



Available at www.sciencedirect.com



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



# Evolution of polydom-like molecules: Identification and characterization of cnidarian polydom (*Cnpolydom*) in the basal metazoan *Hydractinia* <sup>☆</sup>

Ryan S. Schwarz<sup>a,\*</sup>, Thomas C.G. Bosch<sup>b</sup>, Luis F. Cadavid<sup>c</sup>

<sup>a</sup>Department of Biology, University of New Mexico, Albuquerque, NM 87131-0001, USA

<sup>b</sup>Zoological Institute, Christian-Albrechts-University Kiel, Olshausenstrasse 40, 24098 Kiel, Germany

<sup>c</sup>Department of Biology and Institute of Genetics, Universidad Nacional de Colombia, Cr. 30 #45-08, Bogota, DC, Colombia

Received 12 February 2008; received in revised form 14 March 2008; accepted 18 March 2008

## KEYWORDS

SVEP1;  
Pentraxin;  
Invertebrate;  
Innate immunity;  
Cnidaria;  
Gene;  
Defense;  
Immune challenge

## Summary

End sequencing of random BAC clones from a *Hydractinia symbiolongicarpus* (Cnidaria: Hydrozoa) genomic library revealed a gene across a ~37.5kb region of the *H. symbiolongicarpus* genome sharing highest sequence identity and domain architecture to mammalian polydom that we in turn named cnidarian polydom (*CnPolydom*). Sharing all eight domain types characteristic of polydom and organized in a similar 5'–3' manner, *CnPolydom* was predicted to contain three additional domain types: PAN, FA58C, and CUB that are characteristic of *CnPolydom*. Expression analysis of *CnPolydom* from *H. symbiolongicarpus* (*Hysy-CnPolydom*) showed upregulation in response to bacterial and primarily fungal challenges, with transcripts produced specifically by a subset of interstitial stem cells (i-cells) and/or neural cells throughout the ectodermal tissue layer of feeding polyps (gastrozooids). This is the first description of a polydom-like molecule outside of Mammalia and provides evolutionary perspective on the ancestral structure and role of this pentraxin family clade.

Published by Elsevier Ltd.

## Introduction

It has become increasingly clear that members of the ancestral animal phylum Cnidaria have evolved surprisingly complex genomes and provide novel insights into the extent of genetic conservation and divergence across the metazoan lineages [1–3]. Thus, we can gain comparative insight into the evolution of animal immune systems, and the individual building blocks from which they are comprised, by assessing

<sup>☆</sup>The sequences reported in this paper have been deposited to GenBank database under the accession numbers EF566886, EF566887, and EF566888.

\*Corresponding author. Current address: Animal Parasitic Diseases Laboratory, Building 1042, BARC-East, USDA, 10300 Baltimore Avenue, Beltsville, MD 20705, USA. Tel.: +1 301 504 5596; fax: +1 301 504 5306.

E-mail address: schwarz.ryan@gmail.com (R.S. Schwarz).

the repertoire of immune genes in these ancestral animals having a true tissue level of organization (Eumetazoa). Within the Cnidaria, class Anthozoa and Hydrozoa are the most ancestral of the four currently recognized classes [4,5]. Genes encoding immune-related molecules from members of the Hydrozoa and Anthozoa are recent targets of research to assemble the complexity of basal eumetazoan immune components [6], and the hydrozoan *Hydractinia* has proven an effective study organism in this pursuit [7–9]. These colonial cnidarians typically grow as an encrustation on gastropod shells occupied by hermit crabs and are comprised of many polyps, including gastrozooids (feeding polyps) and gonozooids (reproductive polyps), that are interconnected via a gastrovascular system of stolons embedded in a stolonal mat [10,11]. *Hydractinia* are diploblastic, having an ectodermal and endodermal cell layer separated by an acellular mesoglea. Being the most basal eumetazoans, hydrozoans have true tissue level organization and three stem cell lineages from which all of their differentiated cell types arise: ectodermal, endodermal, and interstitial stem cells [12,13]. *Hydractinia* have become a model, colonial hydrozoan representative of Cnidaria for developmental, reproductive, and immunological research [11].

Cnidarian immune research is beginning to reveal novel immune-related genes in cnidarians [9], in addition to genes that are well conserved through higher metazoan taxa. Such conserved immune-related components include a complement 3-like gene described from the anthozoan coral *Swiftia exserta* [14], known to be central to the opsonic function of the complement system in vertebrates, and in both Hydrozoa and Anthozoa conserved genes containing membrane attack complex/perforin (MAC/PF) domain have been reported [6]. In addition, many components of the Toll/TLR pathway have been identified in both Anthozoa [15] and Hydrozoa [6,16], including transmembrane receptors encoding toll/interleukin 1 receptor (TIR) domains (Hydrozoa and Anthozoa) and multiple leucine rich repeat (LRR) domains (Anthozoa only), intracellular signaling cascade components including MyD88 genes containing the functional DEATH domain, and transcription factors of the Rel/NF- $\kappa$ B and NFAT gene family that initiate effector genes such as antimicrobial peptides. Lacking in many of these studies, however, is assessment of the immunologic relevance of these genes within the Cnidaria.

Pentraxin molecules are a well characterized family of immune components known for their highly conserved structural motif, the pentraxin domain, and for their function as recognition and effector molecules in the innate immune responses of both vertebrates and invertebrates (reviewed in [17,18]). One group of pentraxin molecules includes C-reactive protein (CRP) and serum amyloid P component (SAP), the prototypical short pentraxin proteins of the acute-phase response classically described in mammals and known to bind a variety of microorganisms including fungi, yeast, and bacteria [19,20], as well as common membrane moieties including phosphorylcholine, phosphorylethanolamine, and lipopolysaccharide (LPS) [21–24]. These proteins polymerize and have direct opsonic and complement activation ability [25,26]. A diverse repertoire of CRP-like and SAP-like short pentraxin molecules having similar antimicrobial roles in the hemolymph of

three horseshoe crab taxa have also been well studied [27–29]. A second group of pentraxins, known as the long pentraxins, was first identified with the cloning of human pentraxin 3 (PTX3) [30]. PTX3 is also an acute phase protein and has been linked most importantly to its role in mediating fungal recognition and phagocytosis [31,32].

A new member of the pentraxin family has been recently described from mammals, but its function and the extent of its evolutionary conservation in other taxa is unknown. Polydom was first identified and described by Gilgès et al. in 2000 from mouse [33]. *Polydom* was originally amplified from total RNA of a murine haematopoiesis-sustaining bone marrow stromal cell line (MS-5) using degenerate EGF domain-specific primers [33]. The full-length mouse mRNA encodes a 3567aa protein predicted to include a novel combination of domains including an N-terminal von Willebrand factor A (VWA) domain, 2 hyalin repeat domains (HYR), 10 epidermal growth factor repeats (EGF), 34 complement control protein (CCP) domains, and a single pentraxin domain (PTX) at its core, making polydom a member of the pentraxin family of lectins. Based on this domain composition, *polydom* has recently been given the additional name *SVEP1* for sushi, von Willebrand factor A, epidermal growth factor, pentraxin molecule 1 [34].

Though largely unknown, data on polydom are helping to elucidate potential functional roles of the polydom protein in humans and mice. *Polydom* expression has been shown in several tissue types including skeletal tissues (bone and periosteum) of human and mouse, with protein expression demonstrated on the surface of human stromal cells (osteoblasts) where they act as cell adhesion molecules (CAMs) [34]. Several studies have shown one important location of *polydom* expression is in the placenta of human and mouse. In human, the transcript level of *polydom* from basal plate tissue of the placenta, where allogeneic interaction occurs between maternal and fetal cells, more than doubles once a pregnancy comes to term (37–40 weeks) compared to polydom levels during pregnancy (14–24 weeks) [35]. Although placental expression of *polydom* has been shown to be substantial in these mammals, polydom-like sequences are also predicted from non-placental vertebrate taxa including chicken and zebrafish, as well as invertebrate taxa, suggesting the placental role for polydom is a derived function in mammals.

Using a *Hydractinia symbiolongicarpus* BAC library to randomly examine their genome for unknown genes of potential immunologic relevance, we identified a modular gene comprised of many domain types known for their association with immune molecules including: VWA, EGF, CCP, PTX, PAN, FA58C, and CUB domains. Using primers designed from the genomic sequence we performed RT-PCR and amplified partial cDNAs to confirm transcription of this gene from *H. symbiolongicarpus* and its sister species *Hydractinia echinata*, and have determined the encoded gene's exon-intron structure within the *H. symbiolongicarpus* genome according to the corresponding cloned cDNA and predicted ORFs. The expressed gene, comprised of 12 distinct domain types, shared highest identity to the polydom gene from human ( $E = 1e-99$ ; EMBL CAH74139) and we have in turn named it cnidarian polydom (*CnPolydom*), and identify the gene from *H. symbiolongicarpus* and *H. echinata* as *Hysy-CnPolydom* and *Hyec-CnPolydom*,

respectively. In this paper we report on the characterization of CnPolydom from the basal eumetazoan phylum Cnidaria, and provide the first description of a polydom-like gene from a non-vertebrate organism with immune relevance that is highly conserved in architecture across the Eumetazoa.

## Materials and methods

### Animal cultures

Mature colonies of *H. symbiolongicarpus* used in this study were isolated from gastropod shells occupied by hermit crabs collected at the Atlantic Ocean intertidal zone along the Connecticut and Massachusetts shoreline of the North-eastern United States. Mature colonies of *H. echinata* were collected from hermit crab-occupied gastropod shells out of the North Sea off the northwestern coast of Germany. Colonies were maintained in the laboratory by establishing explants on glass microscope slides and growing them in aquaria filled with artificial seawater (Coralife) at 31–33‰ salinity and ~16 °C temperature. Colonies were removed from aquaria for feeding with brine shrimp (*Artemia salina*) nauplii three times a week and concurrent ~15% water volume replacement of aquarium water.

### BAC library clone identification, isolation, and genomic sequencing

Construction of the *H. symbiolongicarpus* bacterial artificial chromosome (BAC) library, generated by Amplicon Express (Pullman, WA), has been described in detail previously [9]. The male individual used as source tissue for construction of the BAC library, 4117-2, was derived from a *H. symbiolongicarpus* inbreeding program [36,37]. Random end sequencing of BAC clones was performed as a screening process for genes of interest. Individual BAC clones were streaked on LB agar chloramphenicol selective plates (20 µg/ml) and single colony isolates were picked and grown overnight in 5 ml of LB chloramphenicol selective broth (20 µg/ml). DNA was isolated from overnight cultures using the PSIΨClone BAC DNA Kit (Princeton Separations, Inc.; Adelphia, NJ). Sequencing reactions were performed on the purified DNA using BigDye Terminator enzyme v1.1 (Applied Biosystems) using the SP6 and T7 priming regions on either side of the BAC insert site, and sequences were run in an ABI Prism 377 DNA Sequencer (Applied Biosystems). Sequence from the SP6 priming site in BAC clone 84010 revealed a 203 bp ORF with highest identity to a portion of the pentraxin domain from a predicted polydom mRNA in the *Rattus norvegicus* genome (XM\_232929;  $E = 1e-16$ ). Primers were designed from this sequence to amplify a 414 bp pentraxin domain fragment (F: 5'-CCTCCTCCTTTTCCAATTCAACA-3', R: 5'-TGCAAAATCT-GACCAAGCAACAA-3') from BAC clone 84010 gDNA and the product was purified from agarose gel using QIAquick® gel extraction kit (Qiagen; Valencia, CA; USA). The pentraxin amplicon was labeled with digoxigenin (DIG)-dUTP through random priming (Roche Diagnostics; Indianapolis, IN; USA) and used in a hybridization screen of 18,432 clones from the BAC library spotted on nylon membrane (plates HHS1-48; ~25% of the entire BAC library), revealing nine additional clones positive for pentraxin: 2N9, 23A3, 8D21, 16A3, 20P23,

25D24, 28B6, 34I17, and 41D2. BAC clone 2N9 was chosen for shotgun sequencing, performed by Amplicon Express. This involved random shearing of purified BAC clone DNA using the “hydroshear” method, subcloning DNA fragments into plasmid vector pPCRSCRIPT, and sequencing plasmid clone inserts to generate ~4.5 × sequence coverage. The preliminary shotgun assembly into contigs was performed using the automated sequence trimming and assembly software packages PHRED and PHRAP [38,39] and the assembly viewer software Consed/Autofinish [40] (<http://www.phrap.org/>). Gaps in the insert sequence were filled by designing primers ~100 bps from the ends of contigs and performing sequencing reactions on BAC clone 2N9 template to walk the genome until contigs could be joined.

### Sequence analysis and cloning of CnPolydom

Protein motifs within ORFs of the genomic sequence were predicted by conducting searches against the Prosite (<http://www.expasy.org/prosite/>) [41], InterPro (<http://www.ebi.ac.uk/InterProScan/>) [42], Simple Modular Architecture Research Tool (SMART, <http://smart.embl-heidelberg.de/>) [43,44], and protein families (Pfam; <http://www.sanger.ac.uk/Software/Pfam/>) [45] databases. Primers designed from ORFs of the genomic sequence predicted to encode distinct protein domains were used in reverse transcription polymerase chain reaction (Access RT-PCR system kit; Promega, Madison, WI, USA) to amplify CnPolydom cDNA from total RNA isolated using TRIzol® reagent (Invitrogen, Carlsbad, CA, USA). Two regions of the *H. symbiolongicarpus* CnPolydom message were isolated and cloned in this way. A 1656 bp cDNA from the 5' region of the gene was amplified using a forward primer from the VWA domain (5'-CAGGAAGCATGGTTTCTACTGGC-3') and reverse primer from the pentraxin domain (5'-GATGGTGA-AACTCTTGACTGAAATG-3'). A 3366 bp cDNA from the C-terminal region of the gene was amplified using a forward EGF domain primer (5'-GCTATTGTGCTTCAAACCATG-3') and a reverse CUB domain primer (5'-GCGTGACTTGAAT-AAGCCAGTG-3'). RT-PCR using CnPolydom pentraxin domain primers designed from the *H. symbiolongicarpus* sequence (F: 5'-CATTTCAGTCAAGAGTTTACC-3' and R: 5'-GCCTAT-ATCGTGATGACATGATG-3') were also used to amplify the pentraxin domain cDNA from *H. echinata*. cDNA fragments were purified from agarose gel using QIAquick® gel extraction kit (Qiagen) and cloned using pGEM-T Easy vector system (Promega). Plasmid inserts were sequenced using BigDye Terminator v1.1 in an ABI Prism 3100 genetic analyzer (Applied Biosystems).

### Sequence and phylogenetic analyses

Database queries using nucleotide and protein translations of CnPolydom were performed with basic local alignment search tool (BLAST) algorithm [46] through the National Center for Biotechnology Information server (NCBI, <http://www.ncbi.nlm.nih.gov>) as well as cnidarian specific sequence collections at CnidBase (<http://cnidbase.bu.edu/>) [47] and compagen ([www.compagen.org](http://www.compagen.org)) to identify other sequences containing pentraxin domains and to identify

sequences sharing similar domain architecture with *CnPolydom*. Sequences were retrieved and translated to amino acid if necessary using the Translate tool through the Expert Protein Analysis System (Expasy) proteomics server (<http://ca.expasy.org>). Amino acid sequence regions corresponding to pentraxin domains were identified and isolated from each retrieved sequence using SMART (<http://smart.embl-heidelberg.de/>) [43,44] prior to alignment using the protein multiple sequence alignment software (MUSCLE) [48]. Phylogeny reconstruction of pentraxin domains was performed on the MUSCLE alignments using MEGA version 4 [49] with the distance-based method of minimum evolution (ME) and neighbor joining (NJ), in addition to maximum parsimony (MP) method. Bootstrap statistical tests of robustness were performed on the inferred phylogenies.

Protein translations of polydom sequences identified from public sequence databases were annotated via SMART using profile hidden Markov model searches for protein domains and families against the following databases: SMART [43,44], research laboratory for structural bioinformatics protein data bank [50] (RCSB PDB; <http://www.pdb.org/>), structural classification of proteins [51] (SCOP v1.71; <http://scop.mrc-lmb.cam.ac.uk/scop/>), protein families [45] Pfam v21.0; <http://www.sanger.ac.uk/Software/Pfam/>), and SignalP 3.0 (<http://www.cbs.dtu.dk/services/SignalP/>) for signal peptides.

### Southern blotting

Genomic DNA for Southern hybridization analysis was isolated from 2 to 3 day fasted *H. symbiolongicarpus* colonies using a teflon pestle to grind the tissue in homogenization buffer (100 mM Tris-HCl pH8.0; 100 mM NaCl; 200 mM sucrose; 50 mM EDTA; 0.5% SDS) along with 2.5 mg of RNase A (Sigma-Aldrich, St. Louis, MO, USA) to degrade contaminating RNA. Samples were incubated at 37 °C for 1 h with occasional inversions to mix. Nuclease-free proteinase K (Sigma-Aldrich) was added (50 µg) and samples were left at 55 °C for ~3 h with occasional inversion to complete cell lysis. High molecular weight gDNA was extracted from the dissolved tissue using phenol, phenol:chloroform:isoamyl alcohol (25:24:1), and chloroform treatments, with 3 min spins at 5000g for phase separation. The final aqueous phase was treated with 100% isopropyl alcohol+1/10 volume 3M sodium acetate, centrifuged at 5000g for 5 min, and the resulting DNA pellet was washed twice in 70% ethyl alcohol followed by 3-min spins at 5000g. Pellets were air-dried and resuspended overnight on a rotator in nuclease-free water (Sigma-Aldrich) at 4 °C.

Twenty microgram samples of gDNA isolated from colony WH06, WH46, and 35B were digested overnight at 37 °C using 30 units of restriction enzyme (*Bam*HI, *Eco*RI, *Pst*I; Promega) brought to 50 µl total volume with appropriate restriction buffer and nuclease-free water (Sigma-Aldrich) if necessary, followed by heat inactivation of the enzyme at 65 °C for 20 min. gDNA digests were subjected to electrophoresis overnight through 0.9% agarose gel using TBE running buffer, stained with ethidium bromide to visualize the quality of the digest run, then destained in TBE. The gel was soaked in denaturing solution (0.5N NaOH+1.5M NaCl) for 30 min, followed by a rinse in sterile water, then soaked in

neutralizing solution (1M Tris-HCl, pH8.0; 1.5M NaCl) for 30 min before a 12 h transfer to positively charged Immobilon-NY+membrane (Millipore; Bedford, MA; USA) using upward capillary transfer in 20 × SSC buffer. Nylon membranes were dried completely before crosslinking DNA using UV light (20,000 µJ/cm<sup>2</sup>).

A plasmid containing a 1060 bp *Hysy-CnPolydom* cDNA insert was used as template in polymerase chain reaction (PCR) to amplify a cDNA fragment 297 bps in length with primers designed from the first open reading frame of the pentraxin domain of *Hysy-CnPolydom* (CnPTXORF1.3s: 5'-CATTTCAGTCAAGAGTTTCACC-3' and CnPTXORF1as: 5'-AAGCCATCATATACCATCAGC-3'). The PCR product was separated in agarose gel by electrophoresis before extraction and purification using a QIAquick<sup>®</sup> spin column (Qiagen). One microgram of this purified cDNA product was labeled for 20 h with digoxigenin-11-dUTP (DIG-high prime) according to the random primed labeling technique as per manufacturer's protocol (Roche Diagnostics). Nylon membranes were prehybridized at 65 °C for 2 h in 5 × SSC (3M sodium chloride, 0.3M sodium citrate, pH 7.0), 0.1% *N*-lauroylsarcosine, 0.02% SDS and 1% blocking reagent (Roche Diagnostics) before adding the DIG-labeled DNA probe (~780 ng) and hybridizing 12–16 h under high stringency conditions at 65 °C. Post-hybridization membrane washing was conducted twice under low stringency at ambient temperature in 2 × SSC, 0.1% SDS and twice under high stringency at 65 °C in 0.5 × SSC, 0.1% SDS. Chemiluminescent detection was carried out using alkaline phosphatase-conjugated anti-DIG antibody and CSPD alkaline phosphatase substrate (Roche Diagnostics) according to manufacturer's guidelines and exposed to Biomax ML imaging film (Eastman Kodak, Rochester, NY).

### Immune challenge and *Hysy-CnPolydom* quantification by semi-quantitative RT-PCR

Overnight liquid cultures of *Bacillus subtilis*, *Vibrio* sp., and *Saccharomyces cerevisiae* were grown and the live cells were pelleted and resuspended in 50 ml of filter sterilized (.22 µm) artificial seawater (31‰) at the following densities as quantified using a hemocytometer for two experimental replicates:  $2.46 \times 10^9$  and  $7.00 \times 10^7$  cells/ml for *S. cerevisiae*,  $7.78 \times 10^9$  and  $1.05 \times 10^9$  cells/ml for *Vibrio* sp., and  $4.86 \times 10^{10}$  and  $1.045 \times 10^9$  cells/ml for *B. subtilis*. Subcloned colonies from two *H. symbiolongicarpus* individuals, 12B for replicate 1 and 35B for replicate 2, were established and cultured on glass slides in artificial seawater for six time points for each of the three microorganism types in addition to two subcloned colonies for non-challenged controls. These colonies were fasted for 4–5 days before being transferred to 50 ml tubes in sterile artificial seawater and exposed to 1 ml of a resuspended microorganism culture for 1 h, or left unexposed to any microorganism as control. At the end of the hour, colonies were transferred to fresh containers of filter sterilized artificial seawater prior to temporal RNA extractions at 0, 1, 3, 6, 18, and 48 h post-exposure.

Total RNA was extracted using TRIzol<sup>®</sup> reagent (Invitrogen) as per manufacturers instruction, precipitated using 1 part 0.8M sodium citrate, 1.2M sodium chloride to 1 part

100% isopropyl alcohol, and resuspended in nuclease-free water (Sigma-Aldrich). First-strand cDNA was then synthesized for each sample using 2.5 µg of total RNA (Amersham Biosciences, Little Chalfont Buckinghamshire, England) according to manufacturer's protocol, from which 1:10 dilutions were used as starting template in PCR amplification. Optimal cycling conditions for primer sets used to amplify the pentraxin domain (F: 5'-CATTTCAGTCAA-GAGTTTCACC-3', R: 5'-GCCTATATCGTGATGACATGATG-3') from *CnPolydom* and for  $\beta$ -actin (F: 5'-GACAATGGATCTGG-TATGTGC-3', R: 5'-GATGGCAACATACATAGCAGG-3') were empirically determined for 20 µl volume reactions overlaid with mineral oil, from which 10 µl of PCR product were loaded into 2% agarose gel stained with ethidium bromide. Once optimal primer cycling conditions were established for each primer set, these conditions were maintained for all subsequent amplifications. PCR amplification for each 1:10 cDNA dilution sample was optimized to occur just below saturation for target (*CnPolydom*) and control ( $\beta$ -actin) primer sets by adjusting the template volume, such that starting template quantity was equal for target and control PCR amplification. Both primer sets span intronic regions within their target genes, therefore any gDNA contamination present in starting template will not affect interpretation of the cDNA amplicons. Amplicon intensities of *Hysy-CnPolydom* were automatically quantified from each sample, equilibrated to  $\beta$ -actin from the same sample, using Kodak Molecular Imaging Software v. 4.0.3. Relative intensities (i.e. change in expression level after immune challenge) were determined by subtracting the equilibrated intensity of non-challenged *Hysy-CnPolydom* amplicons from post-challenge *Hysy-CnPolydom* amplicons.

### In situ hybridization

A 523 bp cDNA was amplified from *Hysy-CnPolydom* (using the same primers targeting the pentraxin domain as described above for semi-quantitative PCR) and cloned into pGEM-T Easy vector (Promega) from which sense and antisense digoxigenin (DIG)-labeled RNA probes were generated using a DIG RNA labeling kit (Roche Diagnostics) by *in vitro* transcription with SP6 and T7 RNA polymerase according to the manufacturer's protocol. A second set of primers were designed from *Hysy-CnPolydom* to amplify a 326 bp cDNA from the coagulation factor 5/8 domain (F: 5'-CTAGCAAGGCTCTATTCAACACC-3', R: 5'-CGTCTGCTT-CTACTGTCGTGTCGG-3') that was similarly used to generate riboprobes. Whole mount *in situ* hybridizations were performed based on [52] and as described previously [9]. Hybridized specimens were prepared for sectioning by embedding the tissue in 1% agarose/0.85% NaCl dissolved in phosphate buffered saline (PBS) followed with consecutive washes to replace the saline solution with alcohol: 0.85% NaCl saline solution for 30 min at 4 °C, 1:1 saline/100% EtOH for 15 min at room temperature (rt), 70% EtOH  $\times$  2 for 15 min each at rt, 85% EtOH for 30 min at rt, 95% EtOH for 30 min at rt, 100% EtOH  $\times$  2 for 30 min each at rt, 100% xylene  $\times$  2 for 30 min each at rt, 1:1 100% xylene/paraffin for 45 min at 60 °C, 100% paraffin  $\times$  3 for 20 min at 60 °C. Specimens were embedded in the third paraffin wash and allowed to solidify at 4 °C overnight. Specimens were

sectioned to 12 µm using a Microm HM325 microtome and mounted in euparal mounting resin (BioQuip Products, Inc., Rancho Dominguez, CA) for microscopic analysis. Mounted specimens were examined using differential interference contrast (DIC) transmitted light microscopy with a Zeiss Axioskop2 mot *plus* microscope and images were taken using a Zeiss AxioCam HRC digital camera and AxioVision (release 4.5) imaging software.

### GenBank accession numbers

The 78.8 kb sequence fragment encoding the *Hysy-CnPolydom* gene obtained via shotgun sequencing of *H. symbiolongicarpus* BAC clone 2N9 has been annotated and cataloged in GenBank under EF566888. *H. symbiolongicarpus* cDNA sequence and predicted protein translation information presented here for *CnPolydom* has been deposited to GenBank under EF566886. Nucleotide and protein translation sequence for the pentraxin domain of *Hycy-CnPolydom* isolated from sister species *H. echinata*, has been deposited to GenBank under EF566887.

## Results

### Identification of cnidarian polydom

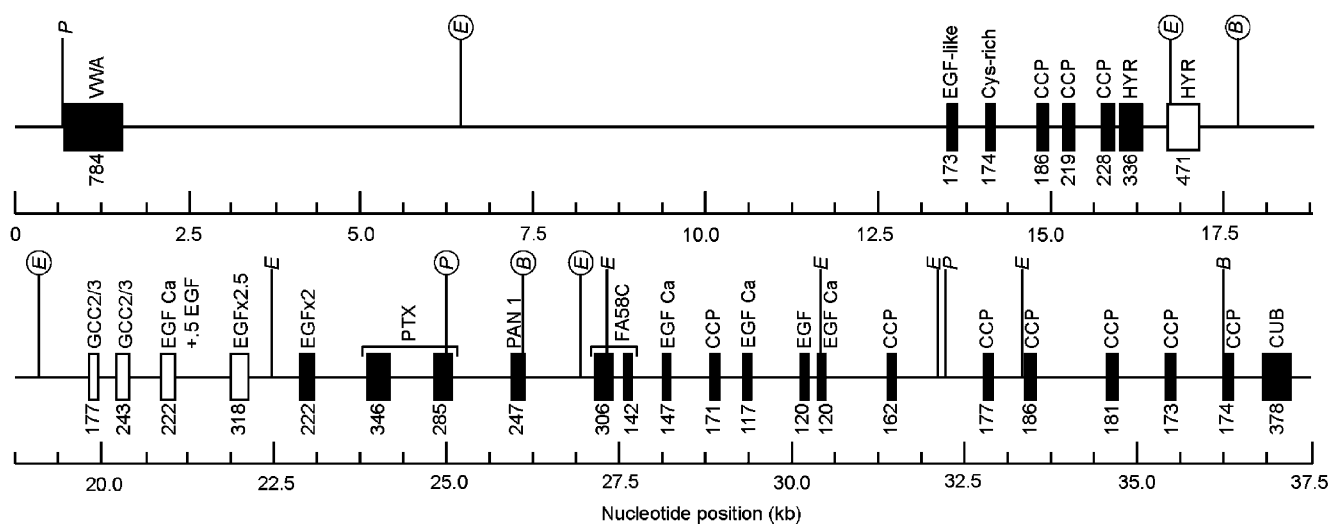
Random end sequencing of *H. symbiolongicarpus* genomic BAC clones was performed to examine their genome in an unbiased way for genes of potential immunologic relevance. In this manner a sequence was identified and predicted to encode an entire 209aa pentraxin domain with highest identity to vertebrate polydom pentraxins (43% identity to *Gallus gallus* predicted protein similar to polydom; XP\_424917;  $E = 8e-42$ ). Pentraxin molecules are well known for their role in immune responses of both invertebrates and vertebrates. We therefore pursued sequencing this region of the genome and quickly found the pentraxin domain, encoded by two ORFs, was in close proximity to other ORFs in the genome encoding additional predicted ligand binding domains, including a PAN 1 and a coagulation factor 5/8 C-terminal (FA58C) domain, as well as EGF domains. Pentraxin molecules contain a single pentraxin domain and are not typically associated with additional domain types. Only two molecules from the pentraxin family have been described as exceptions: (1) polydom from mouse and human [33], which contains 7 domain types associated with a single pentraxin domain and (2) the pentraxin fusion protein described from African clawed frog [53] which contains one eel-Fucolelectin Tachylectin-4 Pentraxin-1 (FTP) lectin domain in tandem with a single pentraxin domain. Though neither of these known molecules contain PAN or FA58C domains as we found in *Hydractinia*, polydom does contain EGF repeats in tandem with the pentraxin domain. We therefore hypothesized that this was a similarly large, polydom-like gene.

Our approach to determine gene architecture was to sequence the *H. symbiolongicarpus* genomic region encoding this polydom-like gene. One quarter of the BAC library (18,432 clones) was screened by hybridizing BAC library membranes with a 414 bp DIG-labeled pentraxin cDNA probe. Nine pentraxin-positive BAC clones were identified,

one of which (2N9) was chosen for shotgun sequencing. The shotgun approach generated 101 kbs of sequence fragments, from which 6 major contigs were formed and subsequently joined by genome walking to obtain a final 78,155 bp contiguous sequence fragment from the *H. symbiolongicarpus* genome that we have cataloged in GenBank under EF566888. The gene was predicted to be encoded by 30 exons across a ~37.5 kb region (Figure 1), comprised of 12 different predicted domain types. Using the blastp algorithm [54] this gene was found to share highest identity with the polydom gene sequence from human ( $E = 1e-99$ ; EMBL CAH74139) and mouse ( $E = 5e-98$ ; GenBank AAG32160). Thus we have named the molecule cnidarian polydom (CnPolydom), and specifically identify the molecule from *H. symbiolongicarpus* as Hysy-CnPolydom. Using primers designed within predicted ORFs of the genomic sequence, a 1656 bp 5' and 3366 bp 3' region of *Hysy-CnPolydom* were amplified via RT-PCR, cloned, and sequenced. With these cDNAs (GenBank EF566886), transcription of 25 out of the 30 predicted exons were confirmed (Figure 1) and the predicted amino acid translations from these cDNAs are shown in Supplementary Figure 1.

Domain prediction analysis was used to identify domains encoded within *Hysy-CnPolydom* (Figure 1), revealing a gene domain architecture sharing important similarity in organization and composition to that described for mouse polydom [33]. Beginning at the N-terminus, SignalP v3.0 (<http://www.cbs.dtu.dk/services/SignalP/>) predicted a 17aa signal peptide (MKVGFLLNLLFLFLTCG) from the first exon that was also predicted to encode an entire VWA domain. Using multiple domain alignments (not shown) a perfect metal ion-dependent adhesion site (MIDAS) was identified within the VWA domain following the pattern D-x-S-x-S...T...TDG [55] (shown in Suppl. Figure 1), which is known to play important roles in mediating diverse ligand binding [56]. Such a MIDAS motif was also identified from the VWA domain of polydom [33]. Following the VWA domain was an

unclassified subfamily member of EGF domains, a single 35aa EGF-like domain encoded within exon 2. No known domains were strongly identified from the 58 residues encoded in the third exon, though it is cysteine-rich with 6 cysteine residues and thus we identified it as Cys-rich domain. The cysteine composition is typical of type 1 EGF domains and the exon showed closest similarity to EGF ( $7.05e+02$ ) and tumor necrosis factor receptor (TNFR) ( $2.47e+03$ ) domains. Gilgès et al. (2000) identified similar low-identity EGF and TNFR features from the corresponding region of mouse polydom [33]. Next were three CCP domains, also known as sushi or short consensus repeat (SCR) domains, each encoded within separate exons (4–6). Proteins that regulate the complement system of vertebrates are entirely or largely comprised by, and thus named for, the CCP domain [57]. Following the three CCP repeats were two hyalin repeat (HYR) domains. HYR domains were first denominated by Gilgès et al. (2000) when describing polydom [33], though they named it for the hyalin protein that is comprised solely of these repeating domains, and was subsequently described as having an immunoglobulin-like fold and involvement in cellular adhesion [58]. We confirmed a portion of the first HYR from exon 7 via cDNA, while the second HYR was predicted from the eighth exon within the genomic sequence. Following the HYRs, four additional exons (9–12) were predicted from the 6704 bp genomic region falling between the 5' and 3' cDNAs. Predicted exons 9 and 10 each encoded a domain containing five conserved cysteine's, known as GCC2/3 domain. Mouse polydom contained 3 such tandem domains, noted for their similarity to thyroglobulin type 2 repeats [33]. Following the GCC2/3 repeats, a total of 6 EGF domain family members were predicted across 3 exons. Predicted exons 11 and 12 contained one EGF-Ca<sup>2+</sup> binding domain and three EGF domains. Following these begins the 3' region of the gene that was confirmed by cDNA. Exon 13 encoded the additional two EGF domains. A 209aa domain encoded within exons 14



**Figure 1** Genomic organization of *CnPolydom* isolated from *Hydractinia symbiolongicarpus*. Filled boxes represent exons confirmed by cDNA while open boxes represent predicted exons. Predicted domain identities are given above the corresponding exons and exon sizes (bp) are given below. Sequence from a single BAC clone (2N9) was used to obtain the 37.5 kb region of the *H. symbiolongicarpus* genome represented here. Restriction sites are indicated as B = BamHI, E = EcoRI, P = PstI with relevant sites for Southern hybridization analysis circled.

and 15 was unequivocally identified as a pentraxin domain ( $E = 8.40e-31$ ). The characteristic pentraxin domain motif H-x-C-x-[S/T]-W-x-[S/T] identified by PROSITE [41], used to specifically identify protein sequence belonging to the pentraxin family, was identified within this domain and is indicated in the *Hysy-CnPolydom* cDNA (Suppl. Figure 1) where all but the cysteine, which was replaced by alanine, are conserved. A substitution of cysteine with alanine was also noted within the pentraxin motif of mouse polydom [33]. A 79aa region was predicted from exon 16 to be a PAN domain module ( $E = 1.50e-04$ ), abbreviated using previous domain names and molecules it was known from including plasminogen/hepatocyte growth factor, apple domains of prekallikrein/coagulation factor XI family, and various nematode proteins [59]. The PAN domain from *Hysy-CnPolydom* had four conserved cysteines typical of the PAN domains from plasminogen and hepatocyte growth factor proteins, and was most similar to PAN 1 of this domain clan as determined by Pfam, thought to be the ancestral form of the domain from which PAN 2 and PAN 3 are derived [60]. Following the PAN domain were 147aa's predicted to be encoded by exons 17 and 18, of which 119 toward the 3' end were predicted to be similar to FA58C domain, also known as discoidin domain ( $E = 2.00e-07$ ). Though FA58C domains are typically about 150 amino acids, which is approximately the size predicted within these two exons of *Hysy-CnPolydom*, the first 27 amino acids toward the 5' region were not conserved and shared no significant homology to any described domain type. However, this 147 amino acid region did encode cysteine residues at each end, as is typical of FA58C domains to form a disulfide bond linking its ends together [61]. Multiple EGF and CCP domains were predicted downstream of the FA58C domain, each encoded within separate exons. Exons 19, 21, and 23 were found to encode calcium-binding EGF domains while exon 22 encoded a non-calcium-binding EGF domain. Exons 20 and 24–29 were each found to encode CCP modules. The C-terminal portion of mouse polydom was similarly shown to be comprised of multiple repeating CCP and EGF domains [33]. A final 102aa domain type was predicted from exon 30 as a CUB domain ( $E = 1.54e-01$ ), named for its occurrence in complement proteins C1r/C1s, urchin EGF-like protein, and bone morphogenetic protein-1 [62]. Twenty-seven residues from this predicted domain were confirmed in the 3' cDNA (Suppl. Figure 1). The CUB domain from *Hysy-CnPolydom* contained only one of the four conserved cysteine residues typical of this domain type. Without these cysteine residues the formation of intra-domain disulfide bridges may not occur, thus the typical beta-barrel structure is unlikely to be formed. A stop codon is predicted at position 54,190 bp in the genomic sequence (EF566888) at the end of exon 30.

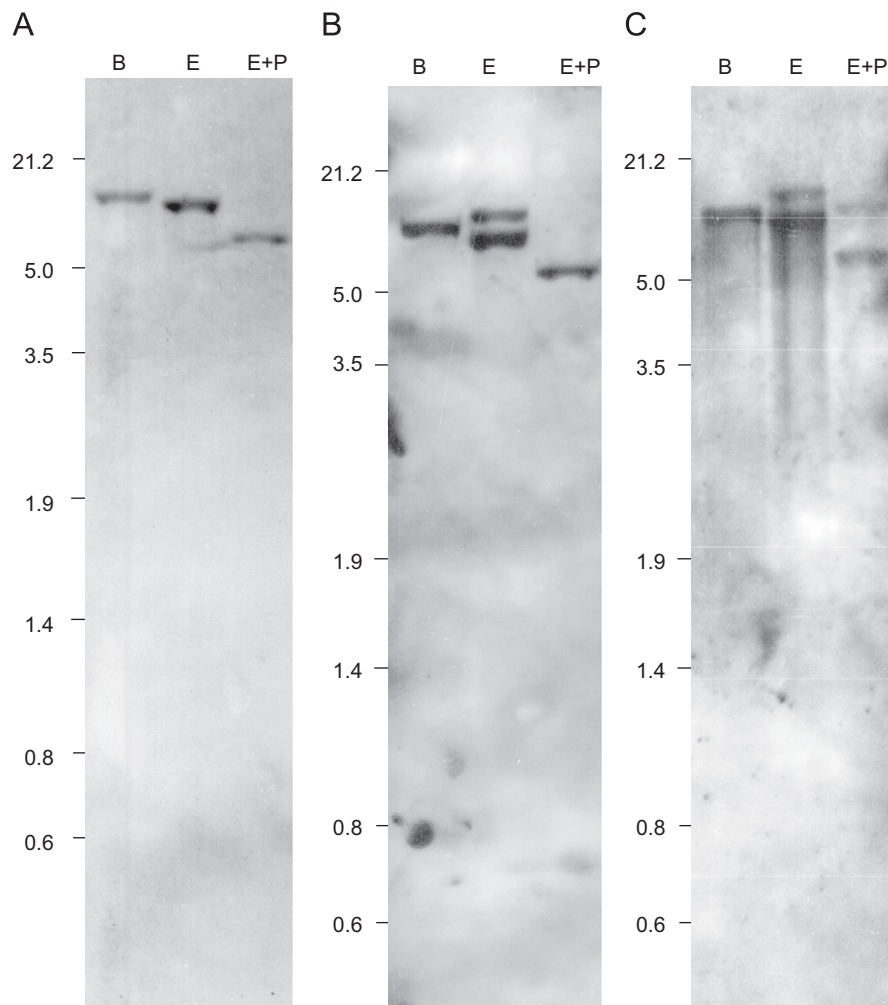
### Hysy-CnPolydom is a single copy gene

The number of *CnPolydom* genes occurring in the *H. symbiolongicarpus* genome was analyzed by Southern hybridization of genomic DNA isolated from 3 different individuals (Figure 2). Genomic DNA was digested using *Bam*HI, *Eco*RI, or *Eco*RI and *Pst*I and subsequently probed using a 297bp region of the first exon encoding the

pentraxin domain. Predicted restriction sites for these enzymes are indicated in Figure 1 based on the genomic sequence of individual 4117-2 used to generate the BAC library, with sites relevant to this Southern hybridization analysis circled. We were not able to perform Southern hybridization analysis on individual 4117-2, however, because this colony died out prior to these experiments. *Bam*HI was predicted to cut at positions 17,701 and 26,127 and generate an 8425bp fragment that would hybridize to the probe. All three individuals showed a single hybridizing band of this expected size (~8.5 kb) from the *Bam*HI digests, consistent with a single *CnPolydom* locus. *Eco*RI was expected to result in a hybridizing fragment of 4421 bp from *Eco*RI restriction sites 22,471 and 26,893, yet none of the three individuals show a band of this size. The restriction site at position 22,471 may be unique to individual 4117-2, explaining the absence of a ~4.4 kb fragment in the *Eco*RI digests of the three individuals analyzed in Figure 2. The next downstream *Eco*RI restriction site at position 19,067 along with site 26,893 would generate a 7825 bp-hybridizing fragment. All three individuals confirmed a hybridizing fragment of this expected size (~7.8 kb) in the *Eco*RI digests. Individual WH06 (Figure 2a) showed probe hybridization to this ~7.8 kb fragment only, consistent with a single *CnPolydom* locus. However, individuals WH46 and 35B (Figure 2b and c) showed probe hybridization to a second fragment at ~10 kb, which may be due to heterozygosity at the *CnPolydom* locus, to a second gene, or to inefficient gDNA enzyme digests. With regard to the later, a 10,201 bp fragment may have been generated if the enzyme cut only at *Eco*RI sites 6478 and 26,893. Finally, a double digest using *Eco*RI and *Pst*I was predicted to cut at positions 24,964 (*Pst*I) and 22,471 (*Eco*RI) resulting in a hybridizing band of 2492 bp. Again, none of the three individuals showed hybridization to a band of this size. Because the *Pst*I restriction site at position 24,964 occurred within an intron, we were able to confirm its presence in the cDNA of individual 35B. Therefore, if we again assume the *Eco*RI restriction site at position 22,471 was unique to individual 4117-2, the next *Eco*RI site at position 19,067 would result in a 5896 bp fragment that should hybridize with the probe. Indeed all three individuals showed a hybridizing band at ~6 kb, which was the only fragment that hybridized in the double digest from individuals WH06 and WH46 (Figure 2a and b). Individual 35B (Figure 2c) had a second hybridizing fragment at ~9 kb which cannot be explained using additional predicted *Eco*RI or *Pst*I restriction sites from the genome of individual 4117-2. Therefore, this band may have been due to an unknown restriction site within 35B or to a second locus. Taken together, these data are most consistent with a single-locus model for the *CnPolydom* gene in *H. symbiolongicarpus*.

### CnPolydom is part of a polydom-specific clade of pentraxin molecules

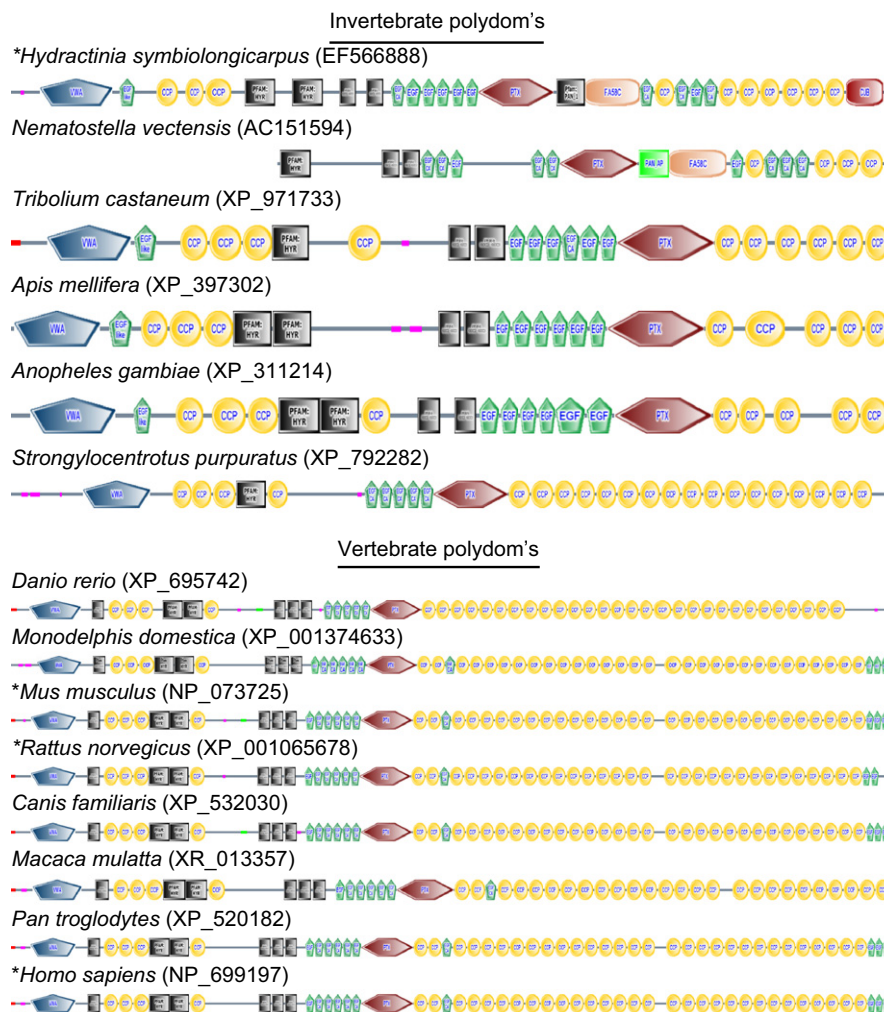
BLAST searches were conducted using *H. symbiolongicarpus* *CnPolydom* sequence in addition to mouse *polydom* and human *polydom* orthologue sequence to identify an additional 13 predicted polydom sequences from public sequence databases. The predicted domain composition of



**Figure 2** Southern blot analysis of *CnPolydom* in 3 individuals of *H. symbiolongicarpus*. Twenty micrograms of genomic DNA from individuals WH06 (a), WH46 (b), and 35B (c) were each digested with *Bam*HI (B), *Eco*RI (E), or both *Eco*RI and *Pst*I (E+P) and probed for *Hysy-CnPolydom* using a 297 bp DIG-labeled region from the first exon encoding the pentraxin domain under high stringency conditions. DNA size marker positions (Roche Diagnostics) run alongside the digests are indicated in kilobases.

these sequences are presented in [Figure 3](#) and highlight the highly conserved gene domain architecture in composition and order from both invertebrate and vertebrate taxa. The two cnidarian sequences, *H. symbiolongicarpus* and *Nematostella vectensis* (starlet sea anemone) are the only polydom-like sequences to contain additional domain types: PAN, FA58C, and CUB. To determine the relationship of *CnPolydom* to all the other members of the pentraxin family, pentraxin domains were retrieved from 33 sequences including neuronal, CRP, SAP, and the known and predicted polydom-like sequences presented in [Figure 3](#), and aligned as amino acid translations with the *CnPolydom* pentraxin domains we isolated from *H. symbiolongicarpus* (*Hysy-CnPolydom*) and *H. echinata* (*Hyec-CnPolydom*) ([Figure 4](#)). The pentraxin domain isolated from *H. echinata* was 99% identical to *H. symbiolongicarpus* and the two are likely orthologs (see also [Figure 5](#)). Sequence identity across the complete pentraxin domain from *H. symbiolongicarpus* *CnPolydom* was most similar to other polydom-like pentraxins including those from chimpanzee, human, chicken, and mouse (42%), starlet sea anemone, zebrafish, Rhesus

monkey, dog, rat, and opossum (41%). Of all the polydom-like pentraxins identified, those from protostome taxa had lowest identity to *Hysy-CnPolydom* pentraxin: from 30% (honey bee) to 36% (tarantula spider). Outside of other polydom-like pentraxin domain sequences, *Hysy-CnPolydom* pentraxin was most similar across the entire predicted domain to the murine neuronal pentraxins I (31%) and II (30%). Three other cnidarian taxa including the stony coral *Acropora millepora*, the brown sea anemone *Metridium senile*, and the freshwater hydroid polyp *Hydra magnipapillata* appeared to encode isolated pentraxin domains, as no other domains were identifiable within the available sequence data. Sequence identity across the available sequence was highest from *H. magnipapillata* (59%), belonging to the same taxonomic class (Hydrozoa) as *Hydractinia*, followed by the anthozoan class members *M. senile* (39%) and *A. millepora* (27–30%). However, *H. magnipapillata* and *M. senile* sequences provided only partial pentraxin domain sequence information (see [Figure 4](#)) and may cause spurious identity scores to the complete *Hydractinia* sequence.



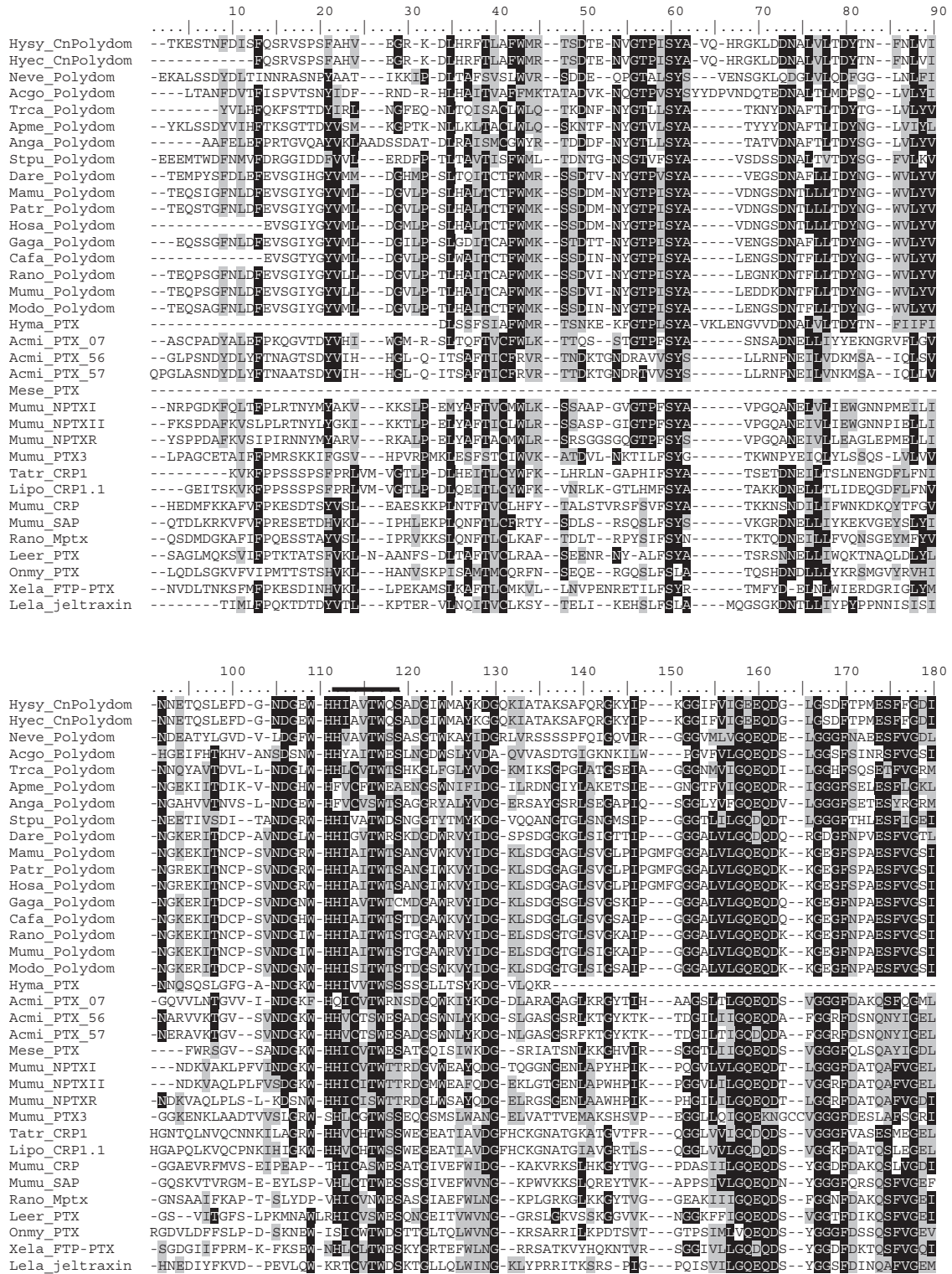
**Figure 3** Predicted domain composition of CnPolydom from *Hydractinia symbiolongicarpus* presented with polydom-like molecules from other metazoan taxa. Polydom molecules have only been described previously from mouse, rat, human, and here *Hydractinia* (indicated by \*); all others are predicted molecules retrieved from public sequence databases. Note the absence of transmembrane regions from all of these molecules. Images of domain composition were generated using SMART (<http://smart.embl-heidelberg.de/>). Accession numbers for the sequences are indicated (see also Suppl. Table 1).

Phylogenetic analyses of the pentraxin domain were performed using the amino acid sequences presented in Figure 4 and additional pentraxin sequences from known and predicted pentraxin molecules. A laminin G domain, the closest protein domain family to the pentraxins [63], retrieved from human neurexin-1-alpha precursor sequence was used to root the tree, having 11% identity to *Hysy-CnPolydom* pentraxin. A representative tree generated by NJ method is presented in Figure 5. Pentraxin molecules fell into five dominant clades: (1) the short pentraxins from vertebrates, including CRP and SAP, (2) the long pentraxin 3, (3) the CRP- and SAP-like short pentraxins from three horseshoe crab genera *Limulus*, *Tachypleus*, and *Carcinoscorpius*, (4) the long neuronal pentraxins, and (5) the polydom pentraxin clade, within which CnPolydom fell (bootstrap value = 78). The *Hydractinia* CnPolydom's clustered most closely with the partial pentraxin sequence from *H. magnipapillata* (bootstrap value = 88) and *N. vectensis* polydom pentraxin (bootstrap value = 73). The phylogenetic

position of the partial pentraxin sequence from the anthozoan *M. senile* was unclear and while the three pentraxin sequences from *A. millepora* were supported as a cluster (bootstrap value = 57), their phylogenetic position among the other pentraxins was also unclear.

### Hysy-CnPolydom transcription is upregulated after microbial challenge

Pentraxin family members are well established as immune-relevant genes substantially upregulated as part of the acute-phase response [65]. We therefore examined the transcript level of *Hysy-CnPolydom* from *H. symbiolongicarpus* in response to 1 h exposures to *Bacillus subtilis* (Gram positive bacteria), a marine *Vibrio* sp. (Gram negative bacteria), or *Saccharomyces cerevisiae* (fungus) from immediately after the 1 h challenge (time 0) to 48 h post-challenge (Figure 6a). Under normal conditions (non-challenged), we



**Figure 4** Alignment of pentraxin domains (as determined by SMART) from *polydom*-like genes and other pentraxin sequences across metazoan taxa. Alignment of amino acid translations was generated using MUSCLE (multiple sequence alignment software) [48]. Highlighted residues correspond to a 48% threshold for identity (black background) and similarity (gray background). Percent sequence identity to *Hydractinia symbiolongicarpus* is indicated at the end of each sequence. The signature pentraxin motif is indicated by the heavy bar. Sequences are referenced by the first two letters of the genus and species and the molecule type from which the domain was isolated. Accession numbers for these sequences and common names for taxa can be found in Suppl. Table 1. Invertebrates: *Hydractinia symbiolongicarpus*, *Hydractinia echinata*, *Nematostella vectensis*, *Acanthoscurria gomesiana*, *Tribolium castaneum*, *Apis mellifera*, *Anopheles gambiae*, *Strongylocentrotus purpuratus*, *Hydra magnipapillata*, *Acropora millepora*, *Metridium senile*, *Tachypleus tridentatus*, *Limulus polyphemus*. Vertebrates: *Leucoraja erinacea*, *Danio rerio*, *Anorchynus mykiss*, *Xenopus laevis*, *Lepidobatrachus laevis*, *Gallus gallus*, *Rattus norvegicus*, *Mus musculus*, *Canis familiaris*, *Macaca mulatta*, *Pan troglodytes*, *Homo sapiens*.

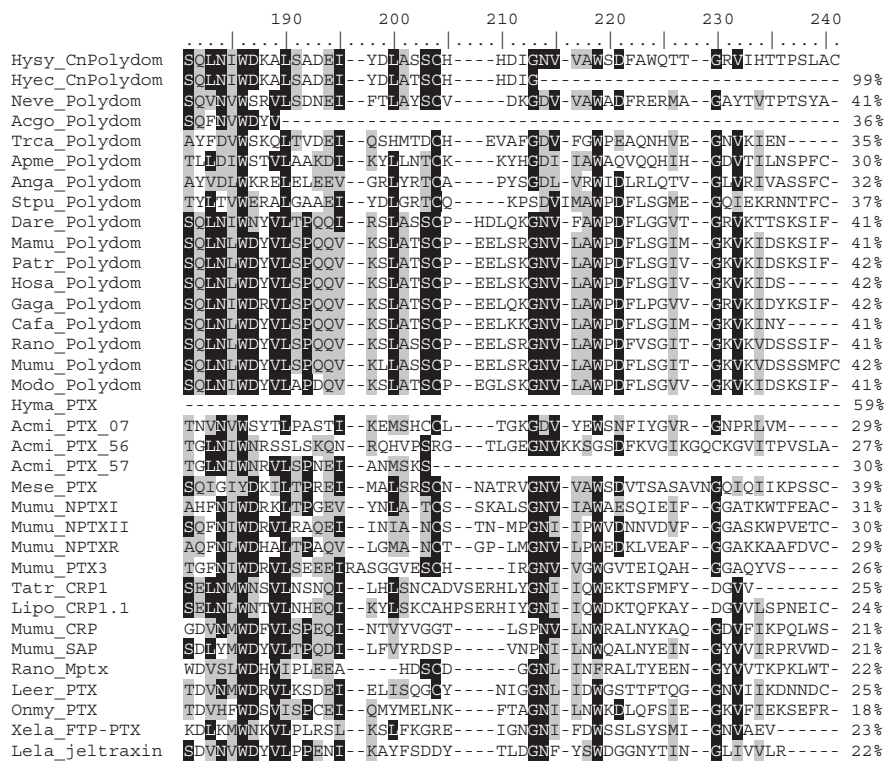


Figure 4 (Continued)

found *Hysy-CnPolydom* to be constitutively expressed at very low levels (Figure 6a and data not shown). After microbial exposure, all challenges induced *Hysy-CnPolydom* transcription above its constitutive level from non-challenged control RNA for at least 48 h, with response to *S. cerevisiae* resulting in the most substantial increase from 1 h to 48 h post-challenge (Figure 6b). Within the 48 h time we assessed, highest transcript levels were seen at 18 h post-challenge for both *S. cerevisiae* (8-fold increase) and *B. subtilis* (3-fold increase), and at 48 h post-challenge for *Vibrio* sp. (2-fold increase).

### Cellular expression of Hysy-CnPolydom

*In situ* hybridization analysis of *Hysy-CnPolydom* showed this transcript was expressed by a distinct subset of cells distributed throughout the body column of adult gastrozooids (feeding polyps), including the tentacles and hypostome, and was not localized to any particular region of the polyp (Figure 7a through c). We determined that some of these cells were rounded to ovoid in shape and ~4–5 μm in diameter seen in Figure 7d, while others had a distinct linear shape and were ~7 μm long by ~1 μm wide as shown in Figure 7e. These cells were smaller than the large, cuboidal ectodermal epithelial cells, and occurred in the interstices of the epithelial cells. Cross-sections of polyps revealed that the cells expressing *Hysy-CnPolydom* were restricted to the ectodermal layer, as shown in Figure 7f and g. The cell morphology, size, and location was consistent with neural cells of the interstitial stem cell lineage and with the undifferentiated interstitial stem

cells themselves as described from several hydrozoans including *Hydractinia* [66–68], *Pennaria* [69,70], and *Hydra* [12,71,72].

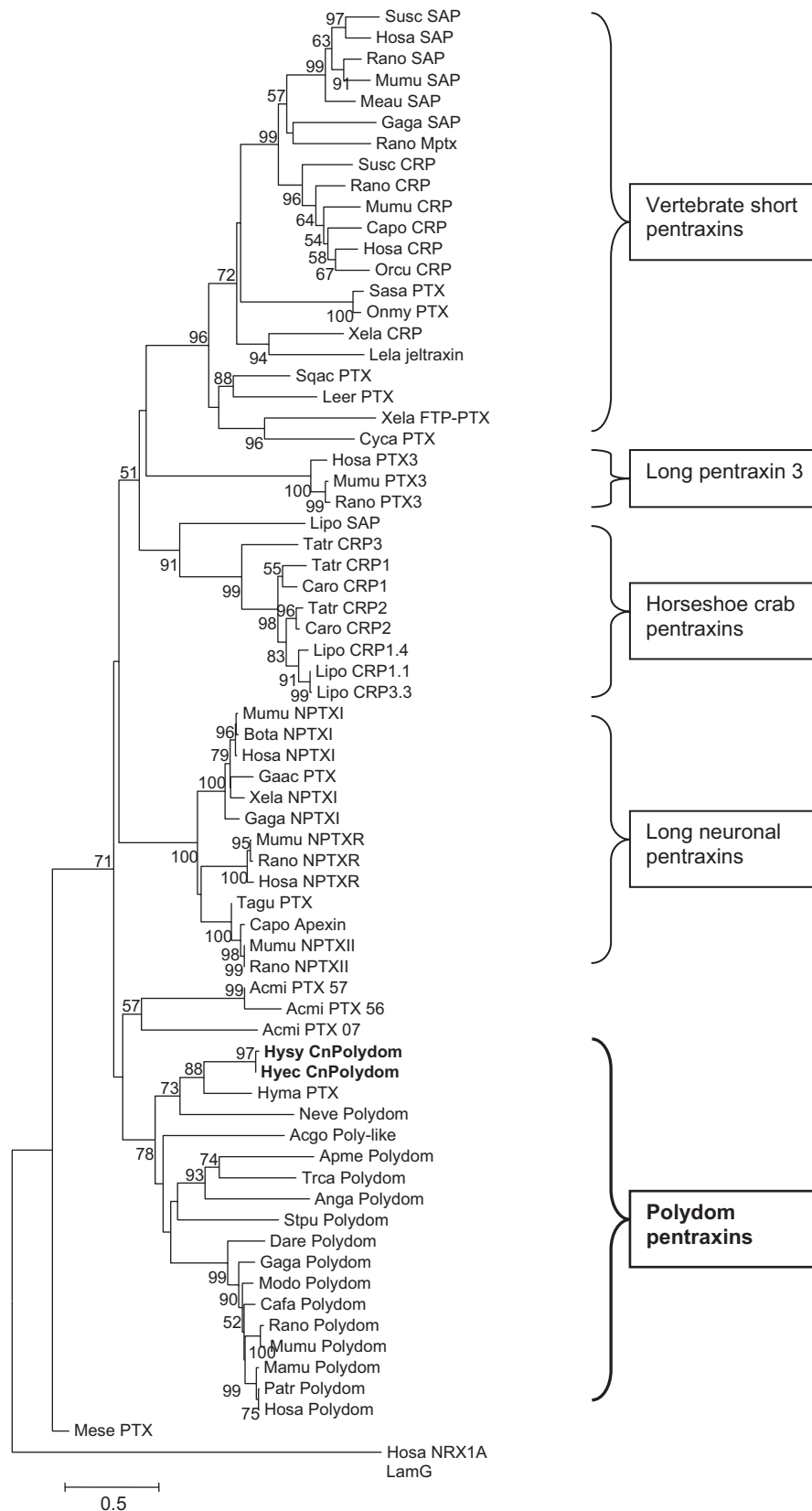
## Discussion

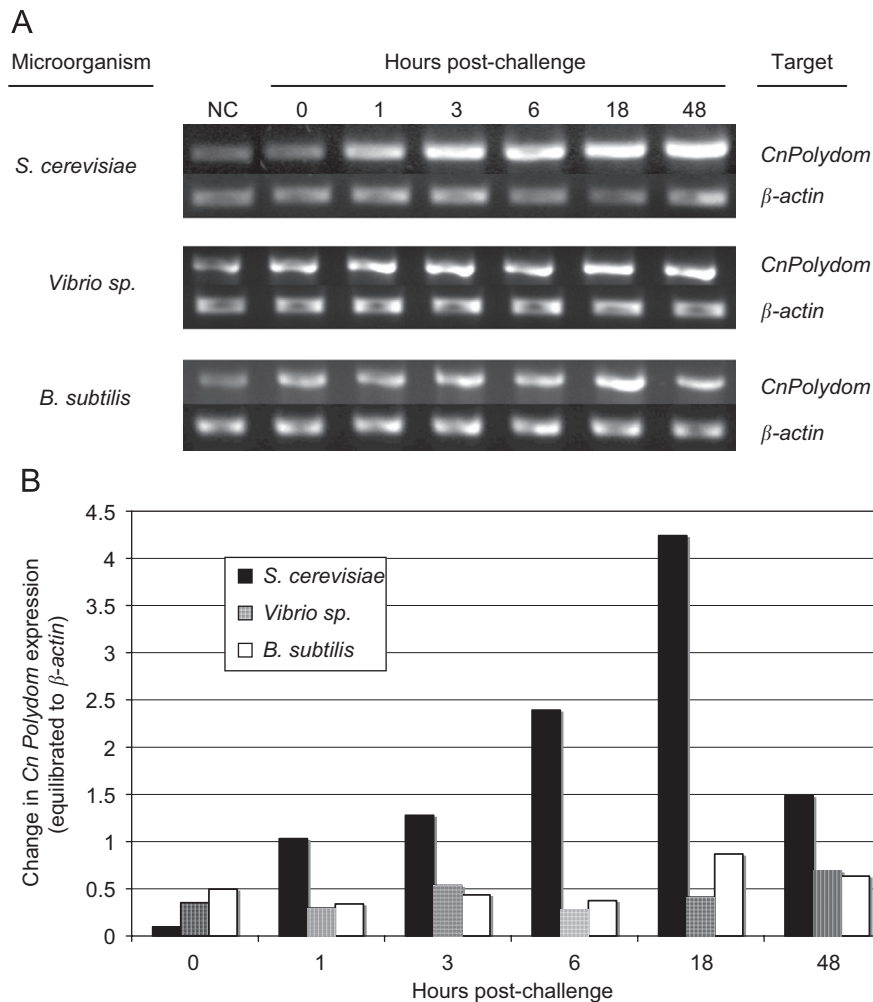
### The novel, modular molecule CnPolydom

Though little is known about the composition of immune genes in Cnidaria, our approach using random end sequencing of BAC clones revealed a multidomain gene sequence in the genome of *Hydractinia symbiolongicarpus* containing multiple immune-related domains including VWA, PTX, CCP, PAN, FA58C, EGF, and CUB. Based on domain composition, sequence and phylogenetic analyses, we determined that this gene is a homolog to the described *polydom* gene from mouse and human (a.k.a. *SVEP1*), and indeed homologous to additional predicted polydom-like sequences from across the Eumetazoa, thusly named *cnidarian polydom* (*CnPolydom*). Though we identified all domain types as described from polydom [33] in an identical order, *Hysy-CnPolydom* uniquely possessed three additional predicted domains: PAN, FA58C (discoidin), and CUB. These domains may offer novel structure and biological function to the cnidarian protein not found in polydom or other polydom-like homologs. The PAN domain is a common domain type in a variety of modular proteins, including coagulation factor protein XI and prekallikrein where it was previously known as the apple domain [59]. The PAN domain has been well established for its direct involvement in the binding ability of proteins including plasminogen [73], hepatocyte growth

factor [74], and coagulation factor XI [75]. The prototype FA58C (“discoidin”) domain was originally described from the discoidin adhesion protein of the slime mold *Dictyoste-*

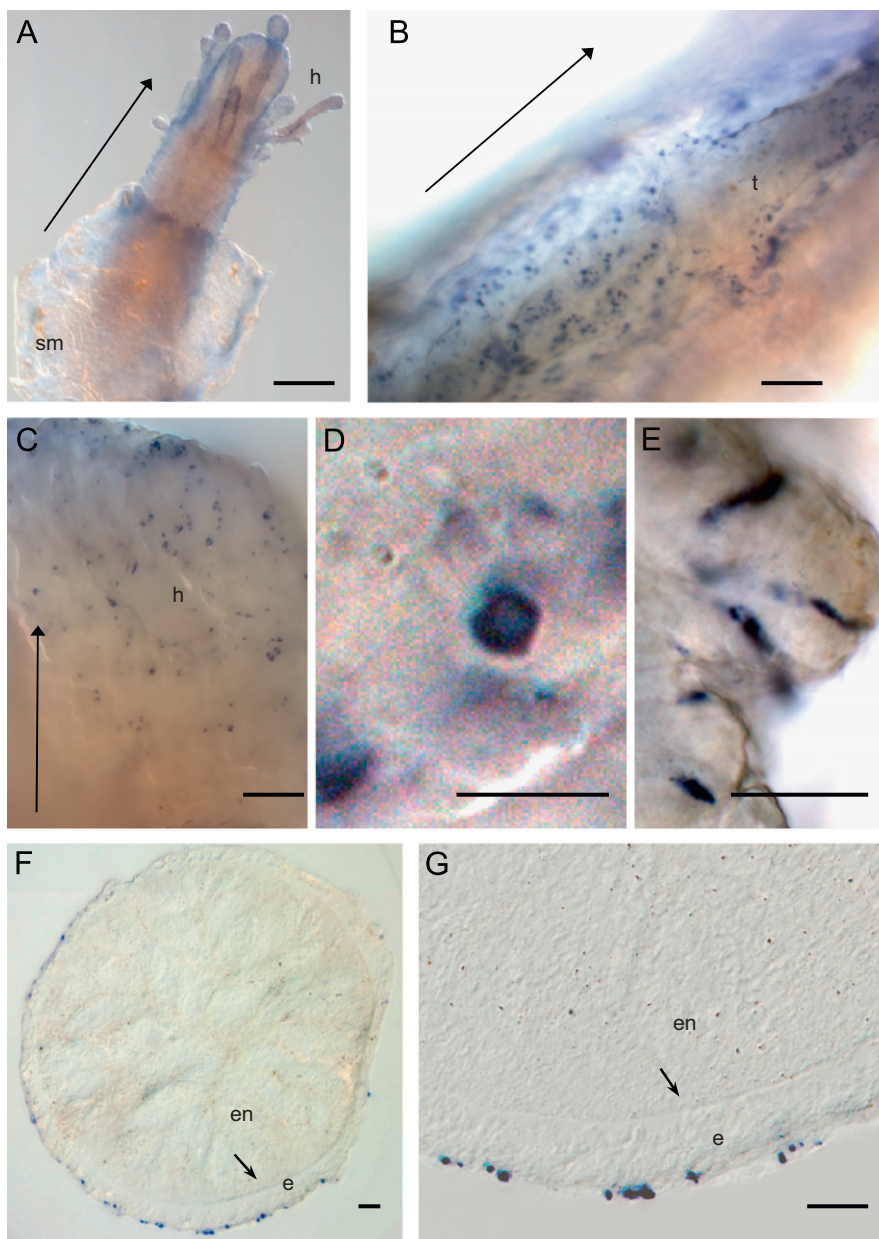
*lium discoideum* [76] and separately from the carboxyl terminus of coagulation factor V and VIII genes [77]. This domain has since been found in a variety of mostly





**Figure 6** Semi-quantitative RT-PCR analysis of *CnPolydom* from microorganism-challenged *H. symbiolongicarpus*. Colonies were exposed for 1 h to either yeast (*Saccharomyces cerevisiae*), Gram-negative bacteria (*Vibrio sp.*), or Gram-positive bacteria (*Bacillus subtilis*), followed immediately by total RNA extraction (time 0) and out to 48 h post-exposure extraction. (a) RT-PCR was performed on first strand cDNA synthesized from the total RNA extractions using primers targeting the pentraxin domain of *Hysy-CnPolydom*. Amplification of  $\beta$ -actin was performed in tandem as a loading control. (b) The change in *Hysy-CnPolydom* expression following challenge experiments relative to non-challenged control and equilibrated against the intensity of  $\beta$ -actin transcript amplification from the corresponding template are tabulated using amplicon intensity calculations from (a). These data are representative of two completely independent experimental replicates.

**Figure 5** Phylogenetic analysis of polydom and other pentraxin molecules, showing the 5 dominant clades that were consistent through multiple phylogenetic reconstruction methods. *CnPolydom* pentraxin domains from *H. symbiolongicarpus* and *H. echinata* nested within the polydom clade and are in bold. Pentraxin domains from amino acid sequences were isolated based on SMART identification [44] and aligned using MUSCLE [48]. Phylogeny reconstruction was performed using MEGA v4 [49]. The tree was generated by Neighbor-Joining distance-based method [64] with Jones-Taylor-Thornton (JTT) matrix using uniform rates among sites and pairwise deletion of gaps. Robustness of the tree was tested using 1000 rounds of bootstrapping; only values  $\geq 50$  are shown. 248 sites were used in 69 sequences. The scale references branch lengths that are proportional to the number of amino acid substitutions in a sequence. The tree is rooted using a laminin G domain sequence as an outgroup, isolated from human neurexin 1 alpha (NRX1A). PTX = pentraxin domain of undefined type, CRP = C-reactive protein, SAP = serum amyloid P, Mptx = mucosal pentraxin, FTP-PTX = eel-fucolelectin tachylectin-4 pentraxin-1 domain fusion protein, PTX3 = pentraxin 3, NPTX = neuronal pentraxin (I or II), NPTXR = neuronal pentraxin receptor. The first two letters of genus and of species are used for taxon abbreviation (see also Suppl. Table 1). Invertebrates: *Hydractinia symbiolongicarpus*, *Hydractinia echinata*, *Hydra magnipapillata*, *Nematostella vectensis*, *Metridium senile*, *Acropora millepora*, *Tribolium castaneum*, *Acanthoscurria gomesiana*, *Apis mellifera*, *Anopheles gambiae*, *Tachypleus tridentatus*, *Limulus polyphemus*, *Carcinoscorpius rotundicauda*, *Strongylocentrotus purpuratus*. Vertebrates: *Leucoraja erinacea*, *Squalus acanthias*, *Gasterosteus aculeatus*, *Danio rerio*, *Oncorhynchus mykiss*, *Cyprinus carpio*, *Salmo salar*, *Xenopus laevis*, *Lepidobatrachus laevis*, *Gallus gallus*, *Taenioptygia guttata*, *Oryctolagus cuniculus*, *Mesocricetus auratus*, *Cavia porcellus*, *Monodelphis domestica*, *Rattus norvegicus*, *Mus musculus*, *Sus scrofa*, *Bos taurus*, *Canis familiaris*, *Macaca mulatta*, *Pan troglodytes*, *Homo sapiens*.



**Figure 7** *In situ* hybridization analysis of *Hysy-CnPolydom* transcripts. Cells expressing *Hysy-CnPolydom* transcripts are stained purple. Whole-mounts of gastrozooid polyps at  $100\times$  (a) and  $1000\times$  (b–e) magnification. Arrow orients the oral–aboral axis of the polyp in (a–c), with the oral end in the direction of the arrowhead. The hypostome (h) region of the gastrozooid and a portion of the translucent stolonal mat (sm) can be seen in (a). The base of a tentacle (t) extends out of the viewing plane from the polyp in (b). Image (c) is focused on the hypostome (h) region (also labeled in (a)). Cell morphology is shown in enlarged images (d) and (e). Twelve micrometer sections of gastrozooid polyp at  $400\times$  (f) and  $1000\times$  (g). The mesoglea (arrow in (f) and (g)) separates ectodermal (e) from endodermal (en) tissue layers. The sense RNA probe produced no hybridization signal. Scale bars =  $100\mu\text{m}$  (a) and  $10\mu\text{m}$  (b–g).

extracellular molecules where it mediates cell adhesion via phospholipid binding [78,79] and carbohydrate binding [80]. FA58C domain has previously been identified in immune molecules of invertebrates including hemocytin from the silk moth *Bombyx mori* [81] and its homolog hemolectin in the fruit fly *Drosophila melanogaster* [82]. FA58C domain is also known to occur in molecules that modulate neural functions [83,84]. CUB domains are known only from extracellular molecules including complement proteins C1s and C1r,

tolloid metalloproteinase, and the scavenger receptor cubilin. This domain has been implicated in protein and carbohydrate ligand binding from a variety of molecules in which they occur [85–87]. Both the FA58C and PAN domains were predicted in *Hysy-CnPolydom* and *Neve-CnPolydom*. The sequence from *N. vectensis*, however, did not cover the complete 5' or 3' encoding region so we were unable to determine if a CUB domain may have also been present in this predicted gene. With the exception of these three

additional domain types in cnidarians, we have determined the domain composition of polydom-like homologs is consistent across all other taxa (Figure 3).

To help elucidate the position of polydom genes among the other pentraxin gene family members, we performed phylogenetic analyses on the pentraxin domains from these genes and showed polydom pentraxins clustered into a distinct clade among the pentraxin family of molecules (Figure 5). As we have noted, pentraxin protein sequences are characterized by the motif H-x-C-x-[S/T]-W-x-[S/T] [41]. In our analysis of polydom-like molecules from multiple vertebrate and invertebrate taxa, we have found that the majority of polydom-like sequences (11 of 17) uniquely possess an alanine instead of a cysteine within this motif. While this change is unique to the polydom pentraxins, it does not alter their homology with the pentraxin family. Therefore, a more refined pentraxin motif is H-x-[C/A]-x-[S/T]-W-x-[S/T], where alanine in the third position of the motif is polydom-specific. The thiol group of the cysteine residue in this motif is known to be involved in disulfide bond formation within the pentraxin domain, required for proper protein folding and stability [88]. However, replacement of cysteine by alanine is a conservative substitution (both have neutral side chains) and may still support a similarly folded, stable protein [89–91]. The only taxa we have identified that do not apparently follow this rule are the protostomes *Tribolium castaneum*, *Apis mellifera*, *Anopheles gambiae*, which possess the typical cysteine residue, the invertebrate echinoderm *Strongylocentrotus purpuratus* (valine), the teleost fish *Danio rerio* (glycine), and the opossum *Monodelphis domestica* (serine).

It is likely polydom gene homologs exist in many other taxa than currently identified. Such large genes encoding a complex variety and repetition of domains are difficult to identify in genomes and obtain complete, accurate transcript sequence from. In addition, it has been noted that the frequently used technique of transcript profiling via cDNA/EST analysis results in relatively short sequence reads, from which it is difficult for gene homology of long transcripts, such as polydom, to be determined [92]. More precise identification of polydom-like sequences can be aided by recognizing particular features common to polydom-like sequences across taxa. Using the ancestral sequence of *CnPolydom* from *H. symbiolongicarpus* and the derived *polydom* ortholog in human, we can determine that the unifying characteristics of polydom-like genes across Eumetazoa should include (1) the requisite inclusion of at least VWA, PTX, HYR, CCP, EGF, and GCC2/3 domains; (2) the dominant polydom-specific pentraxin domain motif H-x-[C/A]-x-[S/T]-W-x-[S/T]; and (3) domain architecture in the following N-to C-terminal order: VWA-[EGF or GCC2/3]-CCP × 3-HYR × 1 or 2-multiple EGF+GCC2/3-PTX—and a tail of several CCP and EGF repeats. Thus, the paucity of information on polydom-like molecules will only be mediated by directed efforts targeting these molecules from a variety of taxa to decipher the extent of evolutionary conservation in structure and function.

### Expression of Hysy-CnPolydom and implications for its role in the immune system of *Hydractinia*

Our semi-quantitative RT-PCR results demonstrated that *Hysy-CnPolydom* transcripts were constitutively expressed

at low levels in *H. symbiolongicarpus* and upregulated after exposure to bacteria (*Vibrio* sp. and *B. subtilis*) and most substantial upregulation in response to the yeast *S. cerevisiae*. Because *S. cerevisiae* is not known to be pathogenic to any organism and they do not form an invasive hyphal stage, it is unlikely that upregulation of *Hysy-CnPolydom* after yeast exposure is due to cellular invasion or damage. It is more plausible to conclude that *Hysy-CnPolydom* upregulation is due to the activation of a signaling cascade initiated by extracellular receptor recognition of stable fungal epitopes on the yeast cell, such as the cell wall components  $\beta$ -1,3 glucan or mannan polymers, known to trigger immune responses in many other taxa [93,94]. Whether *Hysy-CnPolydom* is the recognition molecule in this event or rather comes into play somewhere downstream in the extracellular portion of the recognition/signaling/effector cascade remains unknown. The complexity of multidomain proteins is associated with their ability to mediate interactions between different cellular functional systems, and they are particularly common as signaling pathway components [95]. The multidomain architecture of *CnPolydom* may enable the formation of complex protein networks, mediating interactions among multiple functional systems between cell surfaces and/or cell surfaces with the external matrix.

As given in our results, *Hysy-CnPolydom* was predicted with high probability to encode a signal peptide based on SignalP analysis. *Polydom* from mouse and human were also found to have a signal peptide sequence, as did all other predicted vertebrate polydom sequences and one of the identified invertebrate sequences, that of the flour beetle *Tribolium castaneum*. Thus, it seems polydom-like proteins are directed by signal peptides to the endomembrane system, from where they may ultimately be expressed to the cell's exterior. Previous analysis of human polydom, examining domain composition and using phylogenetic reconstruction based on the VWA domain [55] supported the conclusion that human polydom is part of the multidomain extracellular matrix (ECM) protein group. Shur et al. [34,96] have shown that polydom is in fact expressed at the cellular membrane of bone marrow stromal cells and of several breast carcinoma cell lines.

It has been noted in the past that distinguishing interstitial stem cells from neural cells in histological specimens of hydroids is particularly difficult [72], and indeed we are currently unable to definitively show whether one or both of these cell types express *Hysy-CnPolydom*. However, two lines of evidence support the conclusion that fully differentiated neurons putatively express *Hysy-CnPolydom*. First, *in situ* hybridization for *Hysy-CnPolydom* stained cells within the tentacles and hypostome of polyps (Figure 7). Interstitial stem cells and committed stem cell precursors are largely absent from the hypostome and tentacle regions, which would suggest the cells in these regions expressing *Hysy-CnPolydom* are most likely to be fully differentiated neurons. Second, the cell morphology of at least some of the cells are clearly narrow, elongate, and extend to the surface of the epithelium (Figure 7c) that is a cell morphology unique to hydrozoan sensory cells. Interstitial stem cells of hydroids are firmly

established as a multipotent cell lineage maintained for the life of the animal and capable of differentiating into four dominant classes of cell types including neurons, nematocytes, secretory cells, and gametes [67,69,97,98]. Since the only putative differentiated cell type expressing *Hysy-CnPolydom* is neurons, expression within undifferentiated stem cells may indicate neural precursory stem cells.

We have shown here that fungal cues in particular cause substantial upregulation of *Hysy-CnPolydom*. Sedentary colonial cnidarians such as *Hydractinia* are immobile once the planula larva settles and metamorphoses into a polyp and are thus at particular risk of an invasive fungal attack. Indeed fungal pathogens of sedentary colonial cnidarians have been well described, inflicting strong evolutionary costs via mortality and decreased reproductive success [99,100]. Gorgonian sea fans (Class: Anthozoa) mount an inducible immune response [101] against the pathogenic fungus *Aspergillus sydowii* that includes both melanization [100] and production of an unknown antifungal agent [101,102]. However, the molecule(s) involved in recognizing, signaling, and effecting this response are unknown. While the presence of a *CnPolydom*-like gene in these sea fans has not been assessed, a *CnPolydom*-like gene is present within the genome [103] of another member of Anthozoa, the starlet sea anemone *N. vectensis* (Neve-CnPolydom). Based on its presence in members of two cnidarian classes, Anthozoa (*N. vectensis*) and Hydrozoa (*H. symbiolongicarpus* and *H. echinata*), it is likely that many other taxa within Cnidaria encode a *CnPolydom*-like gene and may similarly express it in response to microbial encounters as we have found to be the case in *H. symbiolongicarpus*.

From these data we conclude that *Hysy-CnPolydom* is part of an extracellular signaling or effector cascade most likely acting as an important molecular bridge between cells and the extracellular matrix in response to bacterial and fungal epitopes. These data contribute to our understanding of the origins of the pentraxin family of lectins and the structural and functional evolution of polydom-like molecules throughout the Eumetazoa. The role this gene plays in the life history of *Hydractinia* can help establish its relationship to homologous genes in other cnidarians and in other eumetazoan taxa.

## Acknowledgments

We thank Linda Hodes Villamar, Tuhama Rihani, Cheng Man-Lun, and the University of New Mexico's Molecular Biology Facility for laboratory assistance and Georg Hemmrich for access to *Hydra* sequence databases. The Kiel aquarium in Germany generously provided specimens of *H. echinata*. Funding for this work was provided by an NSF award to LFC (IBN0315968) and UNM Biology department awards to RSS. This work was supported in part by Grant Number P20 RR18754 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and by a grant from the Deutsche Forschungsgemeinschaft (DFG/SFB617-A1) to TCGB.

## Appendix A. Supplementary materials

The online version of this article contains additional supplementary data. Please visit doi:10.1016/j.dci.2008.03.007

## References

- [1] Kortschak RD, Samuel G, Saint R, Miller DJ. EST analysis of the cnidarian *Acropora millepora* reveals extensive gene loss and rapid sequence divergence in the model invertebrates. *Curr Biol* 2003;13:2190–5.
- [2] Ball EE, Hayward DC, Saint R, Miller DJ. A simple plan—cnidarians and the origins of developmental mechanisms. *Nat Rev Genet* 2004;5:567–77.
- [3] Technau U, Rudd S, Maxwell P, Gordon PMK, Saina M, Grasso LC, et al. Maintenance of ancestral complexity and non-metazoan genes in two basal cnidarians. *Trends Genet* 2005; 21:633–9.
- [4] Bridge D, Cunningham CW, Schierwater B, DeSalle R, Buss LW. Class-level relationships in the phylum cnidaria: evidence from mitochondrial genome structure. *Proc Natl Acad Sci USA* 1992;89:8750–3.
- [5] Collins AG. Phylogeny of MEDUSOZOA and the evolution of cnidarian life cycles. *J Evol Biol* 2002;15:418–32.
- [6] Miller DJ, Hemmrich G, Ball EE, Hayward DC, Khalturin K, Funayama N, et al. The innate immune repertoire in cnidaria—ancestral complexity and stochastic gene loss. *Genome Biol* 2007;8:R59.
- [7] Mali B, Möhrlein F, Frohme M, Frank U. A putative double role of a chitinase in a cnidarian: pattern formation and immunity. *Dev Comp Immunol* 2004;28:973–81.
- [8] Mali B, Soza-Reid J, Frohme M, Frank U. Structural but not functional conservation of an immune molecule: a tachylectin-like gene in *Hydractinia*. *Dev Comp Immunol* 2006;30:275–81.
- [9] Schwarz RS, Hodes-Villamar L, Fitzpatrick KA, Fain MG, Hughes AL, Cadavid LF. A gene family of putative immune recognition molecules in the hydroid *Hydractinia*. *Immunogenetics* 2007;59:233–46.
- [10] Hauenschild Cv. Genetische und entwicklungsphysiologische untersuchungen über intersexualität und gewebeverträglichkeit bei *Hydractinia echinata* flem. Wilhem Roux' *Archiv* 1954;147:1–41.
- [11] Frank U, Leitz T, Müller WA. The hydroid *Hydractinia*: a versatile, informative cnidarian representative. *BioEssays* 2001;23:963–71.
- [12] Bode HR. The interstitial cell lineage of hydra: a stem cell system that arose early in evolution. *J Cell Sci* 1996;109:1155–64.
- [13] Müller WA, Teo R, Frank U. Totipotent migratory stem cells in a hydroid. *Dev Biol* 2004;275:215–24.
- [14] Dishaw LJ, Smith SL, Bigger CH. Characterization of a C3-like cDNA in a coral: phylogenetic implications. *Immunogenetics* 2005;57:535–48.
- [15] Sullivan JC, Kalaitzidis D, Gilmore TD, Finnerty JR. Rel homology domain-containing transcription factor in the cnidarian *Nematostella vectensis*. *Dev Genes Evol* 2007;217:63–72.
- [16] Zheng L, Zhang L, Lin H, McIntosh MT, Malacrida AR. Toll-like receptors in invertebrate innate immunity. *Invertebrate Survival J* 2005;2:105–13.
- [17] Gewurz H, Zhang XH, Lint F. Structure and function of the pentraxins. *Curr Opin Immunol* 1995;7:54–64.
- [18] Garlanda C, Bottazzi B, Bastone A, Mantovani A. Pentraxins at the crossroads between innate immunity, inflammation, matrix deposition, and female fertility. *Annu Rev Immunol* 2005;23:337–66.
- [19] Hind CR, Collins PM, Baltz ML, Pepys MB. Human serum amyloid P component, a circulating lectin with specificity for

- the cyclic 4,6-pyruvate acetal of galactose. Interactions with various bacteria. *Biochem J* 1985;225:107–11.
- [20] Szalai AJ. The antimicrobial activity of C-reactive protein. *Microb Infect* 2002;4:201–5.
- [21] Abernethy TJ, Avery OT. The occurrence during acute infections of a protein not normally present in the blood. I. Distribution of the reactive protein in patients' sera and the effect of calcium on the flocculation reaction with C-polysaccharide of *Pneumococcus*. *J Exp Med* 1941;73:173–82.
- [22] Defrutos PG, Dahlback B. Interaction between serum amyloid P component and C4b binding protein associated with inhibition of factor I-mediated C4b degradation. *J Immunol* 1994;152:2430–7.
- [23] de Haas CJ, van der Tol ME, van Kessel KP, Verhoef J, van Strijp JA. A synthetic lipopolysaccharide-binding peptide based on amino acids 27–39 of serum amyloid P component inhibits lipopolysaccharide-induced responses in human blood. *J Immunol* 1998;161:3607–15.
- [24] Pepys MB, Hirschfield GM. C-reactive protein: a critical update. *J Clin Invest* 2003;111:1805–12.
- [25] Mold C, Baca R, Du Clos TW. Serum amyloid P component and C-reactive protein opsonize apoptotic cells for phagocytosis through Fc gamma receptors. *J Autoimmun* 2002;19:147–54.
- [26] Nauta AJ, Daha MR, van Kooten C, Roos A. Recognition and clearance of apoptotic cells: a role for complement and pentraxins. *Trends Immunol* 2003;24:148–54.
- [27] Armstrong PB, Swarnakar S, Srimal S, Misquith S, Hahn EA, Aimes RT, et al. A cytolytic function for a sialic acid-binding lectin that is a member of the pentraxin family of proteins. *J Biol Chem* 1996;271:14717–21.
- [28] Iwaki D, Osaki T, Mizunoe Y, Wai SN, Iwanaga S, Kawabata S. Functional and structural diversities of C-reactive proteins present in horseshoe crab hemolymph plasma. *Eur J Biochem* 1999;264:314–26.
- [29] Ng PML, Jin ZX, Tan SSH, Ho B, Ding JL. C-reactive protein: a predominant LPS-binding acute phase protein responsive to *Pseudomonas* infection. *J Endotox Res* 2004;10:163–74.
- [30] Breviario F, Daniello EM, Golay J, Peri G, Bottazzi B, Bairoch A, et al. Interleukin-1 inducible genes in endothelial cells-cloning of a new gene related to C-reactive protein and serum amyloid-P component. *J Biol Chem* 1992;267:22190–7.
- [31] Garlanda C, Hirsch E, Bozza S, Salustri A, De Acetis M, Nota R, et al. Non-redundant role of the long pentraxin PTX3 in anti-fungal innate immune response. *Nature* 2002;420:182–6.
- [32] Diniz SN, Nomizo R, Cisalpino PS, Teixeira MM, Brown GD, Mantovani A, et al. PTX3 functions as an opsonin for the dectin-1-dependent internalization of zymosan by macrophages. *J Leukocyte Biol* 2004;75:649–56.
- [33] Gilgès D, Vinit MA, Callebaut I, Coulombel L, Cacheux V, Romeo PH, et al. Polydom: a secreted protein with pentraxin, complement control protein, epidermal growth factor and von Willebrand factor domains. *Biochem J* 2000;352:49–59.
- [34] Shur I, Socher R, Hameiri M, Fried A, Benayahu D. Molecular and cellular characterization of SEL-OB/SVEP1 in osteogenic cells *in vivo* and *in vitro*. *J Cell Phys* 2006;206:420–7.
- [35] Winn VD, Haimov-Kochman R, Paquet AC, Yang YJ, Madhusudhan MS, Gormley M, et al. Gene expression profiling of the human maternal-fetal interface reveals dramatic changes between midgestation and term. *Endocrinology* 2007;148:1059–79.
- [36] Cadavid LF, Powell AE, Nicotra ML, Moreno M, Buss LW. An invertebrate histocompatibility complex. *Genetics* 2004;167:357–65.
- [37] Mokady O, Buss LW. Transmission genetics of allorecognition in *Hydractinia symbiolongicarpus* (cnidaria: hydrozoa). *Genetics* 1996;143:823–7.
- [38] Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 1998;8:186–94.
- [39] Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 1998;8:175–85.
- [40] Gordon D. Viewing and editing assembled sequences using consed. In: Baxevanis AD, Davison DB, editors. *Current protocols in bioinformatics*. New York, NY: Wiley; 2004. p. 11.2.1–11.2.43.
- [41] Bairoch A. PROSITE—a dictionary of sites and patterns in proteins. *Nucleic Acids Res* 1991;19:2241–5.
- [42] Apweiler R, Attwood TK, Bairoch A, Bateman A, Birney E, Biswas M, et al. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res* 2001;29:37–40.
- [43] Letunic I, Copley RR, Pils B, Pinkert S, Schultz J, Bork P. SMART 5: domains in the context of genomes and networks. *Nucleic Acids Res* 2006;34:D257–60.
- [44] Schultz J, Milpetz F, Bork P, Ponting CP. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci USA* 1998;95:5857–64.
- [45] Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, et al. The Pfam protein families database. *Nucleic Acids Res* 2004;32:D138–41.
- [46] Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–402.
- [47] Ryan JF, Finnerty JR. CnidBase: the cnidarian evolutionary genomics database. *Nucleic Acids Res* 2003;31:159–63.
- [48] Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;32:1792–7.
- [49] Tamura K, Dudley J, Nei M, Kumar S. MEGA 4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 2007;24:1596–9.
- [50] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. *Nucleic Acids Res* 2000;28:235–42.
- [51] Lo Conte L, Ailey B, Hubbard TJP, Brenner SE, Murzin AG, Chothia C. SCOP: a structural classification of proteins database. *Nucleic Acids Res* 2000;28:257–9.
- [52] Grens A, Gee L, Fisher DA, Bode HR. CnNK-2, an NK-2 homeobox gene, has a role in patterning the basal end of the axis in hydra. *Dev Biol* 1996;180:473–88.
- [53] Seery LT, Schoenberg DR, Barbaux S, Sharp PM, Whitehead AS. Identification of a novel member of the pentraxin family in *Xenopus laevis*. *Proc R Soc London Ser B-Biol Sci* 1993;253:263–70.
- [54] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–10.
- [55] Whittaker CA, Hynes RO. Distribution and evolution of von willebrand/integrin domains: widely dispersed adhesion and elsewhere. *Mol Biol Cell* 2002;13:3369–87.
- [56] Xiong JP, Stehle T, Zhang R, Joachimiak A, Frech M, Goodman SL, et al. Crystal structure of the extracellular segment of integrin  $\alpha V\beta 3$  in complex with an arg-gly-asp ligand. *Science* 2002;296:151–5.
- [57] Reid KB, Day AJ. Structure–function relationships of the complement components. *Immunol Today* 1989;10:177–80.
- [58] Callebaut I, Gilgès D, Vigon I, Moron JP. HYR, an extracellular module involved in cellular adhesion and related to the immunoglobulin-like fold. *Prot Sci* 2000;9:1382–90.
- [59] Tordai H, Bányai L, Patthy L. The PAN module: the N-terminal domains of plasminogen and hepatocyte growth factor are homologous with the apple domains of the prekallikrein family and with a novel domain found in numerous nematode proteins. *FEBS Lett* 1999;461:63–7.
- [60] Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, et al. Pfam: clans, web tools and services. *Nucleic Acids Res* 2006;34:D247–51.

- [61] McMullen BA, Fujikawa K, David EW, Hedner U, Ezban M. Locations of disulfide bonds and free cysteines in the heavy and light chains of recombinant human factor VIII (antihemophilic factor A). *Protein Sci* 1995;4:740–6.
- [62] Bork P. Complement components C1r/C1s, bone morphogenetic protein 1 and *Xenopus laevis* developmentally regulated protein UVS.2 share common repeats. *FEBS Lett* 1991;288:9–12.
- [63] Beckmann G, Hanke J, Bork P, Reich JG. Merging extracellular domains: fold prediction for laminin G-like and amino-terminal thrombospondin-like modules based on homology to pentraxins. *J Mol Biol* 1998;275:725–30.
- [64] Saitou N, Nei M. The neighbor-joining method—a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987;4:406–25.
- [65] Bottazzi B, Garlanda C, Salvatori G, Jeannin P, Manfredi A, Mantovani A. Pentraxins as a key component of innate immunity. *Curr Opin Immunol* 2006;18:10–5.
- [66] Weis VM, Keene DR, Buss LW. Biology of hydractiniid hydroids. 4. Ultrastructure of the planula of *Hydractinia echinata*. *Biol Bull* 1985;168:403–18.
- [67] Plickert G, Krohler M, Munck A. Cell proliferation and early differentiation during embryonic development and metamorphosis of *Hydractinia echinata*. *Development* 1988;103:795–803.
- [68] Gajewski M, Leitz T, Schloßherr J, Plickert G. LWamides from cnidaria constitute a novel family of neuropeptides with morphogenetic activity. *Dev Genes Evol* 1996;205:232–42.
- [69] Martin VJ, Thomas MB. Nerve elements in the planula of the hydrozoan *Pennaria tiarella*. *J Morphol* 1980;166:27–36.
- [70] Martin VJ, Archer WE. Stages of larval development and stem cell population changes during metamorphosis of a hydrozoan planula. *Biol Bull* 1997;192:41–52.
- [71] Burnett AL, Diehl NA. The nervous system of *Hydra*. I. Types, distribution, and origin of nerve elements. *J Exp Zool* 1964;157:217–26.
- [72] David CN. A quantitative method for maceration of hydra tissue. *Dev Genes Evol* 1973;171:259–68.
- [73] Bányai L, Patthy L. Importance of intramolecular interactions in the control of the fibrin affinity and activation of human plasminogen. *J Biol Chem* 1984;259:6466–71.
- [74] Okigaki M, Komada M, Uehara Y, Miyazawa K, Kitamura N. Functional characterization of human hepatocyte growth factor mutants obtained by deletion of structural domains. *Biochemistry* 1992;31:9555–61.
- [75] Ho DH, Badellino K, Baglia FA, Walsh PN. A binding site for heparin in the Apple 3 domain of factor XI. *J Biol Chem* 1998;273:16382–90.
- [76] Poole S, Firtel RA, Lamar E, Rowekamp W. Sequence and expression of the discoidin I gene family in *Dictyostelium discoideum*. *J Mol Biol* 1981;153:273–89.
- [77] Kane WH, Davie EW. Cloning of a cDNA coding for human factor V, a blood coagulation factor homologous to factor VIII and ceruloplasmin. *Proc Natl Acad Sci USA* 1986;83:6800–4.
- [78] Foster PA, Fulcher CA, Houghten RA, Zimmerman TS. Synthetic factor VIII peptides with amino acid sequences contained within C2 domain of factor VIII inhibit factor VIII binding to phosphatidylserine. *Blood* 1990;75:1999–2004.
- [79] Baumgartner S, Hofmann K, Chiquet-Ehrismann R, Bucher P. The discoidin domain family revisited: new members from prokaryotes and a homology-based fold prediction. *Protein Sci* 1998;7:1626–31.
- [80] Rosen SD, Kafka JA, Simpson DL, Barondes SH. Developmentally regulated carbohydrate-binding protein in *Dictyostelium discoideum*. *Proc Natl Acad Sci USA* 1973;70:2554–7.
- [81] Kotani E, Yamakawa M, Iwamoto S, Tashiro M, Mori H, Sumida M, et al. Cloning and expression of the gene of hemocytin, an insect humoral lectin which is homologous with the mammalian von willebrand factor. *Biochim Biophys Acta—Gene Struct Expression* 1995;1260:245–58.
- [82] Goto A, Kumagai T, Kumagai C, Hirose J, Narita H, Mori H, et al. A *Drosophila* haemocyte-specific protein, hemolectin, similar to human von willebrand factor. *Biochem J* 2001;359:99–108.
- [83] Kitsukawa T, Shimono A, Kawakami A, Kondoh H, Fujisawa H. Overexpression of a membrane protein, neuropilin, in chimeric mice causes anomalies in the cardiovascular system, nervous system and limbs. *Development* 1995;121:4309–18.
- [84] Baumgartner SW, Littleton JT, Broadie K, Bhat MA, Harbecke R, Lengyel JA, et al. A *Drosophila* neurexin is required for septate junction and blood–nerve barrier formation and function. *Cell* 1996;87:1059–68.
- [85] Shimell MJ, Ferguson EL, Childs SR, O'Connor MB. The *Drosophila* dorsal–ventral patterning gene *tolloid* is related to human bone morphogenetic protein 1. *Cell* 1991;67:469–81.
- [86] Töpfer-Petersen E, Romero A, Varela PF, Ekhlesi-Hundrieser M, Dostálová Z, Sanz L, et al. Spermadhesins: a new protein family. Facts, hypotheses and perspectives. *Andrologia* 1998;30:217–24.
- [87] Kristiansen M, Korzyraki R, Jacobsen C, Nexø E, Verroust PJ, Moestrup SK. Molecular dissection of the intrinsic factor vitamin B12 receptor, cubilin, discloses regions important for membrane association and ligand binding. *J Biol Chem* 1999;274:20540–4.
- [88] Nelsestuen GL, Ostrowski BG. Membrane association with multiple calcium ions: vitamin-K-dependent proteins, annexins and pentraxins. *Curr Opin Struct Biol* 1999;9:433–7.
- [89] Davidson FF, Loewen PC, Khorana HG. Structure and function in rhodopsin: replacement by alanine of cysteine residues 110 and 187, components of a conserved disulfide bond in rhodopsin, affects the light-activated metarhodopsin II state. *Proc Natl Acad Sci USA* 1994;91:4029–33.
- [90] Marks CB, Naderi H, Kosen PA, Kuntz ID, Anderson S. Mutants of bovine pancreatic trypsin inhibitor lacking cysteines 14 and 38 can fold properly. *Science* 1987;235:1370–3.
- [91] Staley JP, Kim PS. Complete folding of bovine pancreatic trypsin inhibitor with only a single disulfide bond. *Proc Natl Acad Sci USA* 1992;89:1519–23.
- [92] Shida K, Terajima D, Uchino R, Ikawa S, Ikeda M, Asano K, et al. Hemocytes of *Ciona intestinalis* express multiple genes involved in innate immune host defense. *Biochem Biophys Res Commun* 2003;302:207–18.
- [93] Brown G, Gordon S. Immune recognition of fungal  $\beta$ -glucans. *Cell Microbiol* 2005;7:471–9.
- [94] Hohl TM, Rivera A, Pamer EG. Immunity to fungi. *Curr Opin Immunol* 2006;18:465–72.
- [95] Koonin EV, Galperin MY. Sequence–evolution–function: computational approaches in comparative genomics. Boston: Kluwer Academic Publishers; 2003.
- [96] Shur I, Zemer-Tov E, Socher R, Benayahu D. SVEP1 expression is regulated in an estrogen-dependent manner. *J Cell Phys* 2007;210:732–9.
- [97] David CN, Murphy S. Characterization of interstitial stem cells in hydra by cloning. *Dev Biol* 1977;58:372–83.
- [98] Bosch TCG, David CN. Stem cells of *Hydra magnipapillata* can differentiate into somatic cells and germline cells. *Dev Biol* 1987;121:182–91.
- [99] Smith GW, Ives LD, Nagelkerken IA, Ritchie KB. Caribbean sea fan mortalities. *Nature* 1996;383:487.

- [100] Petes LE, Harvell CD, Peters EC, Webb MAH, Mullen KM. Pathogens compromise reproduction and induce melanization in Caribbean sea fans. *Mar Ecol-Progr Ser* 2003;264:167–71.
- [101] Kim K, Harvell CD, Kim PD, Smith GW, Merkel SM. Fungal disease resistance of Caribbean Sea fan corals (*Gorgonia* spp.). *Mar Biol* 2000;136:259–67.
- [102] Kim K, Kim PD, Alker AP, Harvell CD. Chemical resistance of gorgonian corals against fungal infections. *Mar Biol* 2000; 137:393–401.
- [103] Sullivan JC, Ryan JF, Watson JA, Webb J, Mullikin JC, Rokhsar D, et al. StellaBase: the *Nematostella vectensis* genomics database. *Nucleic Acids Res* 2006;34:D495–9.