Developmental and Comparative Immunology (2008) **32**, 608–612

Available at www.sciencedirect.com

ScienceDirect

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

SHORT COMMUNICATION

Toll-like receptor genes (TLRs) from *Capitella capitata* and *Helobdella robusta* (Annelida)

Charis R. Davidson[a], Natalie M. Best[a], Joseph W. Francis[b], Edwin L. Cooper[c], Todd Charles Wood[a,*]

[a]*Division of Natural Sciences, Bryan College, 721 Bryan Drive, Box 7802, Dayton, TN 37321, USA*
[b]*Department of Biology, The Master's College, 21726 Placerita Canyon Road, Santa Clarita, CA 91321, USA*
[c]*David Geffen School of Medicine at UCLA, University of California, Los Angeles, CA, USA*

Received 5 October 2007; received in revised form 30 October 2007; accepted 13 November 2007
Available online 17 December 2007

KEYWORDS
Innate immunity;
Toll-like receptor;
Annelida;
Helobdella;
Capitella

Summary
Toll-like receptors (TLRs) are an important part of the innate immunity system and are found throughout the animal kingdom, but have not yet been reported in annelids. We searched shotgun reads of the genomes of the leech *Helobdella* and polychaete *Capitella* for TLR homologs. We found 105 TLR homologs in *Capitella* and 16 in *Helobdella*. The deduced phylogeny of these sequences, together with TLRs from other animal phyla, reveals three major clades. One clade consists of a mixture of both vertebrates and invertebrates, including sequences from *Capitella* and *Helobdella*, while the other two clades contain only invertebrate TLRs.

Introduction

Toll-like receptors (TLRs) are pattern-recognition receptors (PRRs) of the innate immune system found in both plant and animal kingdoms [1–5]. Toll (for whom TLRs are named) was originally described as a factor involved in dorsoventral determination in *Drosophila* [6] but was also found to activate antimicrobial mechanisms [7]. In *Drosophila*, Toll is activated indirectly by Spaetzle, a soluble PRR [8]. The Spaetzle-activated Toll stimulates the transcription factor Dorsal that causes the release of antimicrobial factors [9]. Homologs of *Drosophila*'s Toll are found in representatives of the three major groups of bilaterians. In invertebrates, TLR or Toll signaling pathway components have been detected in mosquitoes (*Anopheles* and *Aedes*), sea urchins (*Stronglyocentrotus*), tunicates (*Ciona*), the horshoe crab (*Tachypleus*), sponges (*Suberites*), nematodes (*Caenorhabditis*), squid (*Euprymna*), and slime mold (*Dictyostelium*) [10–16].

The study of TLRs and innate immunity has largely focused on two animal phyla: Chordata and Arthropoda

Abbreviations: TLR, toll-like receptor; PRR, pattern-recognition receptor; LRR, leucine-rich repeat.

*Corresponding author. Tel.: +1 423 775 7277; fax: +1 423 775 7330.
E-mail address: todd.wood@bryan.edu (T.C. Wood).

0145-305X/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.dci.2007.11.004

[10]. The recent publication and analysis of the sea urchin genome of phylum Echinodermata has revealed a wealth of data regarding invertebrate innate immunity, including hundreds of TLRs [10,11]. Protostomes that have been analyzed for the presence of TLRs are primarily representatives of the Ecdysozoa superphylum (e.g. *Drosophila*). The genomes of protostome organisms within the Lopotrochozoa superphlyum, such as annelids, have not been analyzed to any great extent for the presence of TLRs or other innate immunity factors [17,18]. The leech *Helobdella* and polycheate *Capitella* are the first lophotrochozoan (and annelid) species selected for whole genome sequence analysis, and this report represents the first analysis of these annelids for the presence of immune receptor genes.

Methods

Genome sequence reads and quality files of *Capitella* and *Helobdella* were obtained from the NCBI Trace Archive (http://www.ncbi.nlm.nih.gov). Protein–DNA searches were conducted using TFASTY [19] with the human TLR-2 sequence (GenBank ID AAC34133) and later with inferred protein sequences from both *Helobdella* and *Capitella*. Traces homologous to the query sequences ($E<10^{-6}$) were masked using CROSS_MATCH and assembled using PHRAP (http://www.phrap.org). Resulting contigs were examined visually for open reading frames using Artemis [20] and automatically using the NCBI ORF finder (http://www.ncbi.nlm.nih.gov/gorf). Inferred *Capitella* TLR protein sequences were further matched against the v. 1.0 release of the genome annotation using BLASTP and TBLASTN (http://genome.jgi-psf.org/Capca1/Capca1.home.html).

Protein sequences inferred to be from active genes (i.e. no stop codons or frameshift mutations in the coding region) were aligned using CLUSTALW [21]. All trees were constructed using the PROTDIST and NEIGHBOR programs of the PHYLIP package [22]. Protein distances were inferred using Veerassamy et al.'s [23] transition probability model. Protein domains were identified automatically using SMART (http://smart.embl-heidelberg.de) [24].

Results

We searched the trace archives of two annelid worms, *Helobdella* and *Capitella*, for homologs of the protein sequence of human TLR-2 (AAC34133) using TFASTY [19]. Both annelids had approximately 3.7 million traces in the NCBI trace archives, representing a ~5-fold coverage of the genome. We found positive matches ($E<10^{-6}$) in 13 traces from *Helobdella* and 485 traces from *Capitella*. Putative *Capitella* and *Helobdella* TLR protein sequences inferred from the highest-scoring alignments were used as queries in further TFASTY searches, expanding the number of traces from Capitella to 694 and from Helobdella to 304 (Table 1).

We assembled the traces from each species using PHRAP. We found 105 contigs in *Capitella* and 16 in *Helobdella*. Analysis of the ORFs revealed one partial TLR gene in each contig. Since many of the inferred protein sequences were

Table 1 TLR signaling pathway proteins in *Capitella* and *Helobdella*.

Protein	*Capitella*		*Helobdella*	
	Match (GenBank trace ID)	Percent identity	Match (GenBank trace ID)	Percent identity
Drosophila proteins				
Cactus	1109683694	32.8	1112642406	36.9
Dorsal	1027947919	64.2	1122215071	40.0
ECSIT	1085544040	50.0	1115381906	49.7
Pelle	1112699057	34.9		
Relish				
Spaetzle				
TAK1				
TRAF1	1028363621	45.7	1112642913	39.5
Tube				
Human proteins				
ECSIT	1085544040	37.3	1115381906	26.1
IKK β				
IκB ε	1074119071	32.8		
IKK α				
IRAK1				
MEKK	1026033689	36.6		
MyD88	1320293219	34.3	1343967404	51.7
NEMO				
NF-κB p105				
NF-κB p65	1028298985	46.2		
TRAF5	1085417814	49.1	1140751855	32.9