-DDYYIDSEDIVSRTQFPESWLWEENDL
                 ...*:.**: :* :*: : *   :    .::  :  * : ::*.************** **

Zebrafish C3-1   CDK----CSKPSKDKVIYLKDSITTWQILAISLSPTHGICVAEPEEIVVFKSFFIDLKVP
Zebrafish C3-2   CDK----CSKPSKNKVTYLKDSITTWQILAISLSPTHGICVAEPEEIVVFKSFFIDLKVP
Carp C3-H1       CDK----CPTPATEKVIYLKDSITTWEILAVSLSPTLGICVAEPEEMVVFKHLFIDLKMP
Carp C3-S        CDK----CAIPTKEKAIYLKDSITTWQILAISLSPTLGICVAEPEEIVVFKSLFIDLKMP
Carp C3-Q1       PTCRETNCGTTSVTKVIYLKDSITTWQILAVSLSPTLGICVAEPEEVVVLKSFFIDLKMP
                        *  .:  *.**********:***:***** *********:**:* :*****:*

Zebrafish C3-1   YSAVRGEQVEIKAIIHNYTPKHLKKVRVEFLETAGVCSAASKKGKYRTTVNVDKDSSVAV
Zebrafish C3-2   YSAVRGEQVEIKAIIHNYTPKHLKKVRVEFLETAGVCSAASKKGKYRATVNVDKDSSVAV
Carp C3-H1       YSAVRGEQLEIRAIIHNYTPN-KQKVRVEFMETEDVCSFASKKGKYRTTVSVEKDSSISV
Carp C3-S        YSAVRGEQLEIRAIIHNYTPK-KQKVRVEFMETEDVCSSASKKGKYRTTVSVDKGSSISV
Carp C3-Q1       YSAVRGEQLEIKAIIHNYSQMNMKKVRVEFMETEDICSSASKKGKYRTTVSVEKDSSISV
                 ********:**:******:    :******:** .:** ********:**.*:*.**::*

Zebrafish C3-1   SFVIVPMTLESQHIEVKASISD--YEDGVRKTLKVVSEGVLTKFRTEKLELNPGKKGEKP
Zebrafish C3-2   SFVIVPMTLESQHIEVKASILD--YTDGVRKTLKVVSEGVLTKFRSQKLELNPAKKGEQP
Carp C3-H1       SYVIIPMTLGNHMIEVKASAYDAIYTDGVRKPLKVVSEGVLIPLHRKNVELNPVKNGEKP
Carp C3-S        SYVIIPMTLGNHMIEVKASAYDSIYTDGVRKPLKVVAEGVLVPLQRQNLELNPAKNGEKP
Carp C3-Q1       SYVIIPMKLGNHDIEVKASAYD--FTDGVRKPLKVVSEGVRSELLNRTLDLNPAKNGEKT
                 *:**:**.* .: ******  *  : *****.****:***   :  ..::*** *:**:.
```

(Continues to the next page)

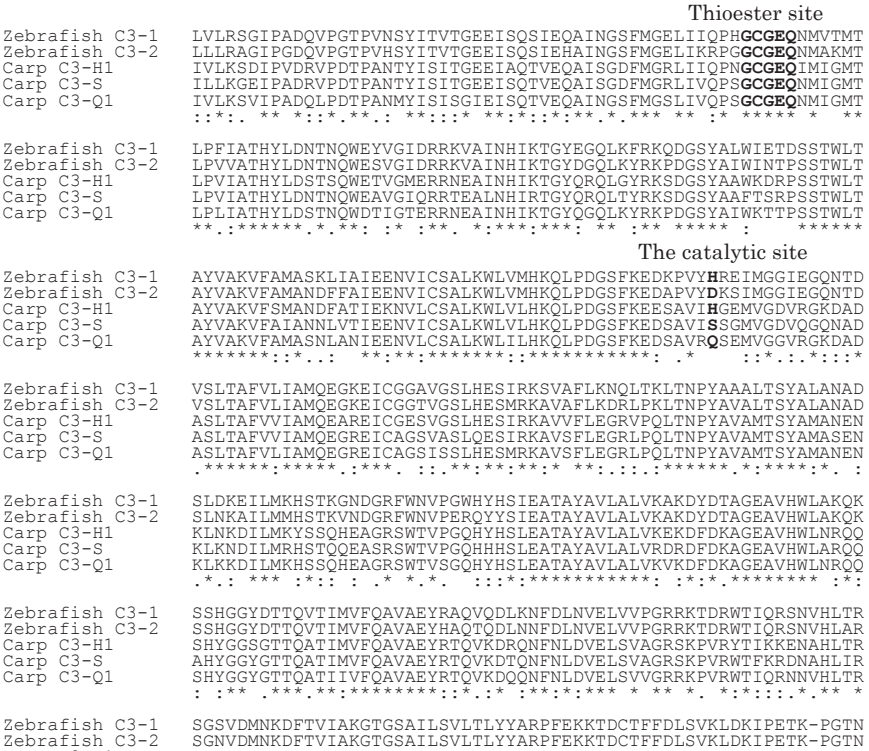

Fig. 1. Alignment of amino acid sequences of zebrafish C3–1 and C3–2 with those of two major C3 isotypes of carp. Signal peptides, and other functional sites such as b–a processing site, C3–convertase cleavage site, the thioester site, and the catalytic site are shown in bold and denoted above the sequence. Asterisks are identical residues shared by all the sequences, and residues with similar side–chains are marked by a single or doublet dots below the sequence.

oxy chain termination method (Sanger *et al.*, 1977) on an Applied Biosystems 377 and Beackman CEQ–8800 sequencers. *De novo* sequences were determined on both strands, following a primer–walking strategy.

Multiple sequence alignment and phylogenetic tree analysis

Multiple alignment of amino acid sequences was generated using CLUSTAL X software (Thompson *et al.*, 1997). The neighbor–joining method (Saitou and Nei, 1987) was employed to obtain phylogenetic trees based on the alignment, and visualized with the Njplot software (Persson, 2000). Bootstrap resampling analysis from 1000 replicates was used to evaluate reliability of given partitioning.

RESULTS

cDNA cloning of C3, C1r/s, B/C2–A, C6, and C7 from zebrafish hepatopancreas

C3 is forms (C3–1 and C3–2)

As a result of RACE–based cloning, two distinct full–length cDNA of zebrafish C3 isoforms, designated C3–1 and C3–2, were obtained after assembly of the overlapping sequence fragments of the RACE products. The full–length cDNA of zebrafish C3–1 molecule is 5170 bp–long containing an open reading frame of 4953 bp encoding 1651 amino acids. Zebrafish C3–2 cDNA consists of 5165 bp including a 4953 bp–long open reading frame that codes for 1651 amino acid.

Multiple alignment of the deduced sequences of the zebrafish C3 isoforms with those of other species shows that several important sites are conserved among the species (Fig. 1). In the zebrafish C3 molecules, the thioester site (CGEQ), a post–translational processing site to produce the typical two–chain structure (RXXR), and a cleavage site for C3–convertase are well conserved in all the sequence analyzed. However, the amino acid residue at the catalytic site of zebrafish C3–1 and C3–2 are histidine and aspartic acid, respectively. Comparing with other teleosts, two zebrafish C3 molecules share the highest identity each other (91%). In silico analysis using Ensemble zebrafish genome browser (http://www.ensemble. Org/Danio_rerio/Info/Index) reveal that the loci of C3–1 and C3–2 locate in tandem on chromosome 1 of zebrafish genomic (data not shown).

By using the neighboring–joining method, a phylogenetic tree of thioester–containing proteins was drawn with sequences of C3, C4 and C5 from various vertebrate species. As shown in Fig. 2, two or more C3 isoforms of each teleost species form respective clusters within each teleost lineage, suggesting that the duplication of C3 isoforms in carp, trout, zebrafish, flounder, spotted wolffish and medaka fish took place independently in the lineages leading to respective species.

C1r/s

Zebrafish C1r/s cDNA has been amplified by RT–PCR. After cloning and sequencing, a complete cDNA sequence of C1r/s molecule was determined to be encoding 661 amino acids. This C1r/s molecule is composed of six domains. From the N–terminus, two CUB (C1r/C1s, Uegf, and bone morphogenetic protein type 1) domains clamps an EGF (epidermal growth factor) domain, followed by two repetitive CCP (complement control protein) modules and finishes by a Trp–SP (Trypsin–like serine pro-

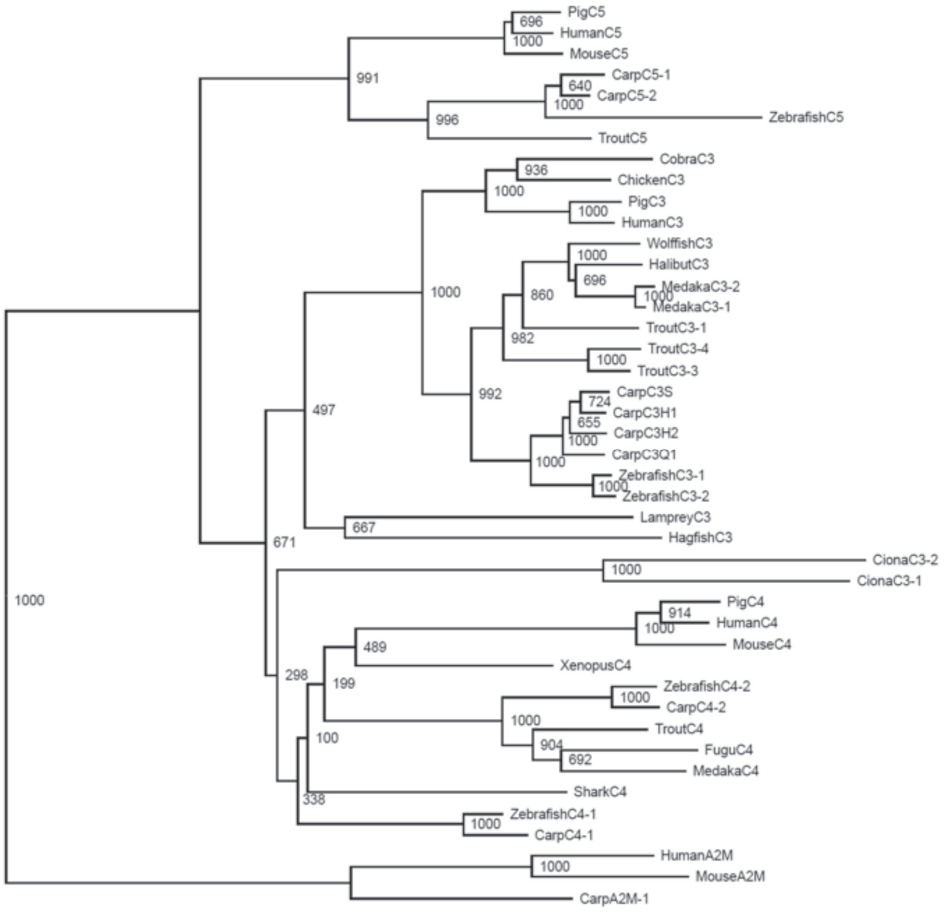

Fig. 2. Neighbor–joining phylogenetic tree of C3, C4, and C5 family members in chordates. Bootstrap percentages less than 900, after 1000 replications, are shown. Abbreviations: A2M, alpha–2–macroglobulin; A1M, alpha–1–macroglobulin; CVF, cobra venom factor.

tease) domain. The triad catalytic residues (His, Asp and Ser) in the Trp–SP domain and cysteine residues through the amino acid sequence in C1r/s molecule are well conserved in zebrafish C1r/s, as revealed in the multiple alignment (Fig. 3).

Zebrafish C1r/s amino acid sequence shared the highest identity with carp C1r/s–B and carp C1r/s–A (71.8% and 70.6%, respectively), followed by trout C1r/s, mouse C1r, human C1r with 46.1%, 33.6%, 33.5% identity, respectively. This C1r/s molecule of zebrafish only shares 30.4% identity with mouse and human C1s.

In the phylogenetic tree of C1r/s and MASP family (Fig. 4), zebrafish C1r/s is in the group of bony fish C1r/s with carp, trout and yellowed perch and this fish C1r/s group forms a cluster with mammalian C1r, separates from C1s and MASP branch.

The gene coding for zebrafish C1r/s locates on chromosome 6 of zebrafish genome (data not shown).

B/C2 isotypes

We obtained two distinct zebrafish B/C2 cDNA sequences similar to B/C2–A1 and B/C2–A3, respectively,

```
                                  / CUB-1
ZebrafishC1rs   ---MDGIHLIICLLWACVNVCECDLAMFGEVSSPQYPDPYPADLQKQWDLEVPQGFQIQL
CarpC1rsA       MFV..RV.......G........EP.............Q....NI...........Y.L..
CarpC1rsB       ---.....V..........L...EP.............Q....NF.E.........Y....


ZebrafishC1rs   TFNHLDIEPSPNCYYDSVSVVSDRKVLGKFCGQNSTDKFHPSDKPILAPGNRLQLLFLTD
CarpC1rsA       ........S..D......T....K.................G.............V....
CarpC1rsB       ...........D...........N.............R...G.......H.....V....

                         / EGF
ZebrafishC1rs   DSNHESHLGFSAFYQAVDTDECSSSSVGIDPPCSQICLNTLGSFLCACQHGYTLQPDNRT
CarpC1rsA       ..........T..F....I........ENG.............Y....Y.....R..Q..
CarpC1rsB       V....L.I..T.......I........EN-A............Y....H...M.R..Q..

                     / CUB-2
ZebrafishC1rs   CILECGGGLHSELEGSISSPGYPDTSPLDLDCVYNISVQPGFMITLNFSQNFHIERVDIQ
CarpC1rsA       .V......VR.....T.....F..........I.T..................VDQ.YS.
CarpC1rsB       .V......V...S..T.....F..V...N...I.T......Y.............Q.YN.


ZebrafishC1rs   GSTCLFHWLQVFVPGKDTHKYCGGKSPGVLNTKSHFVQLKYHTDRYGESRGWSLRYTTER
CarpC1rsA       .ES........S.Q..EPR....V........GT.....E....G..Q.Q....S...Q.
CarpC1rsB       .Q.........S....KPE.............GAN....E.......Q.Q...IH...Q.

                   / CCP1
ZebrafishC1rs   VQCPDPGSVTNGAVTPNFAQYLYRDYIHVRCNPGYKLMRGDIEISRFKSICQSNGQWNVP
CarpC1rsA       ....H..TIG..T...K..............K....I.M.EK...SY..........HLT
CarpC1rsB       ....H..IIG..TI..................M......M.EK..LS.........K.HLT

                        / CCP2
ZebrafishC1rs   LPECKIIDCGAPKPLLNGNFTFISGERNEYQSVIEYQCNEPYYRFGDTPKVRYKCAVDRK
CarpC1rsA       ...............M..D.EL....N...L.....H........K..S.AT........
CarpC1rsB       .............Q....D.Q.....N...L.....R........K..A..T.R...E..

                                      / Trp_SP
ZebrafishC1rs   WAEENNNDVIPPCYPVCGMNTEDLSAGRVFGGKPARSGEIPWQLLHKSSPRGGASLISDY
CarpC1rsA       .TDVS...L..I..........VSFG.............Q.....F..QLR..........
CarpC1rsB       .T.V....I......E......VSFG......MQ..P.Q........MHT..........

                      #                                                    #
ZebrafishC1rs   WALTAAHVVDGYESNSMTWLGGIINASNKSPVIIKTEKIIIHPNYKKVGRDGRQSDYNND
CarpC1rsA       ...........L.NTN.......V.SQDRN..TMEAN......S.QR.PVG.DRKNF...
CarpC1rsB       ...........L.NTT.......TE.R.QN...MEA.......K.QR.SLR.H.TN.D..


ZebrafishC1rs   IALMKMSAMVPLGPNLRPVCLPKKTGEAVKEGMMGTVSGFG-VYNWTVSNVLRYGHVHVY
CarpC1rsA       ...I....R.Q....I......NIISGP.M..K........GFEQGST.EI.....IQE.
CarpC1rsB       ...I....R......I......N..H.S.M..T........GFKEGLT.EI.....IRE.

                                            #
ZebrafishC1rs   SLKSCVSDGLPVSDNMFCAGDDVHGIDSCRGDSGGPLFFPMLGDGTESR-------GPWR
CarpC1rsA       PSEQ..FEDYF..E.......E.KRV...Q.............Y..KEQPYEVRGIVS.G
CarpC1rsB       PSEK..FGKM...K.........EHV...Q.............H.SKEQRYEVRGIVS.G


ZebrafishC1rs   PGACDVRAQFRR-----------------
CarpC1rsA       .AR.GHVSKGYYTKVQNYLGWIEETMANN
CarpC1rsB       .VR.GDVSKAYYTKVENYLGWIEETMANN
```

Fig. 3. Alignment of deduced amino acid sequence of zebrafish C1r/s with carp C1r/s–A and C1r/s–B sequences. Dots show residues identical to that of zebrafish B/C2–A1, and hyphens are indels. Domain names are denoted above the sequences, using the following abbreviations: C1r/C1s, Uegf, and Bmp1–specific domain (CUB), epidermal growth factor domain (EGF), complement control protein module (CCP), and trypsin–type serine protease domain (Trp_SP). The catalytic triad residues (His, Asp and Ser) of Trp_SP domain are denoted by "#".

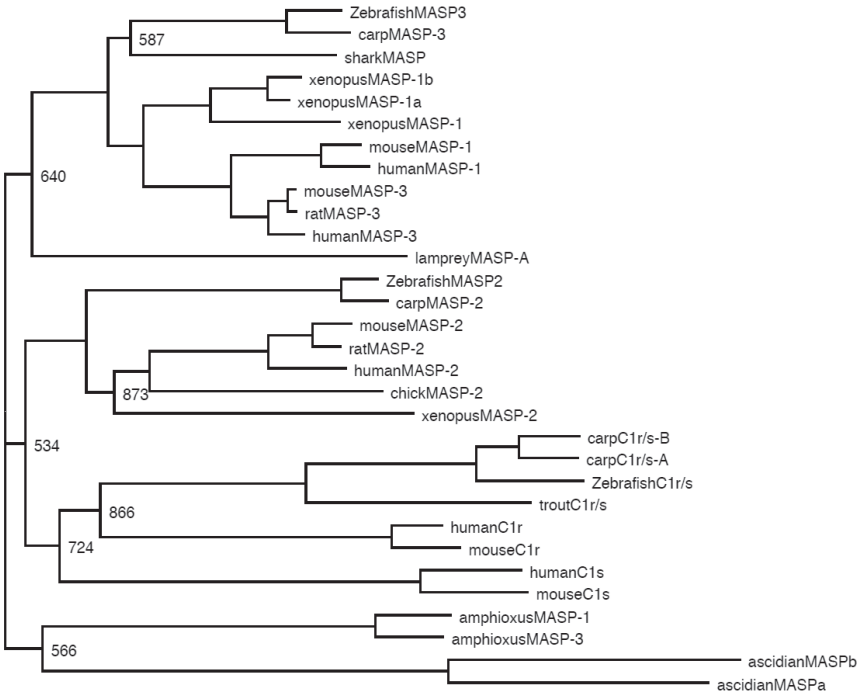

Fig. 4. Neighbor–joining phylogenetic tree of members of C1r/C1s/MASP family in chordates. Bootstrap percentages less than 900, after 1000 replications, are shown.

of carp by RT–PCR. The cDNA of zebrafish B/C2–A1 and B/C2–A3 molecules have the length of 2314 bp and 2356 bp, respectively. Zebrafish B/C2–A1 molecule has a reading frame of 2277 bp encoding for 759 amino acids. Zebrafish B/C2–A3 molecule is constructed by 749 amino acid corresponding to 2247 bp–long coding region.

By the result of multiple alignment and domain analysis, it was found that both B/C2–A isotypes in zebrafish are built up by three CCP modules, one VWA (von Willebrand factor type A) domain and a Trp_SP domain. The triad catalytic residues (His, Asp and Ser) are conserved in the Trp_SP domain in both molecules (Fig. 5). Zebrafish B/C2–A1 shares the highest similarity (64.6%) with carp B/C2–A2 while zebrafish B/C2–A3 shares that one with carp B/C2–A3 (73.3%).

On the other hand, the full–length of zebrafish B/C2–B cDNA includes 2402 bp with an reading frame of 2367 bp encoding 789 amino acids. The primary structure was organized into six domains: four CCP modules, a VWF domain and one Trp_SPc domain. The catalytic triad residues in SP domain are conserved in zebrafish B/C2–B (Fig. 5). The alignment of zebrafish FB/C2B amino acid sequence with other fish species also shows that zebrafish FB–C2B amino acid sequence shares the highest similarity with carp B/C2–B (66.2%), followed by trout B/C2–B (45.4%).

The phylogenetic tree of FB family shows that zebrafish B/C2–A1 and B/C2–A3 forms a cluster carp B/C2–A1 and B/C2–A3, respectively. They form a clade together with other teleost Bf and mammalian Bf, suggesting that B/C2–A lineage represent teleost Bf. On the other hand, zebrafish B/C2–B clustered together with other fish species B/C2–B and mammalian C2 derived from the same ancestor (Fig. 6).

In the zebrafish genome, loci of B/C2–A1 and B/C2–A3 are harbored in tandem on chromosome 21, whereas B/C2–B locus is on chromosome 7 (data not shown).

C6

A full–length zebrafish C6 cDNA was gained and found to be 2775 bp in length with a reading frame 2736 bp encoding 912 amino acids. The deduce amino acid sequence of zebrafish C6 was aligned with trout, human and mouse C6 (Fig. 7). From the alignment, conserved structural back born such as typical mammalian C6–like domain architecture: three thrombospondin domain type 1 (TSP1) domains, a low–density lipoprotein receptor type–A domain (LDLa), two CCP modules, an EGF domain, two factor I membrane attack complex domains (FIMAC) and a MAC/perforin (MACPF) domains. Zebrafish C6 shows the highest amino acid identity (53.9%) with trout C6, followed by human and mouse counterparts with 41.5% and 35.7%, respectively.

In agreement with the alignment result, the phylogenetic tree analysis of C6, C7, C8 and C9 also supported that the sequence cloned here is zebrafish orthologue of mammalian C6 (Fig. 8). Zebrafish C6 locus was found on chromosome 21 (data not shown).

C7–1 and C7–2

Full–length cDNA sequences encoding of two similar

but distinct zebrafish C7–like molecules, designated C7–1 and C7–2. Zebrafish C7–1 (2585 bp) specifies 820 amino acids, while zebrafish C7–2 (2433 bp) 810 residues (data not shown). Both zebrafish C7–1 and C7–2 have typical domains composition of C7: two TSP1 domains, an LDLa domain, a MACPF domain, an EGF domain, two CCPs and two FIMAC domains (Fig. 9). The result of the alignment of deduced amino acid sequences of zebrafish C7–1 and C7–2 with human C7, trout C7, and halibut C7 displays that zebrafish C7–1 and zebrafish C7–2, share 47.3% identity with each other, show the highest identity with trout C7 among other teleost species.

In the phylogenetic tree zebrafish C7–1 and C7–2 are in the branch of C7 with other species (Fig. 8). In zebrafish genome, C7–1 locates in chromosome 21 as does C6, and but C7–2 locus is on chromosome 8 (data not shown).

Identification of genes encoding zebrafish C1qA, C1qB, C1qC, C4–1, C4–2, C5, C8, C9, CD11, CD18, MASP2, MASP3 , fI and P by in silico cloning

As the zebrafish genome database grows increasingly rich in the sequence data with satisfactory accuracy, in silico approach to get entire cDNA sequence of the complement component genes has been becoming promising to get fairly reliable sequence data. Thus we have utilized the genomic and automatically transcribed sequence data from the database to retrieve candidate genes for several

```
                                                                / CCP-1'
ZebrafishB/C2-A1   ------------------------------------------------------------
ZebrafishB/C2-A3   ------------------------------------------------------------
CarpB/C2-A1        ------------------------------------------------------------
CarpB/C2-A2        ------------------------------------------------------------
CarpB/C2-A3        ------------------------------------------------------------
ZebrafishB/C2-B    MHYSTARLSTLLLFISLCEVHSEYDYDYGLAESKSCSLSESISGGTVEFSSGGSVGSRLI
CarpB/C2-B         MHYATVCFVIFFMYISVCEVQSQYDYEYEDNLPKNCTTSESITGGTVEYSNGGAAGSLLI

                                                                      /CCP-1
ZebrafishB/C2-A1   -----------MTSMECGLRLKWLILALICPLIAGAPS----REGSCPEENLDIAGGSFT
ZebrafishB/C2-A3   --------------...ASW.AIFM.TIQS...I...P-----SI...VKDIR.K....S
CarpB/C2-A1        --------------.A.KQQ....M..M..........SMSKGD....K..IN.T..T.V
CarpB/C2-A2        --------------...KQ........M...V......SMSK.DS...K...N.K..T.V
CarpB/C2-A3        --------------...EPWMNLFL..IL....T...-----SV...N...S.S..R.S
ZebrafishB/C2-B    YHCAEGFQPYPIGQRV.SSNGE.EPKVSRVKCEESSDYGDYEDQQKNCSLEVS.KD.RVS
CarpB/C2-B         YHCSAGFEPYPISQKV.SSDGE.KPKVSRVKCEETNDYGDYEETQKNCTLEVS.K...VS

                                                                    / CCP-2
ZebrafishB/C2-A1   LSNGYSDGSYLQYICPDNHYPSISSRRCQ--FGVWTPKASS----RKKAECKKITCPNPR
ZebrafishB/C2-A3   I.K---...SIIFN..E.Y..T.RT...T--K.R.SELPK-----G.RL........D..
CarpB/C2-A1        ...N..H..L.R....NGY...VH..L.--NEH..T.TKM----K.NP..........
CarpB/C2-A2        ..K...H..L.R....NGY...VQ..M..--D.R..S.IKT----..TP..........
CarpB/C2-A3        F.K---...IVR.D.VEGY..T.RI...I--K.K.N.LPK-----.RQP........D..
ZebrafishB/C2-B    Y..EGIE..V.T.H.ETG...FPTAQ.VCGRD.Q.SAMRL.SGKKTLS.V..E.L..AQI
CarpB/C2-B         Y..KGLE..V.T.H.KAG...FPATQ.VCDRD.E.SAMRLLNGQKIL..V..E.L..AQL

                                                                    / CCP-3
ZebrafishB/C2-A1   VLENGEVAPYQERYYINDVTTYSCSSDYKFRGSKVRVCQPNGKWNGSTPICGRDSDHCPD
ZebrafishB/C2-A3   AFL..D.H..SP...V..T.R.F.Q.G.D....ES....A....S........N..Y...
CarpB/C2-A1        .F........K.K..V..T.....N...T....A....K.....S..............
CarpB/C2-A2        .F.....I..K.K..V..T.....H...T....AD...K.....S..............
CarpB/C2-A3        GFK....Y...RQ.FV..T.H...Q.G.D....GT....A....S.G..V...N..Y...
ZebrafishB/C2-B    Q.D..QFW.RRQWLRVGEKQ.F..HEGFVLT..AE.N.THY.G.T.T..V.DDQ.ED.R.
CarpB/C2-B         Q.D..QFW.RRQWFK.GEKQ.F..HEGFVLT..AE.N.TQW.G.T.T..V.DNQ.ED.KN

                                                                  / Unknown
ZebrafishB/C2-A1   PGVPPGSSRTGSIFNIDDEVTYHCDSPLTLIGSKVRKCLDDGQWSGTEPQCYADFTYDTA
ZebrafishB/C2-A3   ....A.TT...NM.H.G.K...R..NK.S.....E.T.Q.N..................P
CarpB/C2-A1        ..........NM..T..K...S.A.........E.V.Q.G..................P
CarpB/C2-A2        ..........NM.....K...R.Q..M........V...G............HY....P
CarpB/C2-A3        ....A......NM.....K...R.ENK.......E.V.Q.N..................P
ZebrafishB/C2-B    ..I...AK.F.HH.R.G.K.R.L.Q.G.DIL..PE.W...SRE...A..R...QYSF.QP
CarpB/C2-B         ..T...AL.S.ER.R.G.K.H.L.Q.G.DML.PSE.H...SKE...A..R...QY...QP

                                                |         / VWF   +++++
ZebrafishB/C2-A1   MEAAEAFGNSLTTTLTVQQGFE-DDQHGKKISLDRG-GKLDIYIAVDASDSIDPKDFGKA
ZebrafishB/C2-A3   E..SD..SS..VSN.QLT.LH.ET..Y....QVHK.-........L.V.....EE..ER.
CarpB/C2-A1        E..S...SS..KSN.A.EKEE-----.Q....T..Q.-.............VEE...DY.
CarpB/C2-A2        E..S...SS..KSN.A.S.QY.KE..Q....TMNQ.-.............E...K...EN.
CarpB/C2-A3        E..SDG.SS..KSN.A.S.QY.GT..Y....RVGK.-........L.V.....EE..E..
ZebrafishB/C2-B    AIV.Q.L.G..SAA.D.SLPDFKKKGQSLGRTIKVEE.R.NVF.LM.T.G..SQDT.QA.
CarpB/C2-B         DAV.Q.L.G...AVMD.SHPDFKKQEQALGRTIRVAE.R.NVF.LL.T.A..S.ES.HL.


ZebrafishB/C2-A1   KKIIKTLIEKISYYEVSPNYEILMFATDVDQIVKMRDFKTNEKARNISKIFEDLDNFNYD
ZebrafishB/C2-A3   .DV.....................L....T.R.IS..E..NGQGK-DLL..IQK.QDYA..
CarpB/C2-A1        .TT..L.L......P.............TP.I..NQ..MQKPT--LTD..KEM.E.T..
CarpB/C2-A2        .TT..M..D.....P.............TR.TS.....N..D...LMN..K...D...E
CarpB/C2-A3        .GV.....................I.....AR..S.....SAQ.N-.LLE.LKR.KDYE.N
ZebrafishB/C2-B    ..A.IE.VR.LDS...NMKFD.VSY.SEPRE..SIMS.NSHDVD----FVLRK.SE.SDE
CarpB/C2-B         .NATIQ.VQ.LDS...TMRFG.ISY.SEAKE..SITNDLSQDVH----YVMRK.HE.SDK


ZebrafishB/C2-A1   KKGDRTGTNIAKLYLKILDSMSLEQVQNKEDFLQTQHVIIVFTDGQANMGGNPKPKVDLI
ZebrafishB/C2-A3   ...QQS.....QA.N..YE..TI.LMT.....KA...IV.M.........S...L..K.
CarpB/C2-A1        ....K.......V.TV.EE..KI.ELN.ATV.SE...I..L.S..H...........EQ.
CarpB/C2-A2        RV.VK.......A.TA..E.IK..ELN.AAI.NE...IV.L.............R...EQ.
CarpB/C2-A3        S..........QA.RS..E..QI..MT...E.KT...IV.M.............R.W..Q.
ZebrafishB/C2-B    VHEN.R..DLS.ALERVYGQLA.LRENK.SH.NE..NIL.IA...HS...P..QIMLNK.
CarpB/C2-B         SH.NKR...LHDALN.VYEELA.LRENKRSH.NE..N...IA...YS...PS.INILPK.
```

(Continues to the next page.)

```
ZebrafishB/C2-A1   KNLVIKNNA---S-RENKLDLYVFGVGKDVKKEDMNGLVSEKKDERHFFKLPDLDEVQNT
ZebrafishB/C2-A3   .S..RQ.S------V.E..E.....L.N..HA..I.D.KTDRAN.KF....KS..DLKE.
CarpB/C2-A1        .R..T..DP---K-..K..........-..NQD.V.....QRDQ.K.....Q..Q...EM
CarpB/C2-A2        .H..T..HP---N-..KN.........D..NQ..I.....QRDQ.KY....K..T...KM
CarpB/C2-A3        .D..K..SP---.EE.EN.......M.D..NA..I.D.KTDRGN.KF....KN.EDL.E.
ZebrafishB/C2-B    RS.LGYKPSSVDHTQ.EL..V...A.....NRK.LTSFA.S..G.K.V.V.Q.YQQLGYV
CarpB/C2-B         R..FGYKSS-VDHTK.EL..V...A..QQ.N.QELQSIA.I......V.V.K.YRQLGLV

                                  / Trp_SP
ZebrafishB/C2-A1   FDLMLDDSTVVGLCGMQQNYD---GSNKRSAYPWLAQLSIAQ---SQISDCMGSLVTSRY
ZebrafishB/C2-A3   ..N.I.EGNS.E...LYKD..DEFE.H..RQ.....KI.--RST-GKN.K....F...SF
CarpB/C2-A1        ..N.I.E..S.....IVWEGLE----...R.F.....IN..R---TKG.N........S.
CarpB/C2-A2        ..D.I.E..S.....IVWEGLE----...R.F.....IN.QHPSA.KG.N......S.S.
CarpB/C2-A3        ..S.I.EG.S.E...LYRD..DGTD.H..QQ.....KI.VTRNN-GK..N.L......SF
ZebrafishB/C2-B    .NQ.IS..-A.TK..IA.EEQSA-ADDVSYTK..HVD.LWGTKT------.R..I.SESM
CarpB/C2-B         .NQ.IS..-A.TK..VA.EEQDK-SETSYTSK..HVDILWGPRT------.R..IL.KSM

                        #
ZebrafishB/C2-A1   ILTAAHCFKEGDTP------DKITVYLEKN-TDVKVEKVFIHPNYSLTAKQSIGIKEFYD
ZebrafishB/C2-A3   ........RVD...------AT..ID..SENQGI..KYIHS..D.NIK..LNL..P.Y.E
CarpB/C2-A1        ............H------.....Q...D-KP...K.YV...Y.N.....QM..Q.Y.E
CarpB/C2-A2        ..................-----.....K...D-LV...KNFTL..K.KPK...H...Q.C.E
CarpB/C2-A3        ........RF...------.R...D.GS.IKGI..KDYIP..R.DIA...KV..P.Y.E
ZebrafishB/C2-B    V......LIKASG----VKATPADIKITHGAGE..AVELIV.SQFNVSGLKDKDV.....
CarpB/C2-B         V......LIKV.SDESVSKASAEDITVQHGNGR..AMELIL..RFD.KGLKDKNV.....

                         #
ZebrafishB/C2-A1   FDVALLQLKTPVKMSVNLRPICLPCTKETNRALKLSDSQGTCEKHEQILLSNELVDAAFT
ZebrafishB/C2-A3   ........EK...........I.......G..R...RE...K..KEL.MTG.T.K...M
CarpB/C2-A1        .....I..EK..DF.ST.....I.......G.....E.E...R...E..M.....E.S..
CarpB/C2-A2        .....I..EKA.E..F......I.......G.....E.E...K...E..MN....E....
CarpB/C2-A3        .....I..EK..I.DLG.....I......SG..R...RE...R.QQEE.M.ADT.K.N.M
ZebrafishB/C2-B    Y.I..IRM.ENITI.RQT.........SS....RMAAGS-..DQ..RV..HL.ETP.H.I
CarpB/C2-B         Y.I..IR.SD-ITI.K.A........SSS....RMAPDS-A.DQ.KNM..HLDETP.Q.I


ZebrafishB/C2-A1   SKMDMEKRSPRKIRRITVKLGKYLDACVEDAKKKE--SKWQTR-RRQLQKISCSGGNQPQ
ZebrafishB/C2-A3   .EVRTKT----EQKD..I.Q..LRQ........AAGMTAEKAT-DIVTDNFL....ID.T
CarpB/C2-A1        .D.ETD-H..KH.KN.IF.....R.........AKGINVENA.-EAVTDNFL....IE.E
CarpB/C2-A2        .E..ND-N..KN.KN.MI.....R.........AKGINVKNA.-EVVTDNFL.....E.K
CarpB/C2-A3        .V.GKSIE----KKQ..I.Q..WR.........ADGITATNAK-DIVTDNFL....IE.T
ZebrafishB/C2-B    .QGTHR-------ADTHIHS.AKREK.T.K.RSVLQENSRA.LTDIITERFM.T..SDRN
CarpB/C2-B         RQGTHR-------SDTRIHI.AKRGS.I.E.RSTIPEHSTASLSEVITDRFM.T..TEKN

                          #
ZebrafishB/C2-A1   RDDVSCKGESGGATHVDKYGRLIQIGVVSWGVKNLCSKAKLDAVSVSDSRDYHINLLRPD
ZebrafishB/C2-A3   T.EIA...D.....F.VPG..VV.V.I......D..K-RSPKPR.E.YT....T..FSAK
CarpB/C2-A1        T...A.........SY.N.K..VF.V.II.....DI.KGSSKKFT.DA......S..FSES
CarpB/C2-A2        T...A...D.....Y.N.K..V..V.........DV.K-E.RRFI.DA......S..FSEK
CarpB/C2-A3        V..IA...D.....F.VPGS.IV.V.I......D..K-NNQKPK.D.HT....TS.FS.E
ZebrafishB/C2-B    THHIT...D....LFLRRRM.YF.VRLL.VIFI----------------------------
CarpB/C2-B         .VQLT...D...PLFLR.RM.YF.VA.....T.QI.DSQTDVKEWPQ.A..F..SVFPLM


ZebrafishB/C2-A1   IQSFLKKHLENDKLGHTLTFLP--------
ZebrafishB/C2-A3   .I.....Y.ADE...TP...M---------
CarpB/C2-A1        .R....E.....RI.NP.....KNETVSAG
CarpB/C2-A2        .R....E..----------------------
CarpB/C2-A3        VRE...RY.G.E...TP....---------
ZebrafishB/C2-B    ----------------------------
CarpB/C2-B         PWLKQHL.E.LEF.P---------------
```

Fig. 5. Alignment of deduced amino acid sequences of zebrafish B/C2 isotypes with those of carp orthologs. Dots show residues identical to that of zebrafish B/C2–A1, and hyphens are indels. Domain names are denoted above the sequences, using the following abbreviations: complement control protein module (CCP), von Willebrand factor domain (VWF), and trypsin–type serine protease domain (Trp_SP). The metal ion–dependent adhesion site (MIDAS) is shown by +++++, and the catalytic triad residues (His, Asp and Ser) of Trp_SP domain are denoted by "#".

complement components, and carefully checked authenticity of their annotations by alignment and phylogenetic tree analyses.

C1q

Three subunit polypeptides of C1q, C1qA (NM_001020527), C1qB (NM_001003482) and C1qC (NM_001005976) were retrieved under the accession numbers shown above. The deduced amino acid sequences were aligned with other protein sequences with similarity, and a phylogenetic tree (Fig. 10) was constructed. As shown in the tree, zebrafish C1q A, B, C–chains formed respective clusters with mammalian orthologs, supporting that the sequences retrieved here encode C1q polypeptides incorporated in the complement C1 complex.

MASP2 and MASP3

These are lectin pathway components responsible for complex formation with mannose–binding lectin and for proteolytic activation of the pathway (Endo *et al.*, 1998; Nakao *et al.*, 2006). Two distinct database entries with similarity with MASP2 (XP_001923881) and MASP3 (XM_001341900) were found. In the phylogenetic tree containing MASP/C1r/C1s family members, zebrafish MASP2 and MASP3 form clusters with MASP2 and MASP3, respectively, from other species with considerably high bootstrap percentages (Fig. 4).

C4 isotypes

It has been reported that in teleost, there are two distinct C4 isotypes with substantial divergence (Mutsuro *et al*., 2005). Database search using carp C4–1 and C4–2 as queries resulted in finding of zebrafish C4–1 (XM_689530) and C4–2 (XM_685959) sequences. The clustering pattern in the phylogenetic tree clearly supports their assignment as C4–1 and C4–2 lineages (Fig. 2).

C5

C5–like gene (XR_029537) was retrieved and found to encode a typical two–chain protein similar to mammalian and teleost C5, though the zebrafish sequence is fragmentary because of incomplete database entry. Nevertheless, the zebrafish C5 amino acid sequence contains typical C5–like signatures such as two–chain structure predicted by the presence of the $\beta-\alpha$ –processing site, lack of thioester–bond in its α –chain, confirming the structural conservation of basic molecular architecture of C5. (data not shown). Phylogenetic tree also show a tight clustering of zebrafish C5 with other teleost C5 (Fig. 2).

C8 and C9

C8 and C9 are cytotoxic components to form a membrane–attack complex of the complement system. Both genes were found by Blast search under the following accession numbers: (C8β, XM_001332783, C9, NM_001024435). Among three chains constituting a C8 molecules, only C8β sequence was found in the present search.

Although C8 and C9 share a similar domain composition, the zebrafish sequences were clearly assigned to C8 and C9 based on the precise domain prediction by multiple alignments (data not shown) and also by phylogenetic tree (Fig. 8).

Complement receptor type 3 (CR3) subunits, CD11 and CD18

CR3 is an opsonic receptor that recognizes iC3b fragment of C3. It belongs to the beta2–integrin family and consists of two distinct subunits CD18 and CD11. Zebrafish CD11(XM_681091) and CD18 (XM_680920) sequences found by Blast search using carp CD11 and CD18 as queries were analyzed for their structural conservation based on multiple sequence alignment with respective orthologs. Zebrafish CD11 and CD18 showed high degree of sequence conservation, keeping several important functional sites unchanged (data not shown). Phylogenetic tree analysis gave a strong support for the assignments showing tight clustering with carp homologs in each lineage of CD11 and CD18 (Fig. 11).

Factor I (fI)

A database entry (XM_689643) for fI–like serine protease was found using carp fI as a query and shown to encode well conserved fI–like domain structure (data not shown). The authenticity of the fI identification was strongly supported by phylogenetic tree in which zebrafish and carp fI sequences formed a tight cluster (Fig. 12). From the tree, it was also suggested that carp fI–A and fI–B diverged after zebrafish and carp were separated from their common ancestor.

Properdin

Properdin sequence in the database (XM_679190) was confirmed to retain a well–conserved structural motifs, five repetitions of TSP1 domains (data not shown). Phylogenetic tree analysis provided further support for its grouping into teleost and mammalian properdin family (Fig. 13).

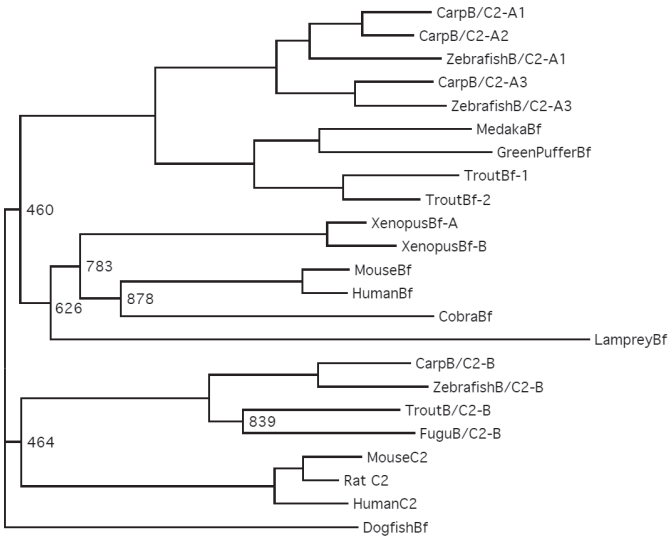

Fig. 6. Neighbor–joining phylogenetic tree of factor B and C2 family members of vertebrates. Bootstrap percentages less than 900, after 1000 replications, are shown. Abbreviations:

```
                                        / TSP
Zebrafish C6   MDCTCILLALLSTLASISPTMGCFCDHYPWSTWSQCTKTCDSGTQSRLRDVQYDDHWFKN
Trout C6       ..R.NS..V..HV.G.VTVSL.......A..S..T..R..NY...E.R.SIR...YYW..
Human C6       -MARRSV.YFILLN.L.NKGQA......A.TQ.TS.S...N......H.QIVV.KYYQE.

                                      / TSP
Zebrafish C6   SCSQLCQIHDNRVCNVEACPINCQLTEFGPWSECPSCAKKSFKIRSVLRPSQFGGADCSQ
Trout C6       ..E....K.ES.A...Q..F.H....DWAN..G.SP....QLRT..L........VE.DA
Human C6       F.E.I.SKQET.E..WQR.....L.GD.....D.DP.IE.QS.V...........QP.TA

                                / LDLR                                    /
Zebrafish C6   SLMEERACHPFKECRIEPLDCKDKFTCDTGRCIHADLQCNDQNDCGDNSDERDCGRIKSV
Trout C6       V.T.D...Y.ST..KM.T.N..-E.Q.GN....SSK.T..K.........KN.DEF.I.
Human C6       P.VAFQP.I.S.L.K..EA...N..R..S....ARK.E..GE..............T.A.

Zebrafish C6   CSNQRMYAFVPGADQIGFGFDAAAEQMRGAVLDNSFMGSNCTVEKQK----RNFYRISAN
Trout C6       .P--VEKRV.....L..N....M..A........M...DT.NLNRSRGSSY.LY..VP..
Human C6       .T--.K.NPI.SVQLM.N..HFL.GEP..E......T.GI.KTV.SS--RTS.P..VP..

Zebrafish C6   IERYEVKVETLEDFQPKE-----------------------------------TKSQKID
Trout C6       V.SF.I...IPD..KQEPQPVYSETVNLASPPVSTRSDSSSGDSSSIWVPIFFVG.RRHH
Human C6       L.NVGFE.Q.A..DLKTDFYKDLTSLG---HNENQQGSFSSQGGSSFSVPIFYSSKRSEN

                                / MACPF
Zebrafish C6   LANEPLYREAVKASQQKDSVFYRVHQVIATSTFKVKSSDLYLSDPFLQFLNSLPLEYNYA
Trout C6       QS.TDTFKQ.IS..KKT..Q.F.....LPV.R.R..D.....TE......T.........
Human C6       INHNSAFKQ.IQ..HK...S.I.I.K.MKVLN.TT.AK..H...V..KA..H......S.

Zebrafish C6   LYRHIFQLFGTHYFSSGTLGGKYDLLFQYDREELKTFGLKESDSEYCLSDDDTLVTFFYN
Trout C6       ...E..KH......A......H....Y..S.L...NS..T.EHTKG..KSESSMFIII.S
Human C6       ..SR..DD......T..S...V....Y.FSS....NS..T.EEAKH.VRIETKKRVL.AK

Zebrafish C6   RHKQRNTCGNISMKTKYEGSMVKASERCITSVQGGRTEFAAALAWEKKGVSPQSTVYTDW
Trout C6       QSSNVVR.SDNT.TQ.H...F.QS..KSFSM.R....GE.......Q..AA.D..T.KN.
Human C6       KT.VEHR.TTNKLSE.H...FIQGA.KS.SLIR...S.YG.......GSSGLEEKTFSE.

                                                          / EGF
Zebrafish C6   IKSTIENPVVINYELLPLVNLVRGISCAVTKRRHFHRALEEYQTEFDSCKCAPCPNNARP
Trout C6       A..L.D..A.VE.....II...K..P..A.....LT...V..LED..............A
Human C6       LE.VK...A..DF..A.I.D...N.P......NNLRK..Q..AAK..P.Q.......G..

                                              / TSP
Zebrafish C6   ALSGTECLCICQTVTYGSNCEKRAPDYTSVEVDGHWSCWSSWSPCDATLKKHRSRTCSNP
Trout C6       V....D.Q.....G...P...Q..Q....EA...Y....GT..A...SM.R..T.E.N..
Human C6       T........V..SG...E....QS...K.NA...Q.G......T....Y.RS.T.E.N..

                                                                  / CCP
Zebrafish C6   APQRGGKPCPGLEKQVEECTISIFQEQNVCINDDAFVTEDSSESRLPPGTSGCAKPLPPA
Trout C6       ..L....L.Q.PAR.E.G.F..........V..ED.A..GRA.G-....VE..PR.KR..
Human C6       .......R.E.EKR.E.D..F..MENNGQPCIN.DEEMKEVDLP-EIEAD...PQ.V..E

                                                                  / CCP
Zebrafish C6   NSHLRINKRQYDYGDHEEIICFTGFELQGFQLIHCLQDGTWEKPKAQCIKKVCSKPSVPD
Trout C6       .....KD..Y.EF..M..FL......ME.Y.F.N.RP....TP.TGM..R....P.A..E
Human C6       .GFI.NE.QL.LV.EDV..S.L....TV.Y.YFR..P....RQGDVE.QRTE.I..V.QE

Zebrafish C6   GMTINPDRMEYKVGSDIMLVCLESGTSPSGRLSYSCGKSLIWEPSIPKDIYCKIDKPYVP
Trout C6       ...LY.SKK....DHSVG.E.T.R.MV..EQGF.T.A...T...PL...LH......F..
Human C6       VL..T.FQRL.RI.ES.E.T.PKGFVVAGPSRYTCQ.N--S.T.P.SNSLT.EK.TLTKL

                 / FIMAC
Zebrafish C6   DSSCKPGEINDGTICVCMTKER-CRGYREDLCVYDAGKETAIMMSLCAFHADRCHGDRLY
Trout C6       .NQ.GR..RHV.SK.I.VAR.T-.LP.KTEF.IFN.EVGSGV...Y.G..SG.....Q.F
Human C6       KGH.QL.QKQS.SE.I..SP.ED.SHHS.....F.TDSNDYFTSPA.K.L.EK.LNNQQL

                                              / FIMAC
Zebrafish C6   FMNNGPCKSDARSLDWAKFRVSVSEQSSVQEPCGSDTCYEWETCSVSK-TCECKMPREYP
Trout C6       ...V...DE.VV..........MAAK.......D........S.TE..-Q.L..V..DC.
Human C6       HFLHIGSCQ.G.Q.E.GLE.TRL.SN.TKK.S..Y....D..K..A.TSK.V.LL.PQCF

Zebrafish C6   KDGKKIYCLKIVRTQTTRSMNLCFMAAMKCSSIEFELQHEGPCAGSYLCTLKL
Trout C6       .E.EHMF.V.LLK..SK..V.......I..RKM...ILN..L.EE.T------
Human C6       .G.NQL..V.MGSSTSEKTL.I.EVGTIR.ANRKM.IL.P.K.LA--------
```

Fig. 7. Alignment of deduced amino acid sequence of zebrafish C6 with trout and human C6. Dots show residues identical to that of zebrafish C6, and hyphens are indels. Domain names are denoted above the sequences, using the following abbreviations: thrombospondin type I (TSP), low–density lipoprotein receptor class A (LDLR), membrane attack complex/perforin (MACPF), epidermal growth factor (EGF), complement control protein (CCP), and factor I MAC module (FIMAC).

DISCUSSION

Zebrafish is one of the teleost species of which genomic sequence database are best accumulated and available, in addition to other genetic data such as chromosomal gene map and various mutants. As for the complement encoding genes, only C3, C4 and factor B sequences have been characterized and published to date (Samonte *et al*., 2002; Gongora *et al*., 1998; Seeger *et al*., 1996). Very recently, several complement component genes have been analyzed for their ontogenic expression and response to LPS challenge in zebrafish, but no isotypic diversity was taken into account and most of the sequences were solely based on the database entry with automatic annotations with less reliability (Wang *et al*., 2008).

It should be noted that C3, Bf and C7 are present as multiple isotypes as reported for other teleost species such as carp and trout. C3–1 and C3–2 showed high degree of sequence similarity but differ at the catalytic site, which is responsible for substrate specificity of the binding reaction of C3 to foreign surface through the thioester site (Dodds and Law, 1998). C3–1 has a well conserved His residue at the catalytic position, on the other hand, C3–2 has an Asp instead, supporting the existence of the His– and non–His–types of C3 isotypes generally in teleost species. An amino acid residue at this position of non–His type C3 is Ser or Gln in carp, Thr in trout, and Ala in medaka fish, suggesting that any amino acid residue with less nucleophilic side chain than imidazole–side chain of His can be positioned here, probably resulting in the same substrate specificity of the thioester. Functional meaning of the existence of the two types of C3 isotypes are still unclear but inferred to be an evolutionary strategy to enhance target–recognition repertoire of C3 molecule (Sunyer *et al*., 1998; Nakao *et al*., 2003).

The two Bf isotypes correspond directly to B/C2–A1 and B/C2–A3, respectively, of carp as judged by phylogenetic tree analysis. In carp, B/C2–A1 is constitutively synthesized mainly in hepatopancreas, whereas expression of B/C2–A3 is mainly in kidney and spleen, where the expression is up–regulated by stimulation with immunostimulants such as β–glucans and sodium alginate (Nakao *et al*., 2003). It would be interesting to examine if such difference in the expression pattern between the two Bf isotypes is also conserved in zebrafish. The chromosomal localization of the B/C2–A1 and B/C2–A3 (both in Chr. 21) indicates that they were generated by tandem gene duplication in a common ancestor of zebrafish and carp.

The presence of two C7 isotypes have been reported only in trout, in which C7–1 and C7–2 share only 47.3% amino acid identity, and a large indel was noted in the most C–terminal domain termed FIMAC domain (Papanastasiou and Zarkadis, 2005), suggesting functional diversity between the two isotypes. In the phylogenetic tree, zebrafish C7–1 and C7–2 formed clusters with trout C7–1 and C7–2, respectively, indicating that the C7–1/C7–2 gene duplication occurred before divergence of trout and zebrafish. In the FIMAC domains of zebrafish C7–1 and C7–2, no large indel was observed in the sequence alignment, but the domain sequences diverge considerably from each other, again suggesting functional differentiation between C7–1 and C7–2 also in zebrafish complement system. The FIMAC domains have been reported to

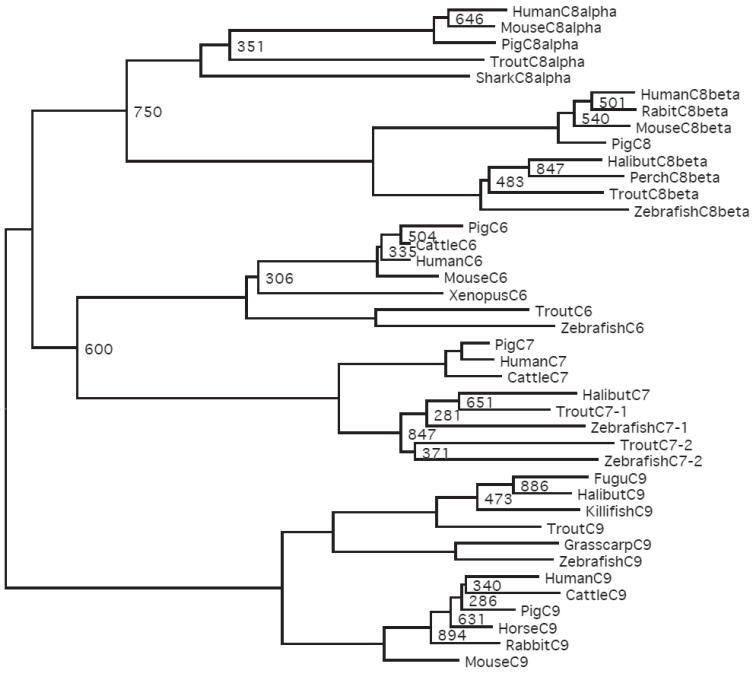

Fig. 8. Neighbor–joining phylogenetic tree of C6, C7, C8, and C9. Bootstrap percentages less than 900, after 1000 replications, are shown.

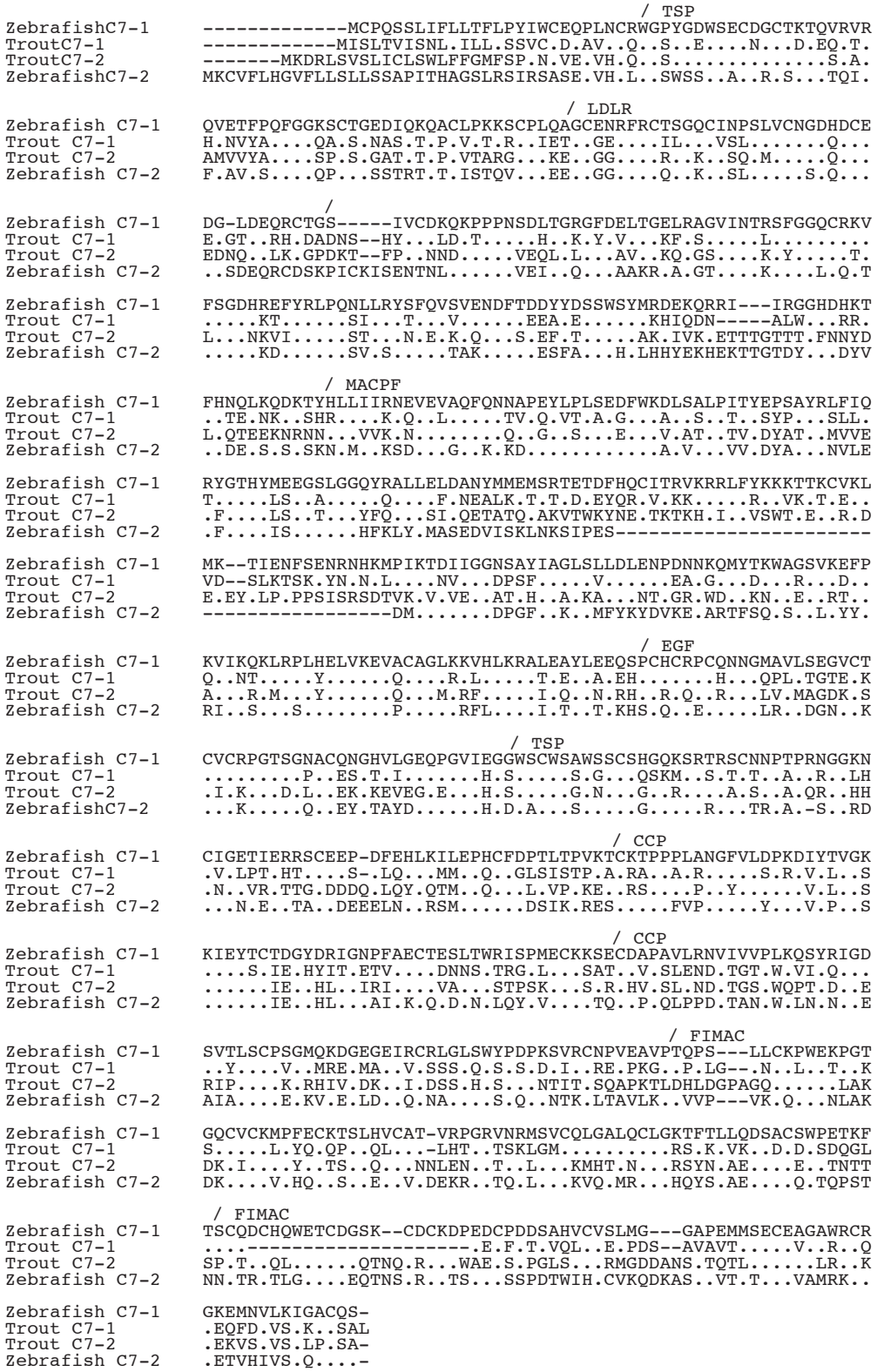

Fig. 9. Alignment of deduced amino acid sequences of zebrafish C7–1 and C7–2 with orthologs from trout. Dots show residues identical to that of zebrafish C7–1, and hyphens are indels. Domain names are denoted above the sequences, using the following abbreviations: thrombospondin type I (TSP), low–density lipoprotein receptor class A (LDLR), membrane attack complex/perforin (MACPF), epidermal growth factor (EGF), complement control protein (CCP), and factor I MAC module (FIMAC).

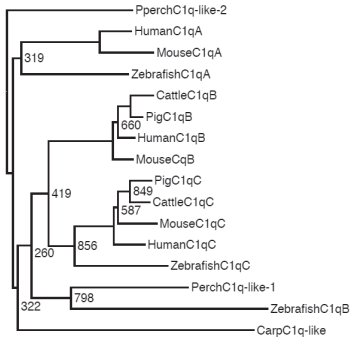

Fig. 10. Neighbor–joining phylogenetic tree of C1q A, B, C–chains and their homologs. Bootstrap percentages less than 900, after 1000 replications, are shown.

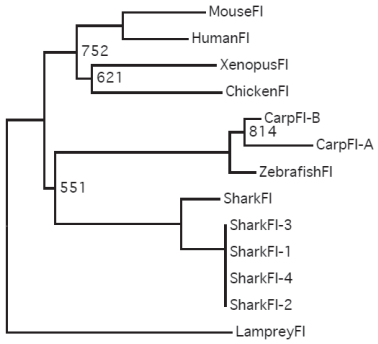

Fig. 12. Neighbor–joining phylogenetic tree of factor I of vertebrates. Bootstrap percentages less than 900, after 1000 replications, are shown.

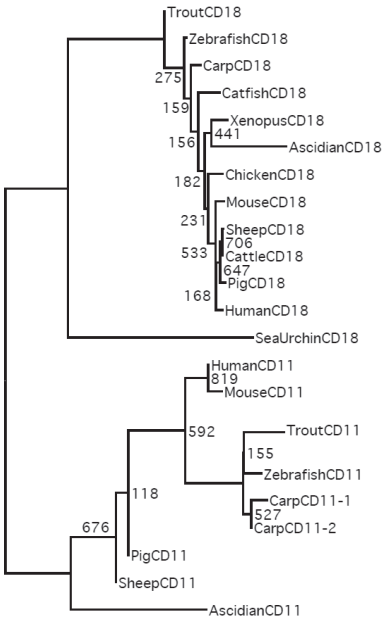

Fig. 11. Neighbor–joining phylogenetic tree of CD11 and CD18 integrin family. Bootstrap percentages less than 900, after 1000 replications, are shown.

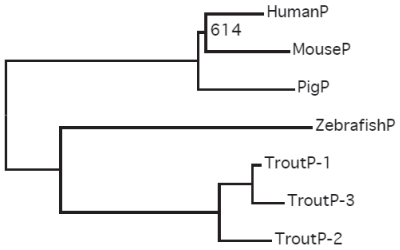

Fig. 13. Neighbor–joining phylogenetic tree of properdin and its homologs. Bootstrap percentages less than 900, after 1000 replications, are shown.

be crucial for interaction of C7 with C5 during the lytic pathway activation in mammalian complement systems (Thai and Ogata, 2004).

Overall, 24 genes encoding zebrafish complement components and their subunits have been identified with a reliable phylogenetic evidence in this chapter. These results allow further functional and genomic analyses of the zebrafish complement components, for example by detailed expression analyses.

A group of genes encoding the complement control proteins, such as the membrane–cofactor protein (MCP) and decay–accelerating factor (DAF) failed to be identified in this study, mainly because of considerable diversity of the domains (SCR domains) that constitute these molecules. Namely, although sequences encoding SCR domains were found by BLAST search, similarity of each SCR domain did not match similarity as a entire molecule, making it difficult to identify them. In addition, many of the SCR–domain–containing protein sequences in the database lacked reliable data around possible transmembrane region, making it ambiguous if they are membrane–bound proteins or soluble proteins, which are important keys for their identification (data not shown). Therefore, members of complement control protein family were not analyzed in deep in the present study. Cloning of their full–length cDNA sequence and more detailed phylogenetic characterization will be needed to identify them.

ACKNOWLEDGEMENT

The authors thank Ms. Miki Shinbara for her assistance in the zebrafish database search and cloning experiments. This study was supported in part by a grant–in–

aid from Japan Society for the Promotion of Science to MN.

REFERENCES

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman 1990 Basic local alignment search tool. *J. Mol. Biol.*, **215**: 403–410

Chomczynski, P. 1993 A reagent for the single–step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. *Biotechniques*, **15**: 532–534

Dodds, A. W. and S. K. Law 1998 The phylogeny and evolution of the thioester bond–containing proteins C3, C4 and alpha 2–macroglobulin. *Immunol. Rev.*, **166**: 15–26

Endo, Y., M. Takahashi, M. Nakao, H. Saiga, H. Sekine, M. Matsushita, M. Nonaka, and T. Fujita 1998 Two lineages of mannose–binding lectin–associated serine protease (MASP) in vertebrates. *J. Immunol.*, **161**: 4924–4230

Gongora, R., F. Figueroa, and J. Klein 1998 Independent duplications of Bf and C3 complement genes in the zebrafish. *Scand. J. Immunol.*, **48**: 651–658

Mutsuro, J., N. Tanaka, Y. Kato, A. W. Dodds, T. Yano, and M. Nakao 2005 Two divergent isotypes of the fourth complement component from a bony fish, the common carp (*Cyprinus carpio*). *J. Immunol.*, **175**: 4508–4517

Nakao, M., Y. Fushitani, K. Fujiki, M. Nonaka, and T. Yano 1998 Two diverged complement factor B/C2–like cDNA sequences from a teleost, the common carp (*Cyprinus carpio*). *J. Immunol.*, **161**: 4811–4818

Nakao, M., J. Mutsuro, M. Nakahara, Y. Kato, and T. Yano 2003 Expansion of genes encoding complement components in bony fish: biological implications of the complement diversity. *Dev. Comp. Immunol.*, **27**: 749–762

Papanastasiou, A. D. and I. K. Zarkadis 2005 Gene duplication of the seventh component of complement in rainbow trout. *Immunogenetics*, **57**: 703–708

Persson, B. 2000 Bioinformatics in protein analysis. *EXS*, **88**: 215–231

Sambrook, J., E. F. Fritsch, and T. Maniatis 1989 *Molecular cloning. A laboratory manual*. 2nd ed. Cold Spring Harbor Laboratory Press

Samonte, I. E., A. Sato, W. E. Mayer, S. Shintani, and J. Klein 2002 Linkage relationships of genes coding for alpha2–macroglobulin, C3 and C4 in the zebrafish: implications for the evolution of the complement and Mhc systems. *Scand. J. Immunol.*, **56**: 344–352

Saitou, N. and M. Nei 1987 The neighbor–joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, **4**: 406–425

Sanger, F., S. Nicklen, and A. R. Coulson 1977 DNA sequencing with chain–terminating inhibitors. *Proc. Natl. Acad. Sci. U S A.*, **74**: 5463–5467

Seeger, A., W. E. Mayer, and J. Klein 1996 A complement factor B–like cDNA clone from the zebrafish (*Brachydanio rerio*). *Mol. Immunol.*, **33**: 511–520

Sunyer, J. O., I. K. Zarkadis, and J. D. Lambris 1998 Complement diversity: a mechanism for generating immune diversity? I*mmunol. Today*, **19**: 519–523

Swaim, L. E., L. E. Connolly, H. E. Volkman, O. Humbert, D. E. Born, and L. Ramakrishnan 2006 *Mycobacterium marinum* infection of adult zebrafish causes caseating granulomatous tuberculosis and is moderated by adaptive immunity. *Infect. Immun.*, **74**: 6108–6117

Thai, C. T. and R. T. Ogata 2004 Complement components C5 and C7: recombinant factor I modules of C7 bind to the C345C domain of C5. *J. Immunol*. **173**: 4547–4552

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins 1997 The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, **25**: 4876–4882

Walport, M. J. 2001 Complement. First of two parts. *N. Engl. J. Med.*, **344**: 1058–1066

Wienholds, E., S. Schulte–Merker, B. Walderich, and R. H. Plasterk 2002 Target–selected inactivation of the zebrafish rag1 gene. *Science*, **297**: 99–102