Chapter 3
Evolution of the Complement System

Masaru Nonaka

Abstract The mammalian complement system constitutes a highly sophisticated body defense machinery comprising more than 30 components. Research into the evolutionary origin of the complement system has identified a primitive version composed of the central component C3 and two activation proteases Bf and MASP in cnidaria. This suggests that the complement system was established in the common ancestor of eumetazoa more than 500 million years ago. The original activation mechanism of the original complement system is believed to be close to the mammalian lectin and alternative activation pathways, and its main role seems to be opsonization and induction of inflammation. This primitive complement system has been retained by most deuterostomes without major change until the appearance of jawed vertebrates. At this stage, duplication of the C3, Bf and MASP genes as well as recruitment of membrane attack components added the classical and lytic pathways to the primitive complement system, converting it to the modern complement system. In contrast, the complement system was lost multiple times independently in the protostome lineage.

Keywords Complement · C3 · Factor B · Gene duplication · Mannan binding protein associated serine protease · Membrane attack complex · Thioester

Abbreviations

ANATO	Anaphylatoxin
Bf	Factor B
CCP	Complement control proteins
CPAMD8	PZP-like A2M domain-containing 8
CUB	C1r, C1s, uEGF and bone morphogenic protein
If	Factor I

M. Nonaka (✉)
Department of Biological Sciences, Graduate School of Science, The University of Tokyo,
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
e-mail: mnonaka@biol.s.u-tokyo.ac.jp

EGFCA   Calcium-binding EGF-like
FIMAC   Factor I membrane attack complex
iTEP   Insect TEP
LDLa   Low-density lipoprotein receptor domain class a
MAC   Membrane attack complex
MACPF   Membrane attack complex/perforin
MASP   Mannan binding protein associated serine protease
MG   Macroglobulin domain
PZP   Pregnancy zone protein
SR   Scavenger receptor Cys-rich
TCC   Terminal complement component
TED   Thioester domain
TEP   Thioester-bond containing protein
Tryp_SPc   Trypsin-like serine protease
TSP1   Thrombospondin type 1 repeats
VWA   Von Willebrand factor type A
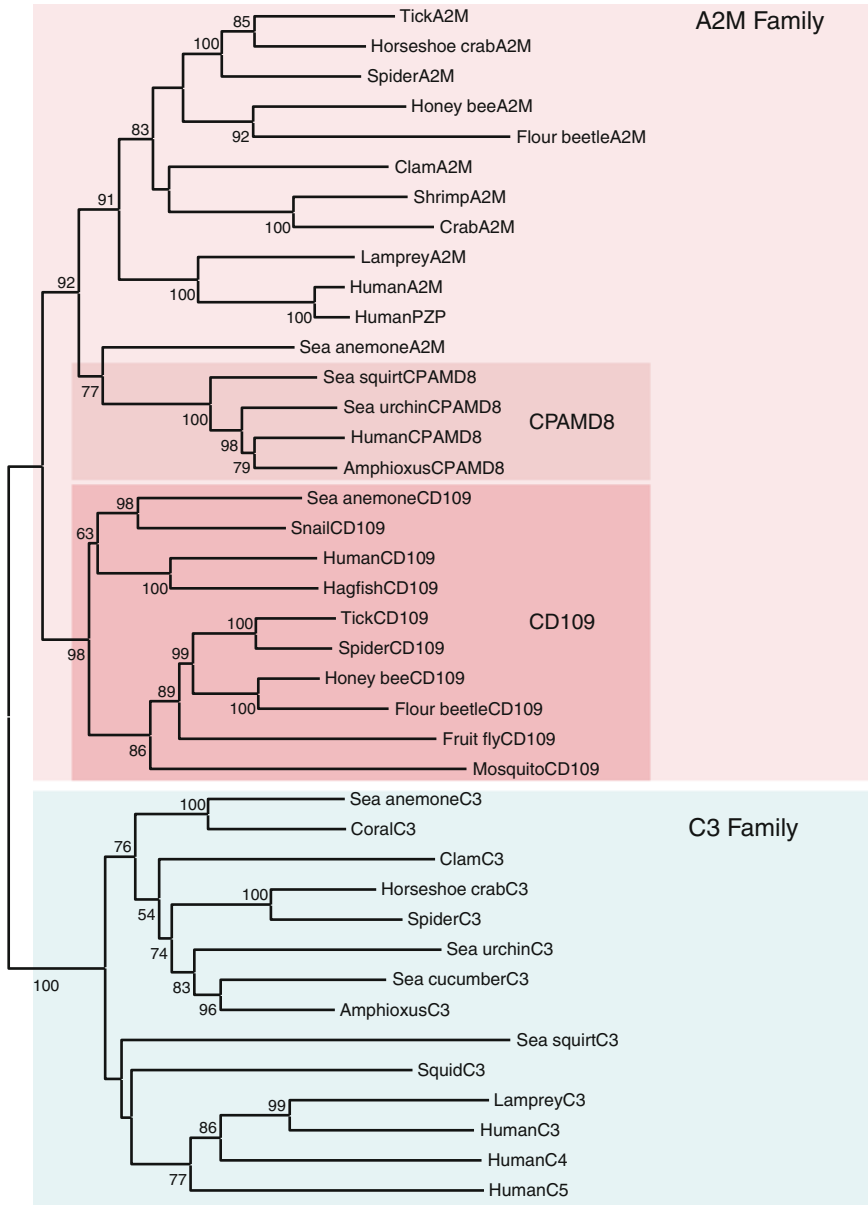
Introduction

The mammalian complement system is composed of more than 30 components present mainly in serum and cell membranes and plays essential roles in innate immunity [1]. The proteolytic activation of the central component C3 (the third component of complement) by the C3 convertases is the pivotal step in complement activation, and most major physiological functions of the complement system are induced by the two activation fragments of C3, the larger C3b and smaller C3a. Upon activation, C3b forms a covalent bond with the surface molecules of microbes using its intrachain thioester bond, and bound C3b functions as a molecular tag for foreign particles enhancing phagocytosis by macrophages and neutrophils. C3b also forms a covalent bond with C3b or C4b (the activation fragment of the fourth component of complement equivalent to C3b) of the C3 convertases (alternative pathway C3bBb and classical pathway C4bC2a), switching their specificity into the C5 convertases. Proteolytic activation of C5 (the fifth component of complement) initiates assembly of the late components, C6 (the sixth component of complement) to C9 (the ninth component of complement), leading to the formation of membrane attack complexes (MACs) which disturb the integrity of cell membranes of microbes. The smaller C3a fragment is also known as an anaphylatoxin and induces inflammation. Most key components of the human complement system including C3 possess unique domain structures, and are classified into five protein families [2]; the C3 family (C3, C4 and C5), Bf (Factor B) family (Bf 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 If (Factor I) family (If only). The origin and evolution of the complement system have been studied by identifying the complement genes possessing these unique domain structures in various eumetazoan species.

Evolution of the TEP Superfamily

C3 family is included in the TEP (thioester-bond containing protein) superfamily whose members are characterized by the unique intrachain thioester bond. Seven members of the TEP superfamily are encoded in the human genome: C3, C4, C5, A2M, pregnancy zone protein (PZP) [3], *CD*109 [4] and the complement 3 and PZP-like A2M domain-containing 8 (*CPAMD*8) [5]. Phylogenetic analysis of many TEP superfamily genes from various eumetazoa indicated the presence of two dichotomous families, C3 and A2M (Fig. 3.1), and this classification was supported by the presence of the ANA (anaphylatoxin) and C345C domains in all members of the C3 family, but never in the A2M family members [6]. The C3 family comprises human C3, C4, C5 and their orthologs of various eumetazoa, whereas the A2M family comprises human A2M, PZP, CD109, CPAMD8 and their orthologs. Accumulation of sequence information concerning various eumetazoa TEPs indicated that the iTEP (insect TEP) once considered unique to insects or arthropods is in fact orthologous to CD109 [6]. A2M, PZP and CPAMD8 are close to each other, and the A2M family is further subdivided into the A2M subfamily and the CD109 subfamily. The basic domain structure of the TEP superfamily members is an eightfold repeat of the macroglobulin (MG) domain. In addition to these eight MG domains, a CUB domain (C1r, C1s, uEGF and bone morphogenetic protein) holding the TED (thioester domain) in the middle is inserted between the seventh and eighth MG domains [7]. Moreover, other specific domains are present in each of the TEP members, such as ANA and C345C (C-terminal of C3, C4 and C5) domains of C3 and the bait domain in A2M.

The evolutionary origin of the TEP superfamily remains to be clarified. However, no TEP gene is present in the published genome information for a sponge, *Amphimedon queenslandica* [8] and a choanoflagellate, *Monosiga brevicollis* [9], suggesting that this gene family arose in the eumetazoan lineage. The three types of TEP genes, C3, A2M and CD109 were identified in two cnidarian sea anemones, *Haliplanella lineate* [10] and *Nematostella vectensis* [11, 12], indicating that differentiation of the TEP genes into C3, A2M and CD109 had completed before the divergence of cnidaria and bilateria. All deuterostome species analyzed thus far such as various vertebrates, urochordate sea squirt [13], cephalochordate lancelet and echinoderm sea urchin [2], possess both the C3 and A2M family members, whereas many protostome genomes deciphered thus far such as fly [14], mosquito [15], honeybee [16], parasitoid wasp [17], aphid [18],

A2M Family

TickA2M
85
100   Horseshoe crabA2M
83   SpiderA2M
Honey beeA2M
92   Flour beetleA2M
91   ClamA2M
ShrimpA2M
100   CrabA2M
92   LampreyA2M
HumanA2M
100
100   HumanPZP
Sea anemoneA2M
77   Sea squirtCPAMD8
Sea urchinCPAMD8   CPAMD8
100
98   HumanCPAMD8
79   AmphioxusCPAMD8

98   Sea anemoneCD109
63   SnailCD109
HumanCD109   CD109
100   HagfishCD109
100   TickCD109
99   SpiderCD109
89   Honey beeCD109
100   Flour beetleCD109
86   Fruit flyCD109
MosquitoCD109

C3 Family

100   Sea anemoneC3
76   CoralC3
ClamC3
100   Horseshoe crabC3
54   SpiderC3
74   Sea urchinC3
83   Sea cucumberC3
96   AmphioxusC3
Sea squirtC3
100   SquidC3
99   LampreyC3
86   HumanC3
HumanC4
77   HumanC5

◀ **Fig. 3.1** Molecular phylogenetic anaylsis of TEP family genes by maximum likelihood method. The evolutionary history was inferred by using the maximum likelihood method based on the Whelan and Goldman + Freq. model. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 40 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 873 positions in the final dataset. Bootstrap percentages with 500 replicates are given. Accession numbers of the used sequences and scientific names of animals are; human (*Homo sapiens*) C3, C4A, C5, A2M, CD109, CPAMD8, and PZP (NP_000055, P0C0L4, AAA51925, P01023, NP_598000, NP_056507 and CAA38255), squid (*Euprymna scolope*) C3 (ACF04700), sea urchin (*Strongylocentrotus purpuratus*) C3 and CPAMD8 (NP_999686 and XP_785018), sea cucumber (*Apostichopus japonicas*) C3 (ADN97000), sea squirt (*Ciona intestinalis*) C3 and CPAMD8 (NP_001027684, and XP_002124325), lamprey (*Lethenteron japonicum*) C3 and A2M (Q00685 and BAA02762), hagfish (*Eptatretus burgeri*) CD109 (BAD12264), horseshoe crab2 (*Tachypleus tridentatus*) C3 and A2M (BAH02276 and BAA19844), amphioxus (*Branchiostoma floridae*) C3 and CPAMD8 (XP_002248496 and XP_002239366), coral (*Swiftia exserta*) C3 (AAN86548), sea anemone (*Haliplanella lineate*) C3, A2M, and CD109 (AB481383, AB481385 and AB481386), fruit fly (*Drosophila melanogaster*) CD109 (*Drme*TEP1) (NP_523578), mosquito (*Anopheles gambiae*) CD109 (*Anga*TEP1) (AAG00600), clam (*Hyriopsis cumingii*) A2M (ABJ89824), clam (*Venerupis decussatus*) C3 (FJ392025), tick (*Ornithodoros moubata*) A2M (AAN10129), tick (*Ixodes scapularis*) CD109 (XP_002409560), snail (*Euphaedusa tau*) CD109 (BAE44110), spider (*Hasarius adansoni*) C3, A2M and CD109 (AB622468, AB622470 and AB622471), shrimp (*Fenneropenaeus chinensis*) A2M (ABP97431), crab (*Eriocheir sinensis*) A2M (ADD71943), honey bee (*Apis mellifera*) A2M (XP_392454) and CD109 (XP_001122599), flour beetle (*Tribolium castaneum*) A2M (EFA07508) and CD109 (XP_972838)

flour beetle [19], and *Caenorhabditis elegans* [20] possessed only the A2M family members. Among protostomes, only horseshoe crab [21, 22], spider [6], tick [23], clam [24] and squid [25] were reported to possess the C3 family, indicating that the loss of the C3 family has occurred multiple times during the protostome evolution [6]. The loss of the C3 family also occurred at least once in cnidaria, since hydra has only the A2M family gene [26]. Although the reason why the C3 family genes were lost so many times during the evolution of cnidarians and protostomes is still to be clarified.

Evolution of the Bf, MASP and If Families

The Bf, MASP and if family members are serine proteases possessing a serine protease domain at their C-termini [27]. The human genome contains two *Bf* family genes, *Bf* and *C2*, four *MASP* genes, *MASP*-1, *MASP*-2, *C1r* and *C1s*, and only one *If* family gene. The gene duplication events which played an important role in establishing a modern complement system like the mammalian one will be discussed below. In the following, the common ancestor genes of *C3*, *C4* and *C5*, *Bf* and *C2*, *MASP*-1, *MASP*-2, *C1r* and *C1s* of invertebrates, will be simply termed as *C3*, *Bf* and *MASP*, respectively. Whereas *Bf* and *MASP* are involved in the complement activation pathways, If has a regulatory function to repress complement activation by degrading activated C3 fragments in the presence of the

cofactor proteins. The domain architectures of these families based on the SMART database (http://smart.embl-heidelberg.de/) are CCP (complement control proteins) × 3, VWA (von Willebrand factor type A), Tryp_SPc (trypsin-like serine protease); CUB (C1r, C1s, uEGF and bone morphogenic protein), EGFCA (calcium-binding EGF-like), CUB, CCP × 2, Tryp_SPc; and FIMAC (factor I membrane attack complex), SR (scavenger receptor Cys-rich), LDLa (low-density lipoprotein receptor domain class a) × 2, Tryp_SPc for the Bf, MASP and If families, respectively. The *Bf* family genes have been identified from most deuterostome species analyzed thus far [2], whereas only a few cnidaria [11] and protostome species belonging to arthropod [22] and mollusk [24] are reported to possess the *Bf* family gene. Actually, the *Bf* family gene is missing from the genome sequences of cnidarian hydra, *C. elegans* and several insect species. Although the only *lophotrochozoa* species reported to possess Bf thus far is a clam, the clam Bf lacks the catalytic Ser residue, suggesting that it should not activate C3 in spite of the conservation of the basic domain structure of Bf [24]. It is therefore suggested that although the common ancestor of cnidaria and bilateria had the *Bf* family gene, it was abandoned multiple times in the cnidarian and protostome lineages. Moreover, the Bf-independent complement activation pathway was well characterized in horseshoe crab, an arthropod [21], although it is suggested that horseshoe crab also has the Bf-dependent activation pathway [28]. At the present moment, information on the *Bf* gene in the protostome is very limited compared to that of the *C3* gene, and it is of interest to find out if most protostome species possessing the *C3* gene also retained the *Bf* gene or not. In contrast to the basic conservation of the domain structure throughout evolution, the primary structure of the serine protease domain of Bf shows a curious evolutionary variability. Compared to the other serine proteases, the serine protease domain of mammalian Bf and C2 has a number of structurally unique features, in particular that the bottom of the S1 pocket has a negative charge at Asp226 instead of the usual Asp189 [29]. All jawed vertebrate Bf and C2 so far analyzed have this unique structure, whereas Bf from lamprey [30], ascidian [31], lancelet [32], sea urchin [33], horseshoe crab [22] and sea anemone [11] has Asp189, not Asp226, like all other serine proteases with the trypsin-type specificity. Thus, the structural specialization of the serine protease domain of Bf/C2 seems to have occurred in the common ancestor of the jawed vertebrate, simultaneously with the appearance of the adaptive immune system. The functional consequences of this structural specialization are, if any, still to be clarified.

   The *MASP* family genes have been reported from cnidarian and deuterostome except for echinoderm [2]. No *MASP* family gene has been identified in the deciphered genome sequences of protostomes. However, no comprehensive search for the *MASP* family gene has been performed in the protostome species known to possess the *C3* gene, and it is still an open question whether the MASP family gene is present in some protostome species or not. The serine protease domain of the human MASP family members are classified into two groups, the usual one termed TCN-type in which the active site serine is encoded by a TCN codon, and the other unique one termed AGY-type which is characterized by an AGY codon for the

active site serine, the absence of the disulfide bond termed the histidine loop, and the absence of introns in the genomic region coding for the serine protease domain [34]. Human MASP-1 belongs to the former group, and human MASP-2, C1r and C1s belong to the latter group. The structure of the human *MASP*-1/-3 gene, which encodes both MASP-1 and MASP-3 by differential usage of the dual serine protease regions, has a usual serine protease region at its 3' end, and the second, intron-less serine protease region just upstream of the usual one [35]. Therefore, the AGY-type MASP is considered to be generated by insertion of the intron-less serine protease region into the TCN-type gene [36]. Whereas only the TCN-type *MASP* gene has been identified from sea anemone [11] and sea squirt [37], the AGY-type *MASP* is present in lancelet [38] and vertebrates. Thus, the timing of the insertion of the intron-less serine protease-encoding region is considered to be before the divergence of cephalochordates and vertebrates, although the entire evolutionary story is still to be clarified.

The If family gene has been identified in all major groups of vertebrates, cyclostomes [39], cartilaginous fish [40], teleosts [41], amphibians [42], reptiles [43], birds [44] and mammals. No If family gene has been reported from protostomes and invertebrate deuterostomes, indicating that the If gene was established in the common ancestor of vertebrates. Thus, the If-dependent regulatory mechanism of complement activation seems to be an innovation in vertebrates. Since unrestricted activation of the complement system could be harmful to the host cells and could lead to depletion of the complement components in a short time, it seems to be essential even for the invertebrate complement system to possess some regulatory mechanism.

Evolution of Terminal Complement Component Genes

The domain structures of five human terminal complement component (TCC) genes defined at the SMART site (http://smart.embl-heidelberg.de/) are; C6, TSP1 (Thrombospondin type 1 repeats)-TSP1-LDLa-MACPF (membrane attack complex/perforin)-TSP1-CCP (Domain abundant in complement control proteins)-CCP-FIMAC-FIMAC; C7, TSP1-LDLa-MACPF-TSP1-CCP-CCP-FIMAC-FIMAC; C8A, TSP1-LDLa-MACPF-TSP1; C8B, TSP1-LDLa-MACPF-TSP1; C9, TSP1-LDLa-MACPF. Thus the TSP1-LDLa-MACPF domain structure is conserved by all five TCC genes, suggesting that they originated from a common ancestor by gene duplication and following modification of the domain structure. However it is still not clear whether the common ancestor of the human TCC genes had a simple domain structure like C9 and new genes were generated by adding extra domains or it had a complex domain structure like C6 and new genes were generated by losing some domains [45]. The genes possessing exactly the same or very similar domain structures as human TCC genes have been identified from all classes of extant jawed vertebrates including cartilaginous fish [46, 47] and teleost [48]. In contrast, no such gene has been identified from lamprey. The previous functional

and biochemical analysis of lamprey serum identified an opsonic complement activity but not cytolytic complement activity. Actually, a natural hemolytic activity was present in the lamprey serum, although the only molecule responsible for this hemolytic activity seems to be a 25 kDa protein without any connection to the complement system [49]. All these results indicate that the cytolytic activity of the complement system was established in the common ancestor of jawed vertebrates. Many C9-like genes have been identified from the genomes of urochordate *Ciona intestinalis* [13] and cephalochordate *Branchiostoma floridae* [50]. A typical domain structure of them is TSP1-LDLa-MACPF, indicating a close evolutionary relationship with the human TCC genes. It is highly probable that they are involved in some cytolytic process. However, it is unlikely that they are activated by the complement system of these animals. Activation of human TCC is initiated by the binding of C6 to the activated C5, C5b. The interaction between the C345c domain of C5 and the two FIMAC domains of C6 has been demonstrated to play an essential role in this binding by biochemical [51] and 3D structure [52] analyses. Since none of the C9 like molecules of urochordates and cephalochordates possesses the FIMAC domain, they seem to function as cytolytic factors independent from their complement system reported to have opsonic and preinflammatory functions. Distribution of the complement genes in various animal groups is summarized in Table 3.1.

Evolutionary Scenario of the Complement System

The tentative evolutionary process of the complement system is schematically shown in Fig. 3.2. The traceable origin of the complement system at the present moment is the primitive complement system composed of C3, Bf and MASP in the common ancestor of eumetazoa. Conservation of the amino acid sequences critical for basic functions of these components between cnidaria and humans suggests that the basic activation mechanism and physiological functions of the primitive complement system were not very different from that of the human complement system. Thus, MASP appears to be the first protease to be activated, which in turn activates Bf. Then Bf activates C3 into the C3a and C3b fragments. C3b covalently tags microbes and enhances phagocytosis, whereas C3a induces inflammation as an anaphylatoxin.

In deuterostomes, this basic complement system is retained by all members analyzed thus far; echinoderm, cephalochordate, urochordate and vertebrate. The marked development of the complement system seems to have occurred in the common ancestor of jawed vertebrates. Gene duplication and following structural and functional specialization seem to play an important role in this process. First, C4, C2 and C1r/C1s, most probably generated by gene duplication from *C3, Bf* and *MASP*, respectively, contributed to establish the classical pathway. The canonical adaptive immunity based on IgSF receptors and MHC is also believed to have been established in the common ancestor of jawed vertebrates, and the establishment of

Table 3.1 Distribution of the complement component genes in various animal groups

	Cnidaria	Protostome		Deuterostome				
		Lophotrochozoa	Ecdysozoa	Echinoderm	Cephalochordate	Urochordate	Agnatha	Gnathostome
C3	±	±	±	+	+	+	+	+
Bf	±	±	±	+	+	+	+	+
MASP	±	–	–	–	+	+	+	+
If	–	–	–	–	–	–	+	+
C9-like	–	–	–	–	+	+	–	+
C6-like	–	–	–	–	–	–	–	+

+ indicates that at least one species of that animal group possesses that complement gene. – indicates that at least one species of that animal group was shown not to possess that complement gene by genome analysis. ± indicates that that animal group contains both type species. TCC like genes with and without the FIMAC domain are termed C6-like and C9-like, respectively
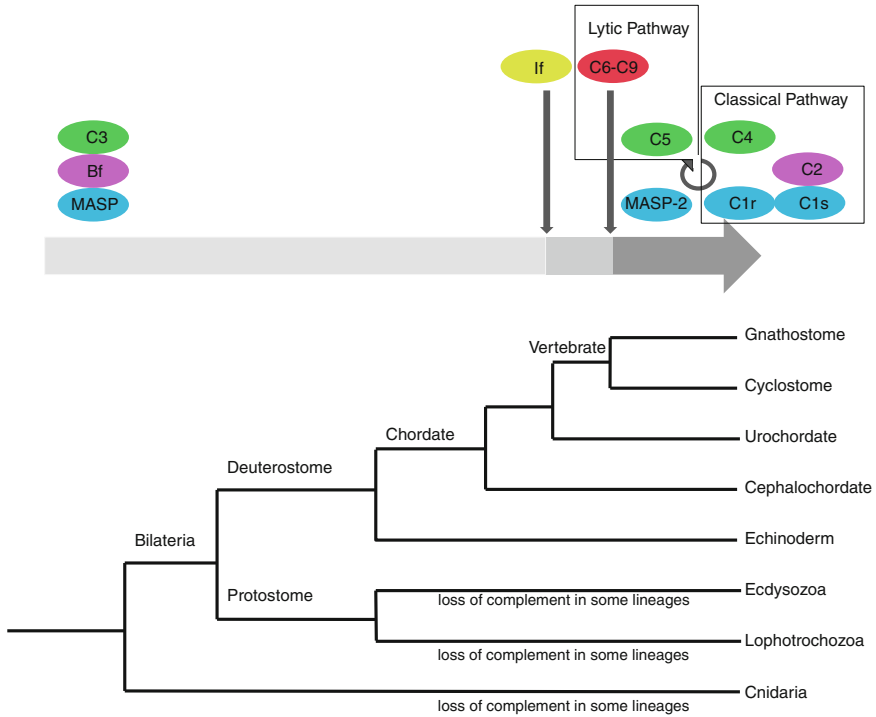
Fig. 3.2 Evolution of the complement system. The bold horizontal arrow represents evolution of the complement system. *Light gray* indicates a primitive complement system composed of C3, Bf and MASP. *Medium gray* indicates an intermediate complement system with the If-dependent regulation mechanism. The *dark gray* indicates a modern complement system found in jawed vertebrates possessing the lytic and classical pathways. Vertical arrows indicate recruitment or innovation of additional components, whereas circular arrow indicates gene duplication events

the classical pathway was epoch-making in the respect that it connected the complement system with adaptive immunity. Second, the gene duplication generating C5 from C3 on the one hand, and the recruitment of the FIMAC domain into the C9-like genes on the other hand connected the preexisting complement system and non-complement cytolytic system.

In contrast to deuterostomes which basically retained the complement system composed of C3, Bf and MASP, the loss of the complement system occurred multiple times in the protostome lineages, and several protostome genome sequences deciphered thus far do not contain any complement genes. Although the reason why the complement system was lost so many times in protostomes is still not clear, in some insects CD109, a paralogue of C3, is reported to be multiplied and play an opsonic role like C3 [53]. Thus, it is possible that some protostomes such as insects developed unique immune mechanisms, making it unnecessary to retain the complement system. Even in horseshoe crab, which retained the complement system, unique specialization is reported compared to the deuterostome

complement system. Thus, factor C originally identified as an LPS-sensitive initiator of hemolymph coagulation stored within hemocytes is used as an initiating protease of the horseshoe crab complement system [21]. In this activation pathway, factor C directly activates C3 without intervention of Bf. Thus, factor C first evolved in the horseshoe crab lineage, and this alteration in the initiating mechanism of the complement system therefore occurred recently on an evolutionary time scale. Thus, as far as the complement system is concerned, the innovations of protostomes seem to be more revolutionary than those of deuterostomes.

References

1. Volanakis JE (1998) Overview of the complement system. In: Volanakis JE, Frank MM (eds) The human complement system in health and disease. Marcel Dekker, New York, pp 9–32
2. Nonaka M, Kimura A (2006) Genomic view of the evolution of the complement system. Immunogenetics 58:701–713
3. Sottrup-Jensen L, Folkersen J, Kristensen T, Tack BF (1984) Partial primary structure of human pregnancy zone protein: extensive sequence homology with human alpha 2-macroglobulin. Proc Natl Acad Sci USA 81:7353–7357
4. Lin M, Sutherland DR, Horsfall W, Totty N, Yeo E, Nayar R, Wu XF, Schuh AC (2002) Cell surface antigen CD109 is a novel member of the alpha(2) macroglobulin/C3, C4, C5 family of thioester-containing proteins. Blood 99:1683–1691
5. Li ZF, Wu XH, Engvall E (2004) Identification and characterization of CPAMD8, a novel member of the complement 3/alpha2-macroglobulin family with a C-terminal Kazal domain. Genomics 83:1083–1093
6. Sekiguchi R, Fujito NT, Nonaka M (2012) Evolution of the thioester-containing proteins (TEPs) of the arthropoda, revealed by molecular cloning of TEP genes from a spider, *Hasarius adansoni*. Dev Comp Immunol 36:483–489
7. Janssen BJ, Huizinga EG, Raaijmakers HC, Roos A, Daha MR, Nilsson-Ekdahl K, Nilsson B, Gros P (2005) Structures of complement component C3 provide insights into the function and evolution of immunity. Nature 437:505–511
8. Srivastava M, Simakov O, Chapman J et al (2010) The *Amphimedon queenslandica* genome and the evolution of animal complexity. Nature 466:720–726
9. King N, Westbrook MJ, Young SL et al (2008) The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. Nature 451:783–788
10. Fujito NT, Sugimoto S, Nonaka M (2010) Evolution of thioester-containing proteins revealed by cloning and characterization of their genes from a cnidarian sea anemone, *Haliplanella lineate*. Dev Comp Immunol 34:775–784
11. Kimura A, Sakaguchi E, Nonaka M (2009) Multi-component complement system of Cnidaria: C3, Bf, and MASP genes expressed in the endodermal tissues of a sea anemone, *Nematostella vectensis*. Immunobiology 214:165–178
12. Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, Kapitonov VV, Jurka J, Genikhovich G, Grigoriev IV, Lucas SM, Steele RE, Finnerty JR, Technau U, Martindale MQ, Rokhsar DS (2007) Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science 317:86–94
13. Azumi K, De Santis R, De Tomaso A, Rigoutsos I, Yoshizaki F, Pinto MR, Marino R, Shida K, Ikeda M, Ikeda M, Arai M, Inoue Y, Shimizu T, Satoh N, Rokhsar DS, Du Pasquier L, Kasahara M, Satake M, Nonaka M (2003) Genomic analysis of immunity in a Urochordate and the emergence of the vertebrate immune system: "waiting for Godot". Immunogenetics 55:570–581

14. Adams MD, Celniker SE, Holt RA et al (2000) The genome sequence of *Drosophila melanogaster*. Science 287:2185–2195
15. Nene V, Wortman JR, Lawson D et al (2007) Genome sequence of *Aedes aegypti*, a major arbovirus vector. Science 316:1718–1723
16. Consortium HGS (2006) Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature 443:931–949
17. Werren JH, Richards S, Desjardins CA et al (2010) Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. Science 327:343–348
18. Consortium (2010) Genome sequence of the pea aphid *Acyrthosiphon pisum*. PLoS Biol 8:e1000313
19. Richards S, Gibbs RA, Weinstock GM et al (2008) The genome of the model beetle and pest *Tribolium castaneum*. Nature 452:949–955
20. Consortium TCeS (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science 282:2012–2018
21. Ariki S, Takahara S, Shibata T, Fukuoka T, Ozaki A, Endo Y, Fujita T, Koshiba T, Kawabata S (2008) Factor C acts as a lipopolysaccharide-responsive C3 convertase in horseshoe crab complement activation. J Immunol 181:7994–8001
22. Zhu Y, Thangamani S, Ho B, Ding JL (2005) The ancient origin of the complement system. EMBO J 24:382–394
23. Buresova V, Hajdusek O, Franta Z, Loosova G, Grunclova L, Levashina EA, Kopacek P (2012) Functional genomics of tick thioester-containing proteins reveal the ancient origin of the complement system. J Innate Immun 3:623–630
24. Prado-Alvarez M, Rotllant J, Gestal C, Novoa B, Figueras A (2009) Characterization of a C3 and a factor B-like in the carpet-shell clam, *Ruditapes decussatus*. Fish Shellfish Immunol 26:305–315
25. Castillo MG, Goodson MS, McFall-Ngai M (2009) Identification and molecular characterization of a complement C3 molecule in a lophotrochozoan, the Hawaiian bobtail squid *Euprymna scolopes*. Dev Comp Immunol 33:69–76
26. Chapman JA, Kirkness EF, Simakov O et al (2010) The dynamic genome of Hydra. Nature 464:592–596
27. Volanakis JE, Arlaud GJ (1998) Complement enzymes. In: Volanakis JE, Frank MM (eds) The human complement system in health and disease. Marcel Dekker, New York, pp 49–81
28. Tagawa K, Yoshihara T, Shibata T, Kitazaki K, Endo Y, Fujita T, Koshiba T, Kawabata S (2012) Microbe-specific C3b deposition in the horseshoe crab complement system in a C2/factor B-dependent or -independent manner. PLoS One 7:e36783
29. Jing H, Xu Y, Carson M, Moore D, Macon KJ, Volanakis JE, Narayana SV (2000) New structural motifs on the chymotrypsin fold and their potential roles in complement factor B. EMBO J 19:164–173
30. Nonaka M, Takahashi M, Sasaki M (1994) Molecular cloning of a lamprey homologue of the mammalian MHC class III gene, complement factor B. J Immunol 152:2263–2269
31. Yoshizaki FY, Ikawa S, Satake M, Satoh N, Nonaka M (2005) Structure and the evolutionary implication of the triplicated complement factor B genes of a *Urochordate ascidian*, *Ciona intestinalis*. Immunogenetics 56:930–942
32. Holland LZ, Albalat R, Azumi K et al (2008) The amphioxus genome illuminates vertebrate origins and cephalochordate biology. Genome Res 18:1100–1111
33. Smith LC, Shih CS, Dachenhausen SG (1998) Coelomocytes express SpBf, a homologue of factor B, the second component in the sea urchin complement system. J Immunol 161:6784–6793
34. Endo Y, Takahashi M, Nakao M, Saiga H, Sekine H, Matsushita M, Nonaka M, Fujita T (1998) Two lineages of mannose-binding lectin-associated serine protease (MASP) in vertebrates. J Immunol 161:4924–4930 (published erratum appears in J Immunol 164:5330)
35. Schwaeble W, Dahl MR, Thiel S, Stover C, Jensenius JC (2002) The mannan-binding lectin-associated serine proteases (MASPs) and MAp19: four components of the lectin pathway activation complex encoded by two genes. Immunobiology 205:455–466

36. Nonaka M, Miyazawa S (2002) Evolution of the initiating enzymes of the complement system. Genome Biol 3:Reviews1001
37. Ji X, Azumi K, Sasaki M, Nonaka M (1997) Ancient origin of the complement lectin pathway revealed by molecular cloning of mannan binding protein-associated serine protease from a urochordate, the Japanese ascidian, *Halocynthia roretzi*. Proc Nat Acad Sci USA 94:6340–6345
38. Endo Y, Nonaka M, Saiga H, Kakinuma Y, Matsushita A, Takahashi M, Matsushita M, Fujita T (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
39. Kimura A, Ikeo K, Nonaka M (2009) Evolutionary origin of the vertebrate blood complement and coagulation systems inferred from liver EST analysis of lamprey. Dev Comp Immunol 33:77–87
40. Terado T, Nonaka MI, Nonaka M, Kimura H (2002) Conservation of the modular structure of complement factor I through vertebrate evolution. Dev Comp Immunol 26:403–413
41. Nakao M, Hisamatsu S, Nakahara M, Kato Y, Smith SL, Yano T (2003) Molecular cloning of the complement regulatory factor I isotypes from the common carp (*Cyprinus carpio*). Immunogenetics 54:801–806
42. Kunnath-Muglia LM, Chang GH, Sim RB, Day AJ, Ezekowitz RA (1993) Characterization of *Xenopus laevis* complement factor I structure–conservation of modular structure except for an unusual insert not present in human factor I. Mol Immunol 30:1249–1256
43. Alfoldi J, Di Palma F, Grabherr M et al (2011) The genome of the green anole lizard and a comparative analysis with birds and mammals. Nature 477:587–591
44. International Chicken Genome Sequencing Consortium (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432:695–716
45. Hobart MJ (1998) Evolution of the terminal complement genes: ancient and modern. Exp Clin Immunogenet 15:235–243
46. Aybar L, Shin DH, Smith SL (2009) Molecular characterization of the alpha subunit of complement component C8 (GcC8alpha) in the nurse shark (*Ginglymostoma cirratum*). Fish Shellfish Immunol 27:397–406
47. Kimura A, Nonaka M (2009) Molecular cloning of the terminal complement components C6 and C8beta of cartilaginous fish. Fish Shellfish Immunol 27:768–772
48. Nakao M, Uemura T, Yano T (1996) Terminal components of carp complement constituting a membrane attack complex. Mol Immunol 33:933–937
49. Nonaka M, Fujii T, Kaidoh T, Natsuume-Sakai S, Yamaguchi N, Takahashi M (1984) Purification of a lamprey complement protein homologous to the third component of the mammalian complement system. J Immunol 133:3242–3249
50. Huang S, Yuan S, Guo L, Yu Y, Li J, Wu T, Liu T, Yang M, Wu K, Liu H, Ge J, Huang H, Dong M, Yu C, Chen S, Xu A (2008) Genomic analysis of the immune gene repertoire of amphioxus reveals extraordinary innate complexity and diversity. Genome Res 18:1112–1126
51. Thai CT, Ogata RT (2004) Complement components C5 and C7: recombinant factor I modules of C7 bind to the C345C domain of C5. J Immunol 173:4547–4552
52. Aleshin AE, DiScipio RG, Stec B, Liddington RC (2012) Crystal structure of C5b-6 suggests structural basis for priming assembly of the membrane attack complex. J Biol Chem 287:19642–19652
53. Levashina EA, Moita LF, Blandin S, Vriend G, Lagueux M, Kafatos FC (2001) Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, *Anopheles gambiae*. Cell 104:709–718