DNA Research **10**, 203–212 (2003)

Identification and Sequence of Seventy-nine New Transcripts Expressed in Hemocytes of *Ciona intestinalis*, Three of Which May Be Involved in Characteristic Cell-cell Communication

Daichi Terajima,[1] Shigeyuki Yamada,[1,2] Ryuji Uchino,[1] Shuntaro Ikawa,[1] Makoto Ikeda,[1,3] Kazuhito Shida,[3] Yoichi Arai,[2] Hong-Gang Wang,[3,4] Nori Satoh,[5] and Masanobu Satake[1,3,*]

Institute of Development, Aging and Cancer, Tohoku University, Sendai 980-8575, Japan,[1] Department of Urology, Graduate School of Medicine, Tohoku University, Sendai 980-8574, Japan,[2] Center for Interdisciplinary Research, Tohoku University, Sendai 980-8578, Japan,[3] H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida 33612, USA,[4] and Department of Zoology, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan[5]

(Received 4 August 2003; revised 8 September 2003)

Abstract

Ascidian is a useful experimental animal for studying body planning principles and host defense mechanisms employed by the phylum chordata. Toward this goal, genome and cDNA/EST projects of *Ciona intestinalis* have been undertaken. Using cDNAs and ESTs derived from *Ciona* hemocytes, we identified 79 possible hemocyte-preferential transcripts and determined the cDNA sequence of each clone. The amino acid sequence of each encoded polypeptide was predicted as well. Among these cDNAs, we identified three transcripts that may be involved in characteristic cell-cell communication in *Ciona*. These transcripts encoded leucine-rich repeat-containing RP105-like, IL-17 receptor/similar expression to FGF-like, and ectodysplasin-like polypeptide of the tumor necrosis factor family, and they are expressed abundantly in hemocytes.

Key words: ascidian; cDNA; hemocytes; cell-cell communication

1. Introduction

Ascidians occupy a unique position in the evolution of deuterostomes. In both ascidian, categorized as urochordata, and amphioxus, categorized as cephalochordata, the development of the notochord is a critical event for induction of the dorsal neural tube. Ascidian is therefore a key animal for understanding the body plan principles that are employed by the phylum chordata, which includes vertebrates. To elucidate these principles, various attempts to screen ascidian mutants[1,2] and to genetically manipulate ascidians[3,4] have been reported. In parallel, multiple genomic studies of this animal are now being performed. For example, a draft genome sequence[5] and a series of cDNA/EST projects of *Ciona intestinalis*, an ascidian species that is used worldwide, have recently been reported. The cDNA/EST projects have covered several different stages of ascidian development, including fertilized eggs,[6] cleavage-stage embryos,[7] tailbud embryos,[8] larvae,[9] young adults,[10] and testis.[11] The total number of collected ESTs has reached 81,637, and they have been grouped into 13,464 clusters[12,13] that probably correspond to independent transcriptional units and most likely cover 85% of the predicted *Ciona* genes.

The host defense mechanisms of ascidian are also interesting. Although elements of acquired immunity have not been found, the innate immune function of ascidian is considered to be a prototype of the corresponding system in vertebrates.[14,15] We recently performed a cDNA/EST project of *Ciona* hemocytes,[16] which play a major role in its host defense mechanism. We have collected 3357 ESTs and grouped them into 1889 clusters. Based on homology searches and domain analyses, 530 clusters were found to be homologous to known genes with some function, and 62 out of 530 clusters represented transcripts involved in cytotoxicity, detoxification, and inflammation. On the other hand, 1359 clusters (72%) did not show significant similarities to known proteins or did not give enough information to speculate with regard to protein function.

Considering the present situation that the draft genome and cDNA/EST sequences are available, one of the next goals in the genome science of ascidian is the full length sequencing of each cDNA clone and their assign-

Communicated by Satoshi Tabata

* To whom correspondence should be addressed. Tel. +81-22-717-8477, Fax. +81-22-717-8482, E-mail: satake@idac.tohoku.ac.jp

ment on the genome sequence. In particular, we have an interest in the features of transcripts which are expressed preferentially in hemocytes, a point not addressed in our previous EST study.[16] In this report, we identified 79 transcripts that appear to be preferentially expressed in *Ciona* hemocytes, sequenced each cDNA clone and predicted its coding sequence. We also examined the expression patterns of several of these transcripts that may play roles in cell-cell communication.

2. Materials and Methods

2.1. In silico *subtraction and cDNA sequencing*

The features of 3357 ESTs derived from a *Ciona* hemocyte cDNA library were published[16] and registered in a db/EST section of the NCBI/GenBank. Their accession numbers are BM959590 through BM962952. We performed a nucleotide (nt)-based BLAST[17] search (blastn) using these 3357 hemocyte-derived ESTs as queries[16] against an EST database containing 81,637 ESTs that represent several developmental stages.[12,13] It must be noted that no expression was detected in hemocytes by the whole mount *in situ* hybridization, in which each of 976 independent clusters was selected randomly from the young adults ESTs and used as a probe.[10] This suggests that the amount of hemocytes is not abundant, if any are present, in the young adult, and indicates that the subtraction procedure described above is reasonable. Thus, we identified 110 ESTs of hemocyte origin that did not show significant homology to any ESTs in the database (score < 100).

The entire sequence of both strands of each cDNA insert was determined by a primer-walking method. We then used these sequences as queries to perform a second blastn BLAST search against the same database of 81,637 ESTs. Out of 110 original sequences, 17 sequences were now found to have corresponding ESTs, but 93 sequences still did not match any EST in the database. We used the Phrap program (provided through the courtesy of Philip Green, University of Washington, Seattle, USA) to assemble several of the 93 sequences into contigs that are presumably derived from identical or alternatively spliced transcripts. Finally, a total of 79 possible hemocyte-preferential transcription units were identified and these sequences were registered in NCBI/GenBank.

2.2. 5′-RACE and RT-PCR

For two of the cDNA clones, cihA3I8 and cihA4F14 (cih is the abbreviation of *Ciona intestinalis* hemocytes), a 5′-RACE method was also performed to extend their 5′ portions. The DNA prepared from a bacterial lysate of a hemocyte cDNA library was used as a template. The primers were set on the cDNA and vector sequences, respectively, and sequences were extended by several 5′-RACE cycles using Ex Taq polymerase (Takara). Thus, this method of 5′-RACE was actually equivalent to that of PCR. The extended sequences were determined by sequencing the PCR products, combined with those derived from the cDNA inserts and were assembled into contiguous sequences.

A semi-quantitative RT-PCR method was performed as described previously.[16] The forward and reverse primers were set on two separate portions of cDNA sequences derived from a different exon. The sequences of forward and reverse primers, respectively, were as follows; 5′-cggaggattcaacagcggacgatt-3′ and 5′-tcgtcacctgccactcaaacggaat-3′ for *ectodysplasin-like*, 5′-gtcacctgtcacaagtgcgtcac-3′ and 5′-ccgatgcgaaggtacagtatgcaac-3′ for *CD40 ligand-like*, 5′-atacctgctcatctgccgtatccct-3′ and 5′-taagggcaggcgaagaaatatggg-3′ for *Fas-ligand-like*, 5′-cggctaacgaatacgaggac-3′ and 5′-acattagattcgctgatgacatatg-3′ for *TNFα-like*, 5′-aatgtccgagtttccagaacg-3′ and 5′-actcacccaactgtagcaaagtctt-3′ for *IL-17R/SEF-like*, 5′-tcgtagattaccatcacagctagcag-3′ and 5′-ccgagtcgaactactgggacaatata-3′ for *RP105-like*, 5′-gttgatgctgacggcaacg-3′ and 5′-tcaatcagcctatggaatga-3′ for *calmodulin*. A fixed amount of cDNA was used for PCR at each of 25, 30, and 35 cycles where one cycle consisted of 30 sec at 94°C, 30 sec at 64°C, and 60 sec at 72°C. The reaction mixture containing Taq DNA polymerase from *Thermus aquaticus* was set up according to the manufacturer's instruction (Sigma) and the reaction products were run through agarose gels. The relative intensity of amplified product compared to the internal DNA marker was determined using the NIH Image 1.63 program (http://rsb.info.nih.gov/nih-image/download.html).

3. Results and Discussion

3.1. Functional classification of cDNA sequences

First, we attempted to determine what polypeptides are encoded by the 79 possible hemocyte-preferential cDNAs. Each sequence was analyzed by a translated BLAST search (tblastx) against the NCBI/GenBank database. The results were classified according to the putative function of each protein. Table 1 shows the classification of each cDNA clone. This classification was originally used in the EST analysis of sea urchin embryos[18] and has been employed in a series of *Ciona* EST projects.[6–11] The cDNAs in groups A, B, and C showed homology to proteins of known function and together represented 52% of the total cDNAs. The cDNAs in group DI showed significant homology to "hypothetical," "unnamed," or "unknown" proteins. In contrast, group DII contained the largest number (26) of cDNAs, which showed no significant homology to any known proteins.

Table 2 shows the gene to which each of the cDNAs in groups A–DI showed the highest homology. The corresponding gene assignment in the draft *Ciona* genome

Table 1. The number of clones as classified into functional groups.

Class			Number of clones
(A) Functions that many kinds of cells use			
A I	Transpotation and binding proteins for ions and small molecules		3
A II	RNA processing,polymerizing,splicing and binding proteins,and enzymes		0
A III	Cell replication,histones,cyclins and allied kinases,DNA polymerases,topoisomerases,DNA modification		0
A IV	Cytoskelton and membrane proteins		4
A V	Protein synthesis co-factors,tRNA synthetases,ribosomal proteins		1
A VI	Intermediary synthesis and catabolism enzymes		5
A VII	Stress response,detoxification and cell defense proteins		1
A VIII	Protein degradation and processing,proteases		2
A IX	Transpotation and binding proteins for proteins and other macromolecules		2
		Subtotal	18
(B) Cell-cell communication			
B I	Signaling receptors,including cytokine and hormone receptors,and signaling ligands		10
B II	Intracellular signal transduction pathway molecules including kinases and signal intermediates		2
B III	Extracellular matrix proteins and cell adhesion		10
		Subtotal	22
(C) Transcription factors and other gene regulatory proteins			1
		Subtotal	1
(D) Miscellaneous			
D I	Not enough information to classify		12
D II	Not significant similarities to known proteins		26
		Subtotal	38
		Total	79

sequence for each cDNA is also provided in Table 2. This assignment should be useful to predict exon-intron boundaries of genes by comparing the cDNA and genomic sequences.

3.2. Putative cDNA coding sequences

Next, we tried to predict the possible coding sequence (CDS) of each cDNA clone (Table 3). The longest reading frame was first determined mechanistically. The amino acid (aa) sequence deduced from such reading frame was then used as a query in an aa-based BLAST search (tblastn) against the NCBI/GenBank database. For each putative CDS, the protein with the highest homology was identical to that which is presented in Table 2. Therefore, the CDS predicted for each clone in groups A–DI most likely represents the authentic protein. There were no similar criteria for the clones in group DII, so only CDSs longer than 100 aa residues were tentatively predicted for each of these clones. Overall, the longest CDS, which was predicted for cihA1M21, was 1970 aa long, and the shortest CDS was 65 aa long, predicted for cihA5D7. The average CDS length was 360 aa.

Each predicted CDS was characterized by the presence or absence of an in-frame upstream termination codon (u.t.c.), an initiating methionine codon (i.m.c.), or a translation termination codon (t.c.) (Table 3). If a CDS has all three elements, it is considered to correspond to an authentic open reading frame (ORF); 22 clones fell into this category. The most common type of CDS (43 clones) had a t.c. but no u.t.c. or i.m.c. This type of CDS most likely represents the carboxy-terminal portion of an ORF. There were five u.t.c. (−), i.m.c. (−), t.c. (−) CDSs and three u.t.c. (+), i.m.c. (+), t.c. (−) CDSs, which probably correspond to the central and amino-terminal portion of ORFs, respectively. The above-described features of various types of CDSs indicate that, although one-quarter of the cDNA clones appear to encode an intact ORF, extension of the cDNA, particularly in the 5′ direction, will be necessary to determine the entire ORF of the remaining clones. Indeed, we have successfully done this for two clones, cihA3I8 and cihA4F14, as described in Materials and Methods.

3.3. Features of 79 hemocyte transcripts

So far, we have identified and sequenced 79 cDNAs, and predicted their CDS. They represent the candidate genes expressed preferentially in *Ciona* hemocytes. One of the characteristics of these cDNAs appear to be the presence of group DII. The genes in DII occupied 33% of total genes analyzed in this study (Table 1). It should be noted that CDSs could be assigned for most of these sequences (Table 3). Since the BLAST search using these CDSs did not hit any significantly homologous gene in the data base, it is likely that these peptide-encoding genes indeed represent new genes.

To confirm the above point, we further searched by BLAST whether the genomes of *Fugu rubripes* and *Danio rerio*, species closest to *Ciona* among vertebrates, contain genes corresponding to those of *Ciona*. As seen in Table 4, the *Ciona* genes in group DII again did not hit any homologous sequence in *Fugu* (nor in *Danio*, data not shown). Therefore, it is highly likely that *Ciona* hemocytes preferentially express a number of unknown genes whose function has not yet been elucidated. It is also

Table 2. Functional characterization of each clone based on a BLAST homology search.

Class	Clone ID	Gene assignment (*Ciona* genome)	Homologous to	Organism	Probability
A I	cihA1E10		degenerin	*Caenorhabditis elegans*	8E-05
	cihA11H14	ci0100139925	solute carrier family 21	*Mus musculus*	5E-35
	cihA11I4	ci0100138296	solute carrier family 22	*Mus musculus*	7E-48
A IV	cihA5I18	ci0100130989	ankyrin repeats	*Caenorhabditis elegans*	4E-50
	cihA6O10	ci0100144414	Huntingtin interacting protein 1	*Homo sapiens*	E-22
	cihA10C1	ci0100135368	ascidian cytoplasmic gelsolin	*Halocynthia roretzi*	E-106
	cihA11J9	ci0100135941	flightless I homolog (Drosophila);	*Homo sapiens*	0
A V	cihA7I1	ci0100154576	eukaryotic translation initiation factor 2	*Mus musculus*	4E-89
A VI	cihA3G13		similar to phospholipase A2, group IVB (cytosolic)	*Mus musculus*	5E-07
	cihA3P11	ci0100139554	sphingomyelin phosphodiesterase 1, acid lysosomal	*Mus musculus*	E-118
	cihA4H9	ci0100150084	Adenylate cyclase, type VI	*Canis familiaris*	E-121
	cihA5G13	ci0100154282	D-2-hydroxy-acid dehydrogenase	*Homo sapiens*	E-23
	cihA11C9	ci0100149204	N-acetylated alpha-linked acidic dipeptidase 2	*Homo sapiens*	8E-27
A VII	cihA8F23	ci0100138088	Glutathione-requiring prostaglandin D synthase	*Oryctolagus cuniculus*	5E-30
A VIII	cihA5E23	ci0100150284	Dcp-1-P1; caspase; caspase-1	*Gallus gallus*	E-23
	cihA6E14	ci0100131776	zinc metalloprotease	*Caenorhabditis elegans*	E-08
A IX	cihA8E11	ci0100143947	similar to Type II membrane protein of ER	*Mus musculus*	6E-83
	cihA9N5	ci0100146822	brain secretory protein SEC10P	*Homo sapiens*	3E-66
B I	cihA1N4	ci0100149595	glycoprotein 330	*Rattus norvegicus*	3E-18
	cihA2K11	ci0100136887	EDG-3	*Takifugu rubripes*	7E-25
	cihA3C10		bA351K23.6 (ectodermal dysplasia 1, anhidrotic)	*Homo sapiens*	4E-05
	cihA3I8	ci0100144783	Toll-7	*Drosophila melanogaster*	3E-22
	cihA4F14	ci0100141150	interleukin 17 receptor	*Mus musculus*	E-11
	cihA5H2	ci0100143261	kappa opioid receptor	*Rattus norvegicus*	2E-25
	cihA7E7	ci0100137028	lectomedin-1 alpha	*Homo sapiens*	6E-86
	cihA8H19		candidate tumor suppressor protein	*Homo sapiens*	E-05
	cihA11J14	ci0100145358	G-protein coupled receptor GRL101 precursor	*Lymnaea stagnalis*	3E-19
	cihA12M13	ci0100131463	receptor tyrosine phosphatase	*Hirudo medicinalis*	E-68
B II	cihA2F13	ci0100147545	similar to ALS2CR17	*Mus musculus*	3E-66
	cihA3K23	ci0100131253	calmodulin	*Oryctolagus cuniculus*	E-36
B III	cihA2N21	ci0100152017	integrin alpha Hr1 precursor	*Halocynthia roretzi*	2E-11
	cihA4L16		CUB domain, von Willebrand factor type A domain	*Caenorhabditis elegans*	0.021
	cihA6E2	ci0100137830	L-selectin precursor	*Pongo pygmaeus*	5E-05
	cihA8L19	ci0100131884	cortical granule lectin	*Xenopus laevis*	6E-31
	cihA8M21		similar to polydom protein	*Homo sapiens*	0.071
	cihA9G6		caspr5 protein isoform 1	*Homo sapiens*	2E-13
	cihA9P2	ci0100140978	HrTT-1	*Halocynthia roretzi*	3E-65
	cihA10E1		brevican soluble core protein precursor	*Xenopus laevis*	0.002
	cihA10G11	ci0100140962	similar to hemicentin	*Homo sapiens*	3E-31
	cihA1M21	ci0100130565	polydomain protein	*Mus musculus*	2E-66
C	cihA1C9	ci0100131764	sirtuin 3	*Homo sapiens*	2E-28
D I	cihA1O20		hypothetical protein	*Thermoplasma acidophilum*	0.019
	cihA1P10	ci0100148292	similar to Hypothetical protein KIAA0233	*Mus musculus*	E-81
	cihA2G5	ci0100130724	KIAA0701 protein	*Homo sapiens*	4E-47
	cihA3N23	ci0100131699	hypothetical protein FLJ14454	*Homo sapiens*	7E-22
	cihA5D7	ci0100138753	HSPC300	*Homo sapiens*	3E-14
	cihA5H10		unknown protein	*Oryza sativa*	E-25
	cihA7M17		similar to hypothetical protein BC018147	*Mus musculus*	8E-17
	cihA8F16		unnamed protein product	*Homo sapiens*	0.086
	cihA8G8	ci0100131233	unnamed protein product	*Homo sapiens*	9E-38
	cihA9G14	ci0100137578	RIKEN cDNA 2810405J04	*Mus musculus*	4E-50
	cihA10L19	ci0100153635	RIKEN cDNA 2810439F02	*Mus musculus*	2E-57
	cihA11M1	ci0100153995	GH09970p	*Drosophila melanogaster*	3E-90

For several cDNA clones, the corresponding gene could not be assigned in the draft *Ciona* genome[5] either due to the lack of gene prediction on the genome sequence or due to the lack of genome sequence itself. A p value less than E-05 was considered to represent a significant homology. Note that five cDNA clones whose p values were more than E-05 were tentatively assigned to the indicated class. They corresponded to cihA4L16, cihA8M21, cihA10E1, cihA1O20, and cihA8F16.

noteworthy that the average aa identity of known *Ciona* and *Fugu* genes in the homologous region in groups A–C was as low as 39%. This suggests that *Ciona* and vertebrates are evolutionarily distant.

Another characteristic feature seen from Table 2 is that most of the genes in groups A–C did not appear to be immediately related to host defense. This was unexpected, since we speculated initially that the genes involved in immunity would be expressed more or less preferentially in hemocytes. However, taking into consideration the present situation whereby the molecular markers of ascidian hemocytes are not known, the genes listed in Table 2 can serve as a basis to search for molecular diagnostic probes of *Ciona* hemocytes.

Table 3. Prediction of coding sequence for each clone.

Class	Clone ID	Accession Number (NCBI/GenBank)	cDNA size (nt)	CDS (nt)	protein (aa)	upstream term. codon	initiat. Met codon	term. codon	oligo(A) tail
A I	cihA1E10	AY261844	679	2-439	146	-	-	+	+
	cihA11H14	AY261834	1,422	1-885	295	-	-	+	+
	cihA11I4	AY261835	1,219	3-1,217	405	-	-	-	-
A IV	cihA5I18	AY261871	1,284	2-1,096	365	-	-	+	-
	cihA6O10	AY261875	300	2-298	99	-	-	-	-
	cihA10C1	AY261827	1,428	1-1,149	383	-	-	+	+
	cihA11J9	AY261837	2,319	1-1,755	586	-	-	+	+
A V	cihA7I1	AY261877	2,539	1-1,845	615	-	-	+	-
A VI	cihA3G13	AY261856	1,456	15-1,454	480	+	+	-	-
	cihA3P11	AY261860	2,155	2-1,801	600	-	-	+	+
	cihA4H9	AY261863	2,527	3-1,322	440	-	-	+	-
	cihA5G13	AY261868	906	1-480	114	-	-	+	+
	cihA11C9	AY261833	537	1-447	149	-	-	+	-
A VII	cihA8F23	AY261882	933	1-597	199	-	-	+	-
A VIII	cihA5E23	AY261867	1,491	3-767	255	-	-	+	+
	cihA6E14	AY261872	835	1-705	235	-	-	+	+
A IX	cihA8E11	AY261880	817	3-716	238	-	-	+	-
	cihA9N5	AY261897	731	3-587	195	-	-	+	+
B I	cihA1N4	AY261846	4,259	1,098-2,537	480	+	+	+	+
	cihA2K11	AY261853	1,293	2-1,156	385	-	-	+	+
	cihA3C10	AY261855	1,348	63-983	307	+	+	+	+
	cihA3I8	AY261857	3,564	46-2,847	934	+	+	+	-
	cihA4F14	AY261862	2,938	62-2,371	770	+	+	+	+
	cihA5H2	AY261870	1,895	13-1,500	496	+	+	+	+
	cihA7E7	AY261876	2,499	148-2,187	680	+	+	+	-
	cihA8H19	AY261884	513	1-414	138	-	-	+	+
	cihA11J14	AY261836	2,318	218-1,963	582	+	+	+	-
	cihA12M13	AY261842	3,144	1,308-2,645	446	+	+	+	+
B II	cihA2F13	AY261850	1,770	3-926	307	-	-	+	+
	cihA3K23	AY261858	935	126-638	171	+	+	+	+
B III	cihA2N21	AY261854	2,024	2-1,669	556	-	-	+	-
	cihA4L16	AY261865	284	3-209	69	-	-	+	-
	cihA6E2	AY261873	1,426	2-994	331	-	-	+	+
	cihA8L19	AY261886	2,018	79-1,818	580	+	+	+	-
	cihA8M21	AY261887	647	2-616	205	-	-	+	+
	cihA9G6	AY261895	1,304	24-950	309	+	+	+	+
	cihA9P2	AY261898	1,514	3-1,514	504	-	-	-	-
	cihA10E1	AY261828	1,413	1-705	235	-	-	+	+
	cihA10G11	AY261829	1,110	2-718	239	-	-	+	+
	cihA1M21	AY261899	6,075	1-5,910	1,970	-	-	+	+
C	cihA1C9	AY261843	895	1-831	277	-	-	+	-
D I	cihA1O20	AY261847	587	190-435	76	+	+	+	-
	cihA1P10	AY261848	2,576	270-1,941	557	+	+	+	-
	cihA2G5	AY261852	4,162	3-3,506	1,168	-	-	+	+
	cihA3N23	AY261859	1,809	3-1,697	565	-	-	+	-
	cihA5D7	AY261866	434	1-195	65	-	-	+	+
	cihA5H10	AY261869	942	1-906	302	-	-	+	+
	cihA7M17	AY261878	1,637	553-948	132	+	+	+	+
	cihA8F16	AY261881	746	3-407	135	-	-	+	+
	cihA8G8	AY261883	2,935	1,413-2,726	438	+	+	+	+
	cihA9G14	AY261894	1,355	84-1,304	407	+	+	+	+
	cihA10L19	AY261831	1,304	3-704	234	-	-	+	+
	cihA11M1	AY261838	1,806	1-810	270	-	-	+	+
D II	cihA8L18	AY261885	2,896	112-2,481	790	+	+	+	-
	cihA8B16	AY261879	2,075	97-1,443	449	+	+	+	-
	cihA9B2	AY261891	1,295	2-1,144	381	-	-	+	+
	cihA2F15	AY261851	1,350	135-1,127	331	+	+	+	-
	cihA1H22	AY261845	1,356	3-1,139	326	-	-	+	+
	cihA9A15	AY261888	1,031	87-1,031	318	+	+	-	-
	cihA9K16	AY261896	1,635	439-1,377	313	+	+	+	-
	cihA10P13	AY261832	1,199	39-968	310	+	+	+	+
	cihA12K2	AY261841	879	1-879	293	-	-	-	-
	cihA9B11	AY261890	1,311	1-678	226	-	-	+	-
	cihA10G7	AY261830	1,173	1-624	208	-	-	+	+
	cihA11O6	AY261840	616	2-514	171	-	-	+	-
	cihA3P6	AY261861	636	1-498	166	-	-	+	-
	cihA9D22	AY261892	1,234	2-460	153	-	-	+	-
	cihA4I22	AY261864	953	6-452	149	+	+	+	+
	cihA2D1	AY261849	443	2-442	146	-	-	-	-
	cihA11N1	AY261839	1,070	1-369	123	-	-	+	+
	cihA9A21	AY261889	522	171-521	117	+	+	-	-
	cihA9D8	AY261893	514	2-340	113	-	-	+	+
	cihA6K13	AY261874	524	2-301	100	-	-	+	+
	cihA2K12	CB556142	932						+
	cihA8I3	CB556151	794						+
	cihA9A13	CB556153	730						-
	cihA9C11	CB556154	2,072						-
	cihA10C11	CB556181	1,751						+
	cihA11M14	CB556139	1,020						-

In this Table, a termination codon is included in the CDS as well as in the number of aa in the predicted protein. The annotated sequences are registered to NCBI/GenBank as an AY series, whereas those registered in a dbEST section of NCBI/GenBank have a heading of CB.

Table 4. Comparison of the *Ciona* genes with the *Fugu* genes.

Ciona intestinalis			*Fugu rubripes*				
Class	Clone ID	Homologous region (aa)	Gene ID	Homologous to	Homologous region (aa)	Probability	aa identities
A I	cihA1E10		No				
	cihA11H14	2-277	FRUP00000156327	PGE1 transporter	357-626	2.00E-28	27%
	cihA11I4	4-377	FRUP00000128936	Organic cation/carnitine transporter 2	17-382	4.00E-48	30%
A IV	cihA5I18	41-260	FRUP00000150913	similar to ankyrin-like protein 1	585-801	2.00E-31	35%
	cihA6O10	7-87	FRUP00000148290	Huntingtin interacting protein 1	676-756	2.00E-24	70%
	cihA10C1	1-382	FRUP00000145498	Gelsolin precursor, plasma	337-716	1.00E-104	49%
	cihA11J9	4-584	FRUP00000145191	Flightless-I protein homolog	609-1230	1.00E-175	51%
A V	cihA7I1	71-539	FRUP00000144195	Eukaryotic translation initiation factor 2-alpha kinase 3 precursor	338-765	6.00E-67	34%
A VI	cihA3G13		No				
	cihA3P11	197-590	FRUP00000157529	Acid sphingomyelinase-like phosphodiesterase 3a precursor	8-395	2.00E-50	33%
	cihA4H9	20-434	FRUP00000143879	adenylyl cyclase type VI	395-827	1.00E-118	54%
	cihA5G13	3-109	FRUP00000144676	Similar to glyoxylate reductase/hydroxypyruvate reductase	189-294	3.00E-17	50%
	cihA11C9	2-142	FRUP00000165024	hypothetical protein	220-366	2.00E-21	42%
A VII	cihA8F23		No				
A VIII	cihA5E23	99-184	FRUP00000131803	CASPASE 8 precursor	13-98	2.00E-11	34%
	cihA6E14		No				
A IX	cihA8E11	1-229	FRUP00000147251	EDEM protein	366-628	1.00E-76	53%
	cihA9N5	3-194	FRUP00000152759	Brain secretory protein hSec10p	416-609	3.00E-65	59%
B I	cihA1N4	9-275	FRUP00000162326	Ensembl_locations(Chr-bp):8-34701068	70-334	4.00E-21	27%
	cihA2K11	34-349	FRUP00000154122	EDG-3	25-315	4.00E-29	28%
	cihA3C10		No				
	cihA3I8	339-766	FRUP00000156964	GARP protein precursor	23-427	1.00E-18	25%
	cihA4F14		No				
	cihA5H2	34-437	FRUP00000132165	Nociceptin receptor	2-285	1.00E-11	20%
	cihA7E7	29-662	FRUP00000154051	lectomedin-3	8-643	3.00E-74	30%
	cihA8H19		No				
	cihA11J14	154-543	FRUP00000136130	Relaxin receptor 1	119-439	9.00E-20	23%
	cihA12M13	1-419	FRUP00000152564	protein tyrosine phosphatase, receptor type, sigma, isoform 2 precursor	546-994	9.00E-51	31%
B II	cihA2F13	1-302	FRUP00000160557	Ensembl_locations(Chr-bp):1-60864675	575-879	2.00E-75	44%
	cihA3K23	30-161	FRUP00000146183	calmodulin	2-130	2.00E-34	54%
B III	cihA1M21	21-503	FRUP00000130024	P-selectin precursor	177-680	2.00E-58	28%
	cihA2N21		No				
	cihA4L16		No				
	cihA6E2		No				
	cihA8L19		No				
	cihA8M21		No				
	cihA9G6	147-305	FRUP00000143188	Caspr5	354-513	9.00E-17	29%
	cihA9P2	260-498	FRUP00000128411	Brain-specific angiogenesis inhibitor 1 precursor	3-267	4.00E-21	26%
	cihA10E1		No				
	cihA10G11	6-229	FRUP00000148438	KIAA0960 protein	534-812	2.00E-15	26%
C	cihA1C9	7-122	FRUP00000142753	SIRTUIN type 3	179-291	3.00E-28	52%
D I	cihA1O20		No				
	cihA1P10	43-498	FRUP00000157667	Hypothetical protein KIAA0233	470-935	1.00E-68	35%
	cihA2G5	1-400	FRUP00000160711	CDNA FLJ20302 fis, clone HEP06648	239-593	1.00E-43	28%
	cihA3N23		No				
	cihA5D7	2-55	FRUP00000149322	HSPC300	15-68	2.00E-13	64%
	cihA5H10		No				
	cihA7M17		No				
	cihA8F16		No				
	cihA8G8	269-435	FRUP00000145559	Hypothetical 66.6 kDa protein	5-181	5.00E-24	37%
	cihA9G14	20-317	FRUP00000163880	DKFZP564F0522 protein	1-308	9.00E-53	38%
	cihA10L19	1-189	FRUP00000165192	None 2810439F02Rik protein	298-495	2.00E-35	36%
	cihA11M1	1-263	FRUP00000150730	NIR3	923-1188	7.00E-83	58%
D II	cihA8L18		No				
	cihA8B16		No				
	cihA9B2		No				
	cihA2F15		No				
	cihA1H22		No				
	cihA9A15		No				
	cihA9K16		No				
	cihA10P13		No				
	cihA12K2		No				
	cihA9B11		No				
	cihA10G7		No				
	cihA11O6		No				
	cihA3P6		No				
	cihA9D22		No				
	cihA4I22		No				
	cihA2D1		No				
	cihA11N1		No				
	cihA9A21		No				
	cihA9D8		No				
	cihA6K13		No				

For *Fugu rubripes*, used was the gene model v3.0 which is located at http://genome.jgi-psf.org/fugu6/fugu6.home.html

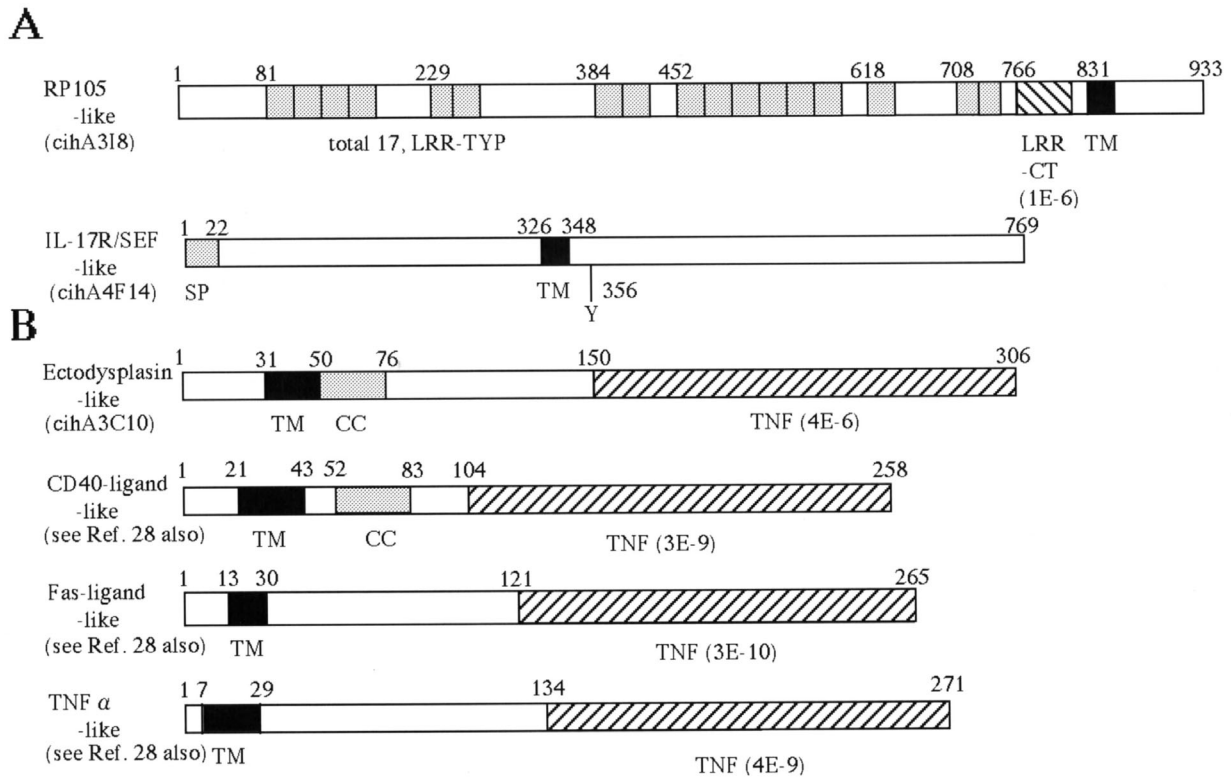

Figure 1. Comparison of cell-cell communication-related proteins. (A) A domain analysis of *Ciona* RP105-like and IL-17R/SEF-like polypeptides. Numbers represent amino acid residues. TM, *trans*-membranous; LRR-TYP, leucine-rich repeat; LRR-CT, leucine-rich repeat carboxy terminal; SP, signal peptide. The probability values shown in parentheses were obtained from the RPS-BLAST search. (B) Comparison of *Ciona* TNF family proteins. The sequences used for the domain analysis of ectodysplasin-like, CD40-ligand-like, Fas-ligand-like, and TNF-alpha-like proteins are derived from cihA3C10 (this study), ciad38f18, ciad58o04, and rciad56b20 (sequence data not shown), respectively. CC, coiled coil; TNF, tumor necrosis factor.

3.4. Genes involved in cell-cell communication

Although the genes directly implicated in host defense were not revealed, three cDNAs-cihA3I8, cihA4F14, and cihA3C10-encode proteins that are involved in cell-cell communication and may reflect characteristics of *Ciona* hemocytes. To analyze them further, we performed a domain analysis for the ORF/CDS predicted for each of these clones using the RPS-BLAST and SMART[19] programs; the results are shown in Fig. 1. The possible biological significance of each of these proteins is discussed below.

3.5. An RP105-like gene

The ORF deduced from clone cihA3I8 predicts a 933-aa polypeptide that shows a significant homology to the *Drosophila* Toll-like receptor (TLR) 7 (probability, 3E-22, Table 2). These two proteins share 27% and 44% overall aa identity and similarity, respectively. Furthermore, a domain analysis revealed the presence of 17 leucine-rich repeats (LRR-TYPs), one LRR-carboxyl-terminal domain (LRR-CT), and a *trans*-membranous region in the cihA3I8-derived protein (Fig. 1A). However, an 80-aa-long presumptive intra-cytoplasmic region appears to lack a Toll/IL-1R (TIR) domain, which is a characteristic feature of TLR family proteins and serves as a binding site for adaptor proteins such as MyD88.[20] This predicted *Ciona* protein is therefore more reminiscent of the mammalian RP105 protein, which is an atypical member of the TLR family.[21] RP105 possesses seven LRR-TYP domains, one LRR-CT domain, and a *trans*-membrane domain, but lacks a TIR domain. It must be noted that the *GARP*-like gene in *Fugu* (Table 4) appears to be a homologue of *Ciona* cihA3I8. The GARP-like protein possesses 15 LRR motifs and a *trans*-membrane region, but lacks a TIR domain again. Detection of a putative TLR/RP105 homologue indicates that *Ciona* hemocytes may play a role in pattern recognition of foreign materials.

3.6. An IL-17R/SEF-like gene

The ORF found in cihA4F14 encodes a polypeptide that is 769 aa residues long. The carboxy-terminal half of the predicted protein shows homology to the cytoplasmic region of mammalian IL-17R (probability, E-11, Table 2, Fig. 1A). Twenty-seven and 40% of the amino

acid residues are identical and similar, respectively, between mammalian IL-17R and cihA4F14. The central portion of the predicted protein is rich in hydrophobic amino acid residues and probably corresponds to a *trans*-membranous region. The extreme amino-terminal 22 residues probably represent a signal peptide sequence. Thus, this polypeptide appears to be a *Ciona* homologue of IL-17R. In mammals, IL-17 is produced from T lymphocytes[22] whereas its receptor is expressed in a variety of cell types, such as fibroblasts and stromal cells, so IL-17 has a broad spectrum of biological effects.[23] In zebrafish, an IL-17R homologue encoded by the *sef* (similar expression to FGF) gene is co-expressed with several types of FGF and can attenuate FGF signaling by interacting with FGF-R1 and FGF-R2.[24] The tyrosine residue located just intra-cellularly of the *trans*-membranous region is conserved between zebrafish SEF and *Ciona* IL-17R/SEF-like and may serve as an adaptor-binding site via its phosphorylation. In contrast, the amino-terminal portion of *Ciona* IL-17R/SEF-like protein, which probably represents an extracellular region of the protein, shows no homology to the corresponding region of mammalian IL-17R or zebrafish SEF. A possible ligand for *Ciona* IL-17R/SEF-like protein cannot currently be predicted. Nevertheless, considering the expression patterns of mammalian IL-17R and zebrafish SEF and the fact that the *Ciona* IL-17R/SEF-like transcript was detected in hemocytes, *Ciona* hemocytes may to some degree be mesenchymal.

3.7. An ectodysplasin-like gene

The cihA3C10 clone harbors an intact ORF that is predicted to encode a 306-aa-long polypeptide. A domain analysis revealed a hydrophobic, *trans*-membranous region at the amino-terminus and a tumor necrosis factor (TNF) domain at the carboxy-terminus (Fig. 1B). TNF domains are characteristic of TNF family proteins including TNF-alpha, CD40-ligand, and Fas-ligand. In mammals, these proteins are produced from monocytes and/or T lymphocytes and are involved in inflammation and apoptosis. However, the protein predicted from cihA3C10 shows highest homology to the *ectodysplasin/tabby* gene product (probability 4E-05, Table 2). Within the TNF domain, 24% and 45% of the amino acid residues are identical and similar, respectively, between mammalian ectodysplasin and *Ciona* ectodysplasin-like proteins. In mammals, mutations in the *ectodysplasin/tabby* gene cause hypoplasticity of epidermal tissues such as hair, teeth, and sweat glands, suggesting that it is involved in epithelium-mesenchyme interactions.[25–27] The significance of *ectodysplasin-like* expression in *Ciona* hemocytes is not presently clear. Perhaps *Ciona* hemocytes retain a more primitive mesenchymal origin than the hematopoietic cells in mammals.

3.8. The TNF gene family

The *Ciona* cDNA/EST database at Kyoto University[12,13] contains transcripts homologous to cihA3C10. Figure 1B illustrates a domain comparison between cihA3C10 and those in the Kyoto database. Interestingly, the sequences of ciad38f18, ciad58o04, and rciad56b20 are more similar to CD40-ligand, Fas-ligand, and TNF-alpha, respectively, based on a BLAST search. This finding, together with the identification of the corresponding gene in the draft *Ciona* genome sequence,[28] demonstrates the existence of an ectodysplasin/TNF-like multigene family in *Ciona* like that found in mammals. Whether *Ciona* hemocytes are involved in the interactions with the epithelium, inflammation, and apoptosis must be elucidated in the future by functional studies.

3.9. Expression profiles of cell-cell communication-related genes

As described above, cihA3I8, cihA4F14, and cihA3C10 were present in a hemocyte-derived collection of ESTs but not in ESTs that were obtained from various stages of *Ciona* development. We therefore decided to examine their expression in several different developmental stages as well as in hemocytes. RNA was prepared from fertilized eggs (E), cleavage-stage embryos (C), gastrulation-stage embryos (G), tailbud embryos (T), tadpole larvae (L), and hemocytes (H). Semi-quantitative RT-PCR analysis was performed as described in Materials and Methods. As seen in Fig. 2, the RP105-like, IL-17R/SEF-like, and ectodysplasin-like transcripts were expressed more abundantly in hemocytes than in other tissues. In addition to these three genes, 76 additional transcripts listed in Table 1 were also selected from hemocyte ESTs by this same procedure. Thus, it is plausible that we may have identified a substantially large number of transcripts that are preferentially expressed in hemocytes.

Finally, we examined the expression profiles of other TNF family genes (Fig. 2). As in the case of ectodysplasin-like protein, the TNF alpha-like and *Fas* ligand-like transcripts were abundantly expressed in hemocytes, suggesting that hemocytes may be involved in inflammation and/or apoptosis. In contrast, the CD40 ligand-like transcript was detected uniformly throughout all developmental stages except for fertilized eggs. This CD40 ligand-like protein may therefore function as a general cell-cell communication molecule.

3.10. Conclusions

By using cDNAs/ESTs derived from hemocytes, we identified and sequenced cDNA clones representing 79 new transcripts. The amino acid sequence of each encoded polypeptide was predicted as well. Three of these transcripts, which encode RP105-like, IL-17R/SEF-like,

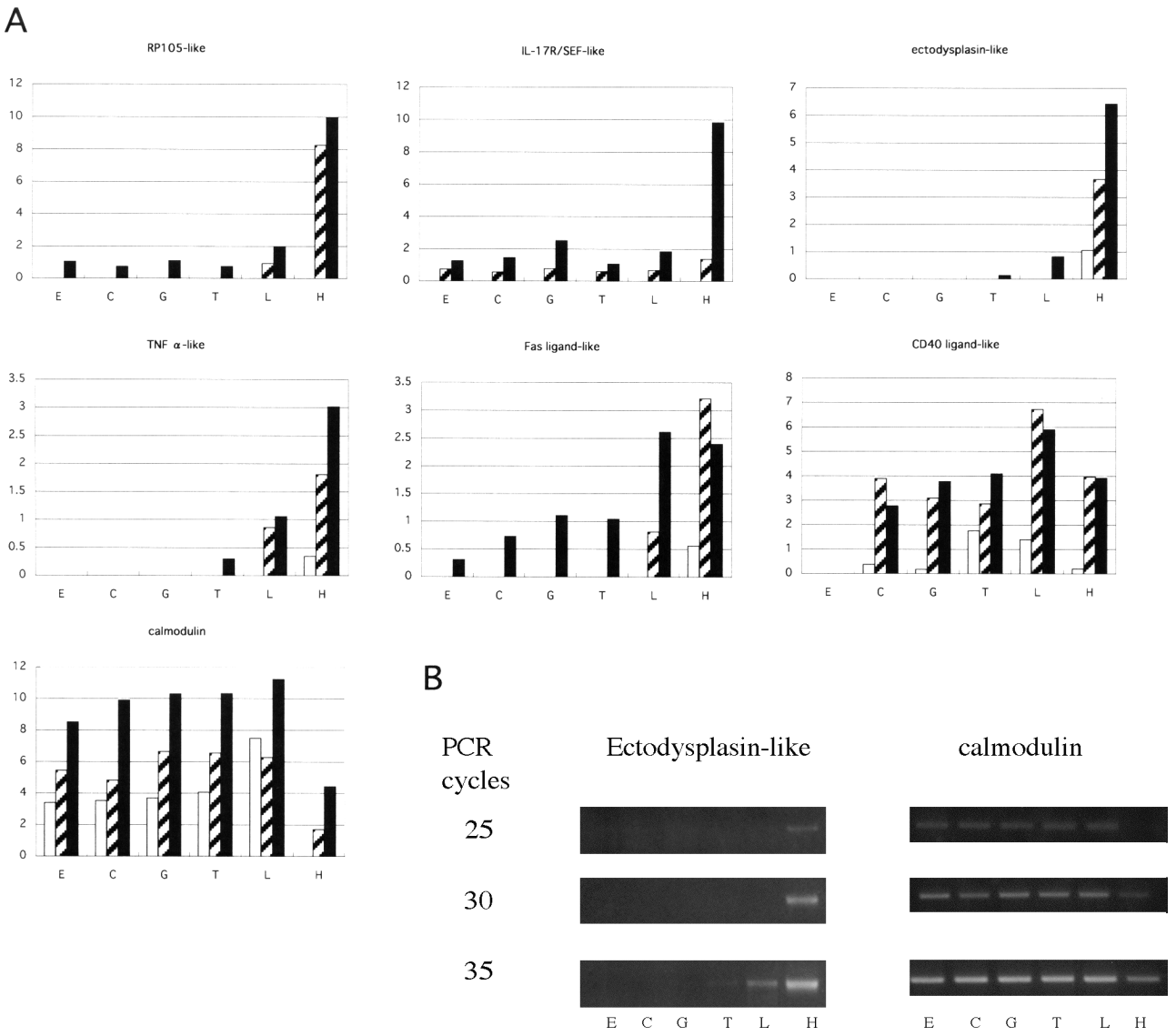

Figure 2. Expression profiles of cell-cell communication-related genes in hemocytes and in various developmental stages of *Ciona*. (A) Semi-quantitative RT-PCR analysis was performed for each gene by changing the number of PCR cycles from 25 (white bar) to 30 (dashed bar), and to 35 (black bar). RNA sources were fertilized eggs (E), cleavage-stage embryos (C), gastrulation-stage embryos (G), tailbud-stage embryos (T), tadpole larvae (L), and hemocytes (H). The relative intensities of the amplified bands were normalized by assuming the internal DNA marker to be 1.0. Note that the scale used for the longitudinal axis varies from panel to panel. The *calmodulin* transcript was used as a reference. (B) Gel electrophoresis profiles of PCR products are shown for the ectodysplasin-like and calmodulin transcripts as representatives of (A).

and ectodysplasin-like (TNF family) proteins, were abundantly expressed in hemocytes and may possibly be involved in cell-cell communication in *Ciona*.

Acknowledgements: We thank K. Azumi and M. Nonaka for helpful discussion and K. Hirayama for providing *Ciona* samples. We also thank M. Kuji for her secretarial assistance. This work was supported by research grants from the Ministry of Education, Culture, Sports, Science and Technology, Japan and from the Mitsubishi Foundation.

References

1. Nakatani, Y., Moody, R., and Smith, W. C. 1999, Mutations affecting tail and notochord development in the ascidian *Ciona savignyi*, *Develop.*, **126**, 3293–3301.
2. Sordino, P., Heisenberg, C-P., Cirino, P. et al. 2000, A mutational approach to the study of development of the protochordate *Ciona intestinalis* (Tunicata, Chordata), *Sarsia*, **85**, 173–176.
3. Satou, Y., Imai, K. S., and Satoh, N. 2001, Action of morpholinos in *Ciona* embryos, *Genesis*, **30**, 103–106.
4. Di Gregorio, A. and Levine, M. 2002, Analyzing gene regulation in ascidian embryos: new tools for new per-

spectives, *Different.*, **70**, 132–139.
5. Dehal, P., Satou, Y., Campbell, R. K. et al. 2002, The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins, *Science*, **298**, 2157–2167.
6. Nishikata, T., Yamada, L., Mochizuki, Y. et al. 2001, Profiles of maternally expressed genes in fertilized eggs of *Ciona intestinalis*, *Dev. Biol.*, **238**, 315–331.
7. Fujiwara, S., Maeda, Y., Shin-I, T. et al. 2002, Gene expression profiles in *Ciona intestinalis* cleavage-stage embryos, *Mech. Dev.*, **112**, 115–127.
8. Satou, Y., Takatori, N., Yamada, L. et al. 2001, Gene expression profiles in *Ciona intestinalis* tailbud embryos, *Develop.*, **128**, 2893–2904.
9. Kusakabe, T., Yoshida, R., Kawakami, I. et al. 2002, Gene expression profiles in tadpole larvae of *Ciona intestinalis*, *Dev. Biol.*, **242**, 188–203.
10. Ogasawara, M., Sasaki, A., Metoki, H. et al. 2002, Gene expression profiles in young adult *Ciona intestinalis*, *Dev. Genes. Evol.*, **212**, 173–185.
11. Inaba, K., Padma, P., Satouh, Y. et al. 2002, EST analysis of gene expression in testis of the ascidian *Ciona intestinalis*, *Mol. Reprod. Dev.*, **62**, 431–445.
12. Satou, Y., Takatori, N., Fujiwara, S. et al. 2002, *Ciona intestinalis* cDNA projects: expressed sequence tag analyses and gene expression profiles during embryogenesis, *Gene*, **287**, 83–96.
13. Satou, Y., Yamada, L., Mochizuki, Y. et al. 2002, A cDNA resource from the basal chordate *Ciona intestinalis*, *Genesis*, **33**, 153–154.
14. Laird, D. J., De Tomaso, A. W., Cooper, M. D. et al. 2000, 50 million years of chordate evolution: seeking the origins of adaptive immunity, *Proc. Natl. Acad. Sci. USA*, **97**, 6924–6926.
15. Azumi, K., De Santis, R., De Tomaso, A. et al. Genomic analysis of immunity in a basal chordate and the emergence of the vertebrate immune system: waiting for Godot, *Immunogenet.*, (in press).
16. Shida, K., Terajima, D., Uchino, R. et al. 2003, Hemocytes of *Ciona intestinalis* express multiple genes involved in innate immune host defense, *Biochem. Biophys. Res. Commun.*, **302**, 207–218.
17. Altschul, S. F., Madden, T. L., Schaffer, A. A. et al. 1997, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucl. Acids Res.*, **25**, 3389–3402.
18. Lee, Y-H., Huang, G. M., Cameron, R. A. et al. 1999, EST analysis of gene expression in early cleavage-stage sea urchin embryos, *Develop.*, **126**, 3857–3867.
19. Schultz, J., Milpetz, F., Bork, P. et al. 1998, SMART, a simple modular architecture research tool: identification of signaling domains, *Proc. Natl. Acad. Sci. USA*, **95**, 5857–5864.
20. Medzhitov, R., Preston-Hurlburt, P., Kopp, E. et al. 1998, MyD88 is an adaptor protein in the hToll/IL-1 receptor family signaling pathways, *Mol. Cell*, **2**, 253–258.
21. Miyake, K., Yamashita, Y., Ogata, M. et al. 1995, RP105, a novel B cell surface molecule implicated in B cell activation, is a member of the leucine-rich repeat protein family, *J. Immunol.*, **154**, 3333–3340.
22. Rouvier, E., Luciani, M-F., Mattei, M-G. et al. 1993, CTLA-8, cloned from an activated T cell, bearing AU-rich messenger RNA instability sequences, and homologous to a herpesvirus saimiri gene, *J. Immunol.*, **150**, 5445–5456.
23. Yao, Z., Spriggs, M. K., Derry, J. M. et al. 1997, Molecular characterization of the human interleukin (IL)-17 receptor, *Cytokine*, **11**, 794–800.
24. Tsang, M., Friesel, R., Kudoh, T. et al. 2002, Identification of Sef, a novel modulator of FGF signaling, *Nat. Cell Biol.*, **4**, 165–169.
25. Ferguson, B. M., Brockdorff, N., Formstone, E. et al. 1997, Cloning of Tabby, the murine homolog of the human EDA gene: evidence for a membrane-associated protein with a short collagenous domain, *Hum. Mol. Genet.*, **9**, 1589–1594.
26. Srivastava, A. K., Pispa, J., Hartung, A. J. et al. 1997, The Tabby phenotype is caused by mutation in a mouse homologue of the EDA gene that reveals novel mouse and human exons and encodes a protein (ectodysplasin-A) with collagenous domains, *Proc. Natl. Acad. Sci. USA*, **94**, 13069–13074.
27. Mikkola, M. L., Pispa, J., Pekkanen, M. et al. 1999, Ectodysplasin, a protein required for epithelial morphogenesis, is a novel TNF homologue and promotes cell-matrix adhesion, *Mech. Dev.*, **88**, 133–146.
28. Terajima, D., Shida, K., Takada, N. et al. 2003, Identification of candidate genes encoding the core components of the cell death machinery in the *Ciona intestinalis* genome, *Cell Death Different.*, **10**, 749–753.