Available online at www.sciencedirect.com
SCIENCE DIRECT
Developmental and Comparative Immunology xx (xxxx) 1–15
Developmental & Comparative Immunology
www.elsevier.com/locate/devcompimm

Towards a molecular code for individuality in the absence of MHC: screening for individually variable genes in the urochordate *Ciona intestinalis*

Konstantin Khalturin*, Ulrich Kürn, Nicole Pinnow, Thomas C.G. Bosch

Zoological Institute, Christian-Albrechts-University Kiel, Olshausenstrasse 40, 24098 Kiel Germany

Received 17 December 2004; revised 27 January 2005; accepted 28 January 2005

Abstract

Urochordates possess several well described allorecognitions systems, the molecular nature of those is not yet understood. A prerequisite for any self-/nonself discimination system is the presence of a group of highly variable molecules, which should vary between individuals. Using suppression subtractive hybridisation (SSH) we surveyed *Ciona intestinalis* for individually variable genes. Our search so far identified two genes, ciS7 and ciMETA2, which both display an unexpected high degree of intra- and interindividual variability, code for secreted proteins, and contain multiple domains suitable for protein–protein interactions. The possible role of these molecules in allorecognition is discussed.
© 2005 Elsevier Ltd. All rights reserved.

Keywords: Urochordata; Innate and adaptive immunity; Somatic variation; MHC; Allorecognition; Phylogeny

1. Introduction

The ability to distinguish self from nonself is one of the basic features at all levels of eumetazoan evolution from cnidarians [1,2] to vertebrates. There are numerous examples of colonial animals, where parts of one colony successfully fuse naturally while pieces of genetically different colonies are rejected. In several animal groups including hydrozoan *Hydractynia echinata* [3] and colonial and solitary tunicates [4–6], it has been shown that this ability is genetically controlled. In addition to allorecognition, many of the invertebrate animals being hermaphrodites possess a genetically controlled system to avoid self fertilization [7]. Despite of the fact that the phenomenon of self/nonself discrimination in Eumetazoa has been noticed more than one hundred years ago [8], knowledge of its molecular basis is limited. The only group of animals where the mechanism of histocompatibility has been elucidated are the vertebrates with their major histocompatibility complex (MHC) based allorecognition system. Outside jawed vertebrates, however, nothing is known about markers for 'self' and 'nonself' nor how the histocompatibility system functions [9].

* Corresponding author. Tel: +49 431 880 4176; Tel: +49 431 880 4747.
E-mail address: kkhalturin@zoologie.uni-kiel.de (K. Khalturin).

0145-305X/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.dci.2005.01.006

UNCORRECTED PROOF

The common ancestors of urochordates and vertebrates have split in evolution around 570 million years ago. On the basis of their monophyletic origin it has been suggested [4,10,11] that their histocompatibility machinery might be similar to that of the higher vertebrates and may represent the ancestral self-/nonself discrimination system. Therefore, one of the urochordates, *Botryllus schlosseri*, became a key model system for deciphering the molecular basis of invertebrate allorecognition and the search for the molecular precursors of MHC [12]. Recently, a putative Fu/HC locus of *Botryllus* has been delineated using the AFLP approach [13,14]. Although 400 kb of the putative FU/HC locus have now been sequenced, the molecular basis of allorecognition in *Botryllus* remains unknown. Moreover, an increasing amount of molecular data including the sequence analysis of *Ciona intestinalis* [15] and *Ciona savignyi* genomes, makes it evident that there are no homologs of immunoglobulins (Igs), T cell antigen receptors (TCRs), major histocompatibility complex (MHC) I and II, and recombination activating genes (RAGs) in jawless vertebrates, urochordates, or any invertebrates [16,17]. Thus, MHC-based allorecognition might be a unique feature of jawed vertebrates. Jawless vertebrates and all invertebrates may use for self-/nonself discimination a not yet understood machinery which is different from that of vertebrates [9].

Direct evidence that there may be major differences in the immune systems between closely related animal groups comes from the recent finding of variable lymphocyte receptors in agnathes [18]. There, a novel system based on rearranging receptors containing leucin rich repeats (LRRs) was identified by using an unbiased functional screening approach. This system differs totally from the mechanism present in vertebrates, which is based on somatic rearrangement of genes coding for proteins with Ig domains. Thus, even in closely related phylogenetic groups, the organisation of immune systems can vary considerably. Different groups of molecules may be responsible for similar functions. This raises the important question whether different groups of invertebrates use a similar molecular machinery for allorecognition or whether each group has developed its own histocompatibility system utilizing taxon specific self-/nonself determinants.

In urochordates, histocompatibility systems have been identified not only in colonial species such as *Botryllus schlosseri* (references in [12]) but also in a number of solitary ascidians. For example, in *Styela plicata* a phenomenon similar to transplantation immunity in vertebrates has been described. Allogenic tunic transplantation in *Styela* causes rejection while isogenic transplants are accepted [19,20]. Several other solitary species such as *Halocynthia rorezi* and *C. intestinalis* have a well documented but not understood mechanism to block self fertilization [7,21–23].

A prerequisite for any histocompatibility system is the presence of highly variable molecules which should vary between individuals. Well known examples include the S-locus proteins in *Brassicaceae* plants which take part in the prevention of self pollination [24,25], the seven-transmembrane receptors and their pheromone ligands in the mating locus of fungi [26], and the MHC antigens in vertebrates.

To search for proteins which might be molecular markers of individuality in *Ciona. intestinalis*, we used a novel and unbiased approach based on suppression subtractive hybridisation (SSH). We compared the somatic transcriptomes of two *Ciona* individuals with the goal to identify cDNAs which are variable between these individuals. As a result we identified two classes of soluble proteins which exhibit a high degree of inter- and intra-individual variability.

2. Materials and methods

2.1. Animals

Adult *C. intestinalis* individuals were obtained from Helgoland Marine Biological Station (Germany) and kept in aquarium with natural sea water at + 12 °C. Animals were starved for two days before mRNA extraction.

2.2. Subtractive hybridisation

Two randomly selected mature *C. intestinalis* individuals (animals ‘12’ and ‘13’) were used for subtractive hybridisation. Tunic was removed and discarded prior to mRNA isolation to avoid possible

UNCORRECTED PROOF

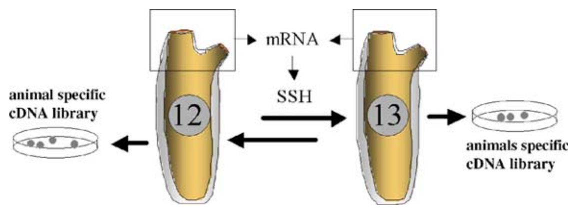

Fig. 1. Outline of the experimental approach used to identify individually variable genes in *Ciona intestinalis*. Suppression subtractive hybridisation (SSH) was performed in both directions, using cDNA pools from two individuals (animals '12' and '13') either as tester or driver. Two cDNA libraries enriched for individual specific transcripts were obtained and 72 clones from each library were sequenced.

contaminations such as epiphytes or bacteria. The tissues collected for subtraction included both siphons, small area of a branchial basket around them, remains of a connective tissue underlying tunic and hemocytes present in this region of the body (see Fig. 1). mRNA was extracted using QuickPrep micro mRNA Purification Kit (Amersham). From each individual (animals '12' and '13', see Fig. 1) 2 μg of mRNA were isolated. The subtractive hybridisation was performed using Clontech PCR-Select cDNA Subtraction Kit (BD Biosciences) according to the user manual. The only difference from the manual was that the hybridisation temperature was increased to 69 °C. Subtractions were performed in two directions using the double stranded cDNAs from individuals '12' and '13'. After the second PCR, resulting subtracted cDNAs for animals '12' and '13' were cloned into pGEM-T vector (Promega), and DH5α *E. coli* were transformed with these libraries.

2.3. Molecular techniques

Nucleic acid isolation, cDNA synthesis, cloning, and sequencing were carried out following standard procedures [27]. Rapid amplification of cDNA ends (RACE) was performed using 'splinkerette' technology described previously [28,29]. Pwo DNA polymerase (Roche) with proofreading activity was used for RT-PCRs to avoid problems with *Taq*-derived nucleotid variations. Fragments of ciS7 were amplified by RT-PCR with primers ciS7_F (CTTAAGCA-TACAGCAGAGGAAAT), ciS7_R1 (AAGTTGACA CCATTACCACGGAAG), ciS7_R2 (ATCGACAG TTGCGTCTTTGTAGAC). ciMETA2 fragments were amplified with primers META2_F (CAGATTG-CATTAACTTTACCTATTC) and META2_R (GAG TTTATGGGGTTAAAGGAAGAC). ciFLRT fragments were amplified with primers FLRT_F (GGTT AATATCTGTCAGGCTTCTG) and FLRT_R (GAA CTTCTTGGTCGATATTTCTC). ciGAPDH was amplified with primer set ciGAPDH_F (CATGA-CAACTGTGCACTCGTACAC) and ciGAPDH_R (GTTCCTGTATCCAAATTCATTGTC).

For Southern blot analysis, 10–15 μg of genomic DNA was digested with *Hind*III, separated on a 0.7% agarose gel and transferred onto a Hybond N+ nylon membrane (Amersham). *Hind*III has no 'star-activity', is not dependent on DNA methylation and in addition restricts *Ciona* DNA frequent enough to be used for restriction fragment length polymorphism (RFLP) analysis. Southern hybridization was performed in the hybridization solution containing 6X SSC, 0.5% SDS, 5X Denhardt solution at 65 °C overnight. High stringency Southern blot hybridisation was performed in the hybridisation solution containing 6X SSC, 50% formamide, 0.5% SDS, 5X Denhardt solution at 45 °C overnight. Washing of Southern blots was performed in 0.2X SSC/0.1% SDS at 60 °C. The positions of the probes are indicated in Fig. 2A.

cDNA clones from subtractive libraries was sequenced on MegaBace 1000 capillary sequencer using DYEnamic ET terminator cycle sequencing kit (Amersham). Sequencing of RT-PCR fragments from different *Ciona* individuals was done using Li-COR 4200 DNA sequencer and e-Seq V2.0 software. All the sequences obtained with Li-COR 4200 were manually verified. All the sequences shown in the parer are submitted to GeneBank under accession numbers AY895020-AY89059, AY899286-AY899290.

2.4. Sequence analysis

Ciona EST database in Kyoto at http://ghost.zool.kyoto-u.ac.jp/indexr1.html and *Ciona* draft genome database at http://genome.jgi-psf.org/ciona4/ciona4.home.html were used. All cDNA fragments obtained by SSH were submitted initially to the BLASTN search against the draft *Ciona* genome sequences. As a result, we could identify assembled genomic regions (referred to as 'scaffolds') where the genes

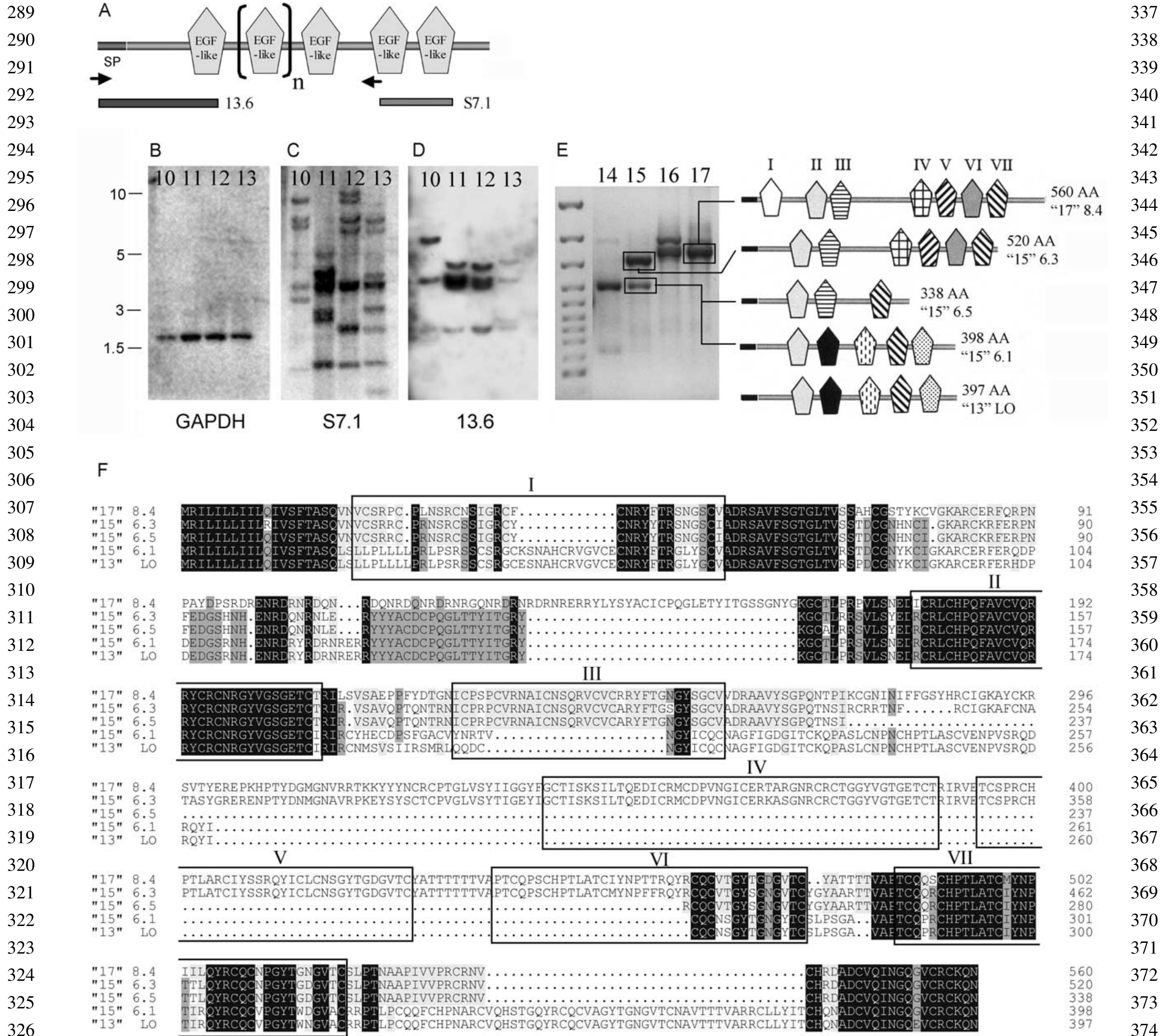

Fig. 2. Inter- and intraindividual variability of ciS7, a gene coding for secreted protein with multiple EGF-like domains. (A) Domain organisation of ciS7 protein, containing variable number of EGF-like motifs. Arrows—the positions of RT-PCR primers; 13.6 and S7.1—probes for 5′ and 3′ region, respectively, used in Southern blot hybridisation; SP—signal peptide. (B) Southern blot hybridisations with *Hind*III restricted genomic DNA from four *Ciona* individuals (animals '10', '11', '12', '13') and GAPDH fragment as a probe. (C) The same blot hybridised with the 3′ region probe S7.1 (AY895058). (D) The same blot hybridised with genomic 5′ region probe 13.6 (AY895048). (E) RT-PCR with cDNAs from four *Ciona* individuals (animals '14', '15', '16', '17') showing ciS7 transcripts which vary in number and length from individual to individual. The stick models show predicted domain organisation of ciS7 proteins, expressed by different individuals. I–VII—EGF-like domains as predicted by SMART. Length of each ORF as well as clone number are shown at the right side of each protein model. Domains which are corresponding to each other are schematically depicted in a similar pattern. (F) Amino acid sequence alignment of RT-PCR fragments from three *Ciona* individuals (animals '17', '15' and '13'). Individual clones are indicated by the animal number in quotation marks, followed by the clone name, e.g. '17' 8.4. The EGF-like domains are boxed according to SMART prediction for the clone '17' 8.4. Amino acid sequences identical in all clones are shown on black background. Genebank accession numbers AY895020-AY895024.

corresponding for the cDNA fragments were located. Thus, the nature of genes to which the subtracted cDNAs belong could be identified.

In addition, nucleotide, protein and translated BLAST search engines at the NCBI server (http://www.ncbi.nlm.gov/BLAST) were used for homology searches in public databases. Simple Modular Architecture Research Tool (SMART, http://smart.embl-heidelberg.de) was used for the annotation of protein domains and the analysis of domain architectures. Protein and DNA sequences were aligned with DNAMAN Version 4.12 software.

3. Results

3.1. Unbiased screening for genes with high degree of interindividual variation

To search for proteins which vary between *Ciona* individuals and, therefore, potentially may be involved in histocompatibility reactions, we used suppression subtractive hybridization (SSH). If there are proteins with amino acid sequences differing considerably between two individuals (similarity less then 75–80%), the corresponding mRNA sequences should be selectively amplified by the SSH procedure. As shown schematically in Fig. 1, subtraction was performed in both directions using the cDNA pools from two *C. intestinalis* individuals (animals '12' and '13' in Fig. 1). As a result of subtractive hybridization, two cDNA libraries were generated and 72 clones of each library were sequenced. Theoretically the cDNA clones from each library should contain transcripts unique for each individual or highly variable between individuals. In reality, however, around 50% of the transcripts in both cDNA libraries contained cDNAs for various ribosomal proteins and endostyle specific transcripts, which had identical sequences in both individuals. Thus, these clones were discarded as false positives.

The library specific for individual '12' contained 11 sequences which in the *Ciona* genome database aligned with different regions of scaffold 61 (see Materials and methods). According to the BLASTX search these cDNAs coded for parts of the ciMETA2 protein [30]. The remaining sequences were singletons or represented by only two cDNA clones with identical sequence. These cDNA clones were not examined further. The library specific for individual '13' contained 16 cDNA clones which gave hits to different areas of scaffolds 549, 757 and 1570 and coded for a protein containing EGF-like motifs. These sequences, however, were not identical to the sequences in the *Ciona* genome database but showed a high degree of nucleotide substitutions. The remaining sequences from individual '13' were again represented by singletons or by only two cDNA clones with identical sequence and therefore were not analysed further.

To get insight into the genomic complexity of the genes described above, Southern blot hybridization was performed with ciMETA2 encoding cDNA clones from individual '12' (AY895054) and EGF-like protein encoding cDNA clones from individual '13' (AY895058). As shown in Figs. 2C and 5B, both genomic regions showed a high degree of restriction fragment length polymorphism (RFLP). The frequent representation of these cDNAs in the subtractive libraries and the complex RFLP patterns lead us to investigate the degree of interindividual variation of the corresponding two genes in more detail.

3.2. Inter- and intraindividual variability of ciS7, a gene containing EGF-like domains

The initial 16 cDNA fragments obtained from individual '13' encoded ORFs with low similarity to EGF-like domains (E> =2e–3) displaying the diagnostic pattern of six conserved cystein residues. The corresponding regions in the *Ciona* genome database (scaffolds 549, 757 and 1570) did not contain any full length ESTs. Therefore, to get more sequence information, 5′ and 3′ rapid amplification of cDNA ends (RACE) with cDNAs from individuals '12' and '13' were performed. As a result, we obtained one sequence from individual 12 and two sequences from individual '13' ('12' S7.1, GeneBank AY899289; '13' HI, GeneBank AY899290; '13' LO, GeneBank AY895024). These cDNAs code for a protein, termed ciS7, with a conserved signal peptide and a variable number of EGF-like domains. Surprisingly, in individual '13' two versions of this protein were identified which differ in length and amino acid sequence ('13' HI and '13' LO in Fig. 3A and B). Thus, the predicted amino acid sequence for ciS7 appears not only to vary

UNCORRECTED PROOF

A

```
"12" S7.1   MRILILLIILQIVSFTASQLSPLFLLLLP.....SSCSRECESNSHCNRIGVCECNRYFTRSYGVCVADRSAVFSGTGLAVRSTDCGYYNCI.GKARCERF  95
"13" LO     MRILILLIILQIVSFTASQLSLLPLLLLPRLPSRSSCSRGCESNAHCRV.GVCECNRYFTRGLYGCVADRSAVFSGTGLTVRSPDCGNYKCI.GKARCERF  99
"13" HI     MRILILLIILQIVSFTASQVNVCSRPC.PLNSRGNSIGRCF............CNRYFTRSNGSCVADRSAVFSGTGLTVSSVHCGSTYKCVGKARCERF  87
"14" T1.1   MRILILLIILQIVSFTASQLSLLPLLLLPRLPSRSSCSRGCESNAHCRV.GVCECNRYFTRGLYGCVADRSAVFSGTGLTVRSPDCGNYKCI.GKARCERF  99
"14" T1.2   MRILILLIILQIVSFTASQVNVCSRPC.PLNSRGNSIGRCF............CNRYFTRSNGSCVADRSAVFSGTGLTVSSAHCGSTYKCVGKARCERF  87
"15" 6.3    MRILILLIILRIVSFTASQVNVCSRRC.PRNSRGSSIGRCY............CNRYYTRSNGSCIADRSAVFSGTGLTVSSTDCGNHNCI.GKARCKRF  86
"15" 6.1    MRILILLIILQIVSFTASQLSLLPLLLLPRLPSRSSCSRGCKSNAHCRV.GVCECNRYFTRGLYSCVADRSAVFSGTGLTVRSTDCGNYKCI.GKARCERF  99
"15" 6.5    MRILILLIILQIVSFTASQVNVCSRRC.PRNSRGSSIGRCY............CNRYYTRSNGSCIADRSAVFSGTGLTVSSTDCGNHNCI.GKARCKRF  86
"16" 7.5    MRILILLIILQIASFTASQVNVCSRPC.PSNSRGNSIGRCF............CNRYFTRSNGSCVADRSAVFSGTGLTVSSAHCGSTYKCVGKARCERF  87
"16" T3.9   MRILILLIILQIVSFTASQVNVCSRPC.PSNSRGNSIGRCF............CNRYFTRSNGSCVADRSAVFSGTGLTVSSAHCGSTYKCVGKARCERF  87
"16" T3.10  MRILILLIILQIVSFTASQVNVCSRPC.PLNSRGNSIGRCF............CNRYFTRSNGSCVADRSAVFSGTGLTVSSAHCGSTYKCVGKARCERF  87
"17" T4.1   MRILILLIILQIVSFTASQVNVCSRRC.PRNSRGSSIGRCY............CNRYYTRSNGSCIADRSAVFSGTGLAVSSTDCGNHNCI.GKARCKRF  86
"17" T4.3   MRILILLIILQIVSFTASQVNVCSGRC.PRNSRGSSIGRCY............CNRYYTRSNGSCIADRSAVFSGTGLAVSSTDCGNHNCI.GKARCKRF  86
"17" T4.9   MRILILLIILQIVSFTASQVNVCSRRC.PRNSRGSSIGRCY............CNRYYTRSNGSCIADRSAVFSGTGLAVSSTDCGNHNCI.GKARCKRF  86
"17" T4.11  MRILILLIILQIVSFTASQVNVCSRPC.PLNSRGNSIGRCF............CNRYFTRSNGSCVADRSAVFSGTGLTVSSAHCGSTYKCVGKARCERF  87
"17" 8.4    MRILILLIILQIVSFTASQVNVCSRPC.PLNSRGNSIGRCF............CNRYFTRSNGSCVADRSAVFSGTGLTVSSAHCGSTYKCVGKARCERF  87
"17" 8.6    MRILILLIILQIVSFTASQVNVCSRPC.PLNSRGNSIGRCF............CNRYFTRSNGSCVADRSAVFSGTGLTVSSAHCGSTYKCVGKARCERF  87
"17" 8.10   MRILILLIILQIVSFTASQVNVCSRPC.PLNSRGNSIGRCF............CNRYFTRSNGSCVADRSAVFSGTGLTVSSAHCGSTYKCVGKARCERF  87

"12" S7.1   ERRDPDEDVSRNHE........NRER.......................TYY...YDCNCPEGLTTYITGRY....KGCTLPRSVLSNEDRCRLCHPQFAV  158
"13" LO     ERHDPDEDGSRNHENRDRYRDRNRER.......................RYY...YACDCPQGLTTYITGRY....KGCTLPRSVLSNEDRCRLCHPQFAV  170
"13" HI     QRPNPAYDPSRDRENRDRNRDQNRDQNRGQNRDQNQDQ....NRDRNRERRYLYSYACICPQGLETYITGSSGNYGKGCTLPRPVLSNEDRCRLCHPQFAV  184
"14" T1.1   ERHDPDGDGSRNHENRDRYRDRNRER.......................RYY...YACDCPQGLTTYITGRY....KGCTLPRSVLSNEDRCRLCHPQFAV  170
"14" T1.2   QRPNPAYDPSRDRENRDRNRDQNRDQNRDR...........NRDRNRERRYLYSYACTCPHGLETYITGSSGNYGKGCTLPRPVLSNEDRCRLCHPQFAV  176
"15" 6.3    ERPNFEDGSHNHE.........NRDQN..................RNLERYY...YACDCPQGLTTYITGRY....KGCTLRRSVLSYEDRCRLCHPQFAV  153
"15" 6.1    ERQDPDEDGSRNHENRDRYRDRNRER.......................RYY...YACDCPQGLTTYITGRY....KGCTLPRSVLSNEDRCRLCHPQFAV  170
"15" 6.5    ERPNFEDGSHNHE.........NRDQN..................RNLERYY...YACDCPQGLTTYITGRY....KGCALRRSVLSYEDRCRLCHPQFAV  153
"16" 7.5    QRPNPAYDPSRDREDRDRNRDQNRDQNRGQNRDQNRDQNRDRNRDRNRERRYLYSYACVCPQGLETYITGSSGNYGKGCTLPRPVLSNEDRCRLCHPQFAV  188
"16" T3.9   QRPNPAYDPSRDREDRDRNRDQNRDQNRGQNRDQNRDQNRDRNRDRNRERRYLYSYACVCPQGLETYITGSSGNYGKGCTLPRPVLSNEDRCRLCHPQFAV  188
"16" T3.10  QRPNPAYDPSRDRENRDRNRDQNRDQNRDQNRDRNRGQNRDRNRDRNRERRYLYSYACICPQGLETYITGSSGNYGKGCTLPRPVLSNEDICRLCHPQFAV  188
"17" T4.1   ERPNFEDGSHNHE.........NRDQN..................RNLERYY...YACDCPQGLTTYITGRY....KGCTLRRSVLSYEDRCRLCHPQFAV  153
"17" T4.3   ERPNLEDGSHNHE.........NRDQN..................RNLERYY...YACDCPQGLTTYITGRY....KGCTLRRSVLSYEDRCRLCHPQFAV  153
"17" T4.9   ERPNFEDGSHNHE.........NRDQN..................RNLERYY...YACDCPQGLTTYITGRY....KGCTLRRSVLSYEDRCRLCHPQFAV  153
"17" T4.11  QRPNPAYDPSRDRENRDRNRDQNRDQNRDQNRDRNRGQNRDRNRDRNRERRYLYSYACICPQGLETYITGSSGNYGKGCTLPRPVLSNEDICRLCHPQFAV  188
"17" 8.4    QRPNPAYDPSRDRENRDRNRDQNRDQNRDQNRDRNRGQNRDRNRDRNRERRYLYSYACICPQGLETYITGSSGNYGKGCTLPRPVLSNEDICRLCHPQFAV  188
"17" 8.6    QRPNPAYDPSRDRENRDRNRDQNRDQNRDQ............NRDRNRERRYLYSYACICPQGLETYITGSSGNYGKGCTLPRPVLSNEDICRLCHPQFAV  176
"17" 8.10   QRPNPAYDPSRDRENRDRNRDQNRDQNRDQNRDRNRGQNRDRNRDRNRERRYLYSYACICPQGLETYITGSSGNYGKGCTLPRPVLSNEDICRLCHPQFAV  188
```

B

Fig. 3. (A) Multiple amino acid alignment of 18 ciS7 clones, amplified from 6 *Ciona* individuals (GeneBank accessions AY895025-AY895042). Individual clones are indicated by the animal number in quotation marks, followed by the clone name, e.g. '12' S7.1. Amino acids identical in all the sequences are shown on black background. (B) Phylogenetic tree showing the divergence of the ciS7 proteins within and between *Ciona* individuals. Proteins from 6 individuals group into three clear loci-ciS7A, ciS7B and ciS7C. *—two clones with identical amino acid sequence isolated from different individuals; triangle—proteins from individuals '15' and '13' showing high similarity in the N-terminal region and divergent further in the C-terminus direction (see Fig. 2F). Scale bar indicates an evolutionary distance of 0.05 amino acid substitution per position in the sequence. The numbers at the nodes are an indication of the level of confidence, given in percentage, for the branches as determined by bootstrap analysis.

between the two individuals but also shows some intraindividual variability. RACE from cDNAs of additional randomly selected individuals and comparisons to the genome and EST data bases indicate that ciS7 is represented by transcripts of variable length ranging from 1.5 to 2.5 kb. As shown in Fig. 2A, ciS7 codes for a protein which consists of a putative signal peptide and a variable number of EGF-like domains.

To examine the degree of variability of ciS7 transcripts between different *Ciona* individuals, we used RT-PCR and primers corresponding to the 5′ region of the predicted protein and a conserved region near the 3′ end (see Fig. 2A). As shown in Fig. 2E, different individuals express multiple versions of ciS7 which vary in number and in length from individual to individual. For example, RT-PCR with cDNA from individual '14' resulted in one single PCR product of 1.2 kb while RT-PCR with cDNA from individuals '15', '16' and '17' resulted in several fragments in the size range of 1.2–1.9 kb (Fig. 2E). Sequence analysis of the obtained PCR fragments indicated individual-specific differences in the amino acid sequences leading to a different number and a different positioning of the EGF-like domains in the corresponding proteins as predicted by SMART analysis (Fig. 2E and F).

Amino acid sequence alignment of five proteins from individuals '15', '17' and '13' shows a high degree of amino acid substitutions as well as domain deletions and insertions along the predicted protein sequence (Fig. 2E). The signal peptide sequence as well as several EGF-like domains are conserved (domains II and VII, Fig. 2E). All five ciS7 proteins, therefore, appear to have their own unique amino acid sequence (Fig. 2F).

To analyse the variability of this gene in greater detail, the 5′ region of ciS7 was amplified by RT-PCR from six randomly sampled *Ciona* individuals with primers ciS7_F and ciS7_R2. The RT-PCR products were cloned and sequenced. Most surprisingly, as shown in Fig. 3, 13 different protein sequences were obtained from the six individuals. Most of the individuals express several different versions of ciS7, indicating intra-individual variability. In addition, the amino acid sequences vary between individuals. Two sequences (marked by asterisks in Fig. 3B) were found to be identical between two different individuals. Most of the sequence differences are due to single amino acid substitutions, deletions, or insertions of small amino acid stretches (Fig. 3A). The phylogenetic analysis (Fig. 3B) suggests the presence of three distinct ciS7 loci. Interestingly, two of them (ciS7A and ciS7B) correspond to the loci present in scaffolds 549 and 757 (see Fig. 7B) of the draft *C. intestinalis* genome. The amplified regions shown in Fig. 3 represent only about 25% of the full length sequence of the protein. Thus, the degree of variation might be much higher when considering the full-length sequence of the protein. Evidence for that view can be found in Fig. 2F. There, clones '15' 6.1 and 13 LO differ in only 2 amino acids in their N-terminal region, but display variation further towards the C-terminus.

The observed variation may be due to at least three different mechanisms: (i) differential splicing of mRNA precursors, (ii) somatic recombination, or (iii) presence of a gene family with differences in the genomic sequences between individuals. To get first insight into the mechanism and origin of the observed variation, we compared ciS7 genomic and cDNA sequences from two individuals (Fig. 4). The 5′ region of ciS7 was PCR amplified from genomic DNA using the same primer pair as used for the amplification of the cDNA fragments shown in Fig. 3. Two different genomic fragments of 869 and 937 bp were amplified from individual '12', while three fragments of 869, 950 and 1030 bp were amplified from individual '13' (data not shown). Fig. 4 shows the alignment of two genomic fragments from individuals '12', two genomic fragments from individuals '13', one cDNA fragment from individual '13' along with two corresponding genomic fragments from the *Ciona* genome database. The alignment shows that the gene in this region consists of 4 exons, separated by 3 introns (Fig. 4). The obtained genomic fragments can be aligned with two genomic regions from the *Ciona* genome database located in scaffold 1570 (ciS7A_NCBI and ciS7B_NCBI in Fig. 4). The sequences taken from the *Ciona* genome database are not identical to the sequences from individuals '12' and '13' but display many single nucleotide polymorphisms (SNPs), as well as numerous insertions and deletions within exons and introns (Fig. 4). Insertions and deletions which are located within the exons cause inter-individual variations in the amino

UNCORRECTED PROOF

Fig. 4. Alignment of four genomic DNA sequences from individuals '12' and '13', cDNA sequence from animal '13' and two genomic DNA sequences of ciS7 gene from *Ciona* genome database. ciS7 12.7, ciS7 12.6—genomic DNA sequences from individual '12'; ciS7 13.3, ciS7 13.6—genomic DNA sequences from individual '13'; ciS7A_NCBI, ciS7B_NCBI—genomic DNA sequences from *Ciona* database; cDNA'13' HI—cDNA fragment from individual '13'. Four exon regions are underlined and marked E1–E4 correspondingly. Genebank accession numbers AY895043-AY895048, AY895027.

acid sequences of the resulting proteins. We note that in all six areas where deletions can be observed within exons, the number of nucleotides is divisible by three. This type of variation, therefore, does not cause any frameshift, but results in production of proteins of different length and amino acid sequence. Furthermore, the data shown in Fig. 4 indicate that two different individuals may have the identical genomic

sequences for this genomic region. For example, sequences '12.6' and '13.3' obtained from individuals '12' and '13' are identical except for one SNP in intron three. cDNA clone '13 HI' from individual '13' was found to be identical to one of the genomic DNA fragments from the same individual (Fig. 4). Thus, the observed variation appears not to be due to RNA editing or alternative splicing, but is caused by variations in the ciS7 genomic sequences between different individuals.

To examine whether the observed intraindividual variability could be caused by somatic recombination events taking place during maturation of *Ciona* blood cells, genomic DNA from somatic tissue and blood was used for PCR amplification. No differences were found in the size and number of amplified genomic fragments obtained from blood DNA and tissue DNA (data not shown). Thus, intraindividual variability in ciS7 appears not to be caused by somatic rearrangement.

Is, therefore, the observed intra- and interindividual variability due to the presence of several different ciS7 genes in each *Ciona* individual producing different proteins? To address this question Southern blot hybridisation was performed using the genomic fragment ciS7 13.6 (AY895054) representing the 5′ region of the gene as probe (see Fig. 4). This probe contains no internal restriction sites for *Hind*III and, therefore, the number of bands should represent the minimal number of alleles of the gene in the genome of the corresponding individual. Fig. 2D shows the result of this Southern hybridisation experiment and indicates that there are three bands in animal '10' and four bands in animals '11', '12' and '13'. All animals have a common bands of about 2 kb in size and the strong hybridisation signals at about 4 kb in animals '11' and '12' represent two closely migrating DNA fragments. Thus, ciS7 is represented by a small gene family and the number of genes varies from individual to individual. Moreover, additional Southern hybridisation experiments with a probe specific for another region of this gene (S7.1 shown in Fig. 2A, AY895058) provide more evidence for a high degree of genetic variability in the ciS7 gene locus. While individuals '11' and '12' have similar RFLP patterns with probe 13.6 covering the 5′ region of ciS7 (Fig. 2D), very different RFLP patterns can be observed when using probe S7.1 which is specific for the region located towards the 3′ end of the gene (Fig. 2C). One can speculate that these genomic differences may be due to intensive reshuffling events of the *Ciona* ciS7 locus.

Taken together, each *Ciona* individual appears to have at least two tightly linked genomic copies of the ciS7 gene which vary in their sequences from individual to individual. Variations in the ciS7 cDNAs and corresponding amino acid sequences within and between different *Ciona* individuals may be based on sequence variations within this locus as well as on the number of ciS7 genes present in each individual. The mechanism causing this genomic variability remains to be elucidated.

3.3. Inter- and intraindividual variability of ciMETA2

Eleven cDNA fragments specific for individual '12' aligned to different regions of scaffold 61 of the draft *Ciona* genome. Their sequences are similar to different parts of the 3′ region of the ciMETA2 gene which contains thrombospondin type 1 (TSP1) domains. ciMETA2 was initially identified by Nakajama et al. [30] as a gene upregulated during metamorphosis in *C. intestinalis*. The published ORF is 813 amino acids long and codes for a secreted protein with a signal peptide, a GGI repetitive region, and three thrombospondin domains close to the C-terminus.

To examine the degree of variability of the ciMETA2 gene locus between different *Ciona* individuals, we performed Southern blot hybridisation experiments. The cDNA fragment from individual '12' obtained in SSH was used as probe ('12' S5.16, AY895059). With low stringent hybridisation conditions (see Materials and methods), multiple bands causing a distinct hybridisation pattern can be detected for each individual (Fig. 5B). Surprisingly, when using the same probe and high stringency hybridisations conditions, no signal can be detected in individual '13' (see Fig. 5C). This may indicate that the sequence which is represented by the probe is specific for individual '12' and is different from that of the individual '13'. Indeed, Fig. 5E shows that ciMETA2 sequences are different in individuals '12' and '13'. Additional evidence that ciMETA2 is a complex gene locus comes from analysis of available genome data. ciMETA2 is presented in the *Ciona*

UNCORRECTED PROOF

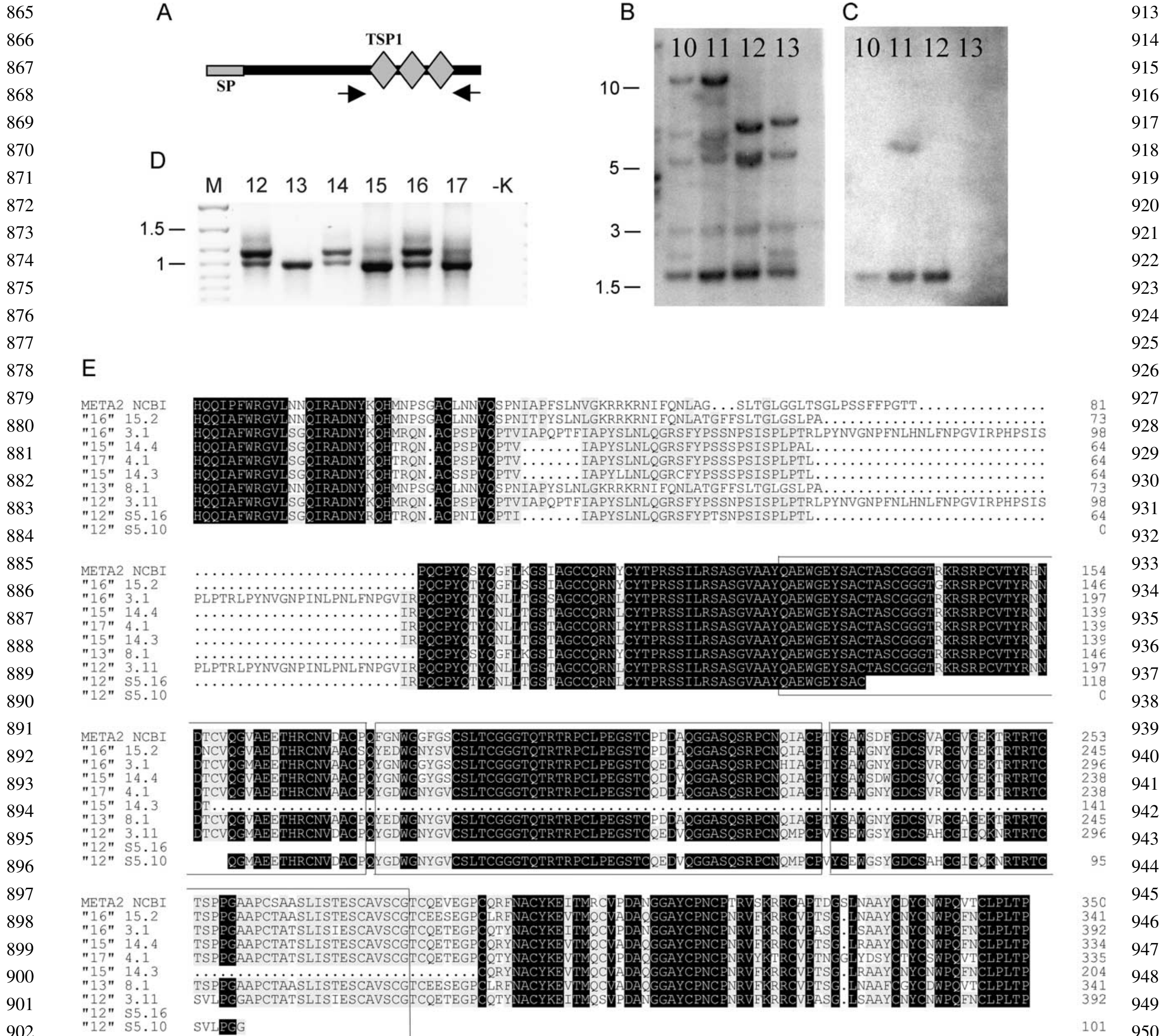

Fig. 5. Inter- and intraindividual variability of ciMETA2 proteins. (A) Domain organisation of the ciMETA2-like gene. SP—signal peptide; TSP1—thrombospondin type 1 domains; Arrows—the positions of RT-PCR primers. (B) Southern blot hybridisations with *Hind*III restricted genomic DNA from four *Ciona* individuals (animals '10', '11', '12', '13') and META2 cDNA fragment from individual '12' as a probe (AY895059). (C) Southern blot hybridisation with the same filter and probe as in B) using high stringency conditions. (D) RT-PCR with cDNAs from six *Ciona* individuals (animals '12', '13', '14', '15', '16', '17') showing variation in number and length of amplified fragments from individual to individual. M—marker. (E) Multiple amino acid alignment of seven ciMETA2 alleles, amplified from 5 *Ciona* individuals and two cDNA fragments from animal '12' from SSH. Individual clones are indicated by the animal number in quotation marks, followed by the clone name, e.g. '16' 15.2. Amino acids identical in all the sequences are depicted by black background; TSP1 domains are boxed; META2 NCBI—amino acid sequence from GeneBank (BAB40595). Accession numbers for the other sequences are AY895049-AY895054, AY899286-AY899288, AY895059.

genome as a gene family of at least seven genes organised in several clusters (Fig. 7A; see also http://genome.jgi-psf.org/ciona4/ciona4.home.html). Scaffold 61 contains a cluster of four genes linked within a region of 25 Kb (Fig. 7A). Scaffold 67 contains a cluster of three META2 genes linked within a region of 13 Kb.

To examine the degree of variability of ciMETA2 transcripts between different *Ciona* individuals, we used RT-PCR and primers corresponding to conserved parts of the 3′ region containing the TSP1 domains. As shown in Fig. 5D, when using cDNAs from individuals ‘12’, ‘14’ and ‘16’, two PCR products can be obtained while in animals ‘13’, ‘15’ and ‘17’ only one PCR product was detected indicating individual specific differences in the amino acid sequences. ciMETA2 fragments obtained from individuals ‘12’, ‘13’, ‘15’, ‘16’ and ‘17’ were sequenced and found to display a number of amino acid substitutions and indels (Fig. 5E). Even within one individual, ciMETA2 proteins display some degree of variation as indicated by clones ‘15’ 14.4 and ‘15’ 14.3 (Fig. 5E). None of the amplified sequences is identical to the published ciMETA2 sequence from the NCBI database. Thus, ciMETA2, similar to ciS7, shows both sequence variation from individual to individual and intraindividual divergence.

3.4. The genomic variability in Ciona is limited to a subset of genes

It is well known that the genetic polymorphism within populations of invertebrates is much higher than that in vertebrates. For example, the whole genome shotgun sequencing project in *C. intestinalis* showed that the overall variation of nucleotide sequences between alleles within one individual is as high as 1.2%, which is nearly 15-fold more than that observed in humans and 3-fold more than that found in puffer fish [15]. In *Ciona savignyi,* genome variation between two haplotypes of one individual seems to be even higher than in *C. intestinalis* with an estimated rate of substitutions rising up to 6%. To exclude the possibility that the observed variability in the ciS7 and META2 loci is within the range of ‘normal’ background interindividual variability, we examined the sequence diversity of two control genes, ciGAPDH (ci0100132109, scaffold 41) and ciFLRT (ci0100131663, scaffold 245) between several *C. intestinalis* individuals. GAPDH codes for a metabolic enzyme and is frequently used for equilibration purposes in RT-PCR or Northern blots. We performed Southern blot hybridisation with the *Ciona* GAPDH fragment (AY89057) as a probe and detected (Fig. 2B) no RFLP polymorphism in this locus between individuals ‘10’, ‘11’, ‘12’, ‘13’. The sequences of several GAPDH cDNA fragments amplified from individuals ‘12’ and ‘13’ were found to be identical. The corresponding amino acid sequence was similar to the sequence in NCBI except for one amino acid substitution (data not shown). The second control gene used, ciFLRT, codes for a transmembrane receptor of unknown function. It consists of a signal peptide, N-terminal LRR and C-terminal LRRs with seven LRR between them, and a fibronectin type III domain in front of the membrane spanning part (see Fig. 6A). This receptor contains several repetitive units (LRR) and is present in two copies, which are linked in the *Ciona* genome (see scaffold 245), much alike the ciS7 genes. Moreover, ciFLRT is similar in structure to the variable lymphocyte receptor (VLR) from lamprey which recently was shown to play an important role in innate immune responses [18]. Thus, ciFLRT appears to be a good control gene to examine overall genetic variability in *Ciona*. RT-PCR with a primer set directed against the N-terminal LRR and C-terminal LRR resulted in fragments of identical size in all four individuals tested (see Fig. 6B). The corresponding sequences from animals ‘14’ and ‘17’ were identical except for three amino acid substitutions (see Fig. 6C). No variations in the LRR positions or any kind of domain reshuffling could be detected. Therefore, although the overall variability in the *Ciona* genome is much higher than that in vertebrates, the variability appears to be limited to certain classes of genes.

4. Discussion

4.1. Suppression subtractive hybridisation (SSH) as method of choice to identify individually variable genes in Ciona

SSH allows a qualitative comparison of transcriptomes between different tissues or between different

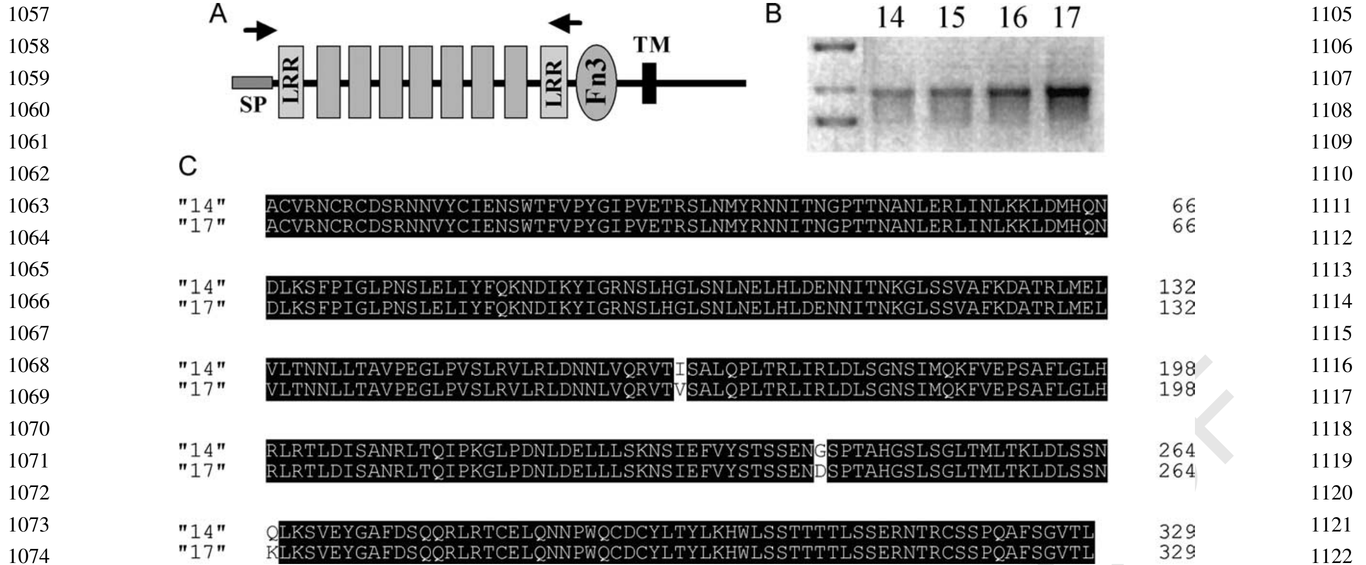

Fig. 6. ciFTLR transmembrane receptor shows minimal variation between *Ciona* individuals. (A) Domain organisation of ciFLRT. Arrows—the positions of RT-PCR primers; SP—signal peptide; LRR—N-terminal and C-terminal leucine rich repeats; Fn3—fibronectin type 3 domain; TM—transmembrane domain. (B) RT-PCR with ciFTLR primers shows products of identical length in all *Ciona* individuals tested; (C) Alignment of consensus amino acid sequences spanning LRR region from individuals '14' and '17' with three substitutions within 329 amino acids. Genebank accession numbers AY895055, AY895056.

developmental stages [31]. More recently, a modification of the SSH procedure has been used for the comparison of whole bacterial genomes to determine the genetic differences between two strains [32]. The complexity of a bacterial genome is comparable to the complexity of the transcriptome of a eukaryotic cell. We, therefore, used the SSH procedure with minor modifications (see Material and methods) to identify transcripts which are variable between two similar somatic tissues from two different *C. intestinalis* individuals. As a result we obtained two classes of proteins of unknown function but with interesting features. Both classes of proteins are (i) variable between individuals and (ii) coded in the *Ciona* genome by several gene copies organized in clusters. One class of individually variable proteins appears to consist of secreted protein with thrombospondin type 1 (TSP1) domain. The second variable class identified by SSH were secreted proteins with multiple EGF-like domains.

Do the variable ciS7 and ciMETA2 proteins take part in *Ciona* allorecognition? Although the question cannot be answered yet, several observations indicate that these proteins are functionally significant and may participate in controlling non-self recognition. (i) The extreme degree of variability is restricted to selected groups of proteins. No evidence could be detected for any considerable variation in genes encoding proteins such as GAPDH or transmembrane receptor ciFLRT. (ii) The EST data as well as our RT-PCR experiments demonstrate that ciS7 and ciMETA2 isoforms are expressed in *Ciona* cell types such as blood cells and/or gametes known to be mediators of various recognition events. (iii) Each *Ciona* individual has several ciS7 and META2 genes and transcribes several mRNAs, coding for similar but not identical proteins belonging to each gene family. According to the RFLP analysis, the number of ciS7 and META2 genes seems to vary from individual to individual. For example, according to the genome data (see Fig. 7A and B), one individual appears to have at least seven different META2 genes and two ciS7 genes. Our RT-PCR data, in addition, show a high degree of amino acid sequence variation in the products of these genes between *Ciona* individuals. Therefore, each *Ciona* individual carries an unique repertoire (haplotype) in the ciMETA2 and ciS7 loci.

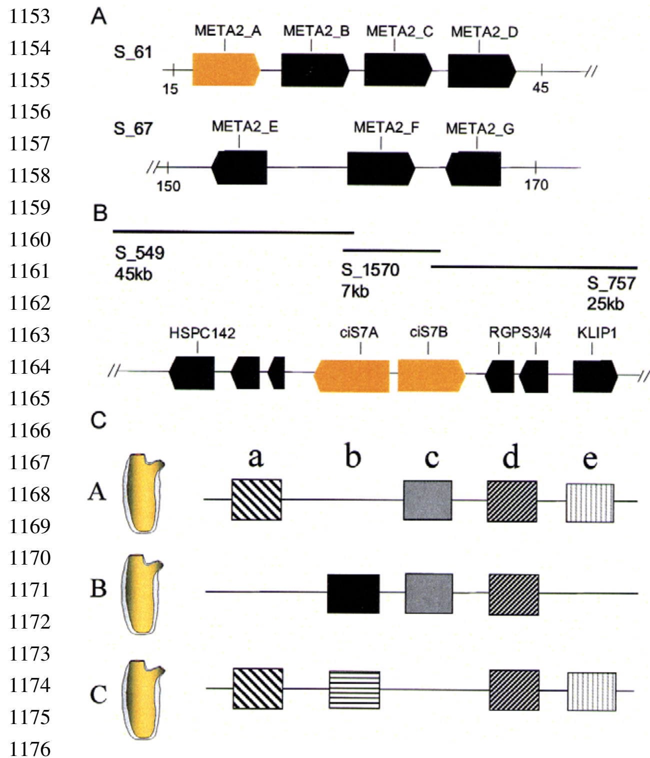

Fig. 7. Schematic representation of the regions in the *Ciona* genome containing ciMETA2 and ciS7 loci. (A) Scaffold 61 contains a cluster of four ciMETA2 genes within 30 kb and scaffold 67 contains a cluster of three ciMETA2 genes linked within 20 kb. (B) Three overlapping scaffolds (S_549, S_1570, S_757) contain two tightly linked ciS7 genes (ciS7A and ciS7B) with opposite transcription directions and one common promoter. (C) Model for the possible organisation of ciMETA2 and ciS7 loci. Each individual carries a unique haplotype, differing from other animals by the number of genes in the locus and sequence variation between alleles. A, B, C—different *Ciona* individuals; a, b, c, d, e—different members of a gene family.

The presence of individual-specific haplotypes in invertebrates was reported previously for genes possibly involved in pathogen defence such as scavenger receptor cystein-rich (SRCR) genes in sea urchin [33]. There, each individual has a cluster of several SRCR genes with the gene number and sequence different from individual to individual. Another example of the highly diversified gene family in invertebrates are the variable region-containing chitin binding proteins (VCBP) in the protochordate *Branchiostoma floridae* (reviewed in [17]). Each *Branchiostoma* individual possess several different genomic copies of VCBP gene, producing proteins with different amino acid sequences [34]. Moreover, the number of VCBP genes and their sequences are different from individual to individual, with 43 different protein sequences isolated from 13 different animals. A similar fascinating example for molecular hypervariability in an invertebrate are fibrinogen-related proteins (FREPs) in hemolymph of *Biomphalaria glabrata* [35,36]. These proteins seem to be involved in internal defense and are diversified at the genomic level by point mutations and recombinatorial processes.

In *Ciona*, the distinct haplotypes in the ciMETA2 and ciS7 loci are most probably established for each individual by crossing over events during gamete maturation and/or fertilization. No evidence could be detected yet for somatic recombination events. Since individuals differ in both number of ciMETA2 and ciS7 genes and in the sequences of the corresponding alleles, we hypothesize that 'individuality' in *Ciona* may be encoded by the haplotype, or in other words, the individual-specific combination of genes in the corresponding genome locus (see Fig. 7C). To some extent this phenomen may resembles the variability of killer cell immunoglobulin-like receptors (KIR) seen in higher vertebrates. Within the human population, for example, KIR haplotypes and genotypes differ greatly both in their gene content and by allelic polymorphism at the individual KIR genes [37].

4.2. Is allorecognition in urochordates based on the same molecular toolkit as fertilization?

In the absence of definitive orthologs of Igs, TCRs, MHCs and RAGs, what is the molecular nature of the urochordate allorecognition system? Despite of considerable efforts and progress in understanding the genetics of the allorecognition by positional cloning in *Botryllus* [13] and more recently in the hydroid *Hydractinia* [3,38], the question remains unanswered yet. Reasonable predictions, however, are that (1) the molecules engaged in self-/nonself recognition must be variable between genetically different individuals; (2) they must be expressed in the cells which execute recognition function; (3) in urochordates similar or genetically linked molecules may control

UNCORRECTED PROOF

allorecognition and block of self-fertilization. Evidence for the latter statement comes from observations in *Botryllus*, where a linkage between the Fu/HC locus and the locus which blocks self fertilization has been demonstrated by genetic experiments [4]. Moreover, in *Halocynthia roretzi*, a solitary tunicate with a well documented block of self-fertilization and cell-based histocompatibility reactions [39], monoclonal antibodies that inhibit allogeneic reactions mediated by hemocytes also inhibit fertilization [22]. Thus, in *Halocynthia* there are receptors and target ligands involved both in self-recognition by somatic cells and self-discrimination by gametes. Recently, one of the receptor molecules blocking self fertilization in *Halocynthia* was identified as viteline coat protein HrVC70 [23]. HrVC70 is a transmembrane receptor expressed in oocytes and consisting of multiple EGF-like domains and one zona pellucida (ZP) domain. The amino acid sequence of HrVC70 varies among genetically different individuals with no identical sequence among several individuals tested. The inter-individual variation in HrVC70 was found to be due to the genomic DNA polymorphisms between individuals which is generated by crossing over events and specific nucleotides substitutions [23]. Currently, HrVC70 represents the only candidate allorecognition molecule known from invertebrates.

One of the classes of variable transcripts identified in our SSH approach, ciS7, encodes secreted proteins with multiple EGF-like domains. We note that in *Caenorhabditis elegans*, a sperm transmembrane protein with multiple EGF-like domains, SPE-9, is required for fertilization [40–42]. Interestingly, sequencing 400 kb of the Fu/HC locus of *Botryllus schlosseri* genome resulted in the identification of a gene with homology to the Notch 4 receptor which contains multiple EGF-like domains [13]. Although it remains to be shown where and when this gene is expressed and whether it is variable between different *Botryllus* individuals, it is intriguing to speculate that its presence in the *Botryllus* Fu/HC locus indicates its involvement in histocompatibility reactions and, more general, that molecules with multiple EGF-like domains are major players in self/nonself recognition reactions in urochordates.

Acknowledgements

We are grateful to the Biological Station at Helgoland for supplying us with *C. intestinalis* and to Matthias Habetha, Ryan Schwarz and Holger Zill for critically reading the manuscript. Supported by the Deutsche Forschungsgemeinschaft (grant DFG BO 848/12,3 to TCGB and KK).

References

[1] Muller WE, Muller I, Zahn RK, Maidhof A. Intraspecific recognition system in scleractinian corals: morphological and cytochemical description of the autolysis mechanism. J Histochem Cytochem 1984;32(3):285–8.

[2] Chadwick-Furman N, Rinkevich B. A complex allorecognition system in a reef-building coral: delayed responses, reversal and nontransitive hierarchies. Coral Reefs 1994;13: 57–63.

[3] Cadavid LF. Self-discrimination in colonial invertebrates: genetic control of allorecognition in the hydroid *Hydractinia*. Dev Comp Immunol 2004;28(9):871–9.

[4] Saito Y, Hirose E, Watanabe H. Allorecognition in compound ascidians. Int J Dev Biol 1994;38(2):237–47.

[5] Raftos DA, Briscoe DA. Genetic basis of allograft rejection in *Styela plicata* (Urochordata, Ascidiacea). J Hered 1990;81: 160–6.

[6] Reddy AL, Bryan B, et al. Integumentary allograft versus autograft reactions in *Ciona intestinalis*: a protochordate species of solitary tunicate. Immunogenetics 1975;1:584–90.

[7] Murabe N, Hoshi M. Re-examination of sibling cross-sterility in the ascidian *Ciona intestinalis*: genetic background of the self-sterility. Zoolog Sci 2002;19(5):527–38.

[8] Bancroft FW. Variation and fusion of colonies in compound ascidians. Proc Calif Acad Sci 1903;3:137–86.

[9] Khalturin K, Pancer Z, Cooper MD, Bosch TCG. Recognition strategies in the innate immune system of ancestral chordates. Mol Immunol 2004;41:1077–87.

[10] Burnet FM. ‘Self-recognition’ in colonial marine forms and flowering plants in relation to the evolution of immunity. Nature 1971;232:230–5.

[11] Scofield VL, Schlumpberger JM, West LA, Weissman IL. Protochordate allorecognition is controlled by a MHC-like gene system. Nature 1982;295:488–502.

[12] Khalturin K, Becker M, Rinkevich B, Bosch TCG. Urochordates and the origin of natural killer cells: identification of a CD94/NKR-P1 related receptor in blood cells of *Botryllus*. Proc Natl Acad Sci USA 2003;100:622–7.

[13] De Tomaso AW, Weissman IL. Initial characterization of a protochordate histocompatibility locus. Immunogenetics 2003;55(7):480–90.

[14] De Tomaso AW, Weissman IL. Construction and characterization of large-insert genomic libraries (BAC and fosmid)

UNCORRECTED PROOF

from the Ascidian *Botryllus schlosseri* and initial physical mapping of a histocompatibility locus. Mar Biotechnol (NY) 2003;5(2):103–15.

[15] Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomaso A, et al. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. Science 2002; 298(5601):2157–67.

[16] Azumi K, De Santis R, De Tomaso A, Rigoutsos I, Yoshizaki F, Pinto MR, et al. Genomic analysis of immunity in a Urochordate and the emergence of the vertebrate immune system: 'waiting for Godot'. Immunogenetics 2003;55(8): 570–81.

[17] Eason DD, Cannon JP, Haire RN, Rast JP, Ostrov DA, Litman GW. Mechanisms of antigen receptor evolution. Semin Immunol 2004;16(4):215–26.

[18] Pancer Z, Amemiya CT, Ehrhardt GR, Ceitlin J, Gartland GL, Cooper MD. Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. Nature 2004; 430(6996):174–80.

[19] Raftos DA, Tait NN, Briscoe DA. Allograft rejection and alloimune memory in the solitary urochordate, *Styela plicata*. Dev Comp Immunol 1987;11:343–51.

[20] Raftos DA. The morphology of allograft rejection in *Styela plicata* (Urochordata: Ascideacea). Cell Tissue Res 1990;261: 389–96.

[21] De Santis R, Pinto MR. Gamete self-discrimination in ascidians: a role for the follicle cells. Mol Reprod Dev 1991; 29(1):47–50.

[22] Arai M, Suzuki-Koike M, Ohtake S, Ohba H, Tanaka K, Chiba J. Common cell-surface antigens functioning in self-recognition reactions by both somatic cells and gametes in the solitary ascidian *Halocynthia roretzi*. Microbiol Immunol 2001;45(12):857–66.

[23] Sawada H, Tanaka E, Ban S, Yamasaki C, Fujino J, Ooura K, et al. Self/nonself recognition in ascidian fertilization: Vitelline coat protein HrVC70 is a candidate allorecognition molecule. Proc Natl Acad Sci USA 2004; 101(44):15615–20.

[24] Hancock CN, Kondo K, Beecher B, McClure B. The S-locus and unilateral incompatibility. Philos Trans R Soc Lond B Biol Sci 2003;358(1434):1133–40.

[25] Kemp BP, Doughty J. Just how complex is the Brassica S-receptor complex? J Exp Bot 2003;54(380):157–68.

[26] Casselton LA. Mate recognition in fungi. Heredity 2002;88(2): 142–7.

[27] Sambrook J, Fritsch EF, Maniatis T. Molecular cloning. A laboratory manual. 2nd ed. NY: Cold Spring Harbor laboratory Press; 1989.

[28] Devon RS, Porteous DJ, Brookes AJ. Splinkerettes—improved vectorettes for greater efficiency in PCR walking. Nucleic Acids Res 1995;23(9):1644–5.

[29] Khalturin K. Molecular analysis of self/nonself recognition in the urochordate *Botryllus schlosseri*. PhD Thesis, University of Kiel http://e-diss.uni-kiel.de/diss_688/

[30] Nakayama A, Satou Y, Satoh N. Isolation and characterization of genes that are expressed during *Ciona intestinalis* metamorphosis. Dev Genes Evol 2001;211(4):184–9.

[31] Diatchenko L, Lau YF, Campbell AP, Chenchik A, Moqadam F, Huang B, et al. Proc Natl Acad Sci USA 1996; 93(12):6025–30.

[32] Akopyants NS, Fradkov A, Diatchenko L, Hill JE, Siebert PD, Lukyanov SA, et al. PCR-based subtractive hybridization and differences in gene content among strains of *Helicobacter pylori*. Proc Natl Acad Sci USA 1998;95(22):13108–13.

[33] Pancer Z. Dynamic expression of multiple scavenger receptor cysteine-rich genes in coelomocytes of the purple sea urchin. Proc Natl Acad Sci USA 2000;97(24):13156–61.

[34] Cannon JP, Haire RN, Schnitker N, Mueller MG, Litman GW. Individual protochordates have unique immune-type receptor repertoires. Curr Biol 2004;14(12):R465–R6.

[35] Loker ES, Adema CM, Zhang SM, Kepler TB. Invertebrate immune systems—not homogeneous, not simple, not well understood. Immunol Rev 2004;198:10–24.

[36] Zhang SM, Adema CM, Kepler TB, Loker ES. Diversification of Ig superfamily genes in an invertebrate. Science 2004; 305(5681):251–4.

[37] Parham P. Killer cell immunoglobulin-like receptor diversity: balancing signals in the natural killer cell response. Immunol Lett 2004;92(1–2):11–13.

[38] Cadavid LF, Powel AE, Nicotra ML, Moreno M, Buss LW. An invertebrate histocompatibility complex. Genetics 2004; 167(1):357–65.

[39] Fuke MT, Numakunai T. Allogeneic cellular reactions between intra-specific types of a solitary ascidian *Halocynthia roretzi*. Dev Comp Immunol 1982;6(2):253–61.

[40] Singson A, Mercer KB, L'Hernault SW. The *C. elegans* spe-9 gene encodes a sperm transmembrane protein that contains EGF-like repeats and is required for fertilization. Cell 1998; 93(1):71–9.

[41] Zannoni S, L'Hernault SW, Singson AW. Dynamic localization of SPE-9 in sperm: a protein required for sperm-oocyte interactions in *Caenorhabditis elegans*. BMC Dev Biol 2003; 3(1):10.

[42] Putiri E, Zannoni S, Kadandale P, Singson A. Functional domains and temperature-sensitive mutations in SPE-9, an EGF repeat-containing protein required for fertility in *Caenorhabditis elegans*. Dev Biol 2004; 272(2):448–59.

UNCORRECTED PROOF