Doctoral Thesis

# Evolution of the complement components with unique domain structure

Ayuko Kimura

Department of Biological Sciences, Graduate School of Science,

The University of Tokyo

December, 2007

# Abbreviations and data depositions

Abbreviations used in this thesis are as follows: A2M, alpha-2-macroglobulin; C3, complement component 3; C6, complement component 6; ds, double strand; fB, complement factor B; fI, complement factor I; FVII; coagulation factor VII, FIX; coagulation factor IX, FX, coagulation factor X; ISH, *in situ* hybridization; MAC, membrane attack complex; MASP, mannan-binding protein-associated serine protease; ML, maximum likelihood; MP, maximum parsimony; NJ, neighbor joining; RACE, rapid amplification of cDNA ends; SP, serine protease; TCC, terminal complement component; TEP, thioester-bond containing protein; and WISH, whole mount ISH.

The sequences reported in this paper have been deposited to the DNA Data Bank of Japan under accession numbers as follows: AB450038~AB540044 for the *N. vectensis* complement component cDNAs, DC611166~DC619952 for lamprey liver ESTs, AB377282~AB377288 for lamprey complement component cDNAs, and AB485948 for shark *C6* cDNA.

# Abstract

To elucidate the evolutionary origin and history of the blood complement system, comprehensive cloning of the five complement gene families, the complement component 3 (C3), factor B (fB), mannan-binding protein-associated serine protease (MASP), complement component 6 (C6), and factor I (fI) families, was performed in agnathan lamprey and cnidarian sea anemone. In additions, for the C6 family which was not found in these animals, chondrichthyes shark was analyzed by RT-PCR using the consensus-degenerate hybrid oligonucleotide primers.

The draft genome search and RACE analysis using cnidarian, *Nematostella vectensis* resulted in identification of the *C3*, *fB* and *MASP* genes. These genes were completely absent in the draft genome sequences of placozoa, porifera, and choanofllagelata, indicating that the multi-component complement system was established in the early stage of eumetazoan evolution before the divergence of Cnidaria and Bilateria. *In situ* hybridization showed the endoderm-specific expression of the identified cnidarian complement genes, indicating that the cnidarian complement system acts mainly in the primitive gut cavity called coelenteron, the only cavity facing endoderm. On the other hand, the liver EST analysis of lamprey, *Lethenteron japonicum*, gave the genes for the C3, fB, MASP, and fI families. However, no evidence for the gene duplication/functional divergence within these families, which was essential for establishing the classical activation pathway of the mammalian complement system, was obtained. RT-PCR analysis using the universal primers for the C6 family genes identified one *C6* gene from shark but none from lamprey, indicating that the unique domain structure of C6 family was established in a common ancestor of the jawed vertebrate.

These results suggest that the complement system comprising at least three components, C3, fB, and MASP, was established in the common ancestor of eumetazoan animals more than 600 million years ago. Function of the primitive complement system was most probably the protection of the gut cavity with primitive circulatory function. Remaining two families, the fI and C6 families, appeared in the primitive vertebrates, before the divergence of cyclostomes for the fI family (more than 500 million years ago), and before the divergence of the Chondrichthyes for the C6 family (more than 400 million years ago), respectively. The gene duplication/functional divergence within each complement gene family, which played essential roles in establishing the sophisticated complement system of jawed vertebrates, was occurred immediately after appearance of all the five complement gene families in the common ancestor of the jawed vertebrates.

# Contents

# General introduction

Mammalians fight with infection through the two cooperative defense systems, innate and adaptive immune systems (1). Germ-line encoded pattern recognition receptors (PRRs) of the innate immune system sense the presence of infection by the pathogen-associated molecular patterns (PAMPs) that are commonly present on many microorganisms but not on the host's own cells (e.g., oligosaccharides on the bacterial cell wall). In contrast, all adaptive immune responses are mediated by lymphocytes, bearing variable cell surface receptor for antigen which is encoded in rearranging gene segments. Innate immunity is a basic, front-line defense system involved in the rapid elimination of pathogens. Almost all pathogens are successfully destroyed by this system, and adaptive immune system is induced only when the innate immune system failed to clear infection. The complement system is one of the major humoral defense systems of mammals, which plays critical roles in the innate immune system, and also *complements* the bactericidal function of immunoglobulins, the central player of the adaptive immune system.

The evolutionary origin of the adaptive immune system based on the antigen recognition by immunoglobulin, major histocompatibility complex (MHC) class I and II molecules, and T-cell receptor is well defined, and only jawed vertebrates possess this system (2). In contrast, the evolutionary origin of the innate immune system comprising the various components and cells, such as phagocytes, lectins, or antimicrobial peptides, is believed to be more ancient, possibly being traced back to the common ancestor of the multicellular animals (3-6). However, there was still fragmental information at the molecular level about the evolutionary origin of the innate immune system.

The mammalian complement system is a highly sophisticated biological reaction

system comprising more than 30 serum or cell surface components (7). Activation of this system occurs by a series of limited proteolytic cascades through three parallel activation pathways called the classical, alternative, and lectin pathways (Fig. 1-1). These pathways merge at the proteolytic activation step of the complement component 3 (C3), and activation fragments of C3 induce various effecter mechanisms for pathogen clearance.

Most key components of the human complement system possess unique domain structures and are classified into five mosaic protein families which were focused in this thesis (Fig. 1-2): C3 family (C3, C4, and C5); fB (factor B) family (fB and C2); MASP (mannan-binding protein-associated serine protease) family (MASP-1, MASP-2, MASP-3, C1r, and C1s); C6 family (C6, C7, C8A, C8B, and C9); and fI (factor I) family (fI only). The C3 family is included in the thioester bond-containing protein (TEP) superfamily together with the non-complement TEPs (e.g., alpha-2-macroblobulin (A2M) and CD109). However, there are several C3 family-specific structural characteristics available for their identification, including the C345c (C-terminal end of C3, C4, and C5) domain at its C-terminal end. Limited proteolysis of C3 family components generates two active fragments (Fig. 1-1). The larger fragment of C3 and C4 covalently tags the pathogens using the intramolecular thioester bond, leading to the enhancement of phagocytosis. In contrast, the smaller fragment acts as an anaphylatoxin and promotes inflammation. The fB, MASP, and fI family components have a serine protease (SP) domain with trypsin-like cleaving specificity at their C-terminal end, and N-terminal domains unique to each family that are considered to be responsible for their substrate specificity (Fig. 1-2). The fB and MASP family components work in the proteolytic activation cascades leading to the

formation of the C3/C5 convertase, whereas fI works in the proteolytic inactivation of the C3/C5 convertase (Fig. 1-1). The C6 family components with the larger activation fragment of C5 are assembled into the membrane attack complex (MAC) on the membrane of pathogens which leads to the induction of cytolysis (Fig. 1-1).

These complement families are thought to have been established by two steps: exon shuffling, which created the unique domain structures of each family, and gene duplication and subsequent functional divergence, which increased the number of members in each family and established the three parallel activation pathways of the mammalian complement system.

In this thesis, I clarified the evolutionary origin and history of the complement system by focusing on these two genetic events. Evolutionary origin of these unique domain structures was clarified by the analysis using six animals located at the key positions in the phylogenetic tree (Fig. 1-3), cnidaria (Part I), placozoa (Part I), porifera (Part I), choanoflagellate (Part I), agnathan (Part II), and chondrichthian (Part III). The exact timing of the gene duplication/functional divergence within the families was elucidated by the comprehensive analysis using agnathan (Part II), which have diverged from the jawed vertebrate lineage in the period between the postulated two round genome duplications (8). In additions, to elucidate the primitive function of the identified complement genes in cnidarian, lacking the cavity specialized for circulation, expression analysis and preliminary protein-level analyses were also performed (Part I).

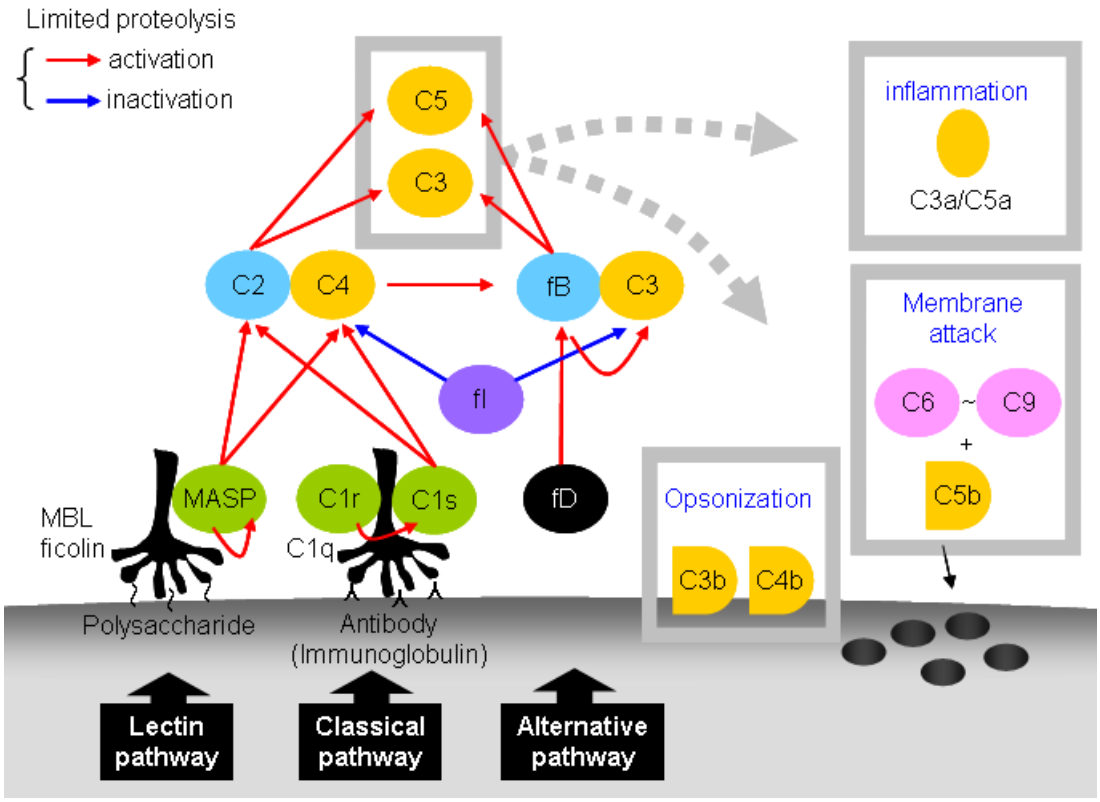



**Fig. 1-1.** Schematic representation of the three activation pathways and three effector functions in the mammalian complement system (1, 7). Only components involved in these reactions are shown here. There are three parallel activation pathways in the mammalian complement system. One is the classical pathway, which is triggered by the binding of complement component C1q to antibody bound to antigen. The second is lectin pathway, which is triggered by complement lectins, mannose-binding lectin (MBL) or ficolin, bound to polysaccharide on the bacterial surface. The third is alternative pathway, which is triggered by the spontaneously activated C3 bound to pathogen surface and by factor D (fD). Five complement gene families are colored as follows: yellow, C3 family; light blue, fB family; yellowish green, MASP family; pink, C6 (or TCC) family; and purple, fI family. Limited proteolysis by the complement serine proteases are indicated by the red (activation) or blue (inactivation) allows.

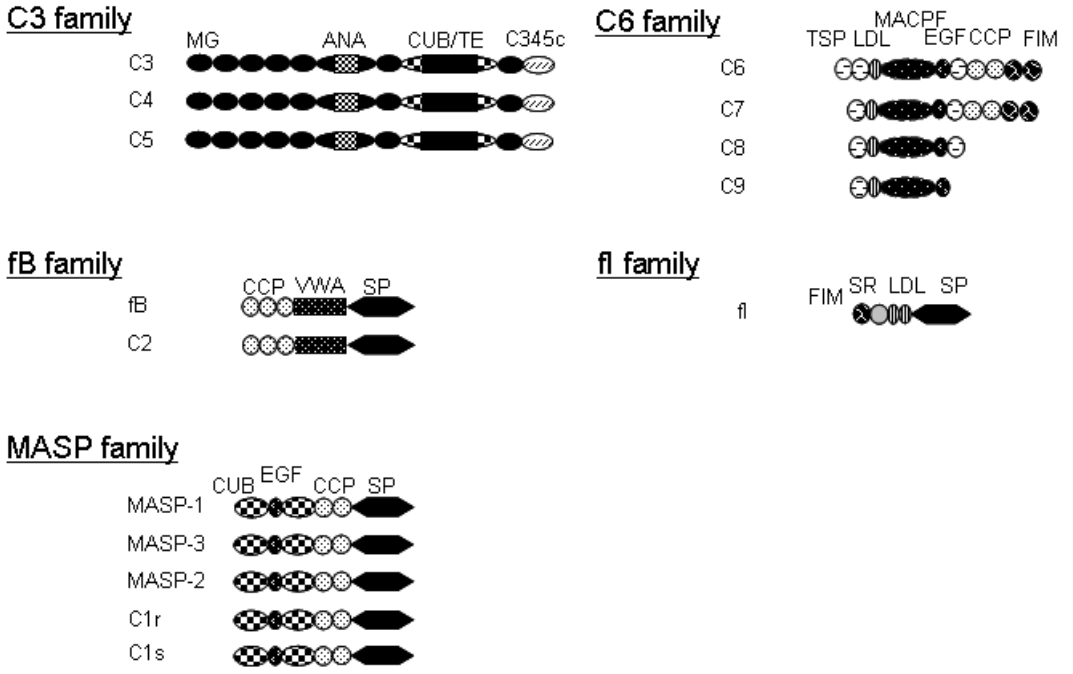



**Fig. 1-2.** Five complement gene families of mammals. Unique domain structures of the five complement gene families are schematically shown here. Abbreviations of domain names are: MG, macroglobulin; MG/ANA, MG domain inserted with anaphylatoxin region; CUB/TE, CUB domain inserted with thioester region; C345c, C-terminal of C3, C4 and C5; CCP, complement control protein; vWA, von Willebrand factor type A; SP, serine protease; CUB, C1r, C1s, uEGF, and bone morphogenetic protein; EGF-like, epidermal growth factor-like; FIM, factor I/MAC; SR, scavenger receptor Cys-rich; LDL, low-density lipoprotein receptor domain class A; TSP, thrombospondin type 1 repeats; and MACPF, MAC/perforin.

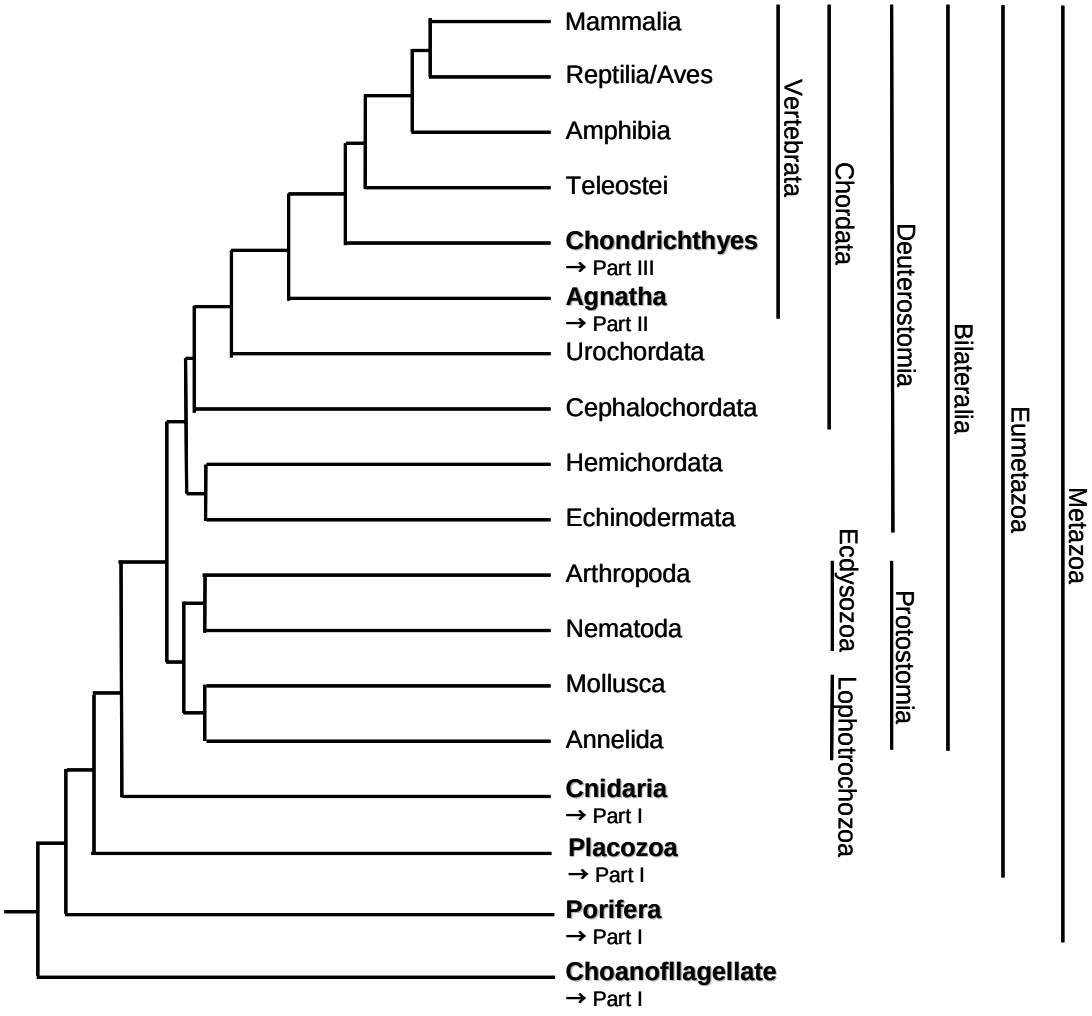



**Fig. 1-3.** Animals discussed in this thesis. The phylogenetic tree was based on the recent phylogenetic analyses (9-12).

**Table 1-1.** List of the accession numbers of the genes used for the phylogenetic tree constructions.

| Animal | Gene | Accession No. |
|---|---|---|
| human | C3 | NP_000055 |
| human | C4 | P0C0L4 |
| human | C5 | AAA51925 |
| human | A2M | P01023 |
| human | CD109 | NP_598000 |
| carp | C3H1 | BAA36619 |
| carp | C4-1 | BAB03284 |
| carp | C5-1 | BAC23057 |
| shark | C3B | unpublished data by Kimura H *et al.* |
| shark | C4 | BAC82347 |
| shark | C5 | unpublished data by Nagumo H *et al.* |
| hagfish | C3 | P98094 |
| lamprey | C3-1 | AAR13241 |
| lamprey | C3-2 | BAA02763 |
| lamprey | C3-3 | AB377282 |
| ascidian (*H. roretzi*) | C3 | BAA75069 |
| ascidian (*C. intestinalis*) | C3-1 | NP_001027684 |
| ascidian (*C. intestinalis*) | C3-2 | CAC85958 |
| sea urchin | C3 | NP_999686 |
| horseshoe crab | C3 | AAQ08323 |
| amphioxus | C3 | BAB47146 |
| coral | C3 | AAN86548 |
| sea anemone (*N. vectensis*) | C3-1 | AB450038 |
| sea anemone (*N. vectensis*) | C3-2 | AB450040 |
| fly | TEP1 | NP_523578 |
| human | C2 | NP_000054 |
| human | fB | AAH04143 |
| frog | fB-A | NP_001081234 |
| frog | C2 | ABB85337 |
| carp | fB/C2-A | BAA34706 |
| carp | fB/C3-B | BAA34707 |
| medaka | fB | NP_001098275 |
| shark | fB-A | AAY55950 |
| shark | fB | BAB63203 |
| lamprey | fB-1 | BAA02763 |
| lamprey | fB-2 | AB377283 |
| ascidian (*H. roretzi*) | fB | AAK00631 |
| ascidian (*C. intestinalis*) | fB-1 | NP_001027973 |
| ascidian (*C. intestinalis*) | fB-2 | NP_001029011 |
| ascidian (*C. intestinalis*) | fB-3 | NP_001027974 |
| horseshoe crab | fB | AAV65032 |
| sea urchin | fB | NP_999700 |

| Animal | Gene | Accession No. |
|---|---|---|
| sea anemone (*N. vectensis*) | fB-1 | AB450041 |
| sea anemone (*N. vectensis*) | fB-2 | AB450043 |
| amphioxus | fB | ABY28382 |
| human | MASP1 | NP_001870 |
| human | MASP2 | O00187 |
| human | MASP3 | AAK84071 |
| carp | MASP | BAA86866 |
| shark | MASP | BAA86867 |
| lamprey | MASP-A | BAC41492 |
| lamprey | MASP-1 | BAC75884 |
| human | C1r | P00736 |
| human | C1s | NP_001725 |
| carp | C1rs-A | BAB17845 |
| carp | C1rs-B | BAB17846 |
| ascidian (*H.roretzi*) | MASPa | BAC41341 |
| ascidian (*H.roretzi*) | MASPb | BAC41342 |
| amphioxus | MASP 3 | BAC75889 |
| amphioxus | MASP 1 | BAC75888 |
| sea anemone (*N. vectensis*) | MASP | AB450044 |
| human | transmembrane serineprotease 7 | NP_001036040 |
| human | fI | NP_000195 |
| mouse | fI | NP_031712 |
| chicken | fI | XP_426329 |
| frog | fI | NP_001095259 |
| fugu | fI | CAF89951 |
| carp | fI-B | BAB88921 |
| zebrafish | fI | AAI29472 |
| shark | fI | BAC01864 |
| lamprey | fI | AB377284 |
| human | VII | AAA51983 |
| human | IX | CAA01607 |
| human | X | P00742 |
| human | PC | NP_000303 |
| rabbit | X | NP_001075485 |
| rat | X | CAA56202 |
| rat | PC | NP_036935 |
| mouse | VII | P70375 |
| mouse | IX | NP_032005 |
| chicken | VII | NP_989773 |
| chicken | IX | Q804X6 |
| chicken | X | NP_990353 |

**Table 1-1.** (Continued) List of the accession numbers of the genes used for the phylogenetic tree constructions.

| Animal | Gene | Accession No. |
|---|---|---|
| chicken | PC | NP_989772 |
| frog | VII | AAI57199 |
| frog | IX | NP_001011223 |
| frog | PC | NP_001080424 |
| snake | X | ABG02403 |
| zebrafish | VII | NP_571894 |
| zebrafish | IX | NP_878288 |
| zebrafish | X | AH56804 |
| fugu | VII | AAO33368 |
| fugu | IX | AAO33372 |
| fugu | X | AAO33371 |
| lamprey | VII | AB377287 |
| lamprey | X1 | AB377285 |
| lamprey | X2 | AB377286 |
| human | PT | AAC63054 |
| chicken | PT | NP_989936 |
| frog | PT | NP_001015797 |
| trout | PT | CAD59688 |
| zebrafish | PT | AAH55596 |
| fugu | PT | AAO33373 |
| hagfish | PT | AAA21620 |
| lamprey | PT | AB377288 |
| human | C6 | AAA59668 |
| human | C7 | AAA51861 |
| human | C8A | AAH20702 |
| human | C8B | NP_000057 |
| human | C9 | AAB51328 |
| rabbit | C8A | NP_001075724 |
| horse | C9 | NP_001075419 |
| wild bore | C7 | NP_999447 |
| mouse | C6 | NP_057913 |
| chicken | C6 | CAX16418 |
| chicken | C7 | XP_424774 |
| chicken | C8A | XP_426667 |
| chicken | C8B | XP_422502 |
| frog | C6 | AAH76972 |
| frog | C7 | NP_001085116 |
| frog | C8A | NP_001005445 |
| frog | C8B | NP_000057 |
| frog | C9 | AAI59018 |

| Animal | Gene | Accession No. |
|---|---|---|
| zebrafish | C6 | NP_956932 |
| zebrafish | C7 | XP_690946 |
| zebrafish | C8A | NP_001003496 |
| trout | C6 | NP_001118093 |
| trout | C7-1 | NP_001118090 |
| trout | C7-2 | NP_001117879 |
| trout | C8A | AAL16647 |
| trout | C8B | NP_001117898 |
| trout | C9 | P06682 |
| fugu | C8A | CAF97618 |
| fugu | C8B | CAF97617 |
| fugu | C9 | AAC60288 |
| flounder | C7 | BAA88899 |
| flounder | C8A | BAA86877 |
| flounder | C9 | BAA86878 |
| shark (*M. manazo*) | C6 | AB485948 |
| shark (*G. cirratum*) | C8A | ABV08807 |
| Chimaera (*C. phantasma*) | C8B | not deposited |
| amphioxus (*B. bercheri*) | TCC | BAB47147 |
| Amphioxus (*B. floridae*) | TCC-1 | XP_002243725 |
| Amphioxus (*B. floridae*) | TCC-2 | XP_002213684 |
| Amphioxus (*B. floridae*) | TCC-3 | XP_002202762 |
| Amphioxus (*B. floridae*) | TCC-4 | XP_002219640 |
| Amphioxus (*B. floridae*) | TCC-5 | XP_002248061 |
| Ascidian (*H. roretzi*) | TCC | AB485949 |
| Ascidian (*C. intestinalis*) | TCC-1 | XP_002126061 |
| Ascidian (*C. intestinalis*) | TCC-2A | XP_002130788 |
| Ascidian (*C. intestinalis*) | TCC-2B | XP_002130679 |
| Ascidian (*C. intestinalis*) | TCC-3A | XP_002128624 |
| Ascidian (*C. intestinalis*) | TCC-3B | XP_002128860 |
| Ascidian (*C. intestinalis*) | TCC-4A | XP_002122361 |
| Ascidian (*C. intestinalis*) | TCC-4B | not deposited |
| Ascidian (*C. intestinalis*) | TCC-4C | XP_002122435 |
| Ascidian (*C. intestinalis*) | TCC-5A | XP_002130807 |
| Ascidian (*C. intestinalis*) | TCC-5B | XP_002124120 |
| Ascidian (*C. intestinalis*) | TCC-5C | XP_002130880 |

# Part I

# Comprehensive cloning and expression analysis of the complement genes in cnidarian sea anemone

**Abstract**

The origin of the complement system is more ancient than that of the adaptive immune system, as shown by the identification of the gene for the C3 in a basic metazoa, cnidarian coral. However, only a few reports on the other complement genes of nonchordates have been published, and the composition of the ancient complement system has not been clarified. Here, I performed comprehensive cloning of the complement genes with characteristic domain structures using a cnidarian, the sea anemone, *Nematostella vectensis*. Partial sequences of the two *C3*, two factor B (*fB*), and one mannan-binding protein-associated serine protease (*MASP*) genes were identified in the draft genome data, and the complete coding sequences of these genes were elucidated by RT-PCR and 5'- and 3'-RACE. In contrast, no *C6* and *factor I* family genes were identified. These cnidarian components shared the unique domain structures and most of the functionally critical amino acid residues with their mammalian counterparts, suggesting the conservation of their basic biochemical functions throughout metazoan evolution. *In situ* hybridization analysis indicated that all five genes are expressed in the tentacles, pharynx, and mesentery in an endoderm-specific manner. These results suggest that the multi-component complement system comprising at least C3, fB, and MASP was established in a common ancestor of Cnidaria and Bilateria more than 600 million years ago to protect the coelenteron, the primitive gut cavity with putative circulatory functions.

## Introduction

Recent completion of the draft genome assembly of the two invertebrate chordates, urochordata ascidian (13) and cephalochordata amphioxus (14), indicated that they have an almost complete set of the complement gene families, the *C3*, *fB*, *MASP*, and *C6*-like genes (15-18). However, only a few reports have focused on the complement genes of nonchordates. So far, only the *C3* and *fB* genes have been identified in the echinodermata sea urchin (19) and arthropoda horseshoe crab (20), and the *C3* gene has been reported in cnidaria coral (21, 22). Identification of the *C3* gene in cnidaria indicates a strikingly ancient origin of the complement system, which predates the divergence of Cnidaria from the Bilateria lineage more than 600 million years ago (23). However, the lack of information about other complement genes in cnidaria limits our gaining a comprehensive view of the ancient complement system.

In animals with three germ layers (triploblast), complement components are secreted into the cavity enclosed by the mesoderm: the body cavity in invertebrates (20, 24, 25) or the blood vessels in vertebrates. Cnidaria with only two germ layers (diploblast) lacks the mesoderm, body cavity, and blood vessels, and has only one cavity called the coelenteron, which is enclosed by the endoderm, and correspond to the gut cavity of triploblast (26). Although the humoral complement components are produced mainly by the vertebrate liver or invertebrate hepatopancreas (20, 25) and blood cells (19, 27), cnidarians lacks direct counterparts of these tissues or cells.

Here, I performed comprehensive cloning and expression analysis of the cnidarian complement genes to clarify the composition and localization of the ancestral complement system, using a cnidarian sea anemone, *Nematostella vectensis*. I also performed the preliminary protein level assay for the three *N. vectensis* complement proteins to clarify their functions.

## Material and Methods

### *Material*

Polyps of *N. vectensis* were kept at 26˚C in 1/3 diluted seawater and fed a few times per week with brine shrimp larvae. Animals were starved more than three days before experiment.

### *Comprehensive search and cloning of the complement genes of N. vectensis*

Genes encoding for the five complement protein families with unique domain structure were BLAST searched in *N. vectensis* using the amino acid sequences of human components as queries. Two peptide sequence collections available in the draft genome database of *N. vectensis* were searched: 1) peptide sequences predicted by Genscan program (28), 2) peptide sequences predicted by the DOE joint genome institute (JGI) using three different algorisms and EST sequences (29). Adult *N. vectensis* cDNA was used as a template for the following PCR amplification. To avoid the intron sequences sometimes mispredicted as exons, first RT-PCR was performed using the primers designed at the region encoding the conserved amino acid sequences, based on the amino acid alignment of the complement components of *N. vectensis* and other animals. The 5' and 3' ends of the identified cDNAs were obtained by the RACE using the SMART RACE cDNA amplification kit (Clontech). When the RACE stopped before the 5' end of the coding sequence, the candidate regions encoding the start codon and signal peptide were searched by eyes within the genome contig harboring the 5'-RACE-amplified sequence, and were confirmed by RT-PCR. Finally, full coding sequences were amplified by RT-PCR using the primers designed at the 5' and 3' untranslated regions. PCR products containing the entire coding region were cloned into

the pCR2.1-TOPO vector (Invitrogen), and eight clones each for these genes were sequenced to detect possible PCR errors and chimeras.

***Comprehensive search for the five complement genes in Porifera, Placozoa, and Choanoflagellate***

BLAST search for the five complement protein families was also performed against the following three draft genome or shotgun sequences of the primitive metazoa or its close unicellular relatives: 1) the draft genome assembly of the choanoflagellate *Monosiga brevicollis* (30) by Joint Genome Institute (JGI), 2) the whole genome shotgun (WGS)/EST sequences of the poriferan *Amphimedon queenslandica* (2,823,539 and 83,040 reads, respectively), available in the comparative genomics platform, Compagen (31), and 3) the draft genome assembly of the placozoan *Trichoplax adhaerens* (12) by Joint Genome Institute (JGI).

***Domain prediction and phylogenetic analysis***

Domain structures were predicted by the SMART program (http://smart.embl-heidel berg.de/). For the C3 families, domains except for the C345c domain, which are not adopted by the SMART or Pfam programs, were predicted by comparing with the primary structure of human C3 by eyes. Multiple alignment of the amino acid sequences was done by ClustalX (32). Based on this alignment, phylogenetic trees were constructed using the full length information, by the neighbor-joining (NJ) method (33) using MEGA4 excluding gaps by pair-wise deletion (34). Pairwise genetic distances were obtained by poisson correction. The reliability for internal branches was assessed by the 1,000 bootstrap replicates.

***RT-PCR***

RNA was isolated using ISOGEN (Nippon gene co. ltd.) from each sample, and 1 μg each of RNA was reverse-transcribed into cDNA. To trace the expression throughout the development, RT-PCR was performed using RNA extracted from the *N. vectensis* at five developmental stages, unfertilized eggs, pre-hatch embryo (0~2 days after fertilization), swimming planula larvae (2 days after fertilization), tetra-tentacle larvae (1 week after fertilization), and adult polyps (1 month after fertilization). Unfertilized eggs were collected from the isolated female polyps, while the others were collected from the pool of the mixed polyps. To examine the possible local expression of the complement genes, adult *N. vectensis* was cut into three pieces, tentacles, pharynx, and body column including mesentery, and RNA was extracted separately. Gene specific primers used for RT-PCR for the five complement genes and actin gene (XP_001637076) were as follows: *Nv (N. vectensis) C3-1* (5': ggctatcccacaagtc, 3': ttgttgtagtagggcgt); *NvC3-2* (5': tacacgctaccagataccataacg, 3': atcgatgactagagagtagacaact); *NvfB-1* (5': cacgtgtaccagataaa ggactacg, 3': gcggttcgtgttatccataagt); *NvfB-2* (5': acgtgtccgtacctgtc, 3': cgtttagtgattgggtttgttttg); *NvMASP* (5': tgtgaccatcactgccataac, 3': tgtgtctcgtagtaagcccggaag); *NvActin* (5': acgggatcgtcactaactgg, 3': aggaaggaaggctggaacat). All PCR reactions were performed by the following parameter; 1 min at 94˚C, 1 min at 50˚C, and 1 min at 72˚C, with optimum cycles before saturation.

***In situ hybridization (ISH) and paraffin section***

Whole mount ISH (WISH) was done based on the previously described protocol for the *N. vectensis* larvae (35) with some modifications, using the adult polyps. Antisense ribonucleotide probes were made by the Digoxigenin (DIG) RNA labeling mix and the

T7 RNA polymerase (Roche). The vector inserted with the complement or control *PaxA* gene (36) sequences were linearized, and was used as a template for RNA synthesis. The synthesized RNA were chemically degraded into about 0.5 kb fragments by treating with 42 mM $NaHCO_3$/63 mM $Na_2CO_3$ solution for 15~18 min at 60˚C. Hybridizations were performed at 44˚C for 41 hours with 250 ng/ml DIG-labeled probes. Signals were visualized using the NBT (4-nitroblue tetrazolium chloride)/BCIP (5-bromo-4-chloro-3-indolyl-phosphate) (Roche) as substrates for the alkaline phosphatase conjugated anti-DIG antibody. Experiments were repeated for several times and representative results were shown. For paraffin section, samples were intensively stained by the prolonged incubation with NBT/BCIP, re-fixed in the 4% paraformaldehyde for 30 min, treated with ethanol, and then with xylen. After incubation in paraffin (paraplast, Sigma) at 65˚C for 20 min for three times, samples were mounted in paraffin block. Sections of 15 μm thick were obtained using a microtome and photographed thorough a bright-field microscope.

***Production of polyclonal antibodies against the recombinant N. vectensis proteins***

The cDNA sequences corresponding to the MG1/MG2/MG3 domains of NvC3-1, the CCP/CCP/VWA domains of NvfB-1, and the CUB/EGF/CUB domain of NvMASP were used for the recombinant protein production. These cDNAs were inserted into the T7/CT/TOPO vector (Invitrogen), and cloned into the pVL1392 transfer vector using the *Sma*I and *Xba*I restriction sites. Recombinant proteins were produced by Sf21 insect cell cotransfected with the recombinant pVL1392 transfer vector and BaculoGold baculovirus DNA using the Baculogold transfer kit (Pharmingen). Recombinant proteins were detected by the SDS-PAGE under reducing condition and western blotting

using the anti-V5 antibody (Invitrogen). Cells were lysed with Lysis buffer (50 mM NaH2PO4, 300 mM NaCl, pH 8.0) supplemented with 1% of Triton X-100 and 1mM of PMSF. Recombinant proteins were collected from the inclusion body with high purity. After washing with 2% TritonX-100 in lysis buffer and solublizing with 8M urea in lysis buffer, they were diluted with equal volume of PBS, emulsified in Frreund's complete adjuvant, and injected into a New Zealand White rabbit. Immunization was repeated for more three times at two weeks intervals, and the rabbit was bled 1 week after the last injection.

***Western blotting***

The western blotting using the raised antisera was performed against the recombinant antigenic protein lysed with 8M urea, or the protein extracted from the body fluid, tentacles, and remaining body column of the adult *N. vectensis* polyps. Body fluid of *N. vectensis* was corrected drop by drop by the gravity force from the cut of body column by scissors. The lysate of recombinant protein and body fluid was added with 2X SDS sample buffer (0.12 M Tris-HCl pH 6.8, 4% SDS, 0.004% BPB, and 10% Sucros). The tissues excised from polyps, the tentacles and body column, were added with 1X SDS sample buffer and homogenized with plastic pestule. After denature at 70°C for 10 min, they were electrophoresed in a SDS-PAGE gel (3% of stacking gel and 7.5% of running gel). Proteins were electrotransfered onto the PVDF membrane, which were then blocked in 5% of skim milk in TBST over night. Membranes were incubated with the antisera diluted 1:500 in TBST for 1 hour, and then with the horseradish peroxidase-conjugated anti-rabbit IgG antibody (Invitrogen) diluted 1:10,000 in TBST, followed by the three times washes, respectively. Signals were visualized using the ECL

detection reagent (Amersham Biosciences). Alternatively, antisera absorbed with the 1/50 diluted recombinant protein were used for western blotting to check the specificity of antisera.

***Lipopolysaccaride (LPS)/zymosan treatment***

Adult polyps of *N. vectensis* were incubated in the 12-well dish containing 1/3 diluted sea water supplemented with three different pathogen-associated molecules or the mixture of them, for three hours at 27 °C. Agents used were as follows: 1) 100 μg/ml of LPS from *Pseudomonas aeruginosa*, 2) 100 μg/ml of LPS from *Eschericia coli*, and 3) 10 mg/ml of zymosan (yeast extract), all of which were obtained from Sigma. As an untreated control, the incubation was performed in the 1/3 diluted sea water only. The expression level of each complement gene was compared between the treated and untreated animals by RT-PCR started from the same amount (1 μg) of RNA and by WISH.

## Results

### *Comprehensive search for the complement genes of N. vectensis*

BLAST search for the complement genes for the five mosaic protein families, C3, fB, MASP, C6, and fI families, against the peptide sequences predicted from the genome data of *N. vectensis* gave partial sequences for two *C3*, two *fB*, and one *MASP* genes. No *C6* and *fI* sequences were found (Fig. 2-1). The MACPF (MAC/fI) domain of the C6 family, which is involved in the membrane attack, was identified in some predicted genes, as reported previously as cnidarian C6-like sequences (22). However, these were discriminated clearly from the authentic C6 family members by the lack of the TSP (thrombospondin type 1 repeats), LDLa (low-density lipoprotein receptor domain class A), CCP (complement control protein), or FIM (fI/MAC) domains. Similarly, the domain set of fI was totally absent from the *N. vectensis* genome. The lack of C6 and fI family genes in Cnidaria is also supported by the absence of the FIM domain in the *N. vectensis* genome, which is unique to these families and involved in the interaction with the C345c domain of the C3 family components (37). Other complement genes were excluded from the current search, because of the lack of unique domain structures available for their identification.

### *Comprehensive search for the five complement gene families in Placozoa, Porifera, and Choanoflagellata*

BLAST search for the five complement gene families against the draft genome data of the porifera, placozoa, and choanoflagellate did not detect any member of these families having the unique domains structures. The *C3* and other thioester-bond containing protein (*TEP*) genes seemed to be absent from the porifera and

choanoflagellata, because the domains comprising the *TEP* genes were totally absent from them (Table 2-1). In contrast, two TEP-like sequences (XM_002111552 and XM_002111553) sharing several domains with the C3 family were found in the genome of placozoan, a primitive metazoan whose phylogenetic position is still under debate (38). However, these TEP sequences showed a higher amino acid identity to the chordate CD109 than to C3, and lacked the C345c domain considered to be specific for the *C3* gene. The C345c domain was present in the placozoa genome, but it is found on the different scaffold from that harboring the TEP genes. In contrast, some domains comprising the other four complement gene families, the CCP, serine protease (SP), C1r/C1s/uEGF/bone morphogenetic protein (CUB), epidermal growth factor-like (EGF-like), TSP, LDLa, and scavenger receptor Cys-rich (SR) domains, were found in all the three animals (Table 2-1). However, sequences encoding the unique domain combinations of the four complement gene families were absent in these animals. Although the final conclusion should wait for the completion of the genome assembly of porifera, these results suggested at least the absence of the central complement component C3 in the primitive metazoans and their close unicellular relatives, indicating the absence of the complement system itself in these animals.

Presence of the *A2M*-like genes were also reported from many gram-negative bacteria (39), and thought to be involved in the protection of the bacterial cell membrane from the attack by the host's proteases. In an archaean *Methanococcoides burtonii* (strain DSM 6242) also, one *A2M*-like sequence (YP_566350) lacking the C345c domain was found in the draft genome sequences (Table 1). Taken together with the complete lack of the domains comprising the *TEP* genes in plant, fungi, and unicellular and primitive multicellular animals, the *TEP* genes of bacteria and archaean

is thought to be resulted from the horizontal gene transfer from the eumetazoan animals.

***Molecular cloning and primary structure analysis of the N. vectensis C3 genes***

BLAST search for the *C3* family genes detected four partial *TEP* sequences, which were identified as two *C3* (2,232 and 689 bp), one alpha-2-macroglobulin (*A2M*) (681 bp), and one *CD109* (2,169 bp) genes by phylogenetic tree analysis (Supplementary Fig. 2-1A). The full coding sequence of the two C3 genes termed *NvC3-1* (1,708 aa) and *NvC3-2* (1,740 aa) were obtained by RACE and RT-PCR. Their predicted amino acid sequences showed 23% identity to human C3 and 41% identity to each other. Two sequences with 99% amino acid identity were detected in a 1:1 ratio for the *NvC3-1* gene from one individual, and these probably represent allelic variants. Because they showed only eight amino acid substitutions at the positions with no obvious functional importance, only one was used in the following analysis.

In the phylogenetic tree constructed based on the alignment of the entire amino acid sequences (Fig. 2-2A), the *C3* family formed a monophyletic group, and vertebrate *C3/C4/C5*s, ascidian *C3*s, and cnidarian *C3*s formed their respective clades supported by 100% bootstrap percentage. However, the relationship among them and other invertebrate *C3*s was not resolved fully, suggesting a complicated evolutionary history of the *C3* genes.

The basic domain structure of the C3 family and the signal peptide for secretion were conserved completely in the *N. vectensis* C3s (Figs. 2-1 and 2-3A). The conservation of the C3a anaphylatoxin region, thioester site (GCGEQ), and catalytic His residue for cleavage of thioester (Fig. 2-3A) suggested that both NvC3-1 and NvC3-2 possess inflammatory and opsonic functions. The typical activation cleavage

site (LXR) by the C3 convertase was also conserved in NvC3-1 but was substituted to RXR in NvC3-2 similar to that observed in coral, horseshoe crab, and ascidian C3s. (20, 21, 25) The presence of the alpha/beta processing site (RXXR) suggested that NvC3-1 and NvC3-2 are processed into two-subunit chains. However, both Cys residues involved in the disulfide linkage between the alpha and beta chains of mammalian C3 (40) were substituted, and no other pair of Cys residues seems to be involved in the interchain linkage. A highly Lys/Arg-rich insertion was found within the MG8 domain of NvC3-1 and NvC3-2, as observed for coral C3. Insertion of the alpha/gamma processing motif (RXXR) into the MG8 domain was also found in the C3s of horseshoe crab (9), amphioxus (5), and lamprey (41), and in the C4s of vertebrates (42). However, the Lys/Arg content of cnidarian C3s was extremely high (~60%), and likely provides a unique, extremely positive charge to this region of *N. vectensis* C3s. The RXXR motif was also found in the Lys/Arg-rich region of NvC3-2 and coral C3, but not in that of NvC3-1.

***Molecular cloning and primary structure analysis of the N. vectensis fB genes***

Two sequences encoding the partial domain set of fB were detected in the draft genome database. The entire coding sequences of the two *fB* genes, termed *NvfB-1* and *NvfB-2*, were elucidated by RACE and RT-PCR. Their deduced amino acid sequences showed 19% and 17% identity with human fB, respectively, and 26% identity to each other. One pair of putative allelic variants showing 99% amino acid identity was found for the *NvfB-1* gene. Only one was used in the following analysis, because none of the four amino acid substitutions was observed at the putative functionary critical sites.

A phylogenetic tree using the entire amino acid sequence was constructed without

including any non-complement genes as an outgroup, because no other protein shares a similar domain structure with fB. The tree topology did not correspond to the phylogenetic relationship of the animals analyzed (Fig. 2-2B), as shown by the strange positions of amphioxus and horseshoe crab *fB*s. Because invertebrate *fB*s have extra CCP, EGF-like, or LDL domains at the N-terminal CCP repeat region, a phylogenetic tree was also constructed using only the vWA and SP domains. However, the obtained topology was completely the same (Supplementary Fig. 2-1B) as the tree shown in Fig. 2-2B. The evolutionary history of *fB* gene also seems not to be straightforward.

The deduced amino acid sequences of NvfB-1 and NvfB-2 contained a possible signal peptide, suggesting that they are secreted proteins. The predicted domain structure of NvfB-1 was the same as that of mammalian fBs, whereas NvfB-2 had two extra CCP domains at the N-terminal region (Figs. 2-1 and 2-3B). The *NvfB-1* is the first invertebrate *fB* gene found to have completely the same domain structure as vertebrate fBs. This result supports the idea that the domain structure of vertebrate fB with only three CCP domains at its N-terminus is an ancestral feature of fB, whereas the N-terminal region of invertebrate fBs experienced domain duplication or shuffling.

The *N. vectensis* fBs were expected to have trypsin-like proteolytic activity, based on the presence of the three catalytic triad residues (H, D, S) and the Asp residues at the bottom of the S1 pocket in the SP domain (Fig. 2-3B). The conservation of the activation cleavage site (RX) between the CCP and vWA domains common to all fB/C2 analyzed thus far suggested that they are cleaved by the trypsin-type SPs at this site. The $Mg^{2+}$ binding site in the vWA domain, which is involved in the allosteric conformational change of the mammalian fB/C2, was also found.

Structural analysis of the mammalian fB/C2 indicated that they have the unique SP

domain, whose S1 pocket conformation is distinct from that of other trypsin-type SPs. The evolutionary analysis of fBs suggests that this structural specialization of mammalian fB/C2 occurred in the common ancestor of the jawed vertebrates (43). In accord with this evolutionary scenario, I found that the SP domain of *N. vectensis* fBs did not share any unique structure with that of jawed vertebrate fBs. Rather, *N. vectensis* fBs retained canonical features of trypsin-type SPs, both the Asp residue at the bottom of the S1 pocket and the activation cleavage site (Arg-X) at the N-terminus of the SP domain. The disulfide bridge linking the two chains processed at the canonical cleavage site was also conserved in both. Therefore, it is indicated that the activated *N. vectensis* fBs comprise the SP domain and the vWA domain, and that these are linked together by a disulfide bond. The vertebrate fB/C2-specific disulfide bridge within the SP domain was also absent from the *N. vectensis* fBs. These canonical features were also found in the fBs of ascidians (44), sea urchin (19), lamprey (45), and horseshoe crab (20), although the horseshoe crab fB secondarily lost the canonical cleavage site at the N-terminus of the SP domain. Reflecting these structural features, the phylogenetic tree using the SP domains from various trypsin-type SPs formed two clades: one comprising only the jawed-vertebrate fB/C2s and the other comprising the lamprey and invertebrate fBs and other non-complement SPs (Supplementary Fig. 2-1C).

***Molecular cloning and primary structure analysis of the N. vectensis MASP gene***

One *MASP* sequence with the complete domain set of MASP was detected in the genome database. The full coding sequence of the *MASP* gene termed *NvMASP* (686 aa) was cloned successfully by RACE and RT-PCR, and the predicted amino acid

sequence showed 33%, 33%, and 32% identities to the human MASP-1, MASP-2, and MASP-3, respectively.

Phylogenetic analysis of the *MASP* family was performed using the *NvMASP* gene as an outgroup, because all other *MASP* genes reported thus far were from chordates. The tree topology reflected the postulated animal phylogeny completely (Fig. 2-2C).

The unique domain structure of the MASP family was conserved by the *NvMASP* (Figs.2-1 and 2-3C), and the presence of the signal peptide indicated that NvMASP is also a secreted SP. NvMASP had the canonical activation cleavage site of the trypsin-type SPs (RX) at the N-terminal end of the SP domain. In additions, the presence of the catalytic triad residues (H, D, S) and the Asp residue at the bottom of the S1 pocket of the SP domain suggested the trypsin-like activity of NvMASP. The mammalian MASP family members can be classified into two groups according to the structure of their SP domain-encoding regions: 1) the MASP-2/MASP-3/C1r/C1s group with an SP domain with some derived features, such as the lack of His loop, the split exons encoding the SP domain, and the TCN codon for the catalytic Ser (31, 32); and 2) the MASP-1 group with an SP domain similar to that of the other canonical SPs. The *NvMASP* gene retained all the canonical features in its SP domain and was similar to the MASP-1 group but not to the MASP-2/MASP-3/C1r/C1s group.

***Expression analysis of complement genes of N. vecnensis***

WISH analysis of the five complement genes using the adult polyp of *N. vectensis* showed similar expression patterns, which were restricted to the endoderm especially at the apical end of the tentacle cavity and the pharynx (Fig. 2-4B, left). In contrast, the control *PaxA* gene showed an ectoderm-specific pattern, indicating that the staining

patterns of the five complement genes were not non-specific. In the section of the WISH samples, all the complement genes showed strong signals in a broad area of the tentacle endoderm facing the tentacle cavity (Fig. 2-4B, right). In addition, signals for the five genes were also observed in certain cells in the multiple-layered pharynx endoderm and mesentery (46). No signals were detected in the ectoderm or gelatinous layer between the two germ layers called mesoglea, which contains amoebocyte-like cells in some cnidarian species (26). For the *PaxA* gene, cell-type specific signals were observed at the edge of the mesentery and the tentacle ectoderm (Fig. 2-4B), a pattern that differed greatly from that of the complement genes. No signal was detected by the sense probe or without probe staining (data not shown).

RT-PCR analysis of the *N. vectensis* at five developmental stages showed that the *C3* genes were expressed in all these stages, whereas the expression of *fB* and *MASP* genes started from the pre-hatch embryo, which contains the blastula and gastrula (Fig. 2-5, left). Together with the ISH results, this difference in expression timing suggests that the expression of the *fB* and *MASP* genes starts simultaneous with the development of the two germ layers, whereas that of the *C3* genes precedes it. The expression levels of the five genes reached a maximum in the hatched swimming larva, and the high expression level continued to the adult polyp. RT-PCR using the RNA purified from each tissue revealed the expression of all complement genes in the tentacles, pharynx, and body column including the mesentery (Fig. 2-5, right), an expression pattern that is consistent with the ISH result.

In additions, treatment of the *N. vectensis* polyps with the mixture of three pathogen-derived molecules seemed to induce the expression of complement genes, espatially for the *NvC3-1* and *NvfB-1* genes (Fig. 2-6, right). The LPS derived from *P.*

*aeruginosa* had the highest potential to induce these two genes (Fig. 2-6, left). Because expression levels were only compared using different individuals, whether these results detected the induction of gene expression or just detected the individual variations remain unclear. However, WISH using the LPS/zymosan-treated and untreated animals clearly showed signals spread over the entire endodermal layer (Fig. 2-7).

***Protein-level analysis using the polyclonal antibodies raised against the recombinant NvC3-1, NvfB-1, and NvMASP proteins***

Rabbit antisera raised against the insect cells-produced recombinant proteins using the partial sequences of NvC3-1, NvfB-1, and NvMASP (Fig. 2-8A) successfully detected these recombinant proteins (Fig. 2-8B). Specificity of antiserum was determined by the western blotting using the antisera absorbed with different recombinant proteins (negative control, Fig. 2-8B, left lane), or that absorbed with the same recombinant protein used as an antigen (Fig. 2-8B, right lane).

Western blotting using these polyclonal antisera detected multiple bands in all the *N. vectensis* protein extracts tested, the tentacles, body column, and body fluid (Fig. 2-8C). Anti-NvfB-1 antiserum detected bands with corresponding size of the precursor (78 kDa) and processed light chain (51 kDa) of NvfB-1. In contrast, the anti-NvC3-1 beta chain and anti-NvMASP antiserum detected multiple bands, which did not corresponded to the deduced molecular weight of beta chain of NvC3-1 (73 kDa) or that of the full-length or heavy chain of MASP (75 and 48 kDa). However, the ~120 kDa band detected by anti-NvC3-1 antisera was obviously induced by LPS/zymosan treatment (Fig. 2-8), in accord with the result obtained by the RT-PCR/WISH analysis (Figs. 2-6 and -7), implying the possibility of the post-translational modification of NvC3-1.

Nevertheless, all the attempts for purifying these antisera-positive bands were unsuccessful, due to difficulty for collecting the enough amount of proteins from the body fluid or for solublizing these proteins from the tissue homogenates.

**Discussions**

The identification of the *C3, fB,* and *MASP* genes from Cnidaria indicated the unexpectedly ancient origin of the multi-component complement system, which can be traced back to before the divergence of Cnidaria from the Bilateria (Triploblastica) lineage more than 600 million years ago in the Precambrian era (23). In contrast, a similar search using the draft genome/EST database of the non-eumetazoan animals, porifera and choanofllagelata, did not detect not only any member of the five complement gene families but also the any domains of TEP genes. Thus, it is likely that the complement system was established in the common ancestor of eumetazoan animals by the de novo creation of the *C3/TEP* gene and the establishment of the domain composition of *fB* and *MASP* genes by the shuffling of the preexisting domains.

The identified cnidarian complement components retained most of the domain structures and amino acid residues shown as functionally critical in mammals, suggesting that these components conserved the basic biochemical functions throughout metazoan evolution. All five complement components of *N. vencensis* retained the putative activation cleavage site susceptible to the trypsin-type SP. On the other hand, NvfB-1, NvfB-2, and NvMASP retained all the amino acid residues within the SP domain essential for the trypsin-type proteolytic activity. In addition, all domains required for the binding of mammalian fB to C3 were retained in NvfBs, suggesting that NvC3s is the physiological substrate of NvfBs. Similarly, NvMASP is expected to be the first enzyme of the cascade, although the physiological substrates of the human MASP-1, which shares structural characteristics with NvMASP, is still debated (47-49). Together with the colocalization of their mRNA in the same tissues, this evidence suggests that these cnidarian components form the primitive proteolytic network similar

to the lectin and alternative pathways of the vertebrate complement system. The finding of the MASP gene in nonchordate is particularly important, indicating the ancient origin of the lectin pathway as old as that of alternative pathway. The possible presence of the upstream initiation factors for the three mammalian complement activation pathways, properdin, factor D, MBL, C1q, or ficollin, are still to be clarified, although these genes are difficult to identify from the domain structures, which are too simple and ubiquitous. The absence of *C6* and *fI* family genes in Cnidaria is consistent with the expected evolutionary origin of these families in the vertebrate lineage (50) and indicates the absence of the cytolytic and fI-dependent regulatory pathways in the primitive complement system of cnidaria.

The most prominent feature found uniquely in cnidarian C3 is the highly KR-rich insertion in the MG8 domain. It is plausible that this highly cationic patch, which is probably exposed on the molecular surface of the cnidarian C3s (42), interacts with the negatively charged molecules on microbes, as do some antimicrobial peptides (51). Another curious feature found only in cnidarian C3s is the lack of a disulfide bridge linking the alpha and beta chains (40), suggesting that the two subunit chains of the mature cnidarian C3 are held together only by noncovalent forces. Cnidarian fB and MASP also retained some ancestral features, which were lost in the jawed vertebrate orthologoues. The *N. vectensis* fBs, together with some invertebrates and lamprey fBs retained the canonical activation cleavage site at the N-terminal end of the SP domain and the Asp residue at the bottom of the S1 pocket, both of which were lost secondarily in the jawed vertebrate fB/C2 (43). In most trypsin-type SPs, cleavage at the canonical site induces a dynamic conformational change in the SP domain leading to transformation of the oxyanion hole and the substrate binding site into an active form

(52). Therefore, the loss of the canonical cleavage site and the drastic structural reorganization of the SP domain, including the substitution of the Asp residue in the S1 pocket, is likely to have occurred simultaneously in the jawed vertebrate fB/C2, creating a novel activation mechanism (53). Similarly, the cnidarian MASP, together with invertebrate MASPs and vertebrate MASP-1, shared some canonical features of the SP domain, which were lost in the vertebrate MASP-2, MASP-3, C1r, and C1s, indicating that the ancestral MASP was an MASP-1 type molecule.

While retaining the several ancestral features, the cnidarian *C3* and *fB* genes showed the closest phylogenetic relationship to those of amphioxus, which is believed to be the most basal chordates, rather than to those of other nonchordates. These results suggest that the ancestral amino acid sequences of these complement genes are well conserved in cnidarians and chordates. In addition, recurrent loss of the complement genes seems to have occurred in protostomes, as indicated by the complete absence of the complement genes in the genome of *Caenorhabditis elegans* or *Dorosophila melanogaster*. Our results indicate that the complement genes are no exception to the postulated tendency of the cnidarian genome (29), closer resemblance to that of vertebrates rather than those of protostomes.

The ISH analysis indicated that the cnidarian *C3, fB,* and *MASP* genes are expressed in the three endodermal tissues of adults, the endoderm of the tentacle and pharynx, and the mesentery. This result is compatible with a previous report on *C3* gene expression in the undifferentiated endodermal cells of the coral embryo and larva (22). All endodermal cells in the tentacle expressed these genes, whereas these genes were expressed only in particular cells, possibly the secretary cells or phagocytes in the pharynx endoderm and mesentery. These results indicate that the five components are

secreted into the coelenteron, the only cavity facing these endodermal tissues. As indicated by ISH analysis, the expression level of these five genes was higher in the tentacle endoderm than in the mesentery or pharynx, although the mesentery is the only internal organ of cnidarians, which has multiple functions (26). This result indicates that the primary site for complement action is the tentacle cavity, which is connected with the coelenteron. The tentacle cavity seems to form the partially separated space within the coelenteron suitable for the storage and concentration of the serum proteins to protect them from the inactivation by digestive enzymes such as trypsin or dilution by sea water. This is the first report implying the critical function of the internal space of the tentacle, whose homologous tissue is not found in other animals.

Supporting this hypothesis, the antigenicities against the recombinant *N. vectensis* complement proteins were also detected in the body fluid and tentacle homogenates. For the identification of these antisera-positive protein bands, MS fingerprint or N-terminal amino acid sequencing using the purified proteins is necessary. For animals developed three germ layers (triploblast), serum complement proteins are highly concentrated in the blood or body fluid, making it easy to purify these chemically unstable proteins only by the several steps of liquid chromatography. In cnidarian, however, it was difficult to collect enough amount of proteins from their sea water-diluted body fluid, and extraction of these proteins from tissue homogenates was also unsuccessful. Using cnidarian with larger size (e.g. jelly fish) seems to be the best way to overcome these problems.

In conclusion, the endodermal cells-specific expression of the cnidarian complement genes indicates the existence of the complement system within the cnidarian gut, the coelenteron. This idea is consistent with the expectation that the cnidarian coelenteron

serves as the digestive tract with an alternative function as the primitive circulatory system. The humoral complement system of vertebrates is thought to have evolved from the primitive complement system within the gut cavity of two germ-layered ancestor similar to cnidarians along with the establishment of the circulatory system by the third germ layer, the mesoderm.

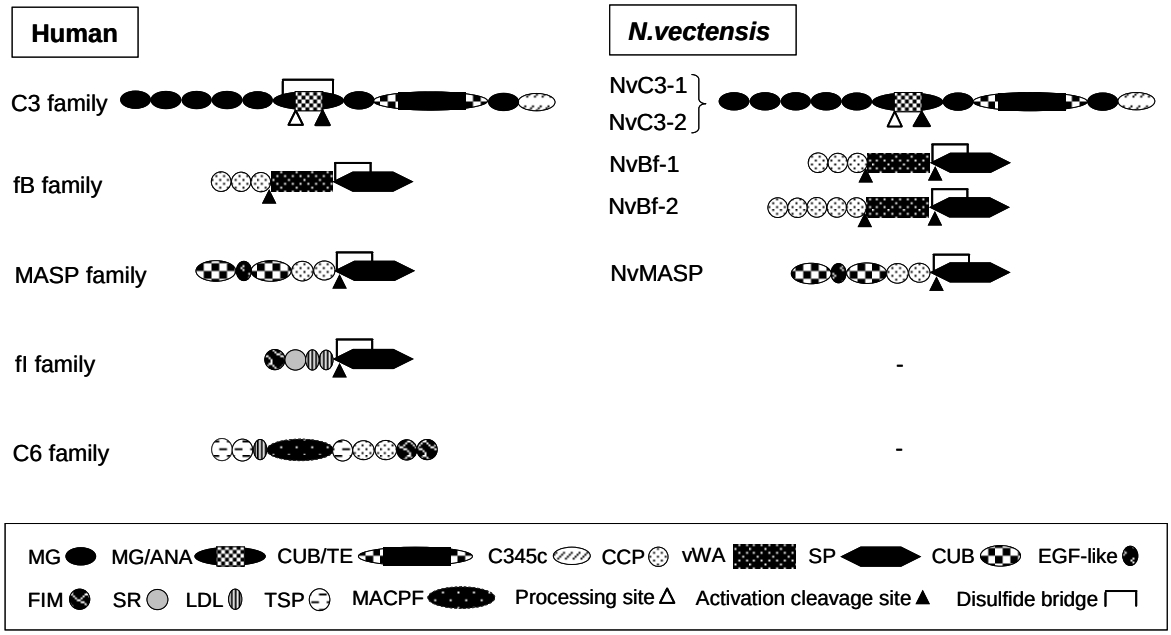



**Fig. 2-1.** Schematic drawing of the unique domain structures of the five complement families of human and that identified from *N. vectensis*. Domain structures of the five complement families of human (left) and *N. vectensis* (right) were compared. For the human C6 family, only the domain structure of C6 is shown, since the C7, C8A, C8B, and C9 lack some of the domains found in C6. Abbreviations of domain names are indicated in the figure legend of Fig. 1-2 (p. 11). The processing site (open triangle), activation cleavage site (solid triangle), and the disulfide bridge linking two subunit chains generated by processing (line) are also indicated.

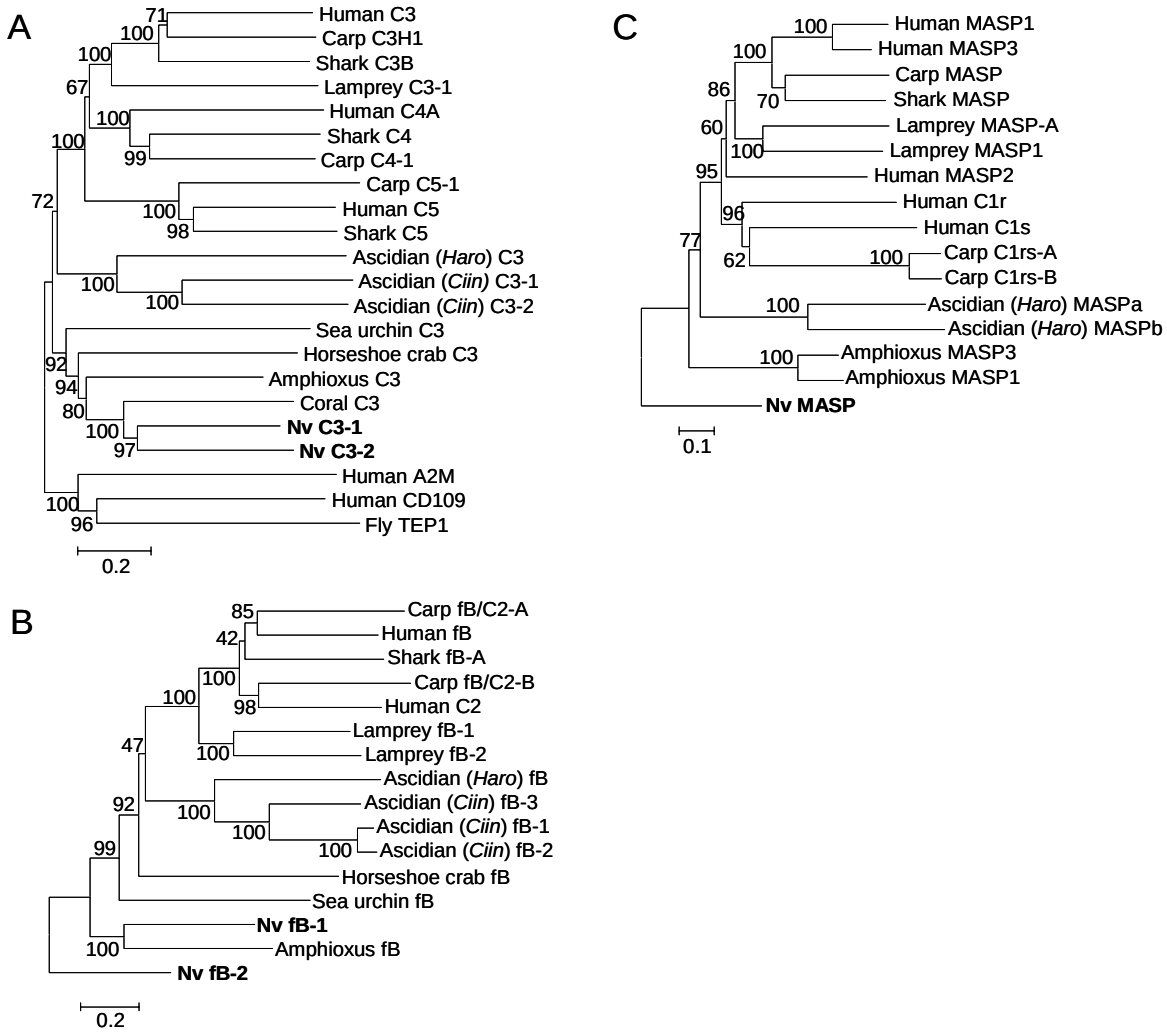



**Fig. 2-2.** Phylogenetic trees of the three complement gene families. Trees were constructed based on the alignment of the full length amino acid sequences of *C3*, *fB*, and *MASP* family genes, using the NJ method excluding gaps by pair-wise deletion. Bootstrap percentages are given. Accession numbers of each entry are listed in table 1-1 (p. 11). (A) Phylogenetic tree of the *C3* family. (B) Phylogenetic tree of the *fB* family. (C) Phylogenetic tree of the *MASP* family.

**Fig. 2-3.** Multiple alignment of the amino acid sequences of the *N. vectensis* complement components with those of representative species. Names and boundaries of domains are shown above the sequences. Abbreviations of the domain names are indicated in the legend of Fig. 1-2. (p. 11) (A) Alignment of NvC3-1 and NvC3-2 with human C3, C4A, and C5, and coral (co) C3. Only one of two putative alleles for *NvC3-1* gene (AB450038) is shown, and eight amino acid substitutions observed in the other allele (AB450039) are shown above the NvC3-1 sequence. Alpha/beta processing site, activation cleavage site, fI cleavage sites, thioester site, catalytic His, and alpha/gamma processing site are boxed. Conserved Cys residues in the C3a anaphylatoxin region are marked (*), and the disulfide bridge between the alpha/beta chains are shown by lines. Signal peptide and KR-rich region are underlined.

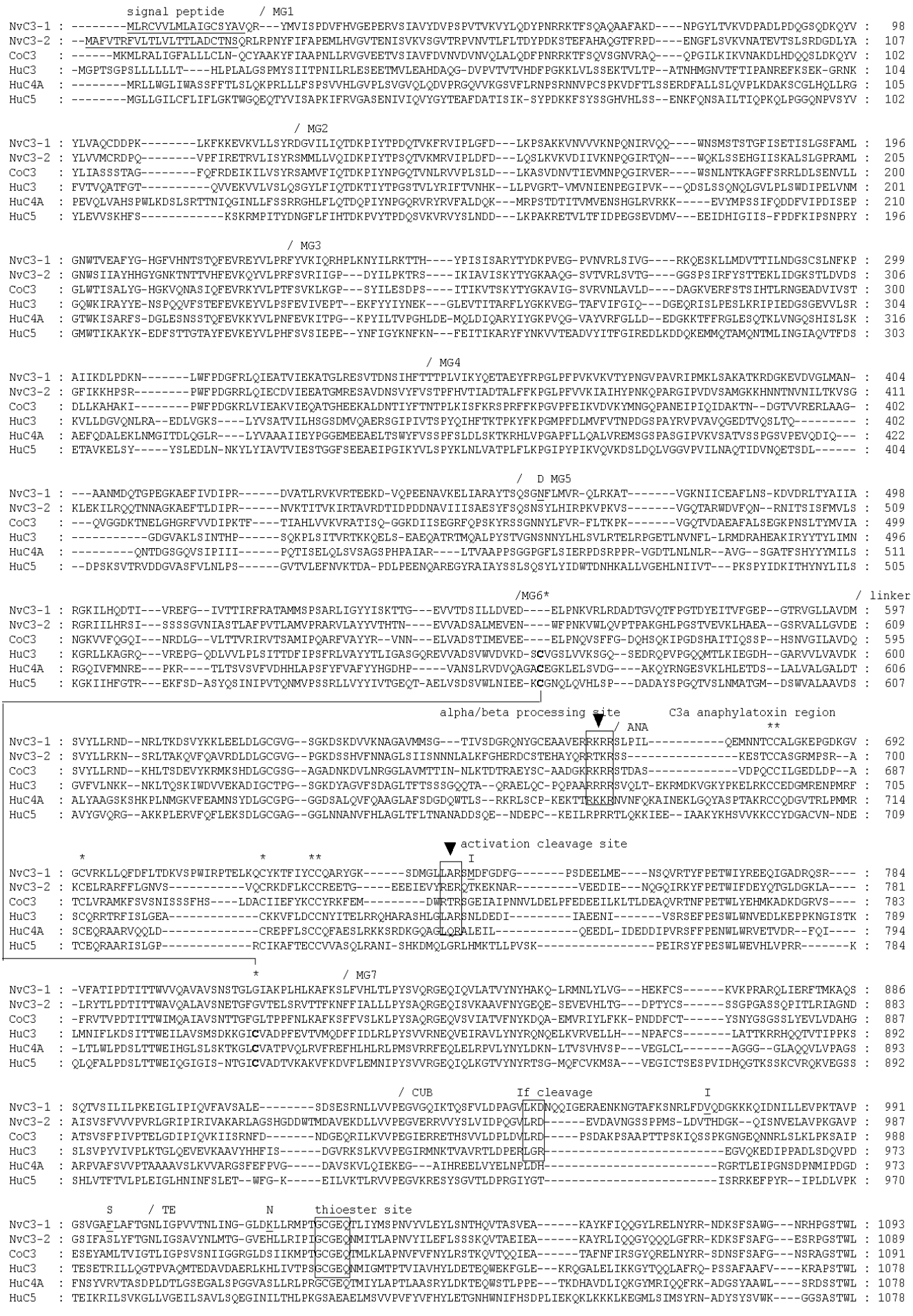

signal peptide
/ MG1
/ MG2
/ MG3
/ MG4
/ D MG5
/MG6*
/ linker
alpha/beta processing site
C3a anaphylatoxin region
/ ANA
activation cleavage site
I
/ MG7
/ CUB
If cleavage
I
S
/ TE
N
thioester site

**Fig. 2-3A.**


**Fig. 2-3A. (Continued)**

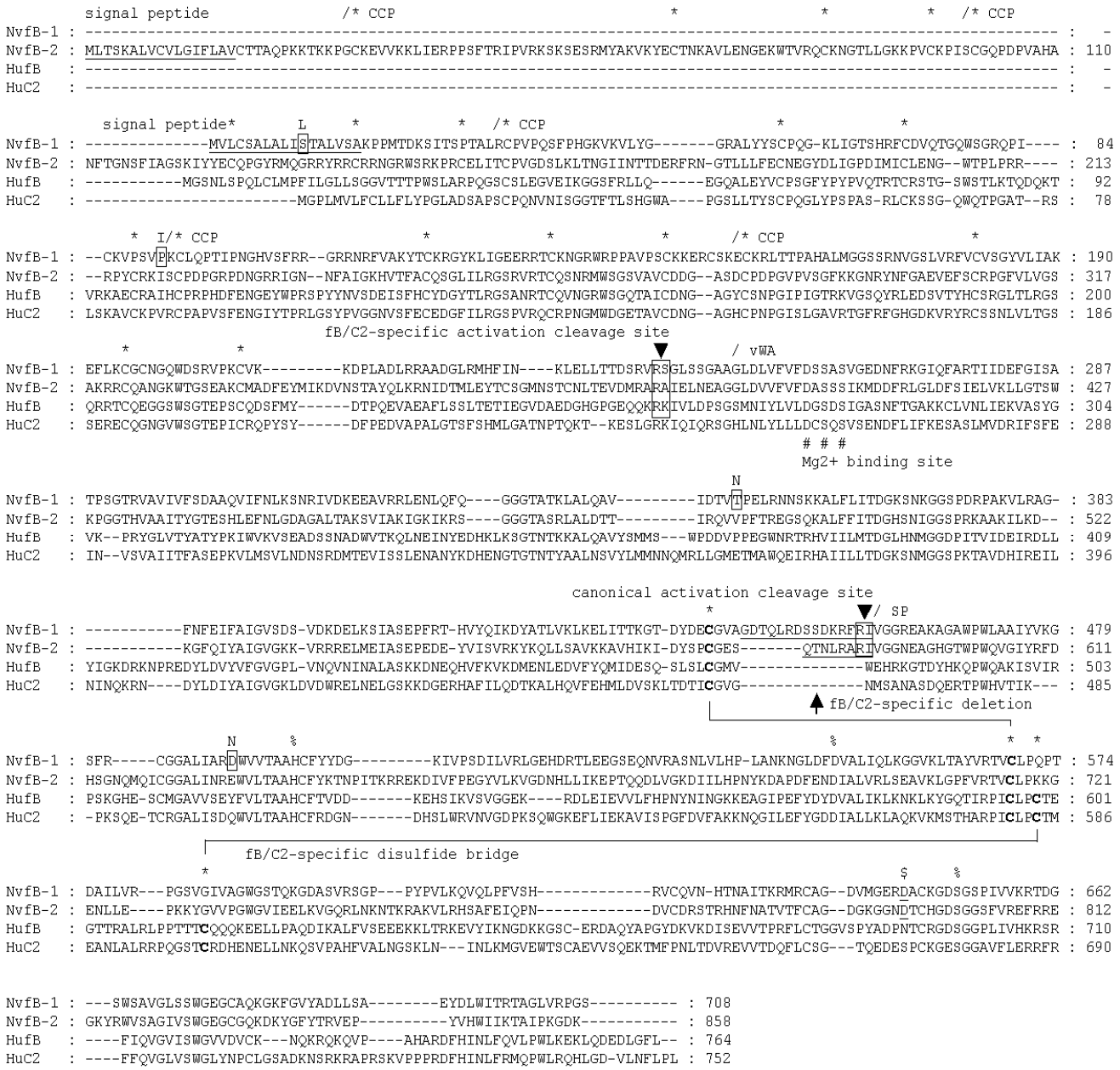



**Fig. 2-3.** (B) Alignment of the NvfB-1 and NvfB-2 with human fB/C2. Only one of two putative alleles for *NvfB-1* gene (AB450041) is shown, and four amino acid substitutions observed in the other allele (AB450040) are shown above the NvfB-1 sequence. The canonical and jawed vertebrate fB/C2-specific activation cleavage sites are boxed. Conserved Cys residues in the CCP domain (*), $Mg^{2+}$ binding site residues (#), catalytic triad residues (%), and Asp residue at the bottom of the S1 pocket of SP domain ($) are marked. Two disulfide bonds, one possibly linking the two chains generated by the canonical cleavage site, and the other unique to the jawed vertebrate fB/C2, are shown by lines. Signal peptide, and the several amino acid residues around the canonical activation cleavage site of NvfB-1 and -2, which are missing from the jawed-vertebrate fB/C2 are underlined.


**Fig. 2-3.** (C) Alignment of NvMASP with the human MASP family components. Activation cleavage site and active site Ser encoded by the TCN codon are boxed. Cys residue conserved in the CUB, EGF-like, and CCP domains (*), catalytic triad residues (%), and Asp residue at the bottom of the S1 pocket of SP domain ($) are marked. Disulfide linkage between the heavy and light chains and His loop are shown by lines. Signal peptide is underlined.

**Fig. 2-4.** *In situ* hybridization for the five *N. vectensis* complement genes. (A) Schematic cross-section structures of the cnidaria and triploblastica. Cnidaria (diploblastica) lacks the mesoderm, body cavity, and blood vessels. (B) WISH for the five complement genes using adult *N. vectensis* polyp. All of the five complement genes showed completely the same expression pattern in the whole-mount samples (left) and in the cross sections of WISH samples at the tentacles, pharynx, and body column including mesentery (right). In whole-mount samples, signals were observed at the endoderm of the tentacle tip (te), pharynx (ph), and mesentery (me) (black arrows). In sections, broad signals were detected in the tentacle endoderm (te.end), and cell-type specific signals were detected in the pharynx endoderm (ph.end) and the middle part of mesentery (me) (black arrows). No signals were detected in the ectoderm (ect) or mesoglea (meso) (white arrows). For the control *paxA* gene, signals were detected at the body colum ectoderm (bc.ect) and tentacle ectoderm (te.ect) in WISH, and at the tentacle ectoderm (te.ect) and the edge of mesentery (me) in sections (black arrows).

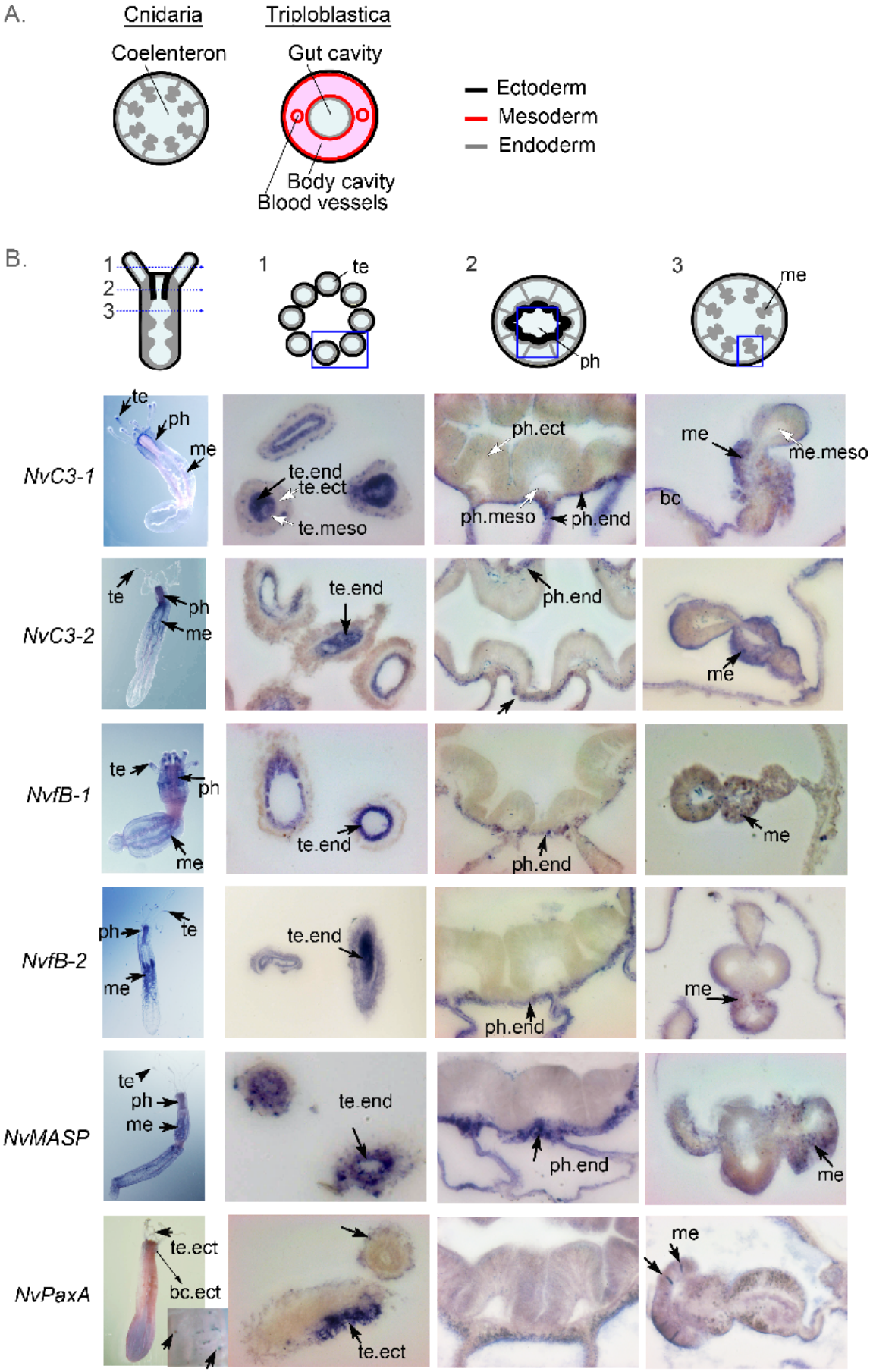

A.
Cnidaria
Triploblastica
Coelenteron
Gut cavity
Body cavity
Blood vessels
Ectoderm
Mesoderm
Endoderm
B.
1
2
3
te
ph
me
NvC3-1
te.end
te.ect
te.meso
ph.ect
ph.meso
ph.end
me.meso
bc
NvC3-2
NvfB-1
NvfB-2
NvMASP
NvPaxA
bc.ect

**Fig. 2-4.**

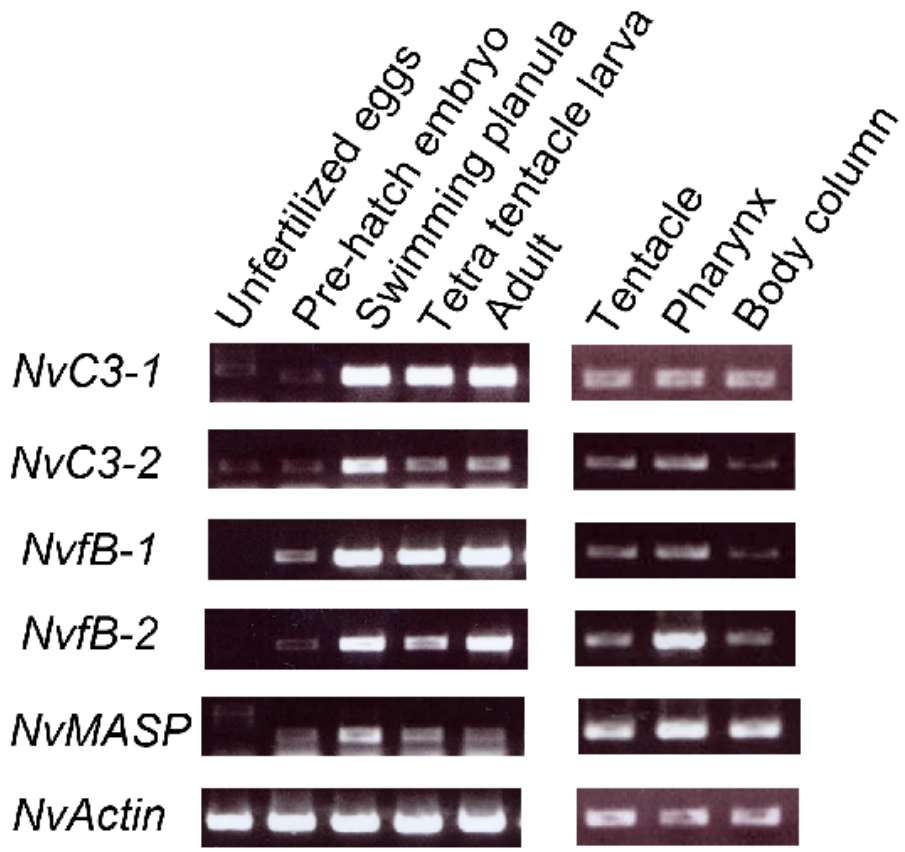



**Fig. 2-5.** Temporal and spatial expression of the five *N. vectensis* complement genes. RT-PCR for the five complement genes were performed using RNA purified from the *N. vectensis* at five different developmental stages (left) or three different parts of adult body (right).

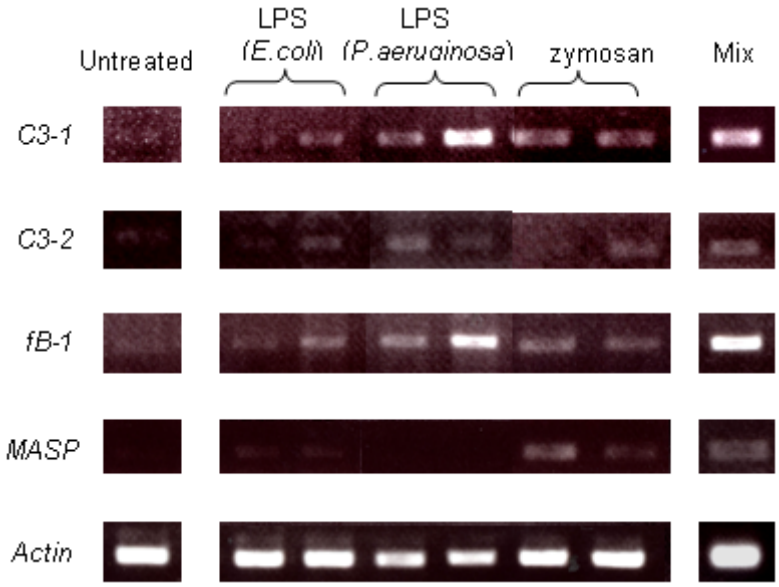



**Fig. 2-6.** RT-PCR using *N. vectensis* treated with pathogen derived molecules. Adult *N. vectensis* polyps treated with 100 μg/ml of LPS from *P. aeruginosa* or *E. coli*, 10 mg/ml of zymosan, or mixture of these agents were subjected to the RT-PCR for the four complement genes. Experiments were not performed for *NvfB-2*.

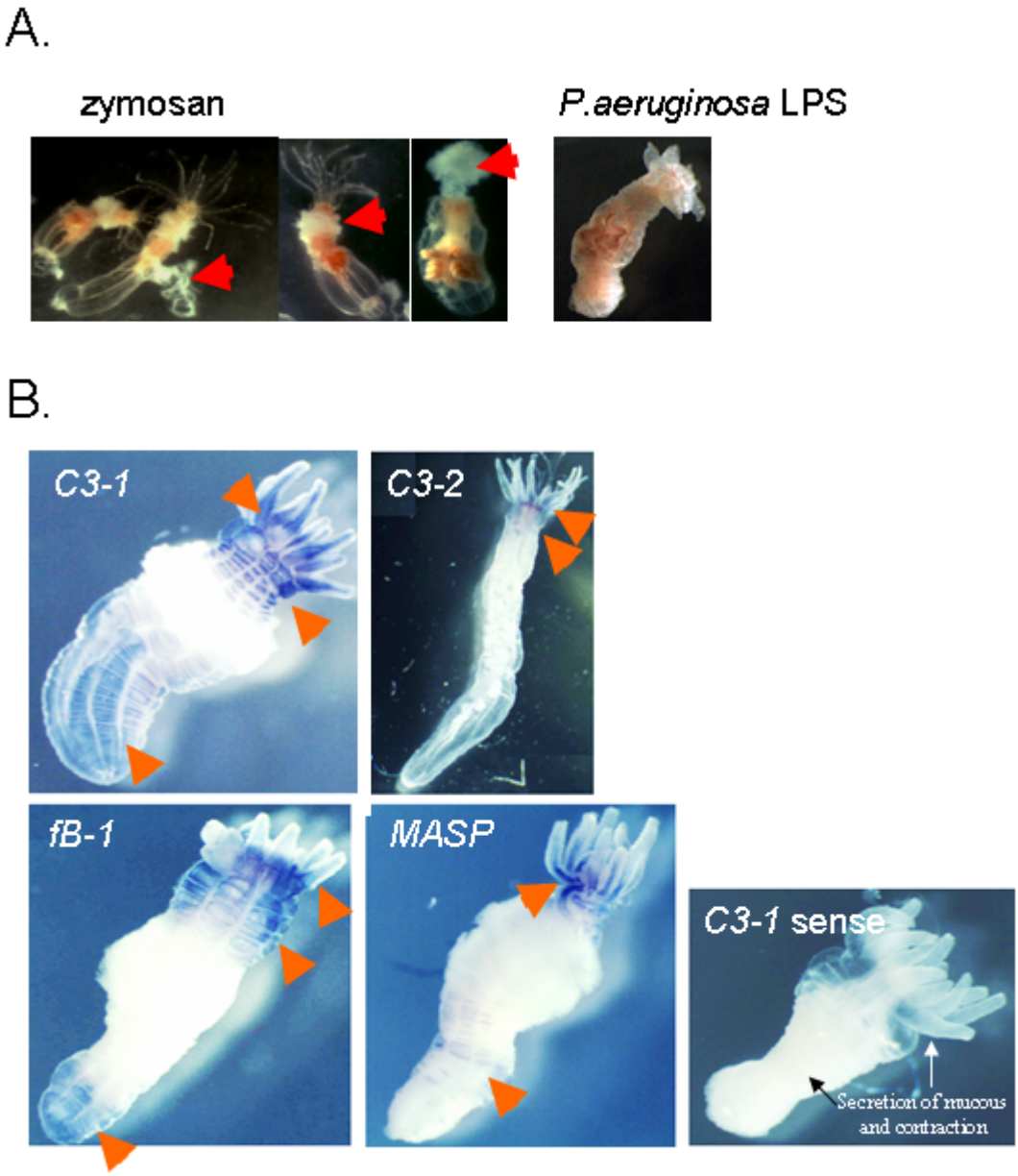



**Fig. 2-7.** WISH after the LPS/zymosan treatment. (A) The *N. vectensis* polyps treated with the *P. aeruginosa* LPS or zymosan showed inflammation-like appearance (e.g. shrinkage of the body and mucous secretion). (B) WISH for the complement genes using *N. vectensis* treated with the mixture of the LPS/zymosan (same amount in Fig. 2-6). Experiments were not performed for *NvfB-2*.

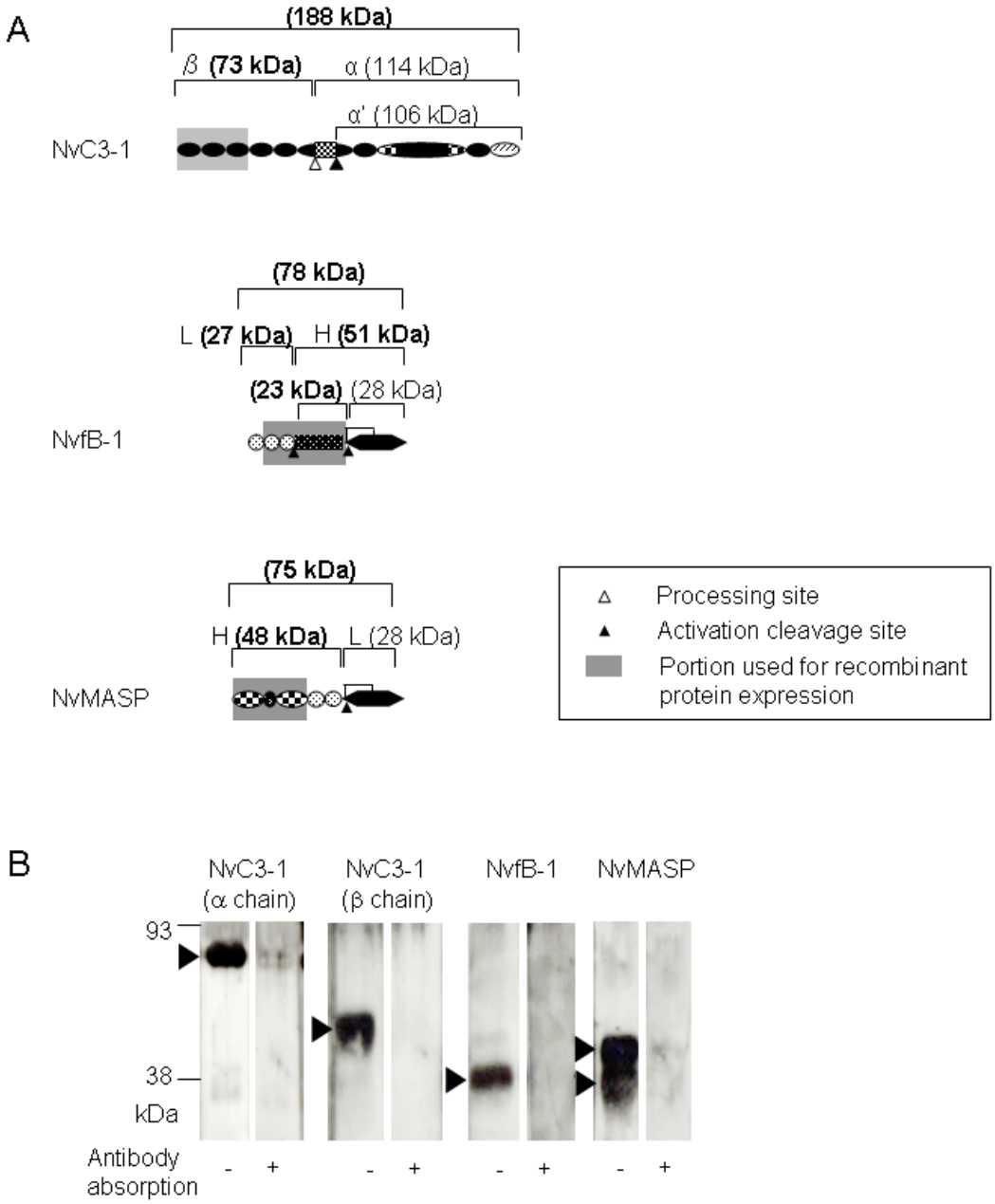



**Fig. 2-8.** Protein-level analysis of the cnidarian complement components. Western blotting using the anti-recombinant *N. vectensis* complement antiserum. (A) Antiserum was produced against the recombinant NvC3-1, NvfB-1, and NvMASP proteins using the paritial deduced cDNA sequences indicated by shadow. Deduced molecular weights (kDa) of the precursor or processed polypeptides are indicated. (B) Western blotting performed against the recombinant NvC3-1 (α/β chains), NvfB-1, and NvMASP proteins. Anti-NvC3-1 (α/β chains), NvfB-1, and NvMASP antisera absorbed with different recombinant proteins still detected original antigen (left), whereas that absorbed with original antigen failed to detect them (right).

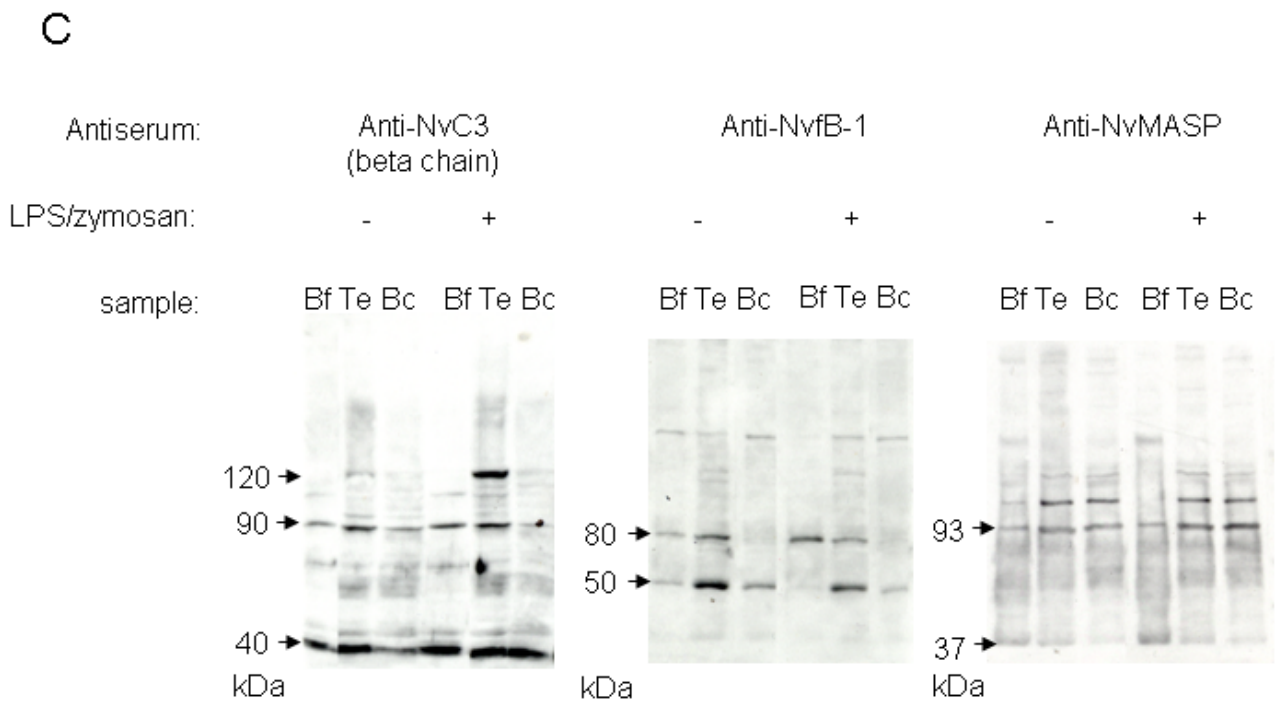


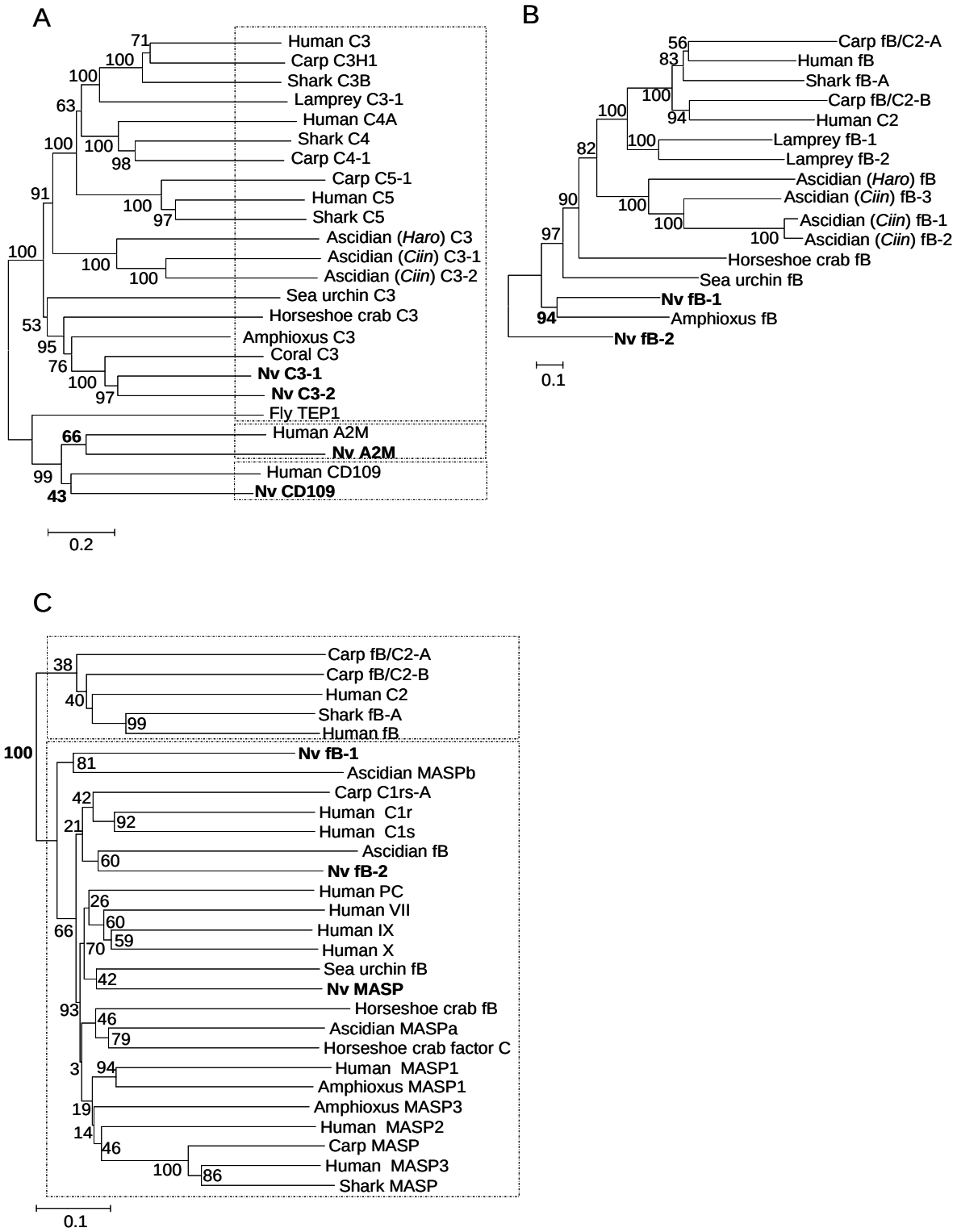

**Fig. 2-8.** Protein-level analysis of the cnidarian complement components (Continued). (C) Western blotting against the proteins extracted from the body fluid (Bf), tentacles (Te), and body column (Bc) of *N. vectensis* polyps with (right) or without (left) the LPS/zymosan treatment was performed. Anti-recombinant NvC3-1 (β chain), NvfB-1, and NvMASP antisera detected multiple bands in all the protein fraction extracted from *N. vectensis*, whereas anti-recombinant NvC3-1 (α/β chains) antisera detected no bands (not shown).

**Table 2-1.** Presense/absence of protein modules comprising the five complement gene families in various taxa.

| | Organizm | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Domain | Deuterostome ¶ | Protostome ¶ | Cnidaria † | Placozoa † | Porifera † | Choano-fllagelata † | Plant ¶ | Fungi ¶ | Archea ¶ | Eubacteria ¶ |
| A2M_N * | + | + | + | - | - | - | - | - | + | + |
| A2M_N_2 * | + | + | + | - | - | - | - | - | + | + |
| ANA * | + | + | + | - | - | - | - | - | - | - |
| A2M * | + | + | + | + | - | - | - | - | + | + |
| A2M_comp * | + | + | + | + | - | - | - | - | + | + |
| A2M_recep * | + | + | + | + | - | - | - | - | + | + |
| C345c | + | + | + | + | - | - | - | - | - | - |
| CCP | + | + | + | + | + | + | + | - | - | + |
| vWA | + | + | + | - | + | + | + | + | + | + |
| SP | + | + | + | + | + | - | + | + | + | + |
| CUB | + | + | + | + | + | - | + | - | - | + |
| EGF | + | + | + | + | + | + | + | + | - | + |
| TSP | + | + | + | + | + | + | + | - | - | + |
| LDLa | + | + | + | + | + | - | + | + | - | + |
| MACPF | + | + | + | - | - | - | + | + | - | + |
| FIM | + | - | - | - | - | - | - | - | - | - |
| (Kazal) | + | + | + | - | - | - | + | + | + | + |
| SR | + | + | + | + | + | + | + | - | - | - |

For the C3 family, domain names defined by pfam (not by SMART) was indicated here. Presence (+)/absence (-) of domains detected by BLAST search† or by SMART/pfam program¶ are shown. Abbraviations of the domain names of C3 family defined by pfam are as follows: A2M_N, A2M N-terminal; A2M_N2, A2M N-terminal 2; ANA, anaphylatoxin/fibulin; A2M, alpha-2-macroblobulin; A2M_comp, A2M complement component; A2M_recep, A2M receptor-binding.

Domains comprising C3 family were basically restricted in eumetazoan animals. Some eubacteria and archean *M. burtonii* also have A2M/CD109-like genes lacking the ANA and C345c domains, however, they are thought to be transferred horizontally from eukaryotes. In contrast, almost all the domains comprising the other complement genes showed ubiquitous distribution in most taxa.




**Supplementary Fig. 2-1.** Phylogenetic trees of the complement genes. Sequences used here are the same with those used in the Fig. 2-2. All trees were constructed by NJ method. (A) Phylogenetic tree of the *C3* family including the partial deduced amino acid sequences of the *NvA2M* and *NvCD109* genes. (B) Phylogenetic tree of the *fB* gene family using only the vWA and SP domains. (C) Phylogenetic tree of the various complement and coagulation SP proteins, using the SP domain sequences.

# Part II

# Comprehensive cloning of the complement genes of agnathan lamprey by the liver EST analysis

**Abstract**

The complement and coagulation systems in mammalian blood are composed of multiple components with unique domain structures, and are believed to be established by exon-shufflings and following gene duplications. To elucidate the evolutionally origin of these two apparently similar reaction systems in vertebrates, liver EST and 5'- and 3'- RACE analyses were performed in lamprey, *Lethenteron japonicum*. For the complement system, the *factor I* cDNA was cloned for the first time outside of the jawed vertebrates. Evidence for the *C3*/*C4*/*C5*, *fB*/*C2* and *MASP-1*/*MASP-2*/*C1r*/*C1s* gene duplications was not found, suggesting that these duplications occurred in the jawed vertebrate lineage. In contrast, the coagulation factors VII and X, prothrombin and protein C-like cDNAs were identified, indicating that duplications among them predated the cyclostome-jawed vertebrate divergence. The genes for terminal complement components (C6 family), coagulation factors XI and XII, or prekallikrein were not found, suggesting that the complement and coagulation systems of an ancestral vertebrate were simpler compared to their mammalian counterparts.

## Introduction

Phylogenetic studies have been accumulating to understand the origin and evolution of the two apparently similar reaction systems, the complement and coagulation systems. Recent completion of genome projects of some teleost (54) and ascidian (13) species made it possible to identify fish or ascidian orthologs of the mammalian complement and coagulation genes, although the teleosts have also experienced their own gene duplications, increasing the numbers of orthologs corresponding to each duplication products found in mammals. These data reveal that the complement and coagulation systems developed by creation of new genes and duplications of the preexisting genes in the vertebrate lineage after the divergence of urochordates (15, 55). However, the lack of detailed information in lower vertebrates makes it difficult to decide whether the significant developments of the complement and coagulation systems occurred before or after the emergence of jawed vertebrates.

The genome analysis of *Ciona intestinalis* indicated that the most complement gene families characterized with unique domain architectures, the *C3*, factor B (*fB*), mannan-binding protein-associated serine protease (*MASP*), and terminal complement component (*TCC* or *C6*) families, are present in the ascidian (15). However, evidence for the *C3/C4/C5*, *fB/C2*, *MASP-1/MASP-2/C1r/C1s* or *C6/C7/C8A/C8B/C9* gene duplications, which probably played important roles in establishing the classical and cytolytic complement pathways, has not been recognized in the ascidian genome. In contrast, evidence for these gene duplications is obvious in the teleost genome, indicating that these gene duplications occurred early in vertebrate evolution (50). It is not clear, however, whether or not these gene duplications predated the emergence of jawed vertebrates.

The coagulation system of higher vertebrates is another proteolytic activation cascade system in blood, involved in the formation of solid clots at the site of injury. Two activation pathways, the contact-activation and tissue-factor pathways, merge at the limited proteolytic activation step of Factor X (FX). Activated FX catalyzes the limited proteolytic activation of prothrombin (PT) into thrombin, which in turn converts fibrinogen into fibrin, the building block of a hemostatic plug. Phylogenetic biochemical studies performed thus far have indicated that blood coagulation involving the basic components, tissue-factor, PT and fibrinogen is present in fish including cyclostomes (56). In addition, sequences for the most coagulation factor genes were identified from the genome of puffer fish, *Takifugu rubripes*, except for the genes in the contact-activation pathway, factor XI (*FXI*), factor XII (*FXII*) and *prekallikrein* (55, 57). In contrast, no obvious orthologs of mammalian coagulation genes have been identified in the urochordate *C. intestinalis* genome (55). These results indicate that most coagulation factor genes, and the coagulation system itself, was generated and developed in the vertebrate lineage before the emergence of bony fish. In addition, identification of the sequence for coagulation factor genes from the lamprey (*Petromyzon marinus*) trace database has just been reported (58), although the entire coding sequences and expression of the predicted genes remain to be clarified.

In this part, I performed EST and 5'- and 3'- RACE analyses using lamprey (*Lethenteron japonicum*) liver mRNA, to obtain comprehensive structural information on the complement and coagulation system genes in cyclostomes, which diverged from the jawed vertebrate lineage approximately 600 million years ago (59).

## Materials and Methods

### *Poly(A)$^+$ RNA extraction*

*L. japonicum* was purchased from a local dealer in Ebetsu, Hokkaido, Japan. RNA was isolated from liver using guanidine thiocyanate method (60), and poly(A)$^+$RNA was isolated by an oligo(dT) cellulose type 7 column (GE Healthcare).

### *Construction of a non-normalized cDNA library*

Five μg of liver poly(A)$^+$ RNA from one individual of lamprey was used for the construction of a directional cDNA library using the pBluesscript II XR cDNA library construction kit (Stratagene). A total of 2,371 colonies with insert sizes ranging from 0.5 to 2 kb were subjected to the sequence analysis.

### *Construction of a normalized cDNA library*

Double stranded-cDNA was synthesized using liver poly(A)$^+$ RNA from 25 individuals of lamprey by the BD SMART$^{TM}$ PCR cDNA synthesis kit (BD Biosciences Clontech). Using the Trimmer cDNA normalization kit (Evrogen), 1.2 μg of cDNA was normalized, which was then cloned into the pCR 2.1 TOPO-TA cloning vector (Invitrogen). A total of 9,619 colonies with insert sizes ranging from 0.5 to 2 kb were subjected to the sequence analysis.

### *Sequence analysis of cDNA clones*

Plasmid DNA was extracted by the Multiscreen-FB plates (Millipore) and used as a template for the DNA sequencing reaction with the BigDye terminator cycle sequencing ready reaction (PE Applied Biosystems). 5'-end of non-normalized cDNA clones was

sequenced using T7 primer, and one end of normalized, non-directional cDNA clones was sequenced using M13R or M13F primer, by a 3100 or 3130 ABI PRISM DNA sequencer (PE Applied Biosystems). Nucleotide sequences of 11,521 clones in total were vector-masked and assembled by EG assembler (http://egassembler.hgc.jp/). Obtained 1,274 contigs and 2,702 singletons were analyzed by the BLAST-X (DNA vs protein) search against the non-redundant protein database at the National Center for Biotechnology Information. 5'- and 3'- Rapid amplification of cDNA ends (RACE) was performed for newly identified complement and coagulation factor cDNAs to obtain the entire coding information using the Smart RACE cDNA amplification kit (Clontech).

***Phylogenetic Analysis***

Multiple alignment of the full length of amino acid sequences was performed using ClustalX (32). Based on the alignment, a phylogenetic tree was constructed by the neighbor-joining (NJ) method (33) using MEGA4 (34), excluding gaps by pair-wise deletion. Pairwise genetic distances were obtained by poisson correction. The degree of support for internal branches was assessed by the 1,000 bootstrap replicates. Additional phylogenetic trees were constructed based on the same alignment by the maximum parsimony (MP) method using MEGA4, not excluding position with gaps, and maximum likelihood (ML) method using Tree-puzzle5.2 (61) by applying the default setting. The NJ, MP and ML trees were compared, and only the internal branches appeared in all three trees were considered to be the consensus internal branches. Only NJ trees are shown with the bootstrap percentages for consensus internal branches in boldface.

## Results

### *EST analysis*

A total of 2,372 and 9,149 clones were analyzed from the non-normalized and normalized cDNA libraries, respectively. After clustering, 1,274 contigs and 2,702 singletons were analyzed by BLAST-X searches. ESTs for more than 50 possible serum protein genes (62) were identified as summarized in Table 3-1. The schematic structures predicted from the identified cDNAs of lamprey complement and coagulation genes with unique domain architecture are summarized in Fig. 3-1.

### *Complement genes*

Thus far, two *C3* (41), one *fB* (45) and three *MASP* (17, 63) genes have been reported from *L. japonicum*, whereas no information has been reported for the *TCC* and *fI* families. The present analysis identified the cDNA for one *fI* gene (Fig. 3-2A) for the first time from outside of jawed vertebrates, and revealed the cDNA for one novel gene in the *C3* (Fig. 3-2B) and *fB* families (Fig. 3-2C), respectively.

### *Factor I gene*

Three cDNA clones were obtained for the *fI* gene (Table 1), and its entire coding sequence was elucidated by 5'-RACE-PCR, revealing that the domain architecture of lamprey fI is exactly the same with that of jawed vertebrate fIs (64), FIM (factor I/membrane attack complex)/SR (scavenger receptor Cys-rich)/LDL (low-density lipoprotein receptor domain class A) /LDL/SP (serine protease) (Figs. 3-1, 3-2A). Since the FIM domain has not been found in the elucidated genomes of invertebrates, and this domain reported in sea urchin was not predicted by SMART or pfam programs, this is

the first report of not only *fI* but also the typical FIM domain from animals other than jawed vertebrates. Alignment of the lamprey fI amino acid sequence with those of jawed vertebrates (Figure 3-2A) showed conservation of most of the invariable amino acids of jawed vertebrate fIs, including Arg-X-X-Arg at the processing site, the Cys residues involved in the disulfide bridge between the heavy and light chains, the His, Asp and Ser residues of the catalytic triad, and the Asp residue at the bottom of the specificity pocket of the SP domain. Upon the phylogenetic tree analysis, lamprey fI formed a cluster with jawed vertebrate fIs, although the phylogenetic relationship within this cluster was not resolved (Figure 3-3A).

***C3 family***

Three cDNA contigs termed *C3-1*, *C3-2*, and *C3-3* were obtained (Table 3-1, Fig. 3-2B) by clustering the ESTs. The *C3-1* contig most probably represents an allelic variant of an already reported *C3* (Q00685) (41), since they share 99% amino acid sequence identity. Similarly, the *C3-2* contig was judged as an allelic variant of another *C3* sequence deposited in the database (BAA02763) based on 97% amino acid sequence identity. In contrast, the *C3-3* contig showed only 88-89% amino acid sequence identity to them. Thus, we tentatively concluded that the *C3-1*, *C3-2* and *C3-3* represent three independent genes rather than allotypes, although the final conclusion should wait for segregation analysis or completion of genome analysis. Amino acid sequence alignment of C3-1, C3-2 and C3-3 showed the perfect conservation of the functionally important residues at the beta-alpha processing site, the C3a region, the C3 convertase cleavage site, and the thioester site (Fig. 3-2B). Upon phylogenetic tree analysis, three lamprey *C3* cDNA contigs formed a clade supported by the highest bootstrap percentage (Fig.

3-3B). The genetic distance among these three lamprey *C3*s was small, suggesting that they arose by recent gene duplications within the lamprey lineage. The lamprey *C3s* together with jawed vertebrate *C3* and hagfish *C3* (65) formed a clade, suggesting that the *C3/C4/C5* gene duplications occurred in the vertebrate lineage before the divergence of the jawed and jawless vertebrates. Human *C3*, *C4*, and *C5* genes are located in the MHC and its paralogous regions, and are believed to be generated by the two-rounds of whole genome duplication (8). The result of our phylogenetic tree analysis is compatible with this hypothesis. However, no *C4* or *C5* cDNA was identified by the present EST analysis of lamprey liver, in spite of identification of 116 *C3* ESTs. Since serum concentrations of human C4 and C5 are about 1/2 and 1/20 of that of C3, respectively (66), it is suggested that there is no *C4* or *C5* gene in lamprey. Supporting this conclusion, the *C4* or *C5*-like sequences are also missing from the draft genome sequence of another lamprey species, *P. marinus* (Genome sequencing center at Washington University, BLAST URL: http://pre.ensembl.org/Petromyzon_marinus/ blastview). These results indicate that the *C3/C4/C5* gene duplications occurred in the jawed vertebrate lineage, and clustering of cyclostome *C3*s with jawed vertebrate *C3*s upon phylogenetic tree analysis may reflect accelerated evolution of *C4* and *C5* during their functional specialization after their gene duplication from *C3*.

***Factor B family***

For the *fB* family, cDNA contigs for one novel and one previously reported *fB* (45) genes named lamprey *fB-2* and *fB-1* were obtained (Fig. 3-2C), and the percent amino acid identity between them is 43% (Table 3-1). Almost all the functionally critical residues were conserved in lamprey fB-2 also, except for the activation cleavage site

between the CCP and vWA domains. As is also found in all the invertebrate fBs (Part I), canonical activation cleavage site of the trypsin-type serine protease was found both in lamprey fB-1 and fB-2. Upon phylogenetic tree analysis, these two contigs formed a clade with the highest bootstrap percentage at the basal position of the vertebrate *fB/C2* cluster, indicating that the *fB/C2* gene duplication occurred in the jawed vertebrate lineage (Fig. 3-3C). However, the exact timing of the *fB/C2* gene duplication in the jawed vertebrate lineage is still not clear, since most internal branches, except for one clustering the *C2* genes of human and frog, were not reproduced in the MP or ML trees. The two lamprey genes are considered to have diverged from a common ancestor of jawed vertebrate *fB* and *C2*, and have duplicated at an early stage of cyclostome evolution.

***MASP family***

For the *MASP* family, one *MASP-1* gene and two *MASP-2/3* type genes termed *MASP-A* and *-B* have been reported from *L. Japonicum* (17, 63). The present EST analysis identified seven *MASP* clones, and all of them showed more than 95% amino acid sequence identity to the *MASP-A* gene (Table 3-1). No *MASP-B* and *MASP-1* cDNA were identified most probably due to the low expression levels of these genes compared to *MASP-A*. Compared to other complement genes, the expression level of the *MASP* genes is also low in human (67). *C1r* or *C1s* cDNAs were not found in the lamprey liver ESTs, and the sequence similar to these genes were also absent from the *P. marinus* genome.

***TCC family***

No *TCC* cDNA was obtained by this EST analysis. In addition, RT-PCR trial to amplify *TCC* cDNA using degenerate primers failed for *L. japonicum* liver mRNA, whereas the same method successfully amplified TCC-like cDNAs from two ascidian species, *Halocynthia roretzi* and *C. intestinalis*, and two shark species, *Mustelus manazo* and *Chimaera phantasma* (see Part III). These results indicate that either the *TCC* gene is absent from the lamprey genome, or is present but not expressed in lamprey liver. BLAST searches failed to detect the TCC-like sequence in the draft assembly of the *P. marinus* genome, supporting the former possibility.

***Other complement genes***

Two lectin genes, *C1q* (68) and *MBL* (mannan-binding lectin) (69), and one complement regulatory gene, *Lacrep* (70), have already been reported from lamprey. Our analysis identified EST sequences corresponding to all these three genes. In addition, ESTs for two novel C1q-type genes were also identified. However, due to the simplicity of the domain architecture of the C1q, consisting of one collagen and one C1q domains, it is difficult to identify them as the classical complement component C1q based only on their primary structure, as was in the case of some invertebrates.

***Coagulation genes***

Human coagulation factors with a unique domain architecture are classified into four families; the vitamin K (VK)-dependent serine protease family (FVII, FIX, FX, protein C (PC) and PT), the FV/FVIII family (FV and FVIII), the prekallikrein/FXI family (prekallikrein and FXI), and the FXII family (only FXII). By this lamprey EST analysis,

cDNA contigs for four genes of the VK-dependent serine protease family, three GLA/EGF/EGF/SP-type and one GLA/KR/KR/SP-type genes, were identified (Table 3-1, Figs. 3-1 and -4). Upon phylogenetic tree analysis (Fig. 3-5A), two of the GLA/EGF/EGF/SP-type cDNAs formed a cluster with the jawed vertebrate *FX*, supported by a bootstrap percentage of 57%, and were termed lamprey *FX1* and *FX2*. The other one, lamprey *FVII*, formed a cluster with teleost *FVII* with a bootstrap percentage of 92%. Although this cluster formed a clade with tetrapod *FVII*s in the NJ and MP tree, this clade was not supported by the ML tree (Supplementary Fig. 3-2A). Another EST coded a SP domain similar to that of human PC. However, our trial to clone the 5' end by 5'-RACE failed, and its entire domain architecture is still unclear. A phylogenetic tree of the *PT* gene with the GLA (domain containing gamma-carboxylated Glu residues)/KR (kringle)/KR/SP domains, drawn using human FX as an outgroup, located lamprey *PT* in the clade of vertebrate *PT* together with the already reported hagfish *PT* (71) (Fig. 3-5B). However, the branching order within this clade was not clear since basal branches were not reproduced in the MP and ML trees (Supplementary Fig. 3-1). A search for the genes encoding the VK-dependent proteases from the *P. marinus* genome database detected at least six scaffolds containing more than two domains out of four domains characteristic to this family, and five of these scaffolds showed an orthologous relationship to the five genes identified from *L. japonicum*. The other scaffold showed no close relationship to any member of this family upon phylogenetic tree analysis (data not shown).

Alignment of amino acid sequences of lamprey and human VK-dependent serine proteases is shown in Fig. 3-4. Functionally important residues such as the Cys residues involved in the interchain disulfide bond of FVII, FX and PC, and the catalytic triad

residues of the SP domain were perfectly conserved. In addition, the Asp residue at the bottom of the specificity pocket of the serine protease domain was also conserved, suggesting that all four lamprey proteases have a trypsin-type cleaving specificity. However, the Arg-Ile residues at the activation cleavage site were substituted by Ser-Ile residues in lamprey FX2. Thus, it is unlikely that lamprey FVII can activate lamprey FX2.

The cDNAs for other three gene families, *FV*/*FVIII*, *FXII* and *FXI*/*prekallikrein* genes were not identified by this lamprey liver EST analysis. Although a systematic search of the *P. marinus* trace database identified FV/FVIII-like sequences (58), none of them contained more than two domains out of the five domains characteristic of this family, leaving the identification of these genes inconclusive.

The other putative coagulation factor cDNAs identified for the first time from cyclostomes were heparin cofactor II with one serpin (serinprotease-inhibitor) domain, and tissue-factor pathway inhibitor with three Knitzu (BPTI/knitzu family of serine protease inhibitors) domains (Table 3-1). However, because of the simple domain organization and the presence of non-coagulation genes sharing the same domain organization in the human genome, identification of these coagulation gene cDNAs is tentative. Orthologous sequences of the four fibrinogen genes, alpha-1, alpha-2, beta and gamma, already reported from *P. marinus* (72), were also identified (Table 3-1).

## Discussion

The present EST analysis identified the following cDNAs of the complement and coagulation systems with unique domain architecture for the first time from lamprey; complement *fI*, coagulation *FVII*, two *FX*s, and *PT*. Except for *PT*, which was already reported from hagfish, the complete primary structures of these genes are reported for the first time from jawless vertebrates. This analysis also helps to clarify the timing of the gene duplication events that played an important role in establishing the sophisticated complement and coagulation systems of jawed vertebrates.

Four of five complement families with characteristic domain architecture were identified from the lamprey liver ESTs. The presence of the *fI* gene in lamprey, possessing exactly the same domain architecture as jawed vertebrate *fI* genes, indicated that the cyclostome complement system is equipped with the fI-dependent regulatory mechanism. The only complement family for which no lamprey liver EST was identified was the *TCC* family. TCC-like genes were already reported from amphioxus (16) and ascidians (15), although they have an incomplete domain structure compared to the jawed vertebrate TCCs, lacking some domains essential for the interaction with other complement components in vertebrates (50). However, a common ancestry of the amphioxus/ascidian TCC-like genes and the jawed vertebrate TCC genes is indicated by the completely conserved order of other several common domains. The failure to detect the TCC-like sequences by the current EST analysis as well as by BLAST searches of the *P. marinus* genome suggests a secondary loss of this family of genes, and the absence of the lytic complement pathway in lamprey, supporting the previous functional study (73). Therefore, the *TCC* genes and lytic complement pathway were established in the common ancestor of the jawed vertebrates, since the shark and all other

gnathostomes have a well developed complement system equipped with the lytic pathway (74). Molecular cloning of the shark *TCC* gene will be described in Part III.

For the other three complement gene families, the *C3*, *fB* and *MASP* families, the present and previous analyses identified multiple lamprey sequences. However, it is not clear whether the multiple copies of these genes were tandem duplication products or genome-wide duplication products, since the scaffolds of the draft genome data of *P. marinus* are not long enough to obtain linkage information. In contrast, the lamprey *fI* seems to be a single copy gene, like all jawed vertebrate *fI* genes analyzed so far, except carp, a tetraploid species (75). The lamprey *C3*, *fB* and *MASP* gene family sequences formed their own cluster in the phylogenetic tree, indicating that the *C3/C4/C5*, *fB/C2* and *MASP-1/MASP-2/C1r/C1s* gene duplications occurred in the jawed vertebrate lineage after the divergence of cyclostomes. Since C4, C2, C1r and C1s are essential components for the antibody-dependent classical pathway, these results strongly support the previous suggestion that the classical pathway was newly established in the common ancestor of jawed vertebrates simultaneously with the appearance of the adaptive immune system based on the immunoglobulin, T-cell receptor, and MHC (50). The putative complement system of lamprey is summarized in Fig. 3-6A. Two activation pathways with innate recognition, the lectin and alternative pathways, and complement regulators are most likely involved in a strictly regulated activation of C3, leading to formation of activation fragments of C3, C3a and C3b, which induce inflammation and opsonization, respectively.

For the coagulation system, the previous studies on cyclostomes identified the PT and fibrinogen genes and proteins (71, 72), indicating that the final step of blood clotting has been conserved throughout the vertebrate evolution. However, it was still

not clear how the activation and regulation steps of blood coagulation evolved. The puffer fish genome contained most orthologs of the mammalian coagulation factor genes, except for the *FXI*, *FXII* and *prekallikrein* genes, all involved in the contact-activation pathway (55, 57), whereas the asicidian genome completely lacked orthologs of the mammalian coagulation factor genes. Neither *FXI*, *FXII* and *prekallikrein* cDNAs nor the *FIX* cDNA were found in the present lamprey EST analysis, suggesting that the entire contact-activation pathway is missing in lamprey. Therefore, this EST analysis of *L. japonicum* indicates that the tissue-factor pathway involving FVII, FX, and tissue-factor is the only activating system of the primitive blood coagulation system of cyclostomes (Fig. 3-6B), consistent with the recent result of genomic analysis of *P. marinus* (58).

This lamprey liver EST analysis indicated that both the complement and coagulation systems of cyclostomes lack some families with characteristic domain architecture such as complement TCC, coagulation prekallikrein/FXI, and FXII, as summarized in Figure 2-1. Since the prekallikrein/FXI family and FXII are also missing from the puffer fish genome, they were probably established in the tetrapod lineage. In addition to establishment of new families, the complement and coagulation systems of jawed vertebrates were augmented via gene duplication. In this respect, the complement and coagulation systems show a sharp contrast. Whereas all gene duplications in the complement system recognized in the mammalian genome occurred in the jawed vertebrate lineage, the coagulation *FVII/FX/PC/PT* gene duplication occurred prior to the divergence of cyclostomes and jawed vertebrates. However, the *C3/C4*, *fB/C2* and *MASP-1/MASP-2/C1r/C1s* gene duplications, which contributed to the establishment of the complement classical pathway, and the *FV/FVIII* and *FIX/FX* gene duplications,

which contributed to the establishment of the coagulation contact-activation pathway, seem to have occurred almost simultaneously at an early stage of jawed vertebrate evolution, possibly connected with the postulated whole-genome duplication events (8).

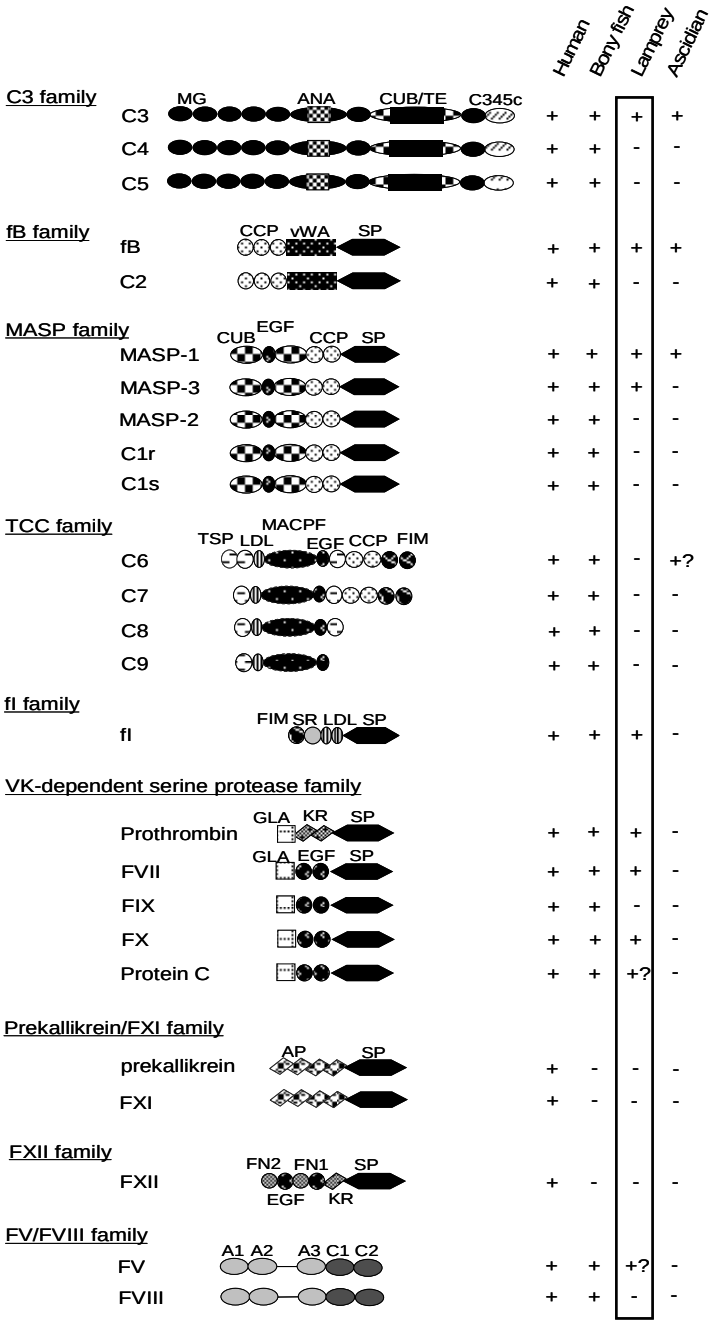



**Fig. 3-1.** Complement and coagulation gene families with characteristic domain structure found in human, bony fish, lamprey and ascidian. Domain structures of the mammalian complement and coagulation gene families are schematically presented at the left side. The presence (+)/ absence (-) of the complement genes (50), and coagulation genes of teleost (55, 57), lamprey (58, 76) and ascidian (55) is presented at the right side. The presence of possible orthologs whose domain structures are either only partially elucidated or imperfect compared to that of human is indicated by '+?'. Abbreviations of domain names are shown in the legend for Fig. 1-2 (p. 11).

**Fig. 3-2.** Alignment of the amino acid sequence of lamprey complement genes with those of representative vertebrate species. Names and boundaries (slash) of the domains are shown above the alignment. Numbers on the right side indicate the amino acid residue number of the rightmost residues. The abbreviations of the domain names are shown in the legend to Fig. 1-2 (p. 11). (A) Alignment of fIs. Three amino acid residues which form the catalytic triad of serine proteases (*), Asp residue located at the bottom of the S1-pocket ($), Cys residues involved in the inter-chain linkage (#), post translational processing site to generate light and heavy chains (%), and Cys residues in the LDL-a domain (c) are indicated. Aligned sequences are; Hu_fI (human fI), Mo_FI (mouse fI), Ch_fI (chicken fI), Fr_fI (frog fI), Ca_fI (carp fI), Sh_fI (shark fI), and La_fI (lamprey fI), which were used for the phylogenetic tree construction in Figure 3-3A. (B) Alignment of C3s. Alpha/beta processing site, activation cleavage site, fI cleavage sites, thioester site, catalytic His, and alpha/gamma processing site are boxed. Conserved Cys residues in the C3a anaphylatoxin region are marked (*), and the disulfide bridge between the alpha/beta chains are shown by lines. Putative signal peptide is underlined. (C) Alignment of fBs. The canonical and jawed vertebrate fB/C2-specific activation cleavage sites are boxed. Conserved Cys residues in the CCP domain (*), $Mg^{2+}$ binding site residues (#), catalytic triad residues (%), and Asp residue at the bottom of the S1 pocket of SP domain ($) are marked. Two disulfide bonds, one possibly linking the two chains generated by the canonical cleavage site, and the other unique to the jawed vertebrate fB/C2, are shown by lines.

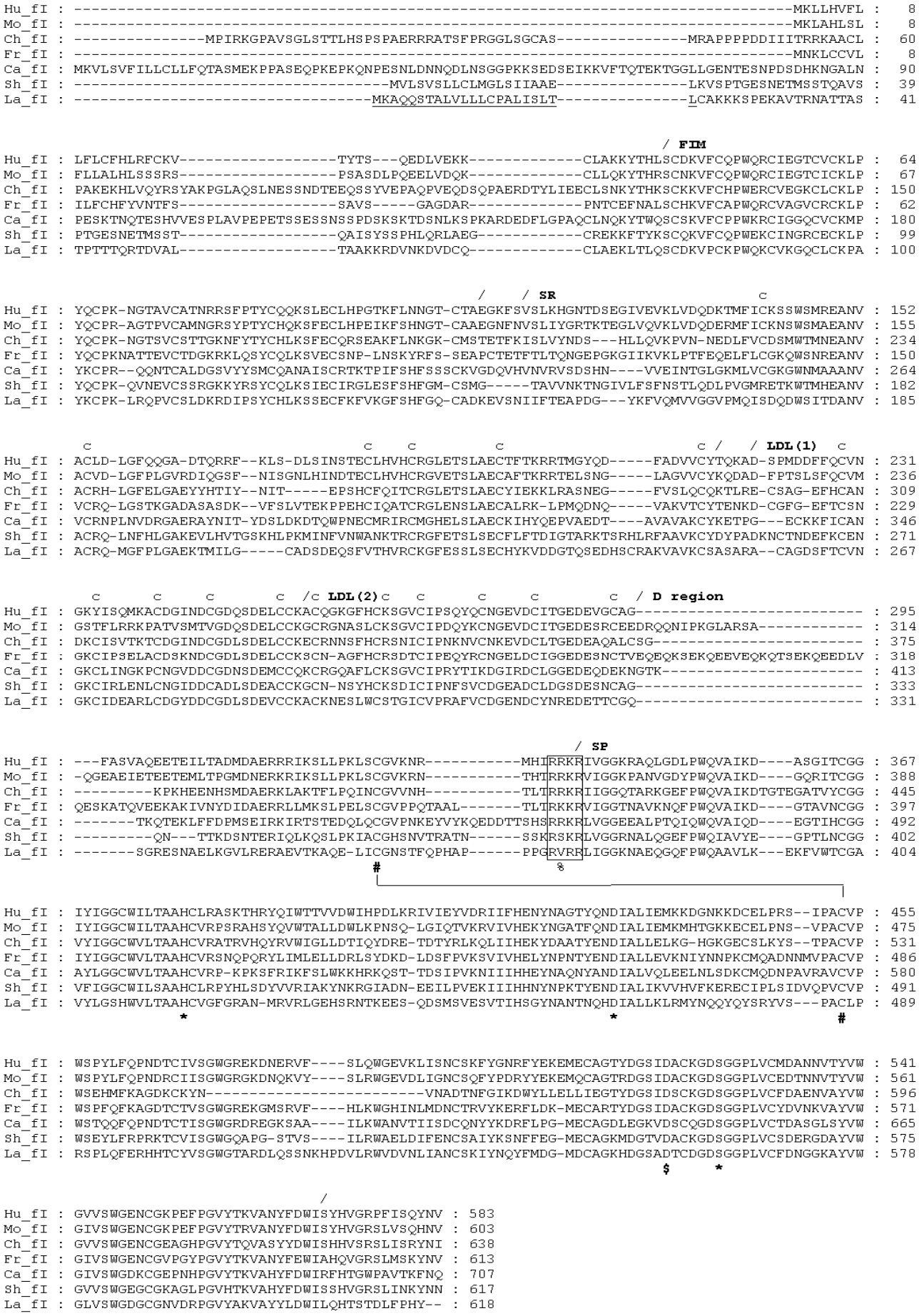



**Fig. 3-2A.**


/ MG1
/ MG2
/ MG3
/ MG4
/ MG5
/ MG6
/ MG7
alpha/beta cleavage site
ANA
activation cleavage site
anaphylatoxin region
/ CUB

**Fig. 3-2B.**

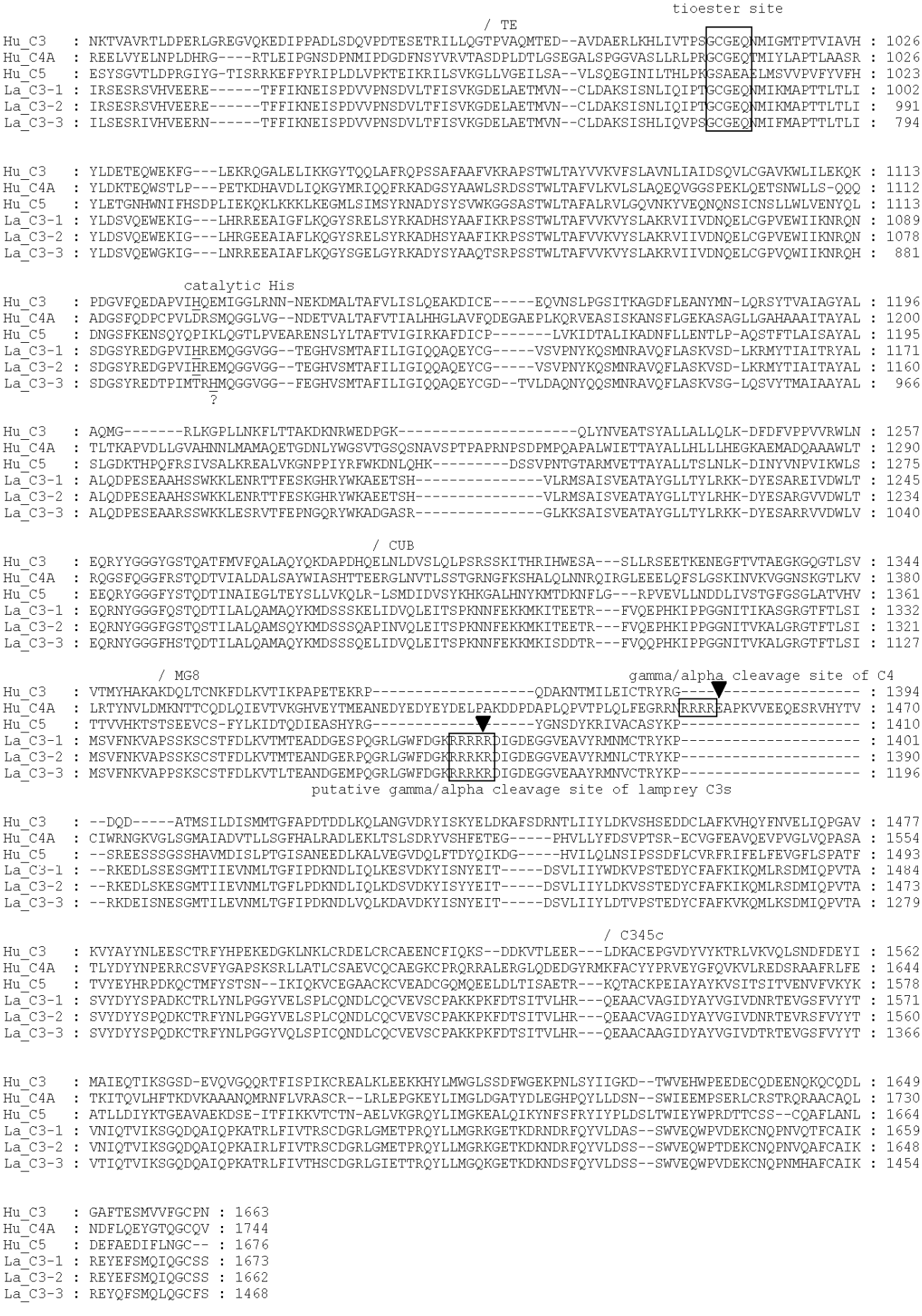



**Fig. 3-2B. (Continued)**

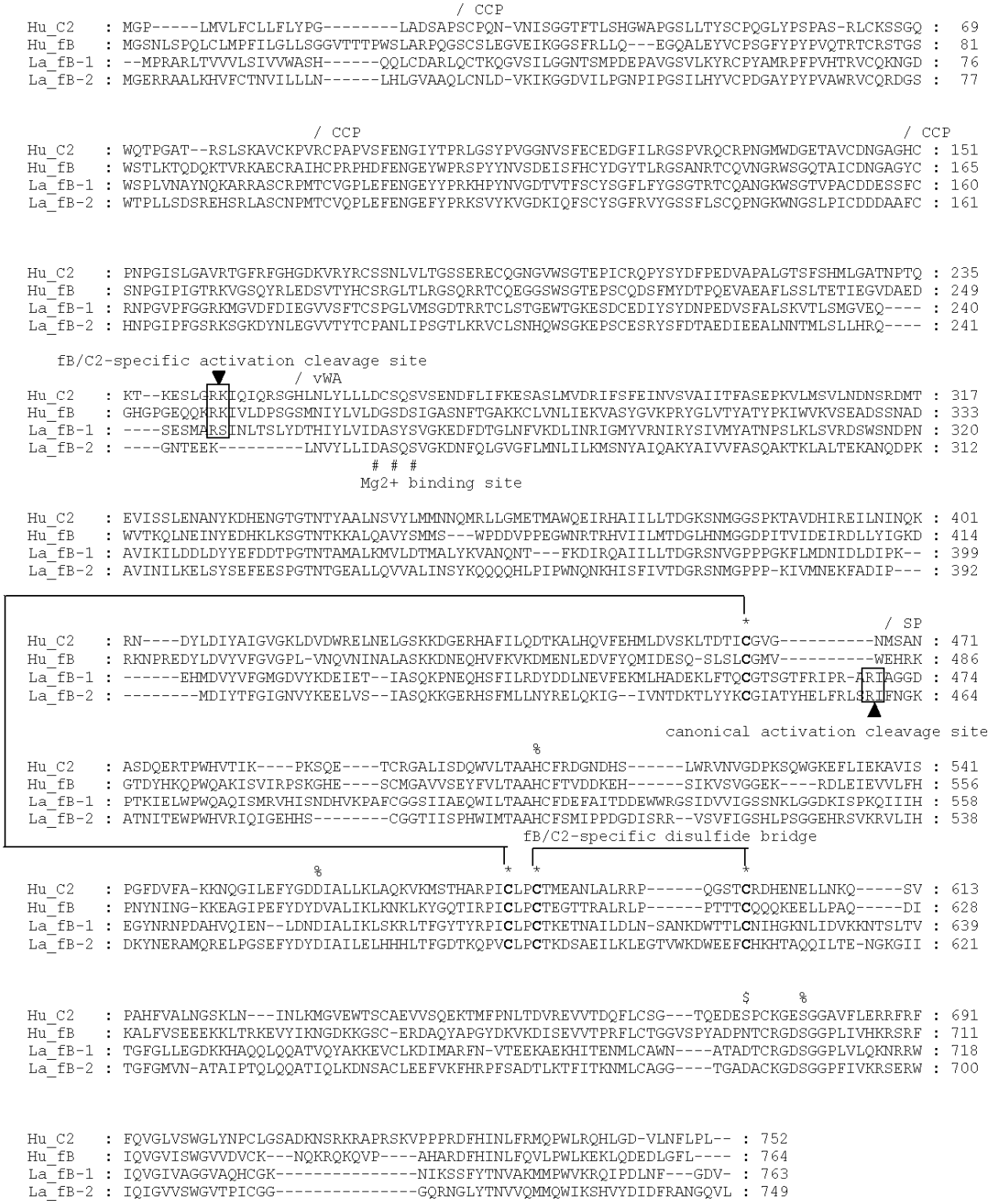

/ CCP
Hu_C2 : MGP-----LMVLFCLLFLYPG--------LADSAPSCPQN-VNISGGTFTLSHGWAPGSLLTYSCPQGLYPSPAS-RLCKSSGQ : 69
Hu_fB : MGSNLSPQLCLMPFILGLLSGGVTTTPWSLARPQGSCSLEGVEIKGGSFRLLQ---EGQALEYVCPSGFYPYPVQTRTCRSTGS : 81
La_fB-1 : --MPRARLTVVVLSIVVWASH------QQLCDARLQCTKQGVSILGGNTSMPDEPAVGSVLKYRCPYAMRPFPVHTRVCQKNGD : 76
La_fB-2 : MGERRAALKHVFCTNVILLLN------LHLGVAAQLCNLD-VKIKGGDVILPGNPIPGSILHYVCPDGAYPYPVAWRVCQRDGS : 77
/ CCP
/ CCP
Hu_C2 : WQTPGAT--RSLSKAVCKPVRCPAPVSFENGIYTPRLGSYPVGGNVSFECEDGFILRGSPVRQCRPNGMWDGETAVCDNGAGHC : 151
Hu_fB : WSTLKTQDQKTVRKAECRAIHCPRPHDFENGEYWPRSPYYNVSDEISFHCYDGYTLRGSANRTCQVNGRWSGQTAICDNGAGYC : 165
La_fB-1 : WSPLVNAYNQKARRASCRPMTCVGPLEFENGEYYPRKHPYNVGDTVTFSCYSGFLFYGSGTRTCQANGKWSGTVPACDDESSFC : 160
La_fB-2 : WTPLLSDSREHSRLASCNPMTCVQPLEFENGEFYPRKSVYKVGDKIQFSCYSGFRVYGSSFLSCQPNGKWNGSLPICDDDAAFC : 161
Hu_C2 : PNPGISLGAVRTGFRFGHGDKVRYRCSSNLVLTGSSERECQGNGVWSGTEPICRQPYSYDFPEDVAPALGTSFSHMLGATNPTQ : 235
Hu_fB : SNPGIPIGTRKVGSQYRLEDSVTYHCSRGLTLRGSQRRTCQEGGSWSGTEPSCQDSFMYDTPQEVAEAFLSSLTETIEGVDAED : 249
La_fB-1 : RNPGVPFGGRKMGVDFDIEGVVSFTCSPGLVMSGDTRRTCLSTGEWTGKESDCEDIYSYDNPEDVSFALSKVTLSMGVEQ---- : 240
La_fB-2 : HNPGIPFGSRKSGKDYNLEGVVTYTCPANLIPSGTLKRVCLSNHQWSGKEPSCESRYSFDTAEDIEEALNNTMLSLLHRQ---- : 241
fB/C2-specific activation cleavage site
/ vWA
Hu_C2 : KT--KESLGRRIQIQRSGHLNLYLLLDCSQSVSENDFLIFKESASLMVDRIFSFEINVSVAIITFASEPKVLMSVLNDNSRDMT : 317
Hu_fB : GHGPGEQQKRKIVLDPSGSMNIYLVLDGSDSIGASNFTGAKKCLVNLIEKVASYGVKPRYGLVTYATYPKIWVKVSEADSSNAD : 333
La_fB-1 : ----SESMARSINLTSLYDTHIYLVIDASYSVGKEDFDTGLNFVKDLINRIGMYVRNIRYSIVMYATNPSLKLSVRDSWSNDPN : 320
La_fB-2 : ----GNTEEK---------LNVYLLIDASQSVGKDNFQLGVGFLMNLILKMSNYAIQAKYAIVVFASQAKTKLALTEKANQDPK : 312
# # #
Mg2+ binding site
Hu_C2 : EVISSLENANYKDHENGTGTNTYAALNSVYLMMNNQMRLLGMETMAWQEIRHAIILLTDGKSNMGGSPKTAVDHIREILNINQK : 401
Hu_fB : WVTKQLNEINYEDHKLKSGTNTKKALQAVYSMMS---WPDDVPPEGWNRTRHVIILMTDGLHNMGGDPITVIDEIRDLLYIGKD : 414
La_fB-1 : AVIKILDDLDYYEFDDTPGTNTAMALKMVLDTMALYKVANQNT---FKDIRQAIILLTDGRSNVGPPPGKFLMDNIDLDIPK-- : 399
La_fB-2 : AVINILKELSYSEFEESPGTNTGEALLQVVALINSYKQQQQHLPIPWNQNKHISFIVTDGRSNMGPPP-KIVMNEKFADIP--- : 392
*
/ SP
Hu_C2 : RN----DYLDIYAIGVGKLDVDWRELNELGSKKDGERHAFILQDTKALHQVFEHMLDVSKLTDTICGVG----------NMSAN : 471
Hu_fB : RKNPREDYLDVYVFGVGPL-VNQVNINALASKKDNEQHVFKVKDMENLEDVFYQMIDESQ-SLSLCGMV----------WEHRK : 486
La_fB-1 : ------EHMDVYVFGMGDVYKDEIET--IASQKPNEQHSFILRDYDDLNEVFEKMLHADEKLFTQCGTSGTFRIPR-ARIAGGD : 474
La_fB-2 : --------MDIYTFGIGNVYKEELVS--IASQKKGERHSFMLLNYRELQKIG--IVNTDKTLYYKCGIATYHELFRLSRIFNGK : 464
canonical activation cleavage site
%
Hu_C2 : ASDQERTPWHVTIK----PKSQE----TCRGALISDQWVLTAAHCFRDGNDHS------LWRVNVGDPKSQWGKEFLIEKAVIS : 541
Hu_fB : GTDYHKQPWQAKISVIRPSKGHE----SCMGAVVSEYFVLTAAHCFTVDDKEH------SIKVSVGGEK----RDLEIEVVLFH : 556
La_fB-1 : PTKIELWPWQAQISMRVHISNDHVKPAFCGGSIIAEQWILTAAHCFDEFAITDDEWWRGSIDVVIGSSNKLGGDKISPKQIIIH : 558
La_fB-2 : ATNITEWPWHVRIQIGEHHS--------CGGTIISPHWIMTAAHCFSMIPPDGDISRR--VSVFIGSHLPSGGEHRSVKRVLIH : 538
fB/C2-specific disulfide bridge
% * * *
Hu_C2 : PGFDVFA-KKNQGILEFYGDDIALLKLAQKVKMSTHARPICLPCTMEANLALRRP------QGSTCRDHENELLNKQ-----SV : 613
Hu_fB : PNYNING-KKEAGIPEFYDYDVALIKLKNKLKYGQTIRPICLPCTEGTTRALRLP------PTTTCQQQKEELLPAQ-----DI : 628
La_fB-1 : EGYNRNPDAHVQIEN--LDNDIALIKLSKRLTFGYTYRPICLPCTKETNAILDLN-SANKDWTTLCNIHGKNLIDVKKNTSLTV : 639
La_fB-2 : DKYNERAMQRELPGSEFYDYDIAILELHHHLTFGDTKQPVCLPCTKDSAEILKLEGTVWKDWEEFCHKHTAQQILTE-NGKGII : 621
$ %
Hu_C2 : PAHFVALNGSKLN---INLKMGVEWTSCAEVVSQEKTMFPNLTDVREVVTDQFLCSG---TQEDESPCKGESGGAVFLERRFRF : 691
Hu_fB : KALFVSEEEKKLTRKEVYIKNGDKKGSC-ERDAQYAPGYDKVKDISEVVTPRFLCTGGVSPYADPNTCRGDSGGPLIVHKRSRF : 711
La_fB-1 : TGFGLLEGDKKHAQQLQQATVQYAKKEVCLKDIMARFN-VTEEKAEKHITENMLCAWN----ATADTCRGDSGGPLVLQKNRRW : 718
La_fB-2 : TGFGMVN-ATAIPTQLQQATIQLKDNSACLEEFVKFHRPFSADTLKTFITKNMLCAGG----TGADACKGDSGGPFIVKRSERW : 700
Hu_C2 : FQVGLVSWGLYNPCLGSADKNSRKRAPRSKVPPPRDFHINLFRMQPWLRQHLGD-VLNFLPL-- : 752
Hu_fB : IQVGVISWGVVDVCK---NQKRQKQVP----AHARDFHINLFQVLPWLKEKLQDEDLGFL---- : 764
La_fB-1 : IQVGIVAGGVAQHCGK---------------NIKSSFYTNVAKMMPWVKRQIPDLNF---GDV- : 763
La_fB-2 : IQIGVVSWGVTPICGG---------------GQRNGLYTNVVQMMQWIKSHVYDIDFRANGQVL : 749

**Fig. 3-2C.**

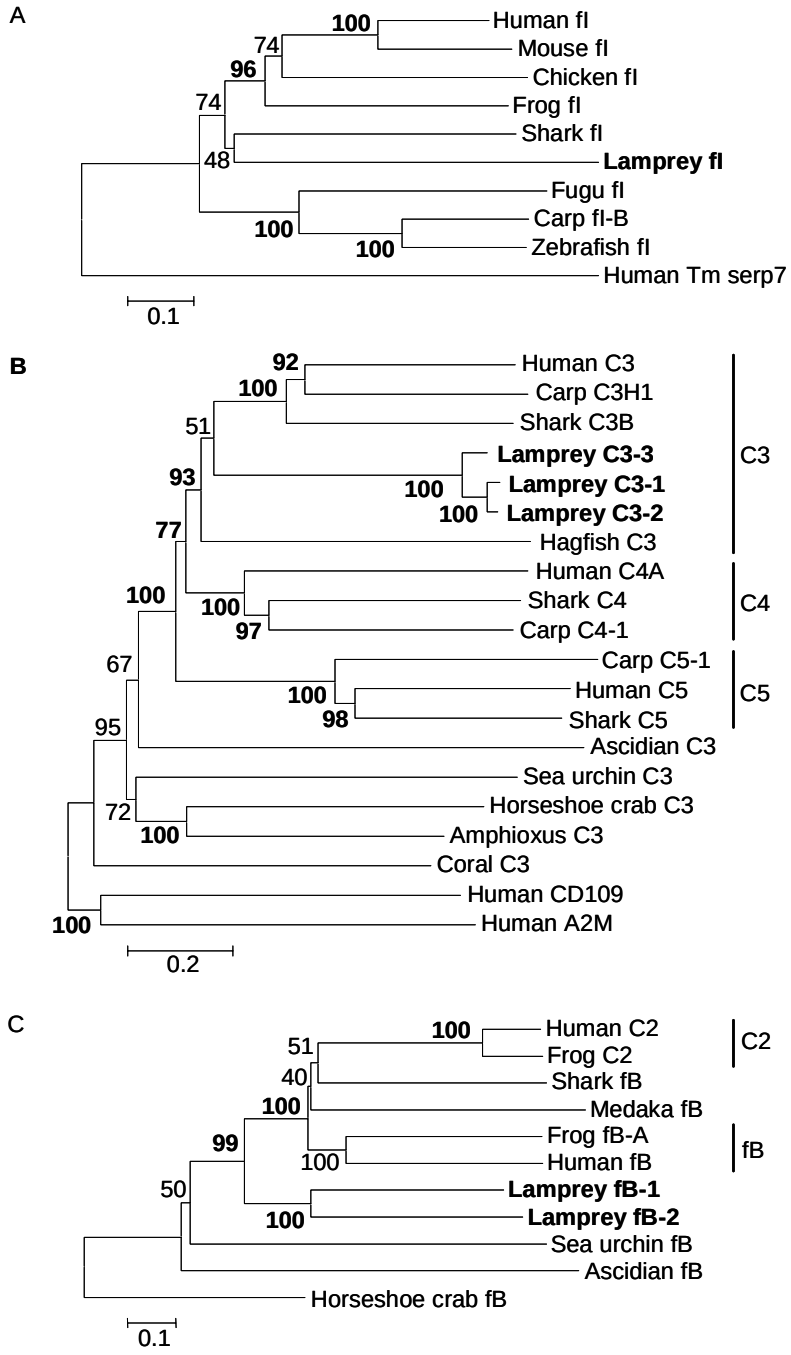



**Fig. 3-3.** Phylogenetic trees of the complement genes with characteristic domain architecture. The NJ tree is shown, and the consensus internal branches reproduced in the MP and ML trees are marked by the bootstrap percentages in boldface. Accession numbers of each entry are listed in Table 1-1 (p. 14). (A) Phylogenetic tree of fI. (B) Phylogenetic tree of C3, C4 and C5. (C) Phylogenetic tree of fB and C2.

**Fig. 3-4.** Alignment of the VK-dependent serine proteases amino acid sequences. Four lamprey (La) and five human (77) sequences were aligned. Names and boundaries (slash) of the domains are shown above for FVII, FIX, FX, and PC or below for PT the alignment. Numbers in the right side indicate the amino acid residue number of the right most residue. Three amino acid residues forming the catalytic triad of the serine proteases (*), two Cys residues conserved within this family and known to be involved in the inter-chain linkage in FVII, FX and PC (#), Asp residue located at the bottom of the S1-pocket ($), Cys residues conserved in the EGF domain (c), and positions of carboxylated Glu residues (γ) are indicated. Post-transcriptional processing and activation cleavage sites for the human components are indicated by the arrows, and activation peptides are underlined. The abbreviations of the domain names are shown in the legend to Fig. 1-2 (p. 11).

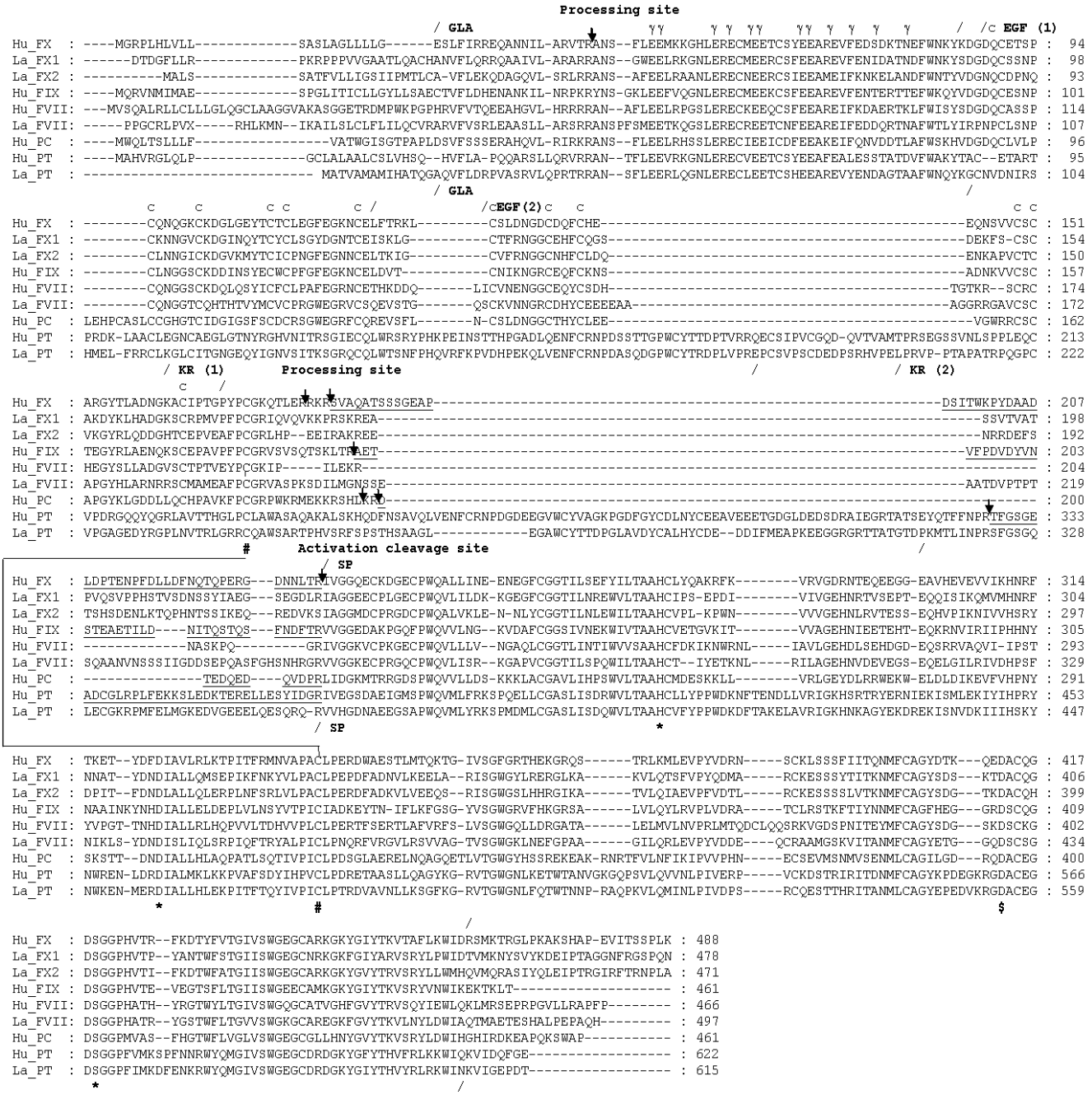

Processing site
/ GLA
/c EGF (1)
Hu_FX : ----MGRPLHLVLL----------SASLAGLLLLG------ESLFIRREQANNIL-ARVTRANS--FLEEMKKGHLERECMEETCSYEEAREVFEDSDKTNEFWNKYKDGDQCETSP : 94
La_FX1 : ------DTDGFLLR----------PKRPPPVVGAATLQACHANVFLQRRQAAIVL-ARARRANS--GWEELRKGNLERECMEERCSFEEAREVFENIDATNDFWNKYSDGDQCSSNP : 98
La_FX2 : ----------MALS----------SATFVLLIGSIIPMTLCA-VFLEKQDAGQVL-SRLRRANS--AFEELRAANLERECNEERCSIEEAMEIFKNKELANDFWNTYVDGNQCDPNQ : 93
Hu_FIX : ----MQRVNMIMAE----------SPGLITICLLGYLLSAECTVFLDHENANKIL-NRPKRYNS-GKLEEFVQGNLERECMEEKCSFEEAREVFENTERTTEFWKQYVDGDQCESNP : 101
Hu_FVII: ---MVSQALRLLCLLLGLQGCLAAGGVAKASGGETRDMPWKPGPHRVFVTQEEAHGVL-HRRRRAN--AFLEELRPGSLERECKEEQCSFEEAREIFKDAERTKLFWISYSDGDQCASSP : 114
La_FVII: -----PPGCRLPVX-----RHLKMN--IKAILSLCLFLILQCVRARVFVSRLEAASLL-ARSRRANSPFSMEETKQGSLERECREETCNFEEAREIFEDDQRTNAFWTLYIRPNPCLSNP : 107
Hu_PC : ----MWQLTSLLLF----------VATWGISGTPAPLDSVFSSSERAHQVL-RIRKRANS--FLEELRHSSLERECIEEICDFEEAKEIFQNVDDTLAFWSKHVDGDQCLVLP : 96
Hu_PT : ----MAHVRGLQLP----------GCLALAALCSLVHSQ--HVFLA-PQQARSLLQRVRRAN--TFLEEVRKGNLERECVEETCSYEEAFEALESSTATDVFWAKYTAC--ETART : 95
La_PT : ---------- MATVAMAMIHATQGAQVFLDRPVASRVLQPRTRRAN--SFLEERLQGNLERECLEETCSHEEAREVYENDAGTAAFWNQYKGCNVDNIRS : 104
/ GLA
/
c EGF(2) c
Hu_FX : --------CQNQGKCKDGLGEYTCTCLEGFEGKNCELFTRKL---------CSLDNGDCDQFCHE----------EQNSVVCSC : 151
La_FX1 : --------CKNNGVCKDGINQYTCYCLSGYDGNTCEISKLG----------CTFRNGGCEHFCQGS----------DEKFS-CSC : 154
La_FX2 : --------CLNNGICKDGVKMYTCICPNGFEGNNCELTKIG----------CVFRNGGCNHFCLDQ----------ENKAPVCTC : 150
Hu_FIX : --------CLNGGSCKDDINSYECWCPFGFEGKNCELDVT----------CNIKNGRCEQFCKNS----------ADNKVVCSC : 157
Hu_FVII: --------CQNGGSCKDQLQSYICFCLPAFEGRNCETHKDDQ-------LICVNENGGCEQYCSDH----------TGTKR--SCRC : 174
La_FVII: --------CQNGGTCQHTHTVYMCVCPRGWEGRVCSQEVSTG-------QSCKVNNGRCDHYCEEEEAA----------AGGRRGAVCSC : 172
Hu_PC : LEHPCASLCCGHGTCIDGIGSFSCDCRSGWEGRFCQREVSFL-------N-CSLDNGGCTHYCLEE----------VGWRRCSC : 162
Hu_PT : PRDK-LAACLEGNCAEGLGTNYRGHVNITRSGIECQLWRSRYPHKPEINSTTHPGADLQENFCRNPDSSTTGPWCYTTDPTVRRQECSIPVCGQD-QVTVAMTPRSEGSSVNLSPPLEQC : 213
La_PT : HMEL-FRRCLKGLCITGNGEQYIGNVSITKSGRQCQLWTSNFPHQVRFKPVDHPEKQLVENFCRNPDASQDGPWCYTRDPLVPREPCSVPSCDEDPSRHVPELPRVP-PTAPATRPQGPC : 222
/ KR (1)
Processing site
/
/ KR (2)
c
Hu_FX : ARGYTLADNGKACIPTGPYPCGKQTLERRKRSVAQATSSSGEAP----------DSITWKPYDAAD : 207
La_FX1 : AKDYKLHADGKSCRPMVPFPCGRIQVQVKKPRSKREA----------SSVTVAT : 198
La_FX2 : VKGYRLQDDGHTCEPVEAFPCGRLHP--EEIRAKREE----------NRRDEFS : 192
Hu_FIX : TEGYRLAENQKSCEPAVPFPCGRVSVSQTSKLTRAET----------VFPDVDYVN : 203
Hu_FVII: HEGYSLLADGVSCTPTVEYPCGKIP-----ILEKR---------- : 204
La_FVII: APGYHLARNRRSCMAMEAFPCGRVASPKSDILMGNSSE----------AATDVPTPT : 219
Hu_PC : APGYKLGDDLLQCHPAVKFPCGRPWKRMEKKRSHLKRD---------- : 200
Hu_PT : VPDRGQQYQGRLAVTTHGLPCLAWASAQAKALSKHQDFNSAVQLVENFCRNPDGDEEGVWCYVAGKPGDFGYCDLNYCEEAVEEETGDGLDEDSDRAIEGRTATSEYQTFFNPRTFGSGE : 333
La_PT : VPGAGEDYRGPLNVTRLGRRCQAWSARTPHVSRFSPSTHSAAGL----------EGAWCYTTDPGLAVDYCALHYCDE--DDIFMEAPKEEGGRGRTTATGTDPKMTLINPRSFGSGQ : 328
# Activation cleavage site
/
/ SP
Hu_FX : LDPTENPFDLLDFNQTQPERG---DNNLTRIVGGQECKDGECPWQALLINE-ENEGFCGGTILSEFYILTAAHCLYQAKRFK--------VRVGDRNTEQEEGG-EAVHEVEVVIKHNRF : 314
La_FX1 : PVQSVPPHSTVSDNSSYIAEG---SEGDLRIAGGEECPLGECPWQVLILDK-KGEGFCGGTILNREWVLTAAHCIPS-EPDI--------VIVGEHNRTVSEPT-EQQISIKQMVMHNRF : 304
La_FX2 : TSHSDENLKTQPHNTSSIKEQ---REDVKSIAGGMDCPRGDCPWQALVKLE-N-NLYCGGTILNLEWILTAAHCVPL-KPWN--------VVVGEHNLRVTESS-EQHVPIKNIVVHSRY : 297
Hu_FIX : STEAETILD----NITQSTQS---FNDFTRVVGGEDAKPGQFPWQVVLNG--KVDAFCGGSIVNEKWIVTAAHCVETGVKIT--------VVAGEHNIEETEHT-EQKRNVIRIIPHHNY : 305
Hu_FVII: ------------NASKPQ---------GRIVGGKVCPKGECPWQVLLLV--NGAQLCGGTLINTIWVVSAAHCFDKIKNWRNL-----IAVLGEHDLSEHDGD-EQSRRVAQVI-IPST : 293
La_FVII: SQAANVNSSSIIGDDSEPQASFGHSNHRGRVVGGKECPRGQCPWQVLISR--KGAPVCGGTILSPQWILTAAHCT--IYETKNL-----RILAGEHNVDEVEGS-EQELGILRIVDHPSF : 329
Hu_PC : ---------------TEDQED----QVDPRLIDGKMTRRGDSPWQVVLLDS-KKKLACGAVLIHPSWVLTAAHCMDESKKLL--------VRLGEYDLRRWEKW-ELDLDIKEVFVHPNY : 291
Hu_PT : ADCGLRPLFEKKSLEDKTERELLESYIDGRIVEGSDAEIGMSPWQVMLFRKSPQELLCGASLISDRWVLTAAHCLLYPPWDKNFTENDLLVRIGKHSRTRYERNIEKISMLEKIYIHPRY : 453
La_PT : LECGKRPMFELMGKEDVGEEELQESQRQ-RVVHGDNAEEGSAPWQVMLYRKSPMDMLCGASLISDQWVLTAAHCVFYPPWDKDFTAKELAVRIGKHNKAGYEKDREKISNVDKIIIHSKY : 447
/ SP
*
Hu_FX : TKET--YDFDIAVLRLKTPITFRMNVAPACLPERDWAESTLMTQKTG-IVSGFGRTHEKGRQS------TRLKMLEVPYVDRN-----SCKLSSSFIITQNMFCAGYDTK---QEDACQG : 417
La_FX1 : NNAT--YDNDIALLQMSEPIKFNKYVLPACLPEPDFADNVLKEELA--RISGWGYLRERGLKA------KVLQTSFVPYQDMA-----RCKESSSYTITKNMFCAGYSDS---KTDACQG : 406
La_FX2 : DPIT--FDNDLALLQLERPLNFSRLVLPACLPERDFADKVLVEEQS--RISGWGSLHHRGIKA------TVLQIAEVPFVDTL-----RCKESSSSLVTKNMFCAGYSDG---TKDACQH : 399
Hu_FIX : NAAINKYNHDIALLELDEPLVLNSYVTPICIADKEYTN-IFLKFGSG-YVSGWGRVFHKGRSA------LVLQYLRVPLVDRA-----TCLRSTKFTIYNNMFCAGFHEG---GRDSCQG : 409
Hu_FVII: YVPGT-TNHDIALLRLHQPVVLTDHVVPLCLPERTFSERTLAFVRFS-LVSGWGQLLDRGATA------LELMVLNVPRLMTQDCLQQSRKVGDSPNITEYMFCAGYSDG---SKDSCKG : 402
La_FVII: NIKLS-YDNDISLIQLSRPIQFTRYALPICLPNQRFVRGVLRSVVAG-TVSGWGKLNEFGPAA------GILQRLEVPYVDDE----QCRAAMGSKVITANMFCAGYETG---GQDSCSG : 434
Hu_PC : SKSTT--DNDIALLHLAQPATLSQTIVPICLPDSGLAERELNQAGQETLVTGWGYHSSREKEAK-RNRTFVLNFIKIPVVPHN-----ECSEVMSNMVSENMLCAGILGD---RQDACEG : 400
Hu_PT : NWREN-LDRDIALMKLKKPVAFSDYIHPVCLPDRETAASLLQAGYKG-RVTGWGNLKETWTANVGKGQPSVLQVVNLPIVERP-----VCKDSTRIRITDNMFCAGYKPDEGKRGDACEG : 566
La_PT : NWKEN-MERDIALLHLEKPITFTQYIVPICLPTRDVAVNLLKSGFKG-RVTGWGNLFQTWTNNP-RAQPKVLQMINLPIVDPS-----RCQESTTHRITANMLCAGYEPEDVKRGDACEG : 559
*
#
$
/
Hu_FX : DSGGPHVTR--FKDTYFVTGIVSWGEGCARKGKYGIYTKVTAFLKWIDRSMKTRGLPKAKSHAP-EVITSSPLK : 488
La_FX1 : DSGGPHVTP--YANTWFSTGIISWGEGCNRKGKFGIYARVSRYLPWIDTVMKNYSVYKDEIPTAGGNFRGSPQN : 478
La_FX2 : DSGGPHVTI--FKDTWFATGIISWGEGCARKGKYGVYTRVSRYLLWMHQVMQRASIYQLEIPTRGIRFTRNPLA : 471
Hu_FIX : DSGGPHVTE--VEGTSFLTGIISWGEECAMKGKYGIYTKVSRYVNWIKEKTKLT---------- : 461
Hu_FVII: DSGGPHATH--YRGTWYLTGIVSWGQGCATVGHFGVYTRVSQYIEWLQKLMRSEPRPGVLLRAPFP-------- : 466
La_FVII: DSGGPHATR--YGSTWFLTGVVSWGKGCAREGKFGVYTKVLNYLDWIAQTMAETESHALPEPAQH--------- : 497
Hu_PC : DSGGPMVAS--FHGTWFLVGLVSWGEGCGLLHNYGVYTKVSRYLDWIHGHIRDKEAPQKSWAP---------- : 461
Hu_PT : DSGGPFVMKSPFNNRWYQMGIVSWGEGCDRDGKYGFYTHVFRLKKWIQKVIDQFGE---------- : 622
La_PT : DSGGPFIMKDFENKRWYQMGIVSWGEGCDRDGKYGIYTHVYRLRKWINKVIGEPDT---------- : 615
*
/

**Fig. 3-4.**

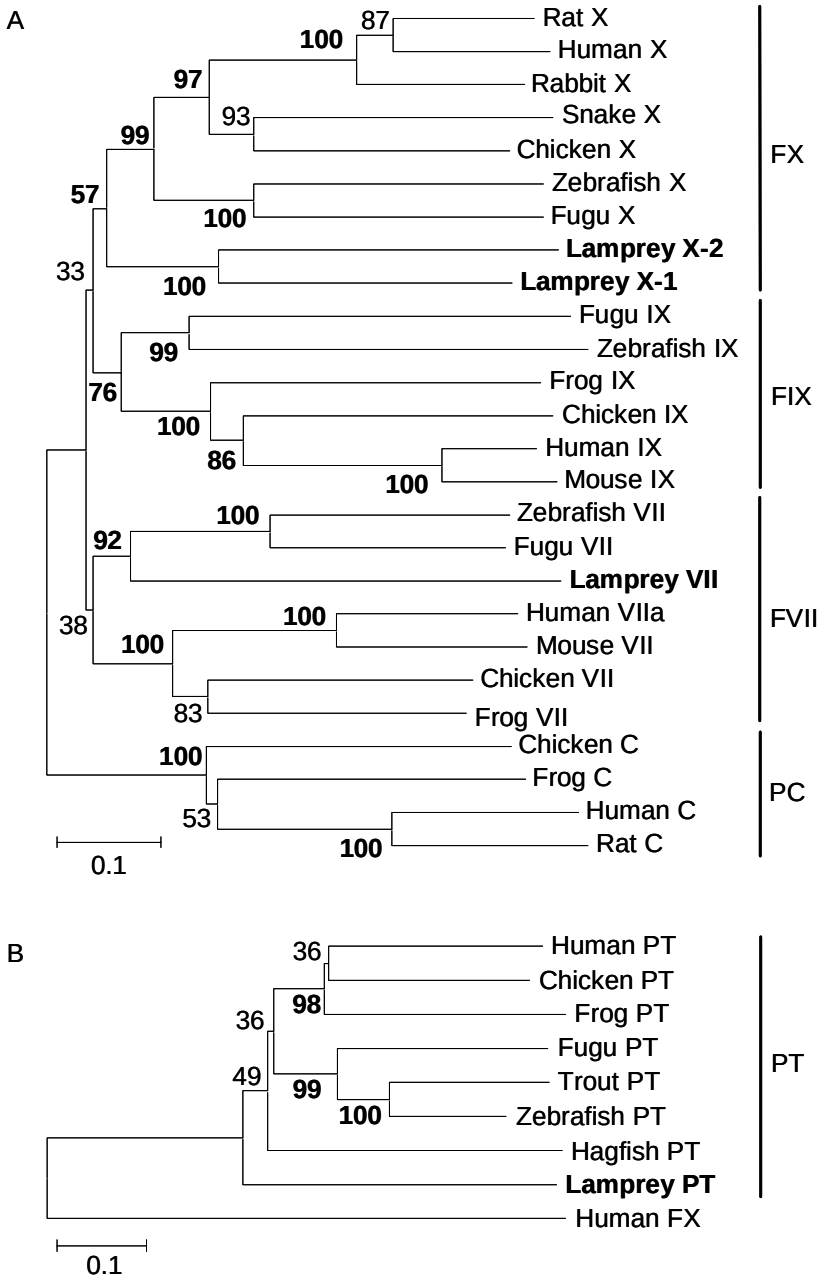



**Fig. 3-5.** Phylogenetic trees of the coagulation factor genes with characteristic domain architecture. The NJ tree is shown, and the consensus internal branches reproduced in the MP and ML trees are marked by the bootstrap percentages in boldface. Accession numbers of each entry are shown in Table 1-1 (p. 14). (A) Phylogenetic tree of FVII, FIX, FX and PC. (B) Phylogenetic tree of PT.

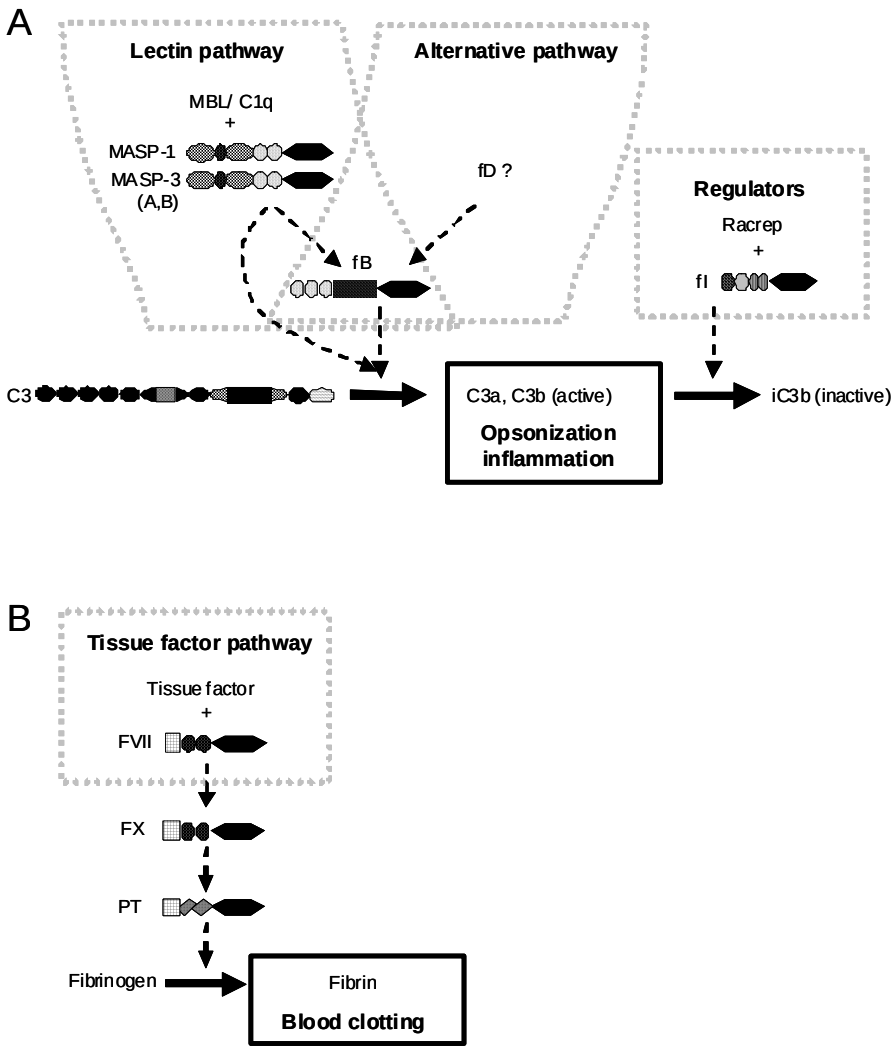
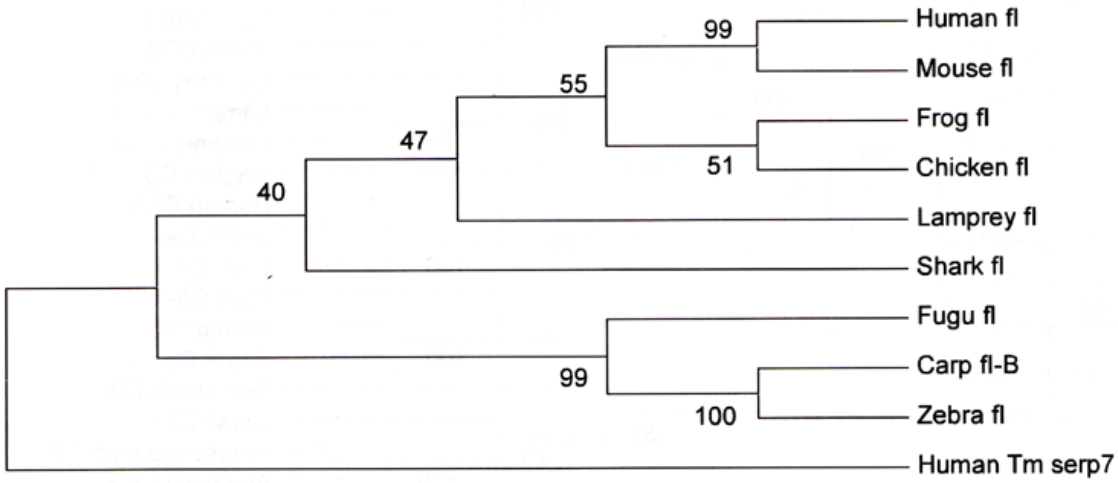
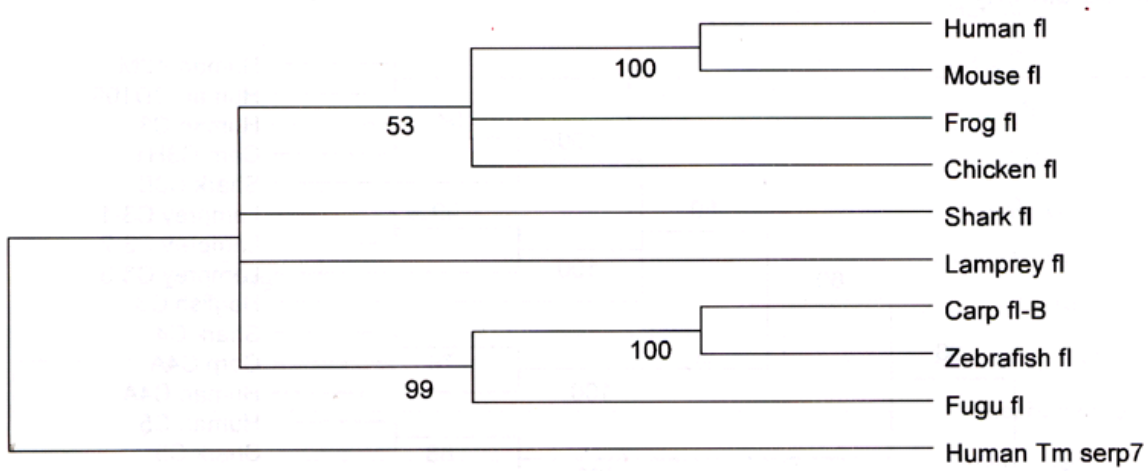



**Fig. 3-6.** Schematic representations of the putative complement and coagulation systems of lamprey. (A) Putative reaction network of the complement system of lamprey. Dotted arrows suggest possible functional interactions analogized from the mammalian complement systems, although there is still no direct experimental evidence. (B) Putative reaction network of the coagulation system of lamprey. The proteolytic cascade was speculated based on the deduced structural features of the lamprey factors.

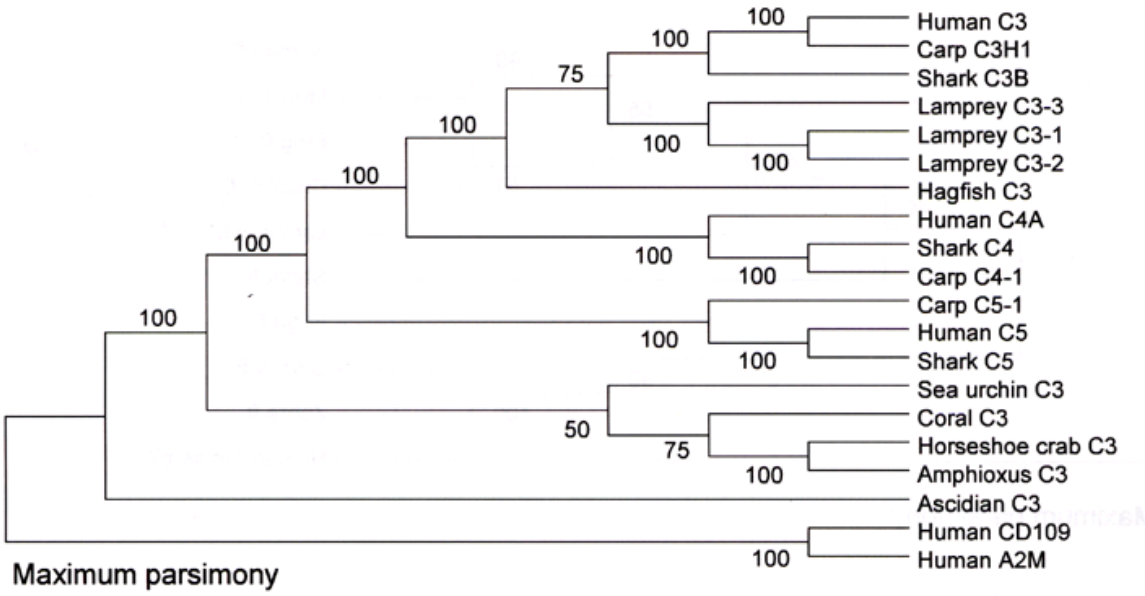


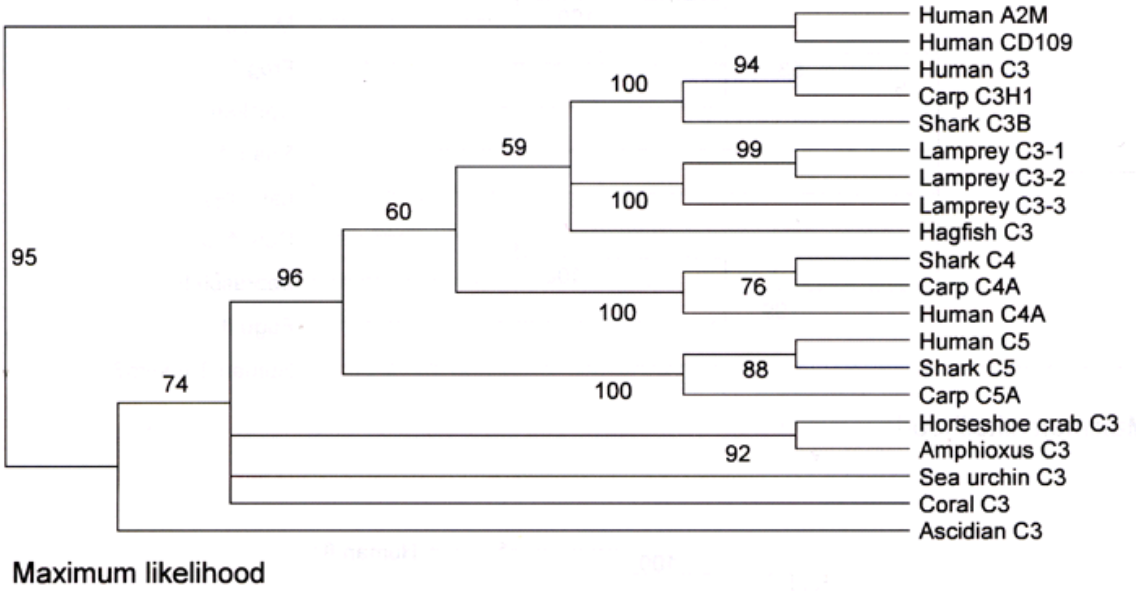

Maximum parsimony



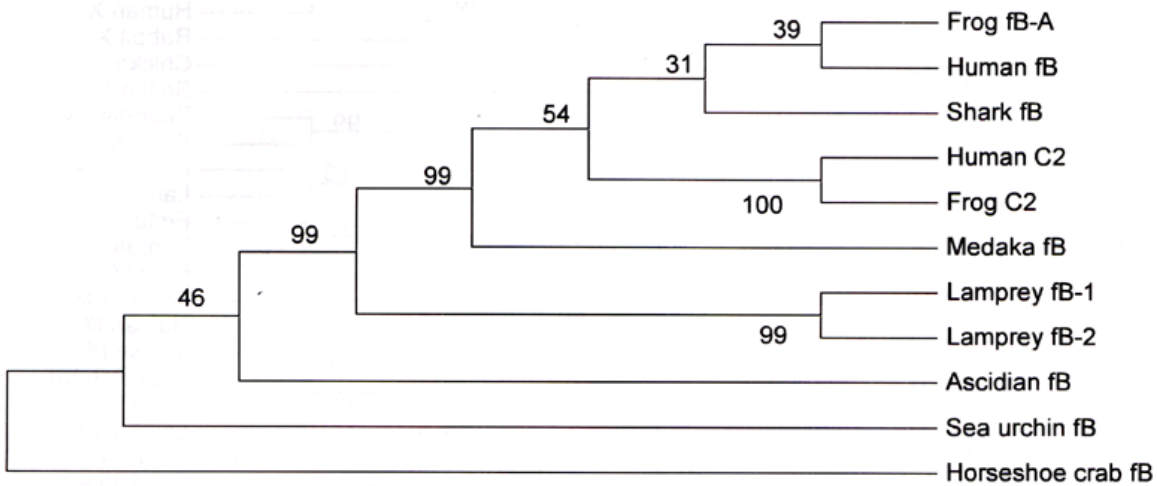

Maximum likelihood

**Supplementary Fig. 3-1.** Phylogenetic trees of the complement genes with characteristic domain architecture, using the maximum parsimony (MP) or Maximum likelihood (ML) methods. Sequences used were same as Fig. 3-3A~C.

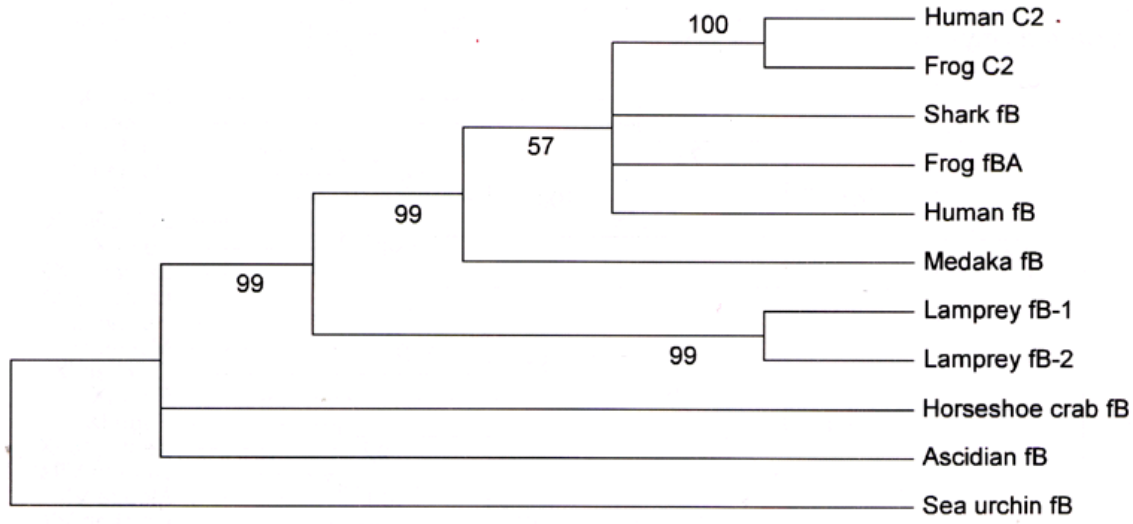



**Supplementary Fig. 3-1. (continued)**

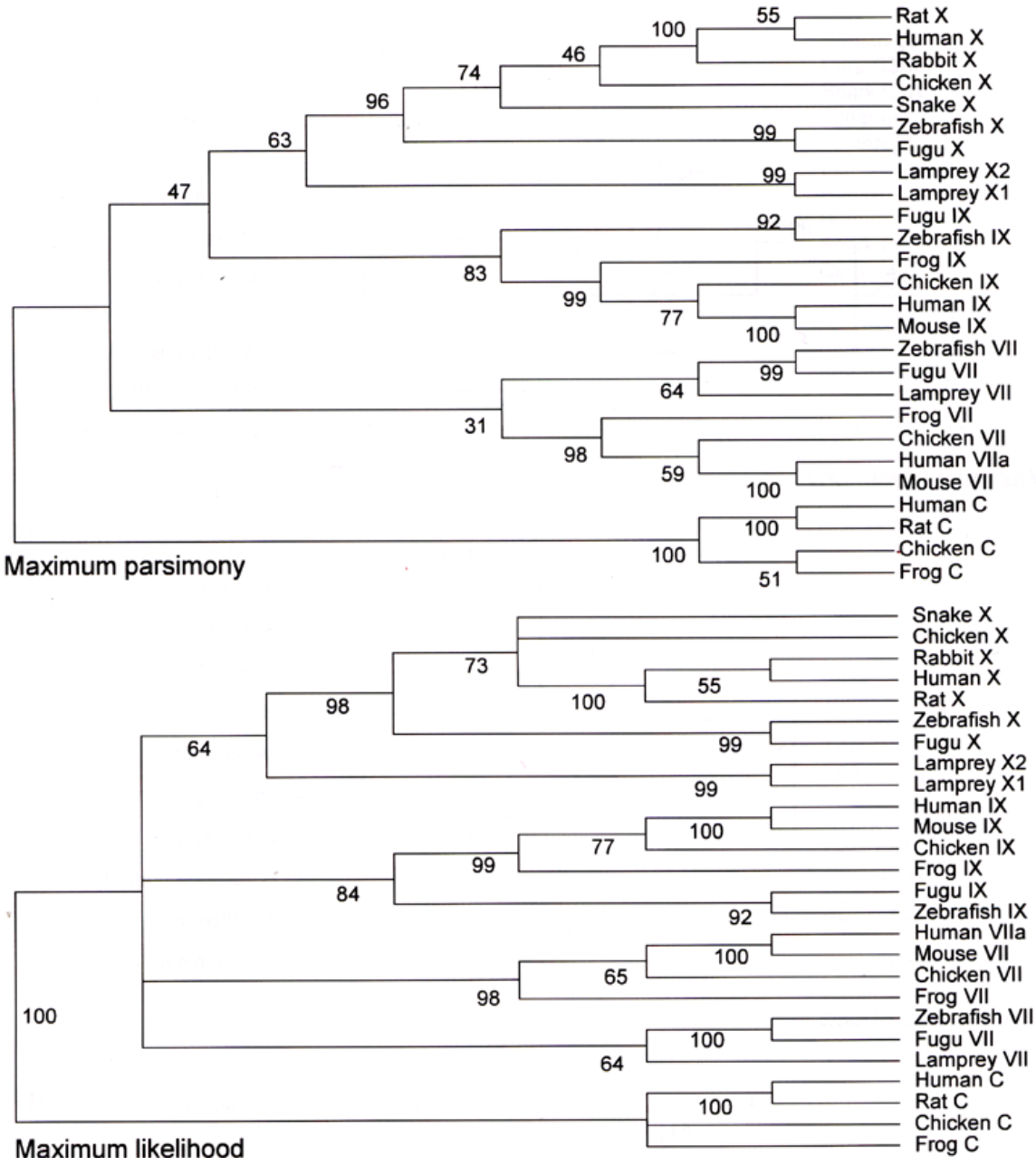



Maximum parsimony

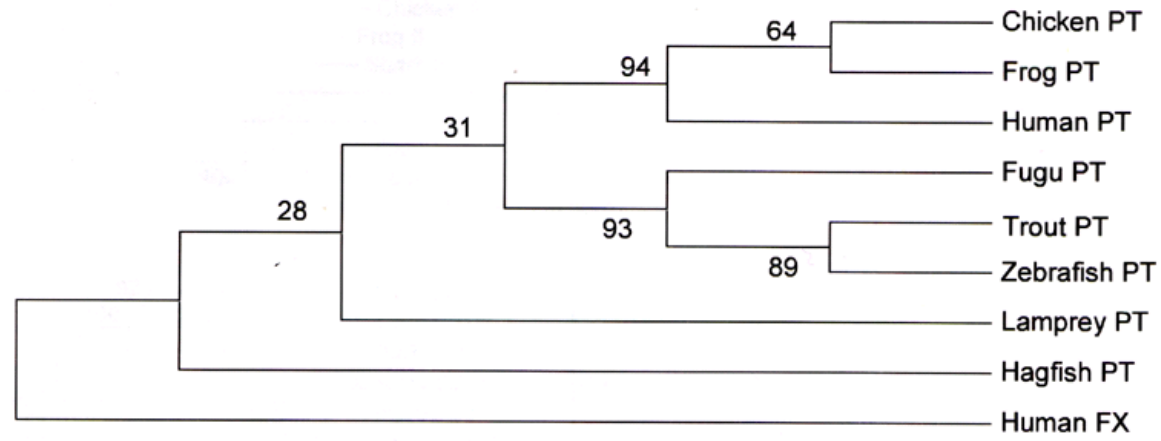



Maximum likelihood

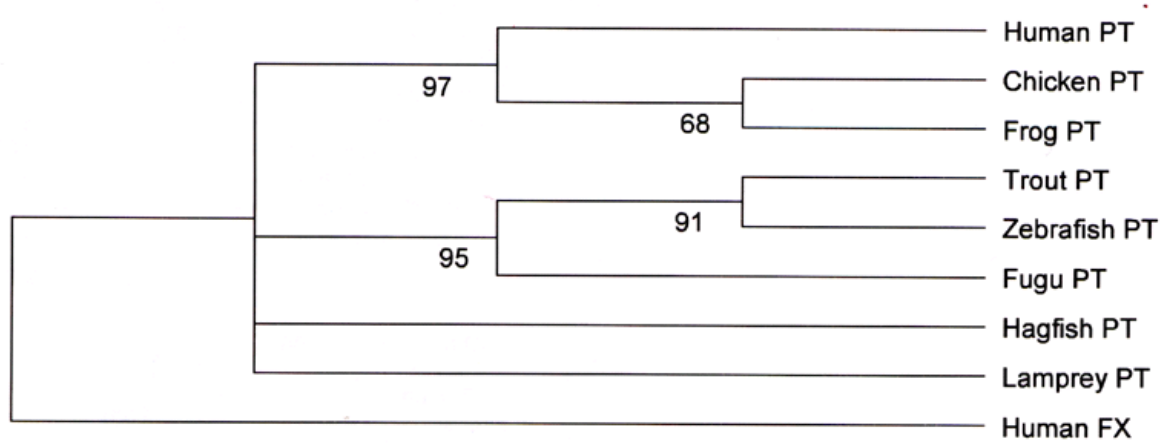
**Supplementary Fig. 3-1. (continued)**




**Supplementary Fig. 3-2.** Phylogenetic trees of the coagulation genes with characteristic domain architecture, using the maximum parsimony (MP) or maximum likelihood (ML) methods. Sequences used were same as Fig. 3-5A, B.




Maximum parsimony




Maximum likelihood

**Supplementary Fig. 3-2. (continued)**

Table 3-1. Lamprey liver ESTs encoding for possible serum proteins

| Accession No.* | Possible serum proteins | Subject | aa identity† | E-value† | Clones‡ |
|---|---|---|---|---|---|
| | | Complement components | | | |
| AB377284 | factor I § | factor I [Triakis scyllium] | 38 | 5.00E-117 | 2 |
| | C3-1 | C3 [Lethenteron japonicum] | 99 | 1.00E-153 | 11 |
| | C3-2 | C3 [Lethenteron japonicum] | 97 | 0 | 92 |
| AB377282 | C3-3 ¶ | C3 [Lethenteron japonicum] | 89 | 0 | 13 |
| | factor B-1 | factor B [Lethenteron japonicum] | 98 | 0 | 12 |
| AB377283 | factor B-2§ | factor B [Lethenteron japonicum] | 42 | 3.00E-171 | 5 |
| | MASP-A (MASP-3 type) | MASP-A [Lethenteron japonicum] | 98 | 2.00E-172 | 7 |
| | MBL-1 | MBL [Lethenteron japonicum] | 100 | 2.00E-13 | 3 |
| DC615635 | MBL-2 | MBL [Lethenteron japonicum] | 70 | 2.00E-20 | 2 |
| | C1q-1 | C1q-like protein [Lethenteron japonicum] | 98 | 1.00E-22 | 4 |
| DC613404 | C1q-2 | C1q-like protein [Lethenteron japonicum] | 41 | 4.00E-34 | 15 |
| DC619952 | C1q-3 | C1qC [Sus scrofa] | 35 | 1.00E-07 | 1 |
| DC612934 | clusterin | clusterin [Numida meleagris] | 37 | 5.00E-15 | 5 |
| | putative CCP-type complement regulator | C4bp/cremp-like [Lethenteron japonicum] | 98 | e-111 | 6 |
| DC614560 | (Racrep) | C4bp/cremp-like [Lethenteron japonicum] | 84 | 2.00E-117 | 1 |
| DC617824 | | C4bp/cremp-like [Lethenteron japonicum] | 74 | 7.00E-86 | 6 |
| DC618623 | | C4bp/cremp-like [Lethenteron japonicum] | 31 | 5.00E-44 | 4 |
| DC615013 | | C4bp/cremp-like [Lethenteron japonicum] | 29 | 2.00E-42 | 6 |
| DC612886 | | C4bp/cremp-like [Lethenteron japonicum] | 64 | 2.00E-35 | 1 |
| DC618342 | | C4bp/cremp-like [Lethenteron japonicum] | 32 | 1.00E-29 | 1 |
| DC612938 | | C4bp/cremp-like [Lethenteron japonicum] | 29 | 9.00E-18 | 21 |
| DC613686 | | C4bp/cremp-like [Lethenteron japonicum] | 29 | 1.00E-09 | 7 |
| DC614819 | | PREDICTED: complement component receptor 2 [Gallus gallus] | 34 | 1.00E-13 | 1 |
| | | Coagulation factors | | | |
| AB377287 | factor VII § | factor X [Oryctolagus cuniculus] | 43 | 2.00E-93 | 1 |
| AB377285 | factor X1 § | PREDICTED: factor X precursor [Monodelphis domestica] | 48 | 2.00E-123 | 1 |
| AB377286 | factor X2 § | PREDICTED: factor X precursor [Monodelphis domestica] | 44 | 2.00E-115 | 4 |
| DC611746 | protein C-like ¶ | PREDICTED: similar to protein C [Monodelphis domestica] | 48 | 9.00E-18 | 1 |
| AB377288 | prothrombin § | prothrombin [Eptatretus stoutii] | 52 | 0 | 12 |
| DC615133 | protein S | PREDICTED: hypothetical protein [Gallus gallus] | 45 | 2.00E-104 | 3 |
| DC619123 | fibrinogen alpha-1 | fibrinogen alpha-1 chain[Petromyzon marinus] | 85 | 0 | 8 |
| DC613720 | fibrinogen alpha-2 | fibrinogen alpha-2 chain[Petromyzon marinus] | 64 | 6.00E-07 | 1 |
| DC611945 | fibrinogen beta | fibrinogen beta chain[Petromyzon marinus] | 98 | 9.00E-95 | 27 |
| DC611387 | fibrinogen gamma | fibrinogen gamma chain[Petromyzon marinus] | 88 | 0 | 95 |
| DC614525 | platelet-activating factor acetylhydrolase | PAF acetylhydrolase[Bos taurus]:BAA06455 | 66 | 1.00E-74 | 1 |
| DC612585 | von Willebrand factor | von Willebrand factor [Branchiostoma belcheri] | 30 | 4.00E-19 | 1 |
| DC617385 | heparin cofactor II | heparin cofactor II [Mus musculus] | 44 | e-106 | 6 |
| DC616379 | tissue factor pathway inhibitor | tissue factor pathway inhibitor 2 [Bos taurus] | 40 | 4.00E-37 | 1 |
| DC619628 | plasminogen | plasminogen[Oryzias latipes] | 46 | 2.00E-79 | 2 |
| DC615102 | plasminogen-binding protein | plasminogen binding protein [Homo sapiens] | 93 | e-146 | 3 |
| | | Other serum proteins | | | |
| DC614557 | 70 kDa peroxisomal membrane protein | 70 kDa peroxisomal membrane protein [Xenopus levies] | 84 | 2.00E-160 | 3 |
| | cytoplasmic actin | cytoplasmic actin [Lethenteron japonicum] | 98 | 2.00E-158 | 2 |
| | alpha-2-macroglobulin | alpha-2-macroglobulin [Lethenteron japonicum] | 99 | 0 | 10 |
| DC612929 | alpha1-microglobulin/bikunin precursor | AMBP protein [Bos taurus] | 31 | 2.00E-20 | 54 |
| DC619468 | apolipoprotein B100 | apo-B100 precursor [Homo sapiens] | 26 | 5.00E-70 | 11 |
| DC611428 | lipoprotein LAL1 | blood plasma apolipoprotein LAL1 precursor [Petromyzom marinus] | 92 | 9.00E-42 | 121 |
| DC612748 | lipoprotein LAL2 | blood plasma apolipoprotein LAL2 precursor [Petromyzom marinus] | 79 | 1.00E-51 | 292 |
| DC611338 | apolipoprotein CIII | PREDICTED: Apo-CIII [Macaca mulatta] | 32 | 0.35 | 86 |
| DC617190 | apolipoprotein B | apolipoprotein B [Gallus gallus] | 30 | 2.00E-25 | 3 |
| DC614136 | carbonic anhydrase | carbonic anhydrase [Petromyzon marinus] | 85 | 3.00E-06 | 1 |
| DC612372 | cathepsin B | cathepsin B [Bos taurus] | 64 | 4.00E-53 | 1 |
| DC613290 | cathepsin D1 | cathepsin D1 [Takifugu rubripes] | 72 | 7.00E-149 | 3 |
| DC615168 | cathepsin A | PREDICTED: cathepsin A [Danio rerio] | 61 | 8.00E-09 | 1 |
| DC613439 | cathepsin | cathepsin [Petromyzon marinus] | 97 | 8.00E-149 | 1 |
| DC617577 | endothelin converting enzyme-1 | PREDICTED: endothelin-converting enzyme [Ornithorhynchus anatinus] | 55 | 1.00E-37 | 1 |
| | hemoglobin 1 | hemoglobin 1 [Lethenteron japonicum] | 100 | 1.00E-78 | 1 |
| | hemoglobin 2 | hemoglobin 2 [Lethenteron japonicum] | 100 | 6.00E-67 | 2 |
| DC612189 | inter-alpha trypsin inhibitor H chain 3 | inter-alpha trypsin inhibitor, heavy chain 3 [Pongo pygmaeus] | 39 | 5.00E-64 | 4 |
| DC615128 | serpin precursor | serpin precursor [Petromyzon marinus] | 45 | 1.00E-85 | 44 |
| DC613254 | serum amyloid A | serum amyloid A type B [Anas platyrhynchos] | 63 | 5.00E-29 | 15 |
| DC613551 | serum lectin | serum lectin precureser [Lethenteron japonicum] | 70 | E-118 | 38 |
| DC615154 | lactate dehydrogenase | lactate dehydrogenase [Petromyzon marinus] | 99 | 0 | 4 |
| DC617434 | selenoprotein P precursor | selenoprotein P isoform 2 [Homo sapiens] | 52 | 7.00E-23 | 16 |
| | plasma albumin precursor | plasma albumin precursor [Lethenteron japonicum] | 98 | 0 | 186 |
| DC612334 | transthyretin | transthyretin [Lampetra appendix] | 99 | 5.00E-76 | 9 |

* Accession number of newly isolated cDNAs of lamprey. For non-singletons, only those of one representative clone are shown.
† Amino acid identities and E-values against the subjects. For non-singletons, those of the contigs are shown.
‡ Number of clones found in this EST analysis.
§ ¶ The newly isolated complement/ coagulation cDNAs with unique domain structure in lamprey, whose entire domain structure was determined(§) or not determined (¶

# Part III.

# Evolutionary origin of the C6 family: Molecular cloning and comparative analysis of the *C6* gene of cartilaginous fish shark

**Abstract**

The terminal components of the mammalian complement system (TCCs), C6, C7, C8A, C8B, and C9, are a group of serum proteins which are involved in the cytolytic killing of the microbial pathogens. The mammalian TCCs share a unique domain structure established by the exon shuffling, and are believed to be diverged by the gene duplications and subsequent deletions of the N- and C- terminal domains. Both the proteins and genes for all the five TCCs have been identified from teleost fish. In contrast, in shark, a cartilaginous fish, only the presence of the multi-step complement cytolytic pathway and the *C8A* gene have been reported. Recent genomic search in invertebrate ascidian and amphioxus also detected multiple C6-like genes. However, they lack the FIM and CCP domains, which are found in the C-terminus of mammalian C6 and C7 and are involved in the interaction with the other complement components, casting a doubt for their involvement in the complement system. Here, I report the molecular cloning of the *C6* gene from a shark, *Mustelus manazo*, predicting exactly the same domain structure as mammalian C6, by the degenerate RT-PCR and RACE. The result suggested that one of the most crucial step in the establishment of the cytolytic complement pathway, addition of the FIM and CCP domains to the primitive TCC, have occurred in the early stage of the jawed vertebrate evolution before the divergence of cartilaginous fish. In additions, taken together with the presence of the *C8A* gene in shark, current isolation of the shark *C6* genes suggested that the duplication and functional divergence within the TCC families have already completed in the early jawed vertebrates.

## Introduction

The mammalian complement system comprises five mosaic protein families established by the exon shuffling and gene duplication/functional diversifications, the complement component 3 (C3), factor B (fB), mannan-binding protein-associated serine protease (MASP), factor I (fI), and terminal complement component (TCC) families (TCC family was called as 'C6 family' in Part I and II for simplicity) (50). Five members of the mammalian TCC families, C6, C7, C8A, C8B, and C9, are plasma proteins involved in the cytolytic activity of the complement system, which is conventionally used for the detection of plasma complement activity. Activation of the complement system through three parallel proteolytic cascades finally leads to the proteolytic activation of C5, which triggers the cytolytic pathway. Activated C5 (C5b) interact with C6 to form the C5b-6 complex possibly through the interaction between the C345c domain of C5 and the FIM (or CCP) domain of C6 at their respective C-terminus ends (37). Subsequent binding of C7 to the C5b-6 complex increases its binding affinity to the lipid bilayer. Then, C8 binds to the C5b-7 complex to further anchor this trimeric complex onto the membrane surface. Finally, using the C5b-8 complex as a scaffold, C9 polymerize to form the channel-like pore called membrane attack complex (MAC) on the lipid bilayer. MAC formation induces the destruction of microbes and blood cells by increasing their membrane permeability. Within the TCC family, C6 possesses the longest domain structure comprising ten domains: the thrombospondin type 1 repeats (TSP)/TSP/low-density lipoprotein receptor domain class A (LDLa)/MAC and perforin (MACPF)/epidermal growth factor-like (EGF)/TSP/complement control protein (CCP)/CCP/ factor I and MAC (FIM)/FIM domains. Other TCCs are believed to be derived from C6 by the gene duplications and

subsequent step-by-step deletions of the N- or C-terminus domains (Fig. 1-2), because the 'random' introns, which are not positioned at the domain boundary, were mostly conserved among the five human TCCs and *C. intestinalis* TCC genes (78, 79). The CCP and FIM domains, which are possibly involved in the linkage between the complement activation and TCC-driven cytolysis, were only found in C6 and C7.

In the phylogenetic aspect, orthologues to the five mammalian TCC genes were found in some teleost fishes. In additions, C5 and five TCC proteins, which have ability to form MAC and induce the hemolysis of sheep and rabbit erythrocytes, were also isolated from rainbow trout (80, 81). In sharks, the presence of the cytolytic complement system and its functional separation into several steps were reported in 1981 by Jensen *et al*. (82). In additions, the membrane lesions formed by shark serum similar to mammalian MAC on the sensitized sheep erythrocytes was observed by electron microscopy. However, the components involved in the MAC formation in shark were still to be clarified except for C8A, whose cDNA was isolated from shark, *Ginglymostoma cirratum* (unpublished data by Smith *et al*.). In invertebrates, multiple *TCC* genes with close phylogenetic relationship to mammalian C6 were found in the draft genome data of urochordate ascidian and cephalochordate amphioxus (14-16). However, all of these invertebrate *TCC* genes lack both the CCP and FIM domains, casting a doubt for their functional homology to mammalian C6, suggesting that the complement system of invertebrates is devoid of the cytolytic pathway.

Here, I performed the RT-PCR using the consensus-degenerate hybrid oligonucleotide primers to obtain the *TCC* genes from cartilaginous fishes shark and chimaera, jawless fish lamprey, and urochordate ascidians to elucidate the evolutionally origin of the cytolytic complement pathway comprising TCCs with the FIM and CCP

domains. The complete protein coding sequence of the shark *C6* gene was identified, and the comparative analysis using the identified shark *C6* gene and primitive *TCC* genes found in invertebrate genomes was performed to reveal the early evolutionary history of the TCC family.

## Materials and Methods

### *Degenerate RT-PCR for the TCC genes using RNA from shark, chimaera, lamprey, and ascidian*

RNA was purified from the liver of shark (*Mustelus manazo*), chimaera (*Chimaera phantasma*), and lamprey (*Lethenteron japonicum*), or the hepatopancreas or blood cells of ascidians (*Halocynthia roretzi* and *Ciona intestinalis*), by the guanidine thiocyanate method (60). cDNA used as a PCR template was synthesized from 1 μg each of total RNA. Consensus-degenerate hybrid oligonucleotide primers for the amplification of the *TCC* genes were designed based on the amino acid alignment of the *TCC* genes of various animals so far identified. The forward primers were C6f1 (5' TGYAAYGGNGAYRAYGAYTG 3', corresponds to the amino acid sequence CNGD(D/N)DC) and C6f2 (5' TGYAAYGGNGAYWAYGAYTGY 3', correspond to the amino acid sequence CNGD(Y/N)DC), and reverse primer was C6r1 (5' TARAARTGNGTNCCRAARTC 3', correspond to the amino acid sequence DFGTHFY). Amplification was performed by the following parameter: denaturing, for 94°C for 1 min, annealing, 42, 44, or 46°C for 30 sec, extension, 72°C for 1min, 30 to 40 cycles. The band of expected size of around 640 bp was gel-purified and subcloned into the pCR2.1-TOPO vector (Invitrogen). For each band, more than twelve clones were sequenced.

### *Cloning of the complete cDNA of the shark C6 gene by rapid amplification of cDNA ends (RACE)*

Based on the obtained partial sequence (588 bp) for the putative *M. manazo C6* gene, gene specific primers and nested gene specific primers were designed to amplify the 5'

and 3' ends of the cDNA, respectively. RACE-PCR was performed using the liver RNA and the SMART RACE cDNA amplification kit (Clontech).

***Domain prediction and phylogenetic analysis***

Domain structures were predicted using the SMART program (http://smart.embl-heidelberg.de/). Multiple alignment of the amino acid sequences was performed by ClustalX software (32), using the *TCC* genes of various animals so far reported. Phylogenetic trees were constructed using the full length information of this alignment, based on the neighbor-joining (NJ) method (33) using MEGA4, excluding gaps by pair-wise deletion (34). Pairwise genetic distances were obtained by poisson correction. The reliability for internal branches was assessed by the 1,000 bootstrap replicates.

## Results

### *Degenerate RT-PCR for the TCC gene using RNA from shark, chimaera, lamprey, and ascidian*

RT-PCR using the consensus-degenerate hybrid oligonucleotide primers designed based on the highly conserved amino acid sequences of the *TCC* genes amplified one each *TCC* sequence from two species of cartilaginous fishes, *M. manazo* and *C. phantasma*, respectively. These partial *TCC* sequences were grouped with the vertebrate *C6* and *C8B* genes by the highest bootstrap percentage, respectively, in the tentative tree using the deduced amino acid sequences (Fig. 4-1A). In lamprey, the same RT-PCR gave multiple bands, but no TCC-like sequence was detected from them. This result is consistent with the complete absence of the *TCC* genes in the liver ESTs of *L. japonicum* and draft genome data of another species of lamprey, *Petromyzon marinus* (see Part II). In ascidians, only one *TCC* sequence was amplified in *H. roretzi*, whereas nine *TCC* sequences were obtained in *C. intestinalis*, which showed one-to-one correspondence to the nine of eleven *TCC* genes predicted from the genome data (15).

### *Cloning and comparative analysis of shark C6 gene*

Nucleotide sequence of *M. manazo* C6 gene which was determined by the 5' and 3' RACE-PCR was 3,047 bp, and encoded 939 amino acids (Fig. 4-2). The deduced amino acid sequence of *M. manazo* C6 had 45%, 25%, 17%, 17%, and 13% of amino acid identity with C6, C7, C8A, C8B, and C9 genes of human, respectively. The domain structure of *M. manazo* C6 predicted by SMART program, the TSP/TSP/LDLa/ MACPF/EGF/TSP/CCP/CCP/FIM/FIM domains, was completely the same with that of vertebrate C6s (Fig. 4-2). The complete protein coding sequence of the only TCC gene

isolated from ascidian *H. roretzi* was also determined by the RACE and cDNA library screening, and predicted only the LDLa/MACPF/EGF domains (data not shown). The TSP domains at the N- and C-terminus, which were also found in the *TCC* genes of amphioxus, were completely missing from the *H. roretzi TCC*. The C-terminus TSP domain was also absent in all the *TCC* genes of another ascidian species belonging to another order, *C. intestinalis*, indicating that the deletion of this domain occurred in the common ancestor of ascidians. Deletions of the N-terminus TSP domains in the *H. roretzi* TCC is expected to have occurred after the divergence of these two orders of ascidians.

Phylogenetic tree constructed using the full length amino acid information of the shark C6 gave basically the same topology to that constructed using the partial sequences (Fig. 4-1B). Vertebrate C6/C7/C8A/C8B/C9 formed a clade supported by 100% of bootstrap percentage, and invertebrate *C. intestinalis* and amphioxus TCC formed another clade supported by 66% of bootstrap percentage. As was the case in Fig. 4-1A, *M. manazo* C6 was located within the vertebrate C6 clade supported with 100% of bootstrap percentage. In contrast, close phylogenetic relationship between the vertebrate C6 and invertebrate C6-like gene was observed only when the phylogenetic tree was constructed including the limited number of invertebrate genes (data not shown). Because the bootstrap percentages supporting these grouping were not high (~50%), the invertebrate TCCs were expected to be diverged before the duplication/functional divergence of the *C6/C7/C8A/C8B/C9* genes. Within the invertebrate TCC clade, amphioxus TCCs form one clade and *C. intestinalis* TCCs form two clades. One of two clades of the *C. intestinalis* TCCs clustered with the amphioxus TCC clade supported by 96% of bootstrap value, indicating that these genes are

descended from one gene and diverged by the lineage-specific duplications. Evidences for the tandem gene duplications, as shown by the tandem location of genes on the same scaffold, were found for almost all the *TCC* genes of *C. Intestinalis* within both clades (unpublished data by Yoshizaki F), but not found for the amphioxus TCC genes (data not shown). The *H. roretzi* TCC grouped with neither invertebrate TCCs nor vertebrate TCCs. Taken together with the lack of TSP domain and gene duplication products and the presence of the putative transmembrane region at the C-terminus, the *H. roretzi* TCC is expected to be a diverged member of the TCC family.

**Discussion**

*TCC* gene which completely conserved the unique domain structure of mammalian C6 was identified for the first time in cartilaginous fish, shark. In contrast, *TCC* gene was completely absent from the two species of lamprey which were used for the current and previous comprehensive analyses. Since TCC-like genes were found in amphioxus and ascidians, these results indicate the secondary loss of the *TCC* genes in jawless vertebrates. Therefore, current result indicates that one of the most crucial steps in the establishment of the cytolytic complement pathway, addition of the FIM and CCP domains to the primitive TCC similar to the invertebrate C6-like molecules, was completed in the early step of the jawed vertebrate evolution before the divergence of cartilaginous fish. Supporting this conclusion, hemolytic activity against rabbit blood cells, which is dependent on the $Mg^{2+}$ ion and C5, was also detected in the plasma from several species of sharks (unpublished data by Nagumo H). In contrast, the hemolytic activity of lamprey plasma was demonstrated to be independent on the complement activation (73), and that of ascidian body fluid was too weak to be compared with that of the mammalian complement system (unpublished data by Miyazawa S and Kimura A). On the other hands, the origin of genes with the TCC-like domain structure is traced back before the divergence of Cephalochordate from the Chordate lineage. Although multiple MACPF domain-containing genes were also found in the draft genome data of sea urchin (83, 84), protostomes (50), and sea anemone (Part I), they lack additional domains found in the vertebrate C6 and C7. Therefore, it is concluded that the domain structure of the amphioxus C6-like genes, the TSP/TSP/LDLa/MACPF/EGF/TSP domains, reflect the common ancestral evolutionary state of the chordate *TCC* genes, and the ascidian and vertebrate TCCs experienced different arrangement of the N- and

C-terminus domains in each lineage. Within the five complement gene families, such trial-and-error modifications of the domain structure is observed only in the TCC families, and the unique domain structure of the C3, fB, and MASP families was perfectly conserved from cnidarian to mammals except for the invertebrate fBs, as was discussed in Part I.

In additions, taken together with the previous identification of the *C8A* gene in shark, the closest phylogenetic relationships and highest amino acid identities between the shark *C6* and chimaera *C8B* genes and those of teleosts/mammals indicated that at least two steps of the gene duplication/functional divergence within the TCC family have occurred in the early jawed vertebrates before the divergence of cartilaginous fish. The 5' and 3' RACE of the chimaera *C8B* gene is now in progress for more detailed analysis using the full coding sequence information. In contrast, multiple *TCC* genes in invertebrate ascidian *C. intestinalis* and amphioxus is concluded to be diverged before the divergence of the *C6/C7/C8A/C8B/C9* genes. They seemed to have experienced the lineage-specific duplications, indicating the advantage of multiple *TCC* genes, irrespective of the complement system.

In conclusion, current finding of shark *C6* and chimaera *C8B* genes provides the first nucleotide-level evidence for the emergence and duplications/functional divergences of the TCCs as a potential effecter arm of the complement system in the early jawed vertebrate before the emergence of cartilaginous fish, more than 400 million years ago.

**Fig. 4-1.** Phylogenetic trees of the C6 family. Trees were constructed based on the alignment of the full length amino acid sequences of C6 family, using the NJ method excluding gaps by pair-wise deletion. Phylogenetic trees including (A) and not including (B) the partial sequence of shark *C8B* gene were shown. Bootstrap percentages are given. Accession numbers of each entry are listed in Table 1-1 (p. 14).

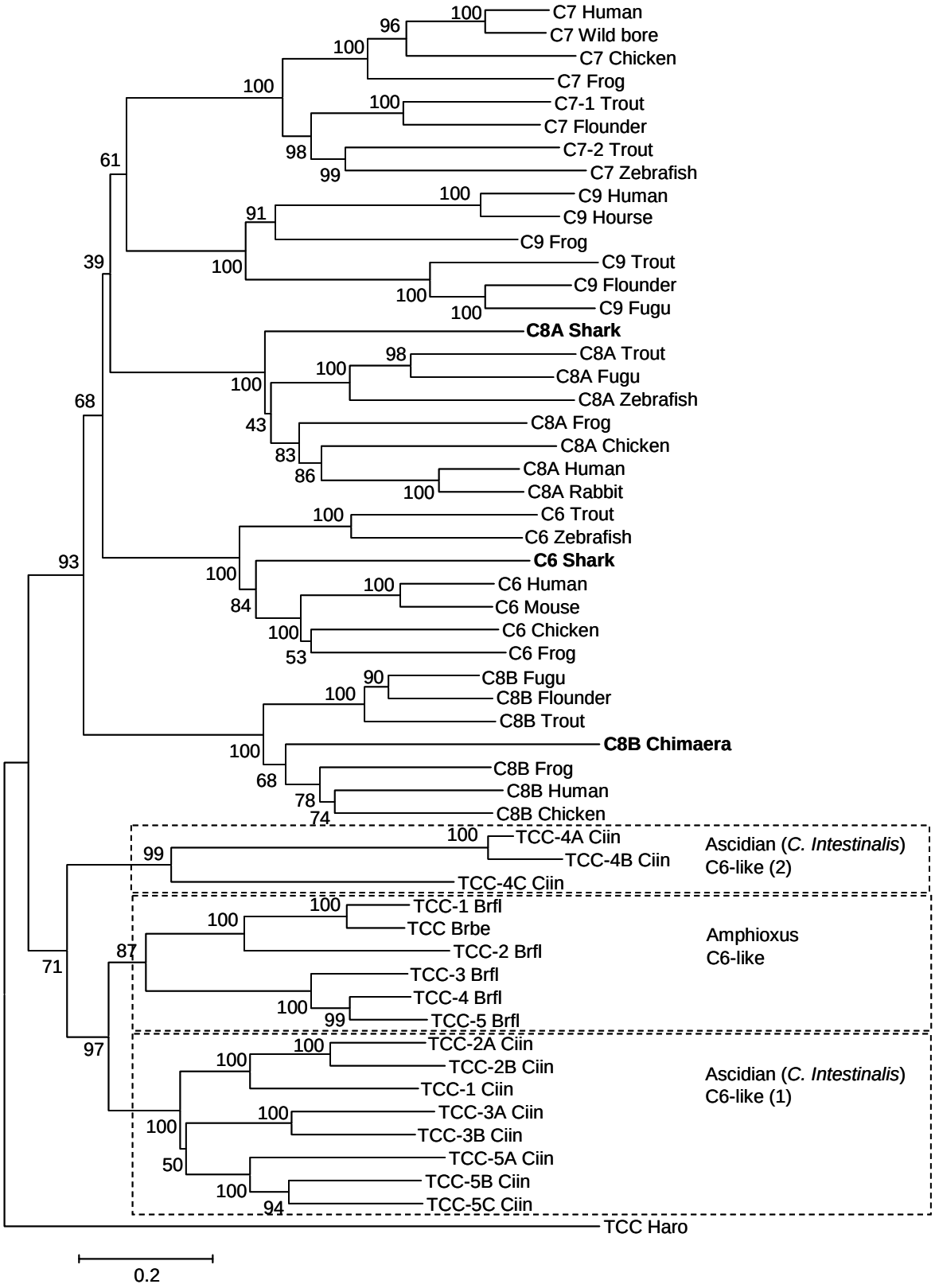

C7 Human
C7 Wild bore
C7 Chicken
C7 Frog
C7-1 Trout
C7 Flounder
C7-2 Trout
C7 Zebrafish
C9 Human
C9 Hourse
C9 Frog
C9 Trout
C9 Flounder
C9 Fugu
C8A Shark
C8A Trout
C8A Fugu
C8A Zebrafish
C8A Frog
C8A Chicken
C8A Human
C8A Rabbit
C6 Trout
C6 Zebrafish
C6 Shark
C6 Human
C6 Mouse
C6 Chicken
C6 Frog
C8B Fugu
C8B Flounder
C8B Trout
C8B Chimaera
C8B Frog
C8B Human
C8B Chicken
TCC-4A Ciin
TCC-4B Ciin
TCC-4C Ciin
Ascidian (C. Intestinalis)
C6-like (2)
TCC-1 Brfl
TCC Brbe
TCC-2 Brfl
TCC-3 Brfl
TCC-4 Brfl
TCC-5 Brfl
Amphioxus
C6-like
TCC-2A Ciin
TCC-2B Ciin
TCC-1 Ciin
TCC-3A Ciin
TCC-3B Ciin
TCC-5A Ciin
TCC-5B Ciin
TCC-5C Ciin
Ascidian (C. Intestinalis)
C6-like (1)
TCC Haro
0.2

**Fig. 4-1A.**

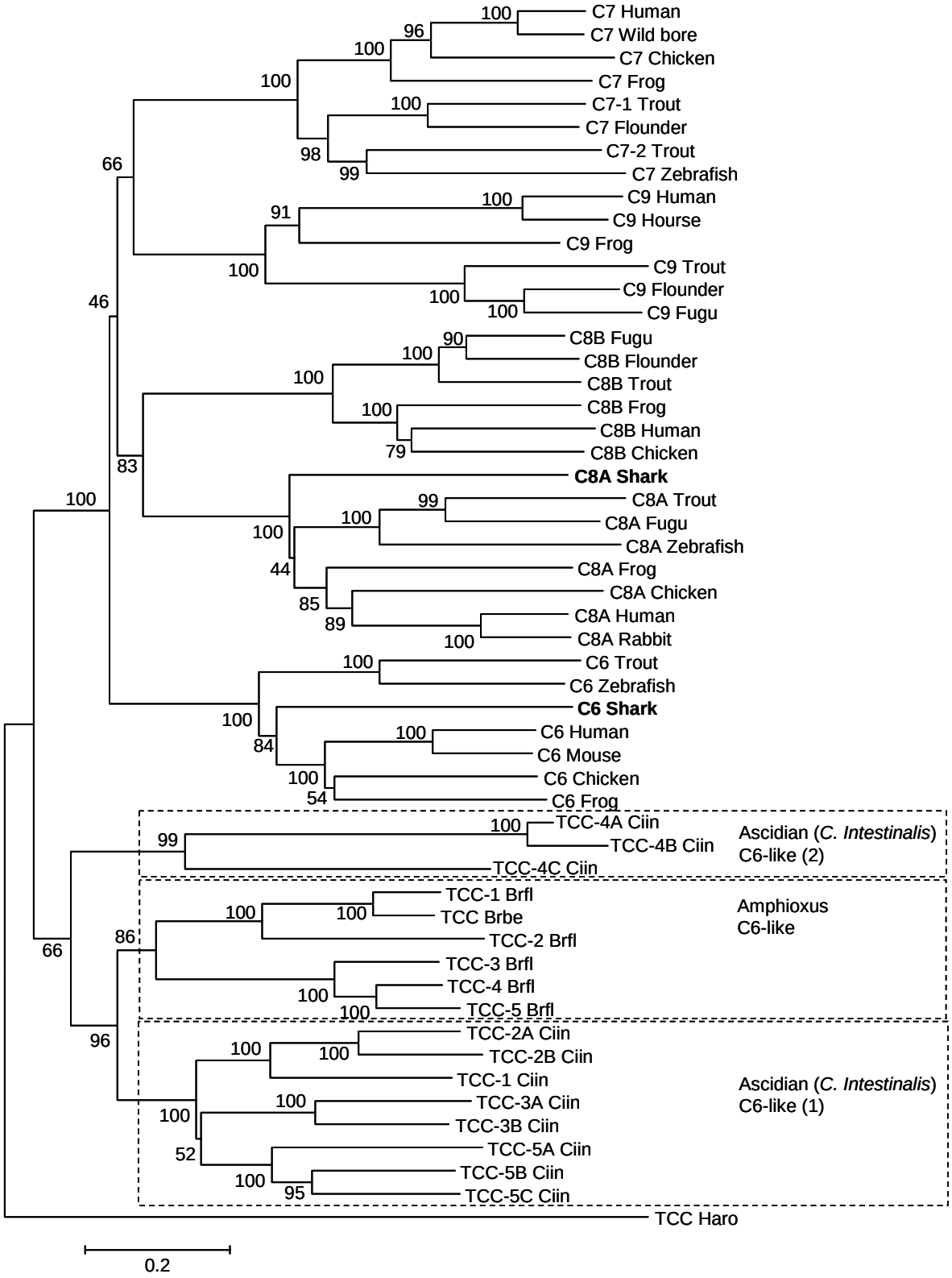

C7 Human
C7 Wild bore
C7 Chicken
C7 Frog
C7-1 Trout
C7 Flounder
C7-2 Trout
C7 Zebrafish
C9 Human
C9 Hourse
C9 Frog
C9 Trout
C9 Flounder
C9 Fugu
C8B Fugu
C8B Flounder
C8B Trout
C8B Frog
C8B Human
C8B Chicken
C8A Shark
C8A Trout
C8A Fugu
C8A Zebrafish
C8A Frog
C8A Chicken
C8A Human
C8A Rabbit
C6 Trout
C6 Zebrafish
C6 Shark
C6 Human
C6 Mouse
C6 Chicken
C6 Frog
TCC-4A Ciin
TCC-4B Ciin
TCC-4C Ciin
Ascidian (C. Intestinalis) C6-like (2)
TCC-1 Brfl
TCC Brbe
TCC-2 Brfl
TCC-3 Brfl
TCC-4 Brfl
TCC-5 Brfl
Amphioxus C6-like
TCC-2A Ciin
TCC-2B Ciin
TCC-1 Ciin
TCC-3A Ciin
TCC-3B Ciin
TCC-5A Ciin
TCC-5B Ciin
TCC-5C Ciin
Ascidian (C. Intestinalis) C6-like (1)
TCC Haro
0.2

**Fig. 4-1B.**


**Fig. 4-2.** Amino acid alignment of the shark C6 and human C6 family components. Names and boundaries of domains are shown above the sequences. Abbreviations of the domain names are indicated in the legend of Fig.1-2 (P.11).

# Conclusions

The current study indicated that the evolutionary origin of the complement system, which is composed of at least three components, C3, fB and MASP, is very ancient, dating back to a common ancestor of Eumetazoa before the divergence of Cnidaria and Bilateria. Since none of the five complement gene families was found in the genome of porifera and choanofllagelata, the complement system seems to be unique to eumetazoan animals, having tissues and cavity separated from environment. The primary structure of the three complement genes was highly conserved from cnidarian to mannmal, indicating the astonishing universality of the basic machinery of the complement activation cascades throughout the eumetazoan evolution. In additions, the comprehensive genomic analysis performed in deuterostome invertebrates, cepharochrdate and urochordate, indicated that the basic compositions of the complement system, comprised of C3, fB, and MASP, has been conserved for long time. In contrast, frequent 'complete losses' of these complement genes were also observed in many invertebrates (e.g. protostome: *Drosophila melanogaster* and *Caenorhabditis elegans*, cnidarian: *Hydra magnipapillata*). These results indicated that the complement system was in the transition period during the invertebrate evolution, waiting for the next stage of evolution which has occurred in the primitive vertebrate.

*In situ* hybridization of the cnidarian complement genes provide some crues for the functions of the primitive complement system of two germ-layered animals. Endodermal-tissue specific expression of the cnidarian complement genes suggested that the coelenteron, primitive gut cavity with putative circulatory function, is a main site of action for the primitive complement system.

The 'second stage' evolution leading to the modern complement system of

mammals seems to have occurred in the primitive vertebrate. First, the fI appeared before the divergence of the cyclostomes and jawed vertebrates, and then, the C6 family appeared before the divergence of the cartilaginous fish and bony fish. Just after the appearance of all the five complement gene families, the gene duplication/functional divergence within each family might have occurred in the primitive vertebrate before the divergence of cartilaginous fish. As a result, two potential equipments were added to the complement system of jawed vertebrates: the antibody-dependent activation pathway, and the cytolytic pathway. Thus the duplication of complement genes seems to have occurred simultaneously with the emergence of antigen recognition molecules of adaptive immune system (immunoglobulin, MHC class I and II molecules, and T-cell receptor), suggesting the presence of a common underlying genetic events. This 'second stage' evolution of the complement system might have converted the primitive system into more sophisticated and potential reaction network, which is closely related to the other immune defense systems. Once established, the sophisticated complement system with three parallel activation cascades have become a universal defense system for all the higher vertebrates including mammals.

# Acknowledgments

I thank to all the people supporting my research by helping my experiments and giving helpful discussions. Thanks are for Kazuyo Takada, Yumi Umehara, Eri Sakaguchi, Hiroyuki Nagumo, and Misaki Takada for their help for the purification and sequencing of plasmid clones of the lamprey liver ESTs. I am also grateful to Dr. Toshitaka Fujisawa for providing *N. vectensis* polyps and helpful discussions about WISH data, Dr. Masayuki Hatta for helpful discussions about WISH data, Dr. Mark Q. Martindale for providing the plasmid clone of the *NvPaxA* gene, Dr. Sumiko Taira for teaching the Parrafin section of samples, Dr. Takeshi Fukuhara for helpful discussions about C6 genes of ascidian, and Dr. Mariko Kondo for helpful discussions throughout this research. I also thank to Eri Sakaguchi, who worked together on the complement genes of *N. vectensis*, and Kazuho Ikeo, the coauthor of the paper about the lamprey liver EST. I also want to express my heartfelt gratitude for the members of The Laboratory of Molecular Comparative Immunology. Finally, special acknowledgment is due to Dr. Masaru Nonaka for his great guidance throughout the doctor's course.

# References


1. Murphy, K., P. Travers, and M. Walport. 2007. *Janeway's immunobiology*. Garland Sciennce, Yew York ans London.
2. Litman, G. W., M. K. Anderson, and J. P. Rast. 1999. Evolution of antigen binding receptors. *Annu Rev Immunol 17:109-147*.
3. Bayne, C. J. 1990. Phagocytosis and non-self recognition in invertebrates. *Bioscience 40:723*.
4. Ausubel, F. M. 2005. Are innate immune signaling pathways in plants and animals conserved? *Nat Immunol 6:973-979*.
5. Gura, T. 2001. Innate immunity. Ancient system gets new respect. *Science 291:2068-2071*.
6. Iwanaga, S., and B. L. Lee. 2005. Recent advances in the innate immunity of invertebrate animals. *J Biochem Mol Biol 38:128-150*.
7. Volanakis, J. E. 1998. Overview of the complement system. In *The human complement system in health and disease*. J. E. Volanakis, and M. M. Frank, eds. Marcel Dekker, Inc., New York, p. 9-32.
8. Kasahara, M. 2007. The 2R hypothesis: an update. *Curr Opin Immunol 19:547-552*.
9. Hedges, S. B., J. E. Blair, M. L. Venturi, and J. L. Shoe. 2004. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evol Biol 4:2*.
10. Blair, J. E., and S. B. Hedges. 2005. Molecular phylogeny and divergence times of deuterostome animals. *Mol Biol Evol 22:2275-2284*.
11. Blair, J. E., and S. B. Hedges. 2005. Molecular clocks do not support the Cambrian explosion. *Mol Biol Evol 22:387-390*.
12. Srivastava, M., E. Begovic, J. Chapman, N. H. Putnam, U. Hellsten, et al. 2008. The Trichoplax genome and the nature of placozoans. *Nature 454:955-960*.
13. Dehal, P., Y. Satou, R. K. Campbell, J. Chapman, B. Degnan, et al. 2002. The draft genome of Ciona intestinalis: insights into chordate and vertebrate origins. *Science 298:2157-2167*.
14. Holland, L. Z., R. Albalat, K. Azumi, E. Benito-Gutierrez, M. J. Blow, et al. 2008. The amphioxus genome illuminates vertebrate origins and cephalochordate biology. *Genome Res 18:1100-1111*.
15. Azumi, K., R. De Santis, A. De Tomaso, I. Rigoutsos, F. Yoshizaki, et al. 2003. Genomic analysis of immunity in a Urochordate and the emergence of the vertebrate immune system: "waiting for Godot". *Immunogenetics 55:570-581*.
16. Suzuki, M. M., N. Satoh, and M. Nonaka. 2002. C6-like and C3-like molecules from

the cephalochordate, amphioxus, suggest a cytolytic complement system in invertebrates. *J Mol Evol 54:671-679*.

17. Endo, Y., M. Nonaka, H. Saiga, Y. Kakinuma, A. Matsushita, et al. 2003. Origin of mannose-binding lectin-associated serine protease (MASP)-1 and MASP-3 involved in the lectin complement pathway traced back to the invertebrate, amphioxus. *J Immunol 170:4701-4707*.
18. He, Y., B. Tang, S. Zhang, Z. Liu, B. Zhao, et al. 2008. Molecular and immunochemical demonstration of a novel member of Bf/C2 homolog in amphioxus Branchiostoma belcheri: Implications for involvement of hepatic cecum in acute phase response. *Fish Shellfish Immunol 24:768-778*.
19. Smith, L. C., C. S. Shih, and S. G. Dachenhausen. 1998. Coelomocytes express SpBf, a homologue of factor B, the second component in the sea urchin complement system. *J Immunol 161:6784-6793*.
20. Zhu, Y., S. Thangamani, B. Ho, and J. L. Ding. 2005. The ancient origin of the complement system. *Embo J 24:382-394*.
21. Dishaw, L. J., S. L. Smith, and C. H. Bigger. 2005. Characterization of a C3-like cDNA in a coral: phylogenetic implications. *Immunogenetics 57:535-548*.
22. Miller, D. J., G. Hemmrich, E. E. Ball, D. C. Hayward, K. Khalturin, et al. 2007. The innate immune repertoire in cnidaria--ancestral complexity and stochastic gene loss. *Genome Biol 8:R59*.
23. Ayala, F. J., and A. Rzhetsky. 1998. Origin of the metazoan phyla: molecular clocks confirm paleontological estimates. *Proc Natl Acad Sci U S A 95:606-611*.
24. Clow, L. A., D. A. Raftos, P. S. Gross, and L. C. Smith. 2004. The sea urchin complement homologue, SpC3, functions as an opsonin. *J Exp Biol 207:2147-2155*.
25. Nonaka, M., K. Azumi, X. Ji, C. Namikawa-Yamada, M. Sasaki, et al. 1999. Opsonic complement component C3 in the solitary ascidian, Halocynthia roretzi. *J Immunol 162:387-391*.
26. Fautin, D. G., and R. N. Mariscal. 1991. Chapter 6 Cnidaria: Anthozoa. In *Microscopic Anatomy of Invertebrates*, Vol. 2. W. H. Frederick, and E. R. Edward, eds. Academic Press, New York, p. 267-357.
27. Marino, R., Y. Kimura, R. De Santis, J. D. Lambris, and M. R. Pinto. 2002. Complement in urochordates: cloning and characterization of two C3-like genes in the ascidian Ciona intestinalis. *Immunogenetics 53:1055-1064*.
28. Sullivan, J. C., J. F. Ryan, J. A. Watson, J. Webb, J. C. Mullikin, et al. 2006. StellaBase: the Nematostella vectensis Genomics Database. *Nucleic Acids Res 34:D495-499*.

29. Putnam, N. H., M. Srivastava, U. Hellsten, B. Dirks, J. Chapman, et al. 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science 317:86-94*.

30. King, N., M. J. Westbrook, S. L. Young, A. Kuo, M. Abedin, et al. 2008. The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. *Nature 451:783-788*.

31. Hemmrich, G., and T. C. Bosch. 2008. Compagen, a comparative genomics platform for early branching metazoan animals, reveals early origins of genes regulating stem-cell differentiation. *Bioessays 30:1010-1018*.

32. 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*.

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

34. Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol 24:1596-1599*.

35. Scholz, C. B., and U. Technau. 2003. The ancestral role of Brachyury: expression of NemBra1 in the basal cnidarian Nematostella vectensis (Anthozoa). *Dev Genes Evol 212:563-570*.

36. Matus, D. Q., K. Pang, M. Daly, and M. Q. Martindale. 2007. Expression of Pax gene family members in the anthozoan cnidarian, Nematostella vectensis. *Evol Dev 9:25-38*.

37. 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*.

38. Dellaporta, S. L., A. Xu, S. Sagasser, W. Jakob, M. A. Moreno, et al. 2006. Mitochondrial genome of Trichoplax adhaerens supports placozoa as the basal lower metazoan phylum. *Proc Natl Acad Sci U S A 103:8751-8756*.

39. Budd, A., S. Blandin, E. A. Levashina, and T. J. Gibson. 2004. Bacterial alpha2-macroglobulins: colonization factors acquired by horizontal gene transfer from the metazoan genome? *Genome Biol 5:R38*.

40. Dolmer, K., and L. Sottrup-Jensen. 1993. Disulfide bridges in human complement component C3b. *FEBS Lett 315:85-90*.

41. Nonaka, M., and M. Takahashi. 1992. Complete complementary DNA sequence of the third component of complement of lamprey. Implication for the evolution of thioester containing proteins. *J Immunol 148:3290-3295*.

42. Abdul Ajees, A., K. Gunasekaran, J. E. Volanakis, S. V. Narayana, G. J. Kotwal, et al. 2006. The structure of complement C3b provides insights into complement activation and regulation. *Nature 444:221-225*.

43. Terado, T., S. L. Smith, T. Nakanishi, M. I. Nonaka, H. Kimura, et al. 2001. Occurrence of structural specialization of the serine protease domain of complement factor B at the emergence of jawed vertebrates and adaptive immunity. *Immunogenetics 53:250-254*.

44. Yoshizaki, F. Y., S. Ikawa, M. Satake, N. Satoh, and M. Nonaka. 2005. Structure and the evolutionary implication of the triplicated complement factor B genes of a urochordate ascidian, Ciona intestinalis. *Immunogenetics*.

45. Nonaka, M., M. Takahashi, and M. Sasaki. 1994. Molecular cloning of a lamprey homologue of the mammalian MHC class III gene, complement factor B. *J Immunol 152:2263-2269*.

46. Frank, P., and J. S. Bleakney. 1976. Histology and sexual reproduction of the anemone *Nematostella vectensis* Stephenson 1935. *J. nat. Hist. 10:441-449*.

47. Takahashi, M., D. Iwaki, K. Kanno, Y. Ishida, J. Xiong, et al. 2008. Mannose-binding lectin (MBL)-associated serine protease (MASP)-1 contributes to activation of the lectin complement pathway. *J Immunol 180:6132-6138*.

48. Hajela, K., M. Kojima, G. Ambrus, K. H. Wong, B. E. Moffatt, et al. 2002. The biological functions of MBL-associated serine proteases (MASPs). *Immunobiology 205:467-475*.

49. Ambrus, G., P. Gal, M. Kojima, K. Szilagyi, J. Balczer, et al. 2003. Natural substrates and inhibitors of mannan-binding lectin-associated serine protease-1 and -2: a study on recombinant catalytic fragments. *J Immunol 170:1374-1382*.

50. Nonaka, M., and A. Kimura. 2006. Genomic view of the evolution of the complement system. *Immunogenetics 58:701-713*.

51. De Smet, K., and R. Contreras. 2005. Human antimicrobial peptides: defensins, cathelicidins and histatins. *Biotechnol Lett 27:1337-1347*.

52. Page, M. J., and E. Di Cera. 2008. Serine peptidases: classification, structure and function. *Cell Mol Life Sci 65:1220-1236*.

53. Krishnan, V., Y. Xu, K. Macon, J. E. Volanakis, and S. V. Narayana. 2007. The crystal structure of C2a, the catalytic fragment of classical pathway C3 and C5 convertase of human complement. *J Mol Biol 367:224-233*.

54. Aparicio, S., J. Chapman, E. Stupka, N. Putnam, J. M. Chia, et al. 2002. Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. *Science 297:1301-1310*.

55. Jiang, Y., and R. F. Doolittle. 2003. The evolution of vertebrate blood coagulation as viewed from a comparison of puffer fish and sea squirt genomes. *Proc Natl Acad Sci U S A 100:7527-7532*.

56. Doolittle, R. F., and D. M. Surgenor. 1962. Blood coagulation in fish. *Am J Physiol 203:964-970*.

57. Davidson, C. J., R. P. Hirt, K. Lal, P. Snell, G. Elgar, et al. 2003. Molecular evolution of the vertebrate blood coagulation network. *Thromb Haemost 89:420-428*.

58. Doolittle, R. F., Y. Jiang, and J. Nand. 2008. Genomic evidence for a simpler clotting scheme in jawless vertebrates. *J Mol Evol 66:185-196*.

59. Kumar, S., and S. B. Hedges. 1998. A molecular timescale for vertebrate evolution. *Nature 392:917-920*.

60. Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry 18:5294-5299*.

61. Schmidt, H. A., K. Strimmer, M. Vingron, and A. von Haeseler. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics 18:502-504*.

62. Anderson, N. L., M. Polanski, R. Pieper, T. Gatlin, R. S. Tirumalai, et al. 2004. The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Mol Cell Proteomics 3:311-326*.

63. Endo, Y., M. Takahashi, M. Nakao, H. Saiga, H. Sekine, et al. 1998. Two lineages of mannose-binding lectin-associated serine protease (MASP) in vertebrates [published erratum appears in J Immunol 2000 May 15; 164(10):5330]]. *J Immunol 161:4924-4930*.

64. Terado, T., M. I. Nonaka, M. Nonaka, and H. Kimura. 2002. Conservation of the modular structure of complement factor I through vertebrate evolution. *Dev Comp Immunol 26:403-413*.

65. Ishiguro, H., K. Kobayashi, M. Suzuki, K. Titani, S. Tomonaga, et al. 1992. Isolation of a hagfish gene that encodes a complement component. *Embo J 11:829-837*.

66. Lambris, J. D., A. Sahu, and R. A. Wetsel. 1998. The chemistry and biology of C3, C4, and C5. In *The human complement system in health and disease*. J. E. Volanakis, and M. M. Frank, eds. Marcel Dekker, Inc., New York, p. 83-118.

67. Terai, I., K. Kobayashi, M. Matsushita, and T. Fujita. 1997. Human serum mannose-binding lectin (MBL)-associated serine protease-1 (MASP-1): determination of levels in body fluids and identification of two forms in serum. *Clin Exp Immunol 110:317-323*.

68. Matsushita, M., A. Matsushita, Y. Endo, M. Nakata, N. Kojima, et al. 2004. Origin of the classical complement pathway: Lamprey orthologue of mammalian C1q acts as a lectin. *Proc Natl Acad Sci U S A 101:10127-10131*.

69. Takahashi, M., D. Iwaki, A. Matsushita, M. Nakata, M. Matsushita, et al. 2006. Cloning and characterization of mannose-binding lectin from lamprey (agnathans). *J Immunol 176:4861-4868*.

70. Kimura, Y., N. Inoue, A. Fukui, H. Oshiumi, M. Matsumoto, et al. 2004. A short consensus repeat-containing complement regulatory protein of lamprey that participates in cleavage of lamprey complement 3. *J Immunol 173:1118-1128*.

71. Banfield, D. K., D. M. Irwin, D. A. Walz, and R. T. MacGillivray. 1994. Evolution of prothrombin: isolation and characterization of the cDNAs encoding chicken and hagfish prothrombin. *J Mol Evol 38:177-187*.

72. Pan, Y., and R. F. Doolittle. 1992. cDNA sequence of a second fibrinogen alpha chain in lamprey: an archetypal version alignable with full-length beta and gamma chains. *Proc Natl Acad Sci U S A 89:2066-2070*.

73. Nonaka, M., T. Fujii, T. Kaidoh, S. Natsuume-Sakai, N. Yamaguchi, et al. 1984. Purification of a lamprey complement protein homologous to the third component of the mammalian complement system. *J Immunol 133:3242-3249*.

74. Jensen, J. A., E. Festa, D. S. Smith, and M. Cayer. 1981. The complement system of the nurse shark: hemolytic and comparative characteristics. *Science 214:566-569.*

75. Nakao, M., S. Hisamatsu, M. Nakahara, Y. Kato, S. L. Smith, et al. 2003. Molecular cloning of the complement regulatory factor I isotypes from the common carp (Cyprinus carpio). *Immunogenetics 54:801-806*.

76. Kimura, A., K. Ikeo, and M. Nonaka. 2009. Evolutionary origin of the vertebrate blood complement and coagulation systems inferred from liver EST analysis of lamprey. *Developmental and comparative immunology 33:77-87*.

77. Tom Tang, Y., T. Hu, M. Arterburn, B. Boyle, J. M. Bright, et al. 2005. The complete complement of C1q-domain-containing proteins in Homo sapiens. *Genomics 86:100-111*.

78. Hobart, M. J. 1998. Evolution of the terminal complement genes: ancient and modern. *Exp Clin Immunogenet 15:235-243*.

79. Nonaka, M., and F. Yoshizaki. 2004. Primitive complement system of invertebrates. *Immunol Rev 198:203-215*.

80. Nonaka, M., S. Natsuume-Sakai, and M. Takahashi. 1981. The complement system in rainbow trout (Salmo gairdneri). II. Purification and characterization of the fifth component (C5). *J Immunol 126:1495-1498*.

81. Nonaka, M., N. Yamaguchi, S. Natsuume-Sakai, and M. Takahashi. 1981. The complement system of rainbow trout (Salmo gairdneri). I. Identification of the serum lytic system homologous to mammalian complement. *J Immunol 126:1489-1494*.

82. Jensen, J. A., E. Festa, D. S. Smith, and M. Cayer. 1981. The complement system of the nurse shark: hemolytic and comparative characteristics. *Science 214:566-569*.

83. Sodergren, E., G. M. Weinstock, E. H. Davidson, R. A. Cameron, R. A. Gibbs, et al. 2006. The genome of the sea urchin Strongylocentrotus purpuratus. *Science 314:941-952*.

84. Hibino, T., M. Loza-Coll, C. Messier, A. J. Majeske, A. H. Cohen, et al. 2006. The immune gene repertoire encoded in the purple sea urchin genome. *Dev Biol 300:349-365*.