Developmental and Comparative Immunology 50 (2015) 58–67

Contents lists available at ScienceDirect

Developmental and Comparative Immunology

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

Evolution of the complement system in protostomes revealed by *de novo* transcriptome analysis of six species of Arthropoda

Reo Sekiguchi, Masaru Nonaka *

Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Japan

ARTICLE INFO

Article history:
Available online 18 December 2014

Keywords:
Complement C3
Complement factor B
Thioester-containing protein (TEP)
Evolution
Arthropoda

ABSTRACT

To elucidate the evolutionary history of the complement system in Arthropoda, *de novo* transcriptome analysis was performed with six species among the Chelicerata, Myriapoda, and Crustacea, and complement genes were identified based on their characteristic domain structures. Complement C3 and factor B (FB) were identified from a sea spider, a jumping spider, and a centipede, but not from a sea firefly or two millipede species. No additional complement components identifiable by their characteristic domain structures were found from any of these six species. These results together with genome sequence information for several species of the Hexapoda suggest that the common ancestor of the Arthropoda possessed a simple complement system comprising C3 and FB, and thus resembled the alternative pathway of the mammalian complement system. It was lost at least twice independently during the evolution of Arthropoda in the millipede lineage and in the common ancestor of Crustacea and Hexapoda.

1. Introduction

The mammalian complement system comprises more than 30 serum and cell surface proteins, and plays a pivotal role in innate immunity (Volanakis, 1998). Evolutionary studies thus far have indicated that the evolutionary origin of the complement system can be traced back to the common ancestor of the Eumetazoa, because the genes for the central component C3, factor B (FB), the serine protease responsible for C3 activation, and mannan-binding lectin-associated serine protease (MASP) – possibly involved in FB activation – were identified from sea anemones, Cnidaria (Kimura et al., 2009; Putnam et al., 2007), whereas no complement gene is present in the genomes of the sponge, *Amphimedon queenslandica* (Srivastava et al., 2010), or the choanoflagellate, *Monosiga brevicollis* (King et al., 2008). Although the C3 gene has been identified from all deuterostome species analyzed thus far (Nonaka, 2014), earlier genomic analyses showed its absence from the genomes of *Drosophila melanogaster* (Adams et al., 2000) and *Caenorhabditis elegans* (*C. elegans* Sequencing Consortium, 1998), indicating that this gene has been lost in at least some lineages of protostomes. On the other hand, the C3 gene has been reported from several other protostome species, such as the clam (Prado-Alvarez et al., 2009), the squid (Castillo et al., 2009), horseshoe crabs (Ariki et al., 2008; Zhu et al., 2005), the spider (Sekiguchi et al., 2012) and ticks (Buresova et al., 2011; Urbanová et al., 2015), and FB has also been reported from the clam (Prado-Alvarez et al., 2009) and horseshoe crab (Tagawa et al., 2012; Zhu et al., 2005). However, to date, no comprehensive analysis of the complement genes has been performed in protostome species possessing the C3 gene and it is still an unsolved question as to whether the protostome complement system shares complement components other than C3 and FB with deuterostomes.

C3 belongs to the thioester-containing protein (TEP) family; the other members of this family are the nonspecific protease inhibitor alpha-2 macroglobulin (A2M) (Dodds and Law, 1998) and the glycosylphosphatidylinositol (GPI)-anchored protein CD109 (Lin et al., 2002) whose function is still poorly understood. In addition, insect TEP (iTEP) (Blandin and Levashina, 2004) has been reported from the fly and mosquito, which lack C3 and A2M, and certain iTEPs show opsonic activity similar to C3 (Levashina et al., 2001), suggesting that insects compensate for the loss of the complement system by expanding the functions of iTEPs. Later, iTEP was shown to be orthologous to CD109 by extensive phylogenetic analysis (Sekiguchi et al., 2012). Macroglobulin complement-related (Mcr) found from *D. melanogaster* shows overall structural similarity to these TEP family members, although it lacks the thioester site and possesses low density lipoprotein receptor class A (LDLa) domain not found in C3, A2M or CD109/iTEP (Stroschein-Stevenson et al., 2006). Therefore, here we treat the C3, A2M, and iTEP/CD109 subfamilies as authentic members of the TEP family, and Mcr as the closest relative of the TEP family. C3 and pregnancy zone protein (PZP)-like A2M domain-containing 8 (CPAMD8) (Li et al., 2004) thus far found only from deuterostomes are included in the A2M subfamily (Fujito et al.,

* Corresponding author. Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo 113-0033, Japan. Tel.: +81 3 58417589; fax: +81 3 58414990.
E-mail address: mnonaka@bs.s.u-tokyo.ac.jp (M. Nonaka).

http://dx.doi.org/10.1016/j.dci.2014.12.008
0145-305X/

2010). Because all these three subfamily genes have been identified in Cnidarian sea anemones (Fujito et al., 2010; Kimura et al., 2009; Putnam et al., 2007), the gene duplications that generated them, and the subsequent functional diversification seem to have been completed before divergence of the Cnidaria from the Bilateria. The A2M gene has not been identified in some insect genomes sequenced thus far (Adams et al., 2000; Holt et al., 2002; International Aphid Genomics Consortium, 2010), indicating that not only C3 but also A2M has been lost in some of these lineages. In contrast, all three TEP genes have been reported from a spider (Sekiguchi et al., 2012) and ticks (Buresova et al., 2011; Urbanová et al., 2015), so the evolutionary history of TEP genes in the Arthropoda is still unresolved.

The phylum Arthropoda is estimated to contain 5–10 million extant species (Ødegaar, 2000), which are classified into four subphyla, the Chelicerata, Myriapoda, Crustacea, and Hexapoda. Recent molecular phylogenetic studies strongly suggest that the Crustacea subphylum is actually paraphyletic, forming a clade, the Pancrustacea, together with the Hexapoda (Regier et al., 2010). Here we performed RNA sequencing (RNA-seq) analysis of six species belonging to the Chelicerata, Myriapoda, and Crustacea to elucidate the evolution of complement and TEP genes in the Arthropoda.

2. Materials and methods

2.1. Animal collection and isolation of RNA

A sea spider, *Ammothea* sp. (abbreviated below to Amsp), and a sea firefly, *Vargula* sp. (Vasp), were collected at the Misaki Marine Biological Station of the University of Tokyo in Kanagawa, Japan. A jumping spider, *Hasarius adansoni* (Haad), a centipede, *Scolopendra subspinipes* (Scsu), and the millipedes *Niponia nodulosa* (Nino) and *Epanerchodus* sp. (Epsp) were collected in Tokyo, Japan. In the following material, the abbreviations shown in parentheses based on the species names are used as a prefix for each protein name. Total RNA was isolated from the whole body of each animal using ISOGEN (NIPPON GENE Co. Ltd., Tokyo, Japan), except for *S. subspinipes* with its large body, where approximately 10 μg of the gut from several segments was used.

2.2. Sequencing, data processing, and de novo assembly

Construction of cDNA libraries and sequencing were performed by the Beijing Genomics Institute (BGI; Shenzhen, China). The cDNA libraries were constructed using Illumina TruSeq RNA library preparation kits (Illumina Inc., San Diego, CA, USA), and pair-end sequencing (2 × 90 bp) was performed on an Illumina HiSeq 2000 platform. The adaptor sequences were removed from raw reads, and low quality reads (quality value ≤10 for more than 20%) were removed using filter_fq (BGI internal software). High quality reads were obtained and assembled using the Trinity program (Grabherr et al., 2011).

2.3. Gene annotation and estimation of expression levels

To annotate the complement and TEP genes from such massive amounts of sequence data, we used the local Basic Local Alignment Search Tool (BLAST) program. Local BLAST software was downloaded from the NCBI website (http://blast.be-md.ncbi.nlm.nih.gov/Blast.cgi). A database was constructed from the assembled contigs of each animal, and the amino acid sequences of eumetazoan complement component or TEP were used as queries for tBLASTN. To estimate the expression level of the annotated genes, Bowtie (ver. 1.0.0) (Langmead et al., 2009) and eXpress (ver. 1.5.1) (Roberts and Pachter, 2013) softwares were used to calculate the values of 'fragments per kilobase of exon per million mapped fragments' (FPKM).

2.4. Phylogenetic analysis of the complement and TEP genes

The amino acid sequences of the C3, FB, A2M, and iTEP/CD109 proteins of various animal species occupying critical phylogenetic positions for understanding evolution were obtained from the NCBI database (see earlier). These sequences were aligned with the amino acid sequences of these proteins of the six arthropod species deduced in this present study using ClustalX (http://www.clustal.org/clustal2) (Thompson et al., 1997). Some manual corrections of the alignments and following evolutionary analyses were conducted in MEGA6 (http://www.megasoftware.net) (Tamura et al., 2013). Upon model testing, the lowest Bayesian Information Criterion (BIC) score model was used and a phylogenetic tree was constructed using the maximum likelihood (ML) method (Goldman et al., 2000).

3. Results

3.1. Generation of raw reads and assembly of reads

The raw reads generated on the Illumina Hiseq 2000 platform were trimmed for adaptor sequences, and low quality reads were excluded from subsequent analyses. High quality reads were assembled using Trinity software. As shown in Table 1, approximately 50 million high quality reads were obtained for each species, which were assembled into 37,757–155,223 contigs. The mean contig length and N50 values (defined as the summed lengths of scaffolds/contigs from the longest to the shortest, and reaching 50% of the total assembly size) are also shown in Table 1. The values of *Vargula* sp. and *H. adansoni* were low, most probably because of the difficulty in obtaining high quality RNA samples from these species.

3.2. TEP family

To annotate the C3, A2M, and iTEP/CD109 proteins of each species, the amino acid sequences of arthropod and human TEP were used as queries for tBLASTN against local BLAST databases, which consisted of the assembled contigs of each arthropod species. The TEPs identified from each arthropod species are shown in Table 2. A2M and iTEP were identified from all analyzed species, whereas C3 was identified only from *Ammothea* sp., *H. adansoni*, and *S. subspinipes*. The deduced protein sequences are named below as four-character prefixes representing the species name plus protein name. When multiple isotypes were present, numbers are given as suffixes. Thus, the six TEPs identified from *Ammothea* sp. have been designated AmspC3, AmspA2M-1, AmspA2M-2, AmspA2M-3, AmspA2M-4, and AmspiTEP. The entire amino acid sequences were determined for all TEP proteins analyzed here, except for VaspA2M-1, VaspA2M-2, and VaspiTEP. Only partial amino acid sequences were obtained for these three TEPs of *Vargula* sp., presumably because of poor total RNA quality. The deduced amino acid sequences of arthropod TEPs were aligned with eumetazoan TEP sequences by ClustalX using the

Table 1
Summary statistics of *de novo* transcriptome analysis of several arthropods.

	High quality reads	Assembled contigs	Mean contigs length (bp)	N50 value (bp)
Ammothea sp.	46,754,904	47,309	1087	2134
Hasarius adansoni	53,164,698	102,190	635	1266
Scolopendra subspinipes	53,851,246	155,223	1434	3090
Niponia nodulosa	51,298,384	124,129	1961	3754
Epanerchodus sp.	46,902,374	37,757	1283	2557
Vargula sp.	93,520,202	112,077	646	925