A putative nuclear growth factor-like globular nematode specific protein

Ranil Samantha Dassanayake* [1], Witharanage Wasana Prasadini Rodrigo[2], Eric Hamilton Karunanayake[2], Ovitigala Vithanage Don Sisira Jagathpriya Weerasena[2], Naduviladath Vishvanath Chandrasekharan[1]

[1]Department of Chemistry, University of Colombo, Colombo 3, Sri Lanka; [2]Institute of Biochemistry, Molecular Biology and Biotechnology, University of Colombo, Colombo - 3, Sri Lanka; Email: rsdassan@webmail.cmb.ac.lk; Phone: 94-11-2503367; Fax: 94-11-2503367; * Corresponding author

Abstract:
Expressed sequence tags (ESTs) are an effective approach for discovery of novel genes. In the current study, approximately 250 ESTs of the cattle parasitic nematode *Setaria digitata* were examined and a cDNA clone identified whose coding sequence could not be functionally annotated by searching over publicly available genome, protein, EST and STS databases. Here, we report the extensive characterization of this ORF (UP) and its homologues using a bioinformatics approach. Uncharacterized protein (SDUP) of *S. digitata* consists of 204 amino acids with a predicted molecular weight and isoelectric point of 22.8KDa and 9.94, respectively. A search carried out using SDUP over nucleotide, EST and protein databases at NCBI, NEMBASE3 and Parasite Genome Database (PGD) identified homologous counterparts from the human parasitic nematodes *Wuchereria bancrofti* (WB), *Brugia malayi* (BM), *Onchocerca volvulus* (OV), the mouse filarial worm *Litomosoides sigmodontis* (LS), swine parasitic nematodes *Ascaris suum* (AS) and diverged counterparts from the plant parasitic nematode *Meloidogyne hapla* (MH) and free living nematodes *Caenorhabditis elegans* (CE) and *Caenorhabditis briggsae* (CB). Phylogenetic analyses revealed the UPs to be undergoing divergent evolution. A search of the ESTs at PGD showed that UP is expressed in all the stages of BM. Secondary structure analyses of multiply-aligned sequences of homologues using Jpred server indicated UPs to be rich in beta-pleated structures. TMMHH server and beta barrel finder programme indicated, UPs to be neither transmembrane or beta barrels proteins but are likely to be globular proteins. Further, the Motif discovery tool of MEME identified three novel potential motifs for UPS, of which only two are present in CE, CB & MH. Analyses of UPs using Signal IP, TargetP, Psort servers predicted this group of proteins to be devoid of signal peptide cleavage sites, are not mitochondrial targeting peptides but appear to be localized to the nucleus, respectively. Further analyses of the UPs using ScanProsite server for phosphorylation revealed potential sites for cAMP- and cGMP-dependent protein kinase, Protein kinase C and Casein kinase II. Putative functional analysis using ProtFun 2.1 Server indicated UPs to be nonenzymatic, growth factor like protein. Finally, collating all the information derived from bioinformatics analyses, we conclude that the UPs of nematodes are most likely to be expressed at all stages in the life cycle, localized to the nucleus, regulated by phosphorylation, rich in beta-pleated strands and are growth factor like nematode specific proteins.

Keywords: nematodes, *Setaria digitata*, bioinformatics, servers

Background:
Nematodes are the most numerous multicellular animals on earth and nearly 80,000 species have been described under the phylum Nematoda, of which over 15,000 are parasitic and infect humans, domestic animals and food crops [1]. The free-living nematodes on the other hand contribute the largest number of species in Nematoda and are largely present in soil and do not parasitize plants, but are beneficial in the decomposition of organic matter. It has been estimated that there can be as many as 500,000 or more described and undescribed species in the phylum Nematoda [2]. Parasitic nematodes can readily infect plants or animals, although there are basic structural differences in cellular features of their hosts. These two groups of parasitic nematodes may have common properties that are important for parasitizing the host. Parasitic nematodes of animals, especially humans have a profound effect on the economy due to their direct involvement in health and productivity. Over 1 billion people are infected worldwide by parasitic nematodes [3], and the diseases they cause collectively lead to mortality severe morbidity, blindness, anemia, intestinal disease, respiratory disease, and disfigurement of major organs and limbs. On the other hand, plant parasitic nematodes can cause even greater economic losses as they could infect crops that are important for both animals and humans. A report published in 2002 indicated the annual crop loss caused by plant parasitic nematodes to be roughly $80 billion worldwide, with $8 billion in the USA [4]. Although more recent estimates on global crop losses are not available, it may be very much higher with the increase in agricultural productivity. Despite these facts, the biology of the vast majority of parasitic and free living nematode species is poorly understood. In addition to genetic and biological information gleaned from research on these organisms, further advances in these areas could be achieved by unraveling biological information from the genomes of nematodes available in public databases using tools in bioinformatcis. Further, evolutionary relationships and the selective pressures can also be ascertained using the information in genomes of these organisms. In the current study, we have used the latter tools to systematically characterize a group of nematode specific proteins using an ORF we characterized from a parasitic nematode *Setaria digitata.* This nematode is found in the peritoneal cavity of many ungulates including cattle and buffaloes and causes nematodiasis, a neuropathological disorder. In the affected animals mild motor weakness to severe paralysis and death could be brought about by the infective stage larvae (L_3) of *S. digitata* [5].
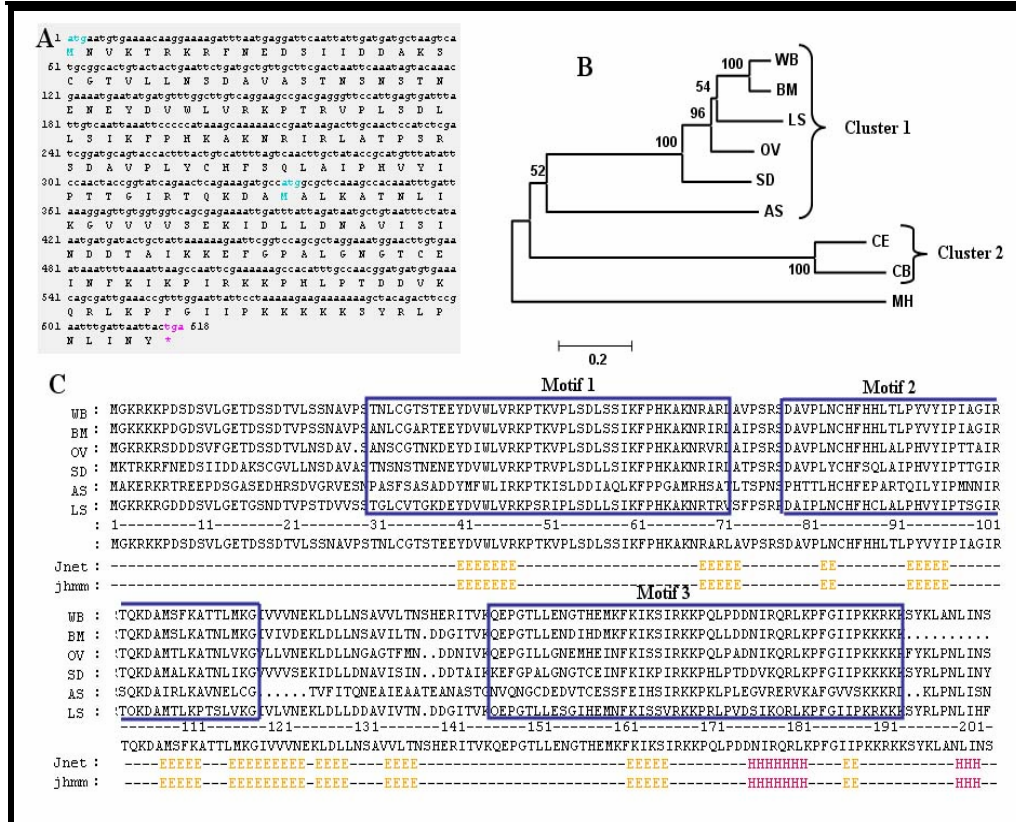
Figure 1: Coding properties of *S.digitata* uncharacterized nucleotide sequence (A), Phylogenetic tree based on uncharacterized sequences of *W. bancrofti* (WB), *B. malayi* (BM), *O. volvulus* (OV), *A. suum* (AS) *S. digitata* (SD), *L. sigmodontis* (LD), *M. hapla* (MP), *C. elegans* (CE) and *C. briggsae* (CB). The lengths of the horizontal lines are proportional to the minimum number of amino acid differences required to join nodes. Vertical lines are for spacing branches and labels (B), and Jpred3 server predicted secondary structures for multiply aligned unknown sequences (Jnet: final secondary structure prediction for query, jhmm - Jnet hmm profile prediction - H , α helices and E , β strands) (C). The boxes in the alignment indicates motifs identified by MEME-Motif discovery tool for uncharacterized sequences using position-specific probability matrices, CE, CB & MH were not included for the alignment due to the vast dissimilarity in sequences.

Methodology:
Construction of cDNA library
A cDNA library of S. digitata was constructed in the vector λ Zap according to manufacturer's instructions (Stratagene). The library was in-vivo excised in *E. coli*, XL1- Blue MRF to give phagemid colonies, which were randomly picked and plasmid DNA was extracted using the alkaline lysis method.

Sequencing
Bidirectional sequencing of randomly picked plasmid DNA was carried out with vector derived primers using Thermo Sequenase TM CY5 Dye Terminator kit and ALF express TM DNA Sequencer (Amershan-Pharmacia, Sweden). A clone designated pSDC13 containing *S. digitata* UP (SDUP) was completely sequenced.

Bioiformatic analysis
Amino acid sequences of *W. bancrofti* and *B. malayi*, were acquired by executing an iterative protein-protein BLAST at NCBI [6] against all non-redundant GenBank CDS translations + RefSeq Proteins + PDB + SwissProt + PIR + PRF protein databases using SDUP as the query sequence. Amino acid sequences of UPs of *O. volvulus*,

A. suum and *M. hapla,* and *Litomosoides sigmodontis* were retrieved by executing an iterative WU-BLAST of the Parasite Genome Database [7] and nucleotide-protein BLAST of NEMBASE3 [8] respectively, using SDUP as the query Sequence. Multiple alignments of nucleotide sequences ware carried out using ClustalW of the BioEdit software program. Molecular Biology programmes and servers used to analyse different properties of proteins are indicated in the results and discussion section.   A phylogenetic tree was constructed by Neighbor- Joining with 100 bootstraps replicates using multiply-aligned sequence with MEGA 3.1 software program.

Discussion:
Nearly 250 randomly picked clones of the cDNA expression library of the cattle parasitic nematode *S. digitata* were sequenced bi-directionally and functional annotation of these cDNAs was attempted by performing a BLAST search over publicly available genomic, protein and domain databases. In the latter process, we identified a cDNA clone to which functional annotation was not possible by searching over databases indicated above. Thus, as an initial step towards the characterization of this clone, it was fully sequenced and the complete ORF

371

(SDUP) was recovered (Figure 1A). The ORF consists of 204 amino acids and its predicted molecular weight and isoelectric point were found to be 22.8Kda and 9.94, respectively. Searching over the non redundant protein sequences database of NCBI to identify similar ORF using algorithm programs Protein-Protein BLAST (BLASTP), Pattern Hit Initiated BLAST (PHI-BLAST), Position-Specific Iterated BLAST (PHI–BLAST) [6] identified homologous sequences from the human filarial parasitic nematode *B. malayi,* and a remarkably diverged sequences covering entire length of SDUP from free living nematodes *C. elegans* and *C. briggsae* with insertions and deletions. A TBLASTX search at NCBI identified a homologous sequence from the human filarial parasitic nematode *W. bancrofti.* Further, searches in Parasite Genomes Database with WU-Blast2 [7], identified sequences with significant similarities from human parasitic nematodes *O. volvulus* (African riverblindness nematode) and *A. suum* (in small intestine of pig), and sequence with a very low similarity with regions of insertions and deletions was identified from plant parasitic nematode *M. hapla* (Root knot nematode) (Table 1 & Figure 1C). Furthermore, a TBLASTX search over the databases in NEMBASE3 [8] identified a homologous sequence from the mouse filarial worm *L. sigmodontis.* Ambiguities in these retrieved sequences were corrected and coding properties were recovered. Taking *S. digitata* ORF as template, ORF of the retrieved sequences were analyzed and refined. Secondary structure analysis of multiply-aligned sequences of UPs with Clustal W using Jpred consensus method for protein secondary structure prediction server [9], indicated these proteins (UP of WB, BM, LS, SD & AS) to be rich in beta pleated structure and with two potential regions for alpha helixes (Figure 1C). Thus, due to the structural similarities in UPs it could be assumed that this group of proteins performs similar functions in these parasitic nematodes. A putative conserved domain (CD) search with an expect value threshold of 0.1 [10] and a motif search [11] did not result in any hits for the UPs. However, when the expected value was increased up to 1 or 10, CD search over the CDD27036PSSMS database identified a region of uncharacterized proteins to have entire length of the conserved domain of RNAP_Rpb7_N_like, which represents the N-terminal ribonucleoprotein (RNP) domain of the Rpb7 subunit of eukaryotic RNA polymerase (RNAP) [12]. Although, it was not possible to assign any known conserved domain or motifs with a significant structural similarity to this group of proteins, the motif discovery tool of MEME [11] identified three potential novel motifs for unknown sequences of WB, BM, LS, SD & AS using position-specific probability matrices and of these only two were revealed in the sequences of CE, CB & MH. On the basis of primary and secondary structures analyses, and motifs present, UPS can be clearly divided into two groups. In spite of tremendous sequence diversity in these two groups, presence of some of putative motifs (Figure 1B) in these two groups suggests structural and functional similarities. To further study the relationships of these proteins, a phylogenetic analysis was carried out and a tree was constructed by neighbor joining method of MEGA using multiply-aligned sequences. After generating several trees, the best tree was selected with bootstrap values. Phylogenetic reconstruction analyses revealed two clusters with WB, BM, LS, OV SD and AS grouped into

one while CE & CB into another, MH is the most diverse. Despite the fact that these members can be phylogenetically grouped, sequence divergence amongst the members coming under cluster 1 (Figure 1B) is remarkable and sequence identities of the members are in the range of 84.3 to 25.4, although, their highest classification level is class Chromadorea. Further, a significant sequence divergence was also seen amongst closely related CE and CB of the genus *Caenorhabditis.* These sequence divergences amongst the counterparts which are also reflected by the lengths of the branches of the phylogenetic tree imply that this group of proteins is undergoing divergent evolution perhaps to perform species-specific function or to adapt to environmental conditions that they live in. For instance, WB, BM, LS and OV are classified under the family Onchocercidae, WB and BM, and LS are human and mouse filarial parasites living in the lymphatic system, respectively. LS causes blindness in humans. Phylogenetic reconstruction analyses revealed strong inclination of LS to group with WB and BM than OV, which is further evident when the phylogenetic tree is generated without either WB or BM. This implies that this protein may be under selection pressure by the microenvironment they lived in than the organism which they parasitize. However, to come to a definite conclusion, more information of UPs sequences from different nematodes is required. A search of ESTs in the PGD [7] resulted in significant hits for different stages in the life cycle of the same parasitic nematodes. For instance, ESTs from all the stages of *B. malayi* (adult male, female, molting L3 larva and infective larva) were observed. Similarly, embryo and adult female of *A. suum,* egg of *M. hapla* , L3 larva of *O. volvulus* and unknown stages of *W. bancrofti* and *C. elegance* were found (Table 1 in supplementary material). This suggests that this protein is apparently expressed in all stages of the life cycle of parasitic nematodes.

In order to understand the properties of this group of unknown proteins they were extensively analyzed using bioinformatics tools available on the web as an initial move to design experiments to characterize these proteins further. Server TargetP 1.1 server [13] was used analyze the cellular localization of these proteins, which predicted these proteins to be non-excretory, non-mitochondrial and non-cytoplasmic. Therefore, the PSORT program [14] which predicts the sub cellular localization sites of proteins from their amino acid sequences was used. Results from the above analysis indicated proteins in this group to be devoid of peroxisomal targeting signal, possible vacuolar targeting motifs, RNA-binding motifs, N-myristoylation patterns, transport motifs from golgi to cell surface, DNA binding motifs and ribosomal protein motifs. However, latter analysis indicated the presence of nuclear localization signals in all UPs. Thus it is likely that they are localized in the nucleus and this finding is further strengthened by the identification of RNAP_Rpb7_N_like domain in UPs, which is generally present in ribonucleoproteins. Eukaryotic cells widely use the phosphorylation of proteins to transmit and integrate signals received from their environment and to regulate cellular functions. The nucleus disintegrates during mitosis. This is mainly driven by the mobilization of lamins and nuclear membranes by cyclin dependent protein kinases and other kinases, such as protein kinase

C, at mitotic sites and is reverted at the end of mitosis by dephosphorylation of such sites [15, 16]. Therefore, this group of the proteins was analyzed for possible sites of phosphorylation for different protein kinases using ScanProsite server [17]. Results revealed potential sites for cAMP- and cGMP-dependent protein kinase, Protein kinase C and Casein kinase II indicating a possible regulatory effect on these proteins by different kinases. Searching for the cellular role of this group of proteins using ProtFun 2.2 server [18] uniformly predicted the UPs to be non enzymatic, cell envelope and growth factor like proteins with higher probabilities of being non-enzymatic and growth factor proteins. However, transmembrane analysis of this group of proteins with the TMHMM Server v. 2.0 [19] indicated with higher certainty that this group of proteins lack regions for transmembrane alpha helical domains. With the revelation that this group of proteins to be rich in beta strands (secondary structure analyses with Jpred3 server), we analysed them using beta barrel finder programme [20], which too indicated the absence of potential regions to form transmembrane beta barrels indicating this group of proteins are likely to be globular proteins.

Conclusion:
Both the structure and function of the nematodes specific proteins described in the current study are truly unknown. However, the expression of this type of protein in all the stages of the nematode life cycle indicates importance. The divergent evolution to which this group of proteins is subjected to indicates that they may be evolving to perform species-specific functions. Taking these facts into consideration and in conjunction with other high probability of bioinformatic predictions, i.e. nuclear localization signals, numerous potential sites for phosphorylation, we hypothesize that this group of proteins is nematode specific, constitutively expressed, nuclear phosphoproteins that are undergoing divergent evolution. In addition, if this group of proteins is in fact growth factors specific to nematodes, as revealed by the bioinformatic analyses, it would have great potential as therapeutic target for the control of disease causes by parasitic nematode, as growth factors are often targeted for drug development. Finally, prior to drawing any conclusion from the bioinformatic predicted characters about this unknown nematode specific unique protein, it should be experimentally tested along the lines that Bioinformatics tools predicted, which we are currently carrying out in our laboratories.

References:
[1] M. Blaxter, _Science_. 282; 2041 (1998) [PMID: 9851921].
[2] http://en.wikipedia.org/wiki/Nematode
[3] S. I. Hirst & L. A. Stapley, _Parasitol. Today_ 16:321(2000) [PMID: 10637575]
[4] http://www.cals.ncsu.edu/pgg/dan_webpage/index.htm
[5] KC Tung et al., _J Vet Med Sci_. 65:977 (2003). [PMID: 14532689]
[6] http://blast.ncbi.nlm.nih.gov/Blast.cgi
[7] http://www.ebi.ac.uk/Tools/blast2/parasites.html
[8] http://www.nematodes.org/
[9] http://www.compbio.dundee.ac.uk/~www-jpred/
[10] http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi
[11] http://meme.sdsc.edu/meme4_1/intro.html
[12] H. Meka et al, _Nucleic Acids Res,_ 33:6435 (2005) [PMID: 16282592].
[13] http://www.cbs.dtu.dk/services/TargetP/
[14] http://psort.ims.u-tokyo.ac.jp/
[15] H.otto et al, _Eur. J. Biochem_. 268, 420 (2001) [PMID: 11168378]
[16] M. Dreger, Biochemistry 38, 9426 (1999) [PMID: 10413518]
[17] http://www.expasy.ch/tools/scanprosite/
[18] http://www.cbs.dtu.dk/services/ProtFun/
[19] http://www.cbs.dtu.dk/services/TMHMM/
[20] http://www.imtech.res.in/raghava/tbbpred/
[21] http://nematode.lab.nig.ac.jp/

Edited by P. Kangueane

Citation: Dassanayake _et al_**,** Bioinformation 3(9): 370-374 (2009)

Supplementary material

Nematode species	Stage/strain	Accession number	Source/Database
O. volvulus	molting L3 larva	BG809035	PGD
W. bancrofti	unknown	AF159281	NCBI
B. malayi	BmL3	AF100558	PGD
	Adult female	AA454307/XM_001897013	PGD/ NCBI
	Adult male	AI784822	PGD
	infective larva	AA089404	PGD
A. suum	embryo	CA954690	PGD
	female gut pAMP1 v2	BQ381163	PGD
M. hapla	egg pAMP1 v1	BM884715	PGD
L. sigmodontis	Adult female	DN556679	NEMBASE3
C. elegans	Bristol N2	NM_064827/ AV197850	NCBI/ NEXTDB
C. briggsae	strain AF16	XM_001666418	NCBI

Table 1: The accession numbers of UP nucleotide sequences of nematode species and their corresponding stages of the life cycle from which nucleotide sequence was characterized, and source/database in which sequences were identified. PDG: Parasite Genome Database, NEXTDB: The Nematode Expression Pattern DataBase [21].