Tropical Biomedicine 31(3): 514–524 (2014)

Novel parasitic nematode-specific protein of bovine filarial parasite *Setaria digitata* displays conserved gene structure and ubiquitous expression

Rodrigo, W.W.[1,2], Dassanayake, R.S.[*1], Weerasena, S.J.[2] and Silva Gunawardene, Y.I.[3]
[1]Department of Chemistry, Faculty of Science, University of Colombo, Colombo 03, Sri Lanka
[2]Institute of Biochemistry, Molecular Biology and Biotechnology, University of Colombo, 90, Cumaratunga Munidasa Mawatha, Colombo 03, Sri Lanka
[3]Molecular Medicine Unit, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka
[*]Corresponding author email: rsdassanayake@chem.cmb.ac.lk
Received 16 November 2013; received in revised form 8 March 2014; accepted 28 March 2014

Abstract. *Setaria digitata* is an animal filarial parasite, which can cause fatal diseases to livestock such as cattle, sheep, goat, buffaloes, horses etc. inflicting considerable economic losses to livelihood of livestock farmers. In spite of this, the biology and parasitic nature of this organism is largely unknown. As a step towards understanding these, we screened the cDNA library of *S. digitata* and identified an open reading frame that code for parasitic nematode-specific protein, which showed a significant homology to functionally and structurally unannotated sequences of parasitic nematodes *Wuchereria bancrofti*, *Brugia malayi*, *Onchocerca volvulus*, *Loa loa* etc., suggesting its role in parasitism. RT-PCR analysis indicated that the *S. digitata* novel gene (*SDNP*) is expressed in adult female and male, and microfilariae. Southern hybridization studies revealed that this gene is a single-copy gene. Sequence analysis of the genomic region obtained from overlapping PCR amplification indicated that the size of the genomic region is 1819 bp in which four exons encoding 205 amino acids were interrupted by three introns of varying lengths of 419, 659 and 123 bp, and also the expansion of the size of the introns of *S. digitata* compared to its orthologues by integrating micro and mini-satellite containing sequence. Sequences around the splice junctions were conserved and agreed with the general GT-AG splicing rule. The gene was found to be AT rich with a GC content of 38.1%. Bioinformatic analysis indicated that the gene structure of *SDNP* and its orthologues is conserved and it expressed ubiqutously in all the stages of nematode's life cycle. Therefore, taking these outcomes together, it can be concluded that *SDNP* is a parasitic nematode-specific, single copy gene having conserved gene structure of four exons interrupted by three introns and that the gene is expressed ubiquitously throughout nematode's life cycle.

INTRODUCTION

Setaria digitata is a cattle filarial parasite that can cause fatal neuropathological disorders known as cerebrospinal nematodiasis in sheep, goat, buffaloes and horses, affecting livestock and inflicting economic losses to people who rely on livestock farming in South East Asia and Far Asia (Tung *et al*., 2003). Furthermore, this nematode can also infect humans and cause abscesses, allergic reactions, enlarged lymph nodes, eye lesions and lung inflammation, showing its gradual adaptation to humans. On the other hand, human lymphatic filariasis caused by *Wuchereria bancrofti* which accounts for 91% of lymphatic filarial infections throughout the world is posing huge threat to human life (Rao, 2005). Owing to several obstacles including development of vector resistance to currently available drugs,

non-availability of vaccines and effective macrofilaricidals, the eradication of lymphatic filariasis caused by *W. bancrofti* has become extremely challenging and thus identification and characterization of parasitic nematode-specific genes will be one of the strategies for the development of more effective drugs and vaccine candidates for elimination of this disease. *Wuchereria bancrofti* cannot be used for this purpose due to the difficulty in procuring adult parasites residing in the lymphatic system of humans and also the lack of culture techniques for nocturnally periodic microfilariae (Murugananthan *et al.*, 2010). *Setaria cervi* is generally used as a model organism to avoid the limitations associated with *W. bancrofti* (Ahmad & Srivastava, 2007) but it is confined to certain geographical regions around the world and is not found in Sri Lanka. Therefore, *Setaria digitata* has been used successfully as a model organism for the study of genomics and the biology of filarial parasites (Murugananthan *et al.*, 2010). *Setaria digitata* closely resembles *W. bancrofti* in many aspects including morphology, histology, antigenicity and response to drugs (John *et al.*, 1995). *Setaria digitata* can be easily obtained from cattle in Sri Lanka, thus it provides an excellent model system for the present study. This study was undertaken as an initial step toward the characterization of a novel parasitic nematode-specific gene isolated from cattle filarial parasite *S. digitata*.

MATERIALS AND METHODS

Construction of *S. digitata* cDNA library

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

Sequencing of the randomly picked colonies

Bidirectional sequencing of plasmid DNA from randomly selected colonies was carried out with vector derived primers using Thermo Sequenase™ CY^5 Dye Terminator kit and ALF express™ DNA Sequencer (Amershan-Pharmacia Biotech, Sweden). A clone designated pSDC13 containing *S. digitata* novel gene (*SDNP*) was selected and fully sequenced and deposited in NCBI database (accession number: GU222920.1).

Bioinformatic analysis of SDNP

Amino acid sequences of *W. bancrofti*, *Brugia malayi* and *Loa loa* were obtained by executing a BLASTP against the databases at NCBI using SDNP as the query sequence. Amino acid sequences of novel parasitic nematode-specific protein of *Onchocerca volvulus*, *Ascaris suum*, a *Meloidogyne hapla* and *Litomosoides sigmodontis* were retrieved by executing WU-BLAST of the Parasite Genome Database (http://www.ebi.ac.uk/Tools/sss/wublast/parasites.html., Martin *et al.*, 2008) and nucleotide-protein BLAST of NEMBASE3 (Parkinson *et al.*, 2004) using SDNP as the query sequence. Multiple alignments of amino acid sequences were carried out using ClustalW in BioEdit program. Subcellular localization of novel parasitic nematode-specific proteins was analyzed using PSORT program (Nakai & Horton, 1999). Alternative splice site prediction was carried out using splice site prediction tool of the neural network of Berkeley Drosophila Genome Project using search organism human and others (Reese *et al.*, 1997).

Collection of *S. digitata* material and extraction of RNA and DNA

Adult *S. digitata* worms were collected from the peritoneal cavity of slaughtered cattle in Sri Lanka and microfilariae were also collected from the same peritoneal fluid by centrifugation. Both the adults and microfilariae were washed with phosphate buffered saline (pH 7.4) to avoid contamination by any

cattle tissue or blood and they were used for RNA/DNA extraction. RNA was extracted from male and female worms [identified by light microscopy according to the morphological features (Rhee *et al.*, 1994 and Kim *et al.*, 2010) and also from microfilariae following crushing with liquid nitrogen and using TRIZOL® reagent (Invitrogen, USA) according to the manufacture instructions and stored at -80ºC until use. Genomic DNA of adult *S. digitata* crushed with liquid nitrogen was extracted using the method described previously (Wijesundera *et al.*, 1996) and stored at -20ºC until use.

Reverse Transcription PCR based *SDNP* expression analysis

Reverse transcription was carried out using 2 µg of total RNA, 0.25 µM of random primers (Promega, USA), 0.5 mM of dNTPs, 4 µl of 1X first strand buffer (Invitrogen, USA), 10 mM DTT, 2 units of RNaseOUT™ (Invitrogen, USA) and 10 units of Super-Script™II RT (Invitrogen, USA) in a total volume of 20 µl and incubating at 42ºC for 50 minutes. PCR amplification was performed to obtain 621 bp fragment using 2.5 µl of cDNA, 0.5 unit of *Taq* DNA polymerase (GenScript, USA), 160 µM of each dNTP (Promega, USA) and 80 pM of each primer [UNF (5' GAT ATG AAT GTG AAA ACA AGG AAA AGA 3') and UNR (5' TCA GTA ATT AAT CAA ATT CGG AAG TCT 3')] in a total volume of 25 µl. A thermal cycler was programmed and run at 94ºC for 2 minutes, followed by 35 cycles of denaturation at 94ºC for 30 seconds, annealing at 50ºC for 30 seconds and extension at 72ºC for 1 minute, with a final extension step of 72ºC for 7 minutes. Samples were loaded on 1% agarose gel with a 100 bp ladder (Promega, USA) as a marker.

Southern blot analysis of *S. digitata* genomic DNA for *SDNP* copy number

Southern blot analysis was carried out using *S. digitata* genomic DNA (5 µg), cleaved separately with fivefold excess of four different restriction enzymes *Bam*H1, *Hind*111, *Eco*R1 and *Pvu*1 and separating them by electrophoresis on a 0.8% agarose gel with a 1.0 kb ladder (GenScript, USA). The resolved DNA was then denatured by soaking the gel for 10 minutes in 0.2 M HCl and for 45 minutes in denaturing solution (1.5 M NaCl, 0.5 M NaOH) and neutralized by soaking for 30 minutes in neutralization solution (1 M Tris (pH 7.4), 1.5 M NaCl). Resulted DNA was transferred to nitro-cellulose membrane (Sigma Aldrich, USA) and hybridized with biotin labeled PCR product (probe) for 16 hours at 65ºC. The membrane was washed with 2x SSC (17.53 g of sodium chloride and 8.82 g of sodium citrate in 1 liter) containing 0.1% SDS for 10 minutes at room temperature and 0.2X SSC containing 0.1% SDS for 15 minutes at 65ºC. Then the membrane was washed two times with 2X SSC and bands were visualized by using the BCIP/NBT substrate of Biotin Chromogenic Detection Kit (Fermentas).

Biotin labeling probe was prepared using plasmid DNA of pSDC13 as a template to amplify the 621 bp PCR product using UNF and UNR primers, and PCR reaction components and parameters discussed elsewhere, except for the addition of 70 µM of Biotin-11-dUTP and 200 µM of GATC[65] mixture into the PCR mix.

Characterization of *SDNP* genomic region

Genomic region of the *SDNP* was amplified by four primer combinations of UNF /UNR, UNF/UNR2 (5'-ATT TTC TCG CTG ACC ACC AC-3'), UNF2 (5'-ATT TTC TCG CTG ACC ACC AC-3')/UNR and UNF2/UNR2 and run on 1% agarose gel to determine the size of the genomic region. Southern hybridization was carried out for the verification of the amplified products. Amplified products of DNA fragments were cloned into T/A cloning vector of pCR®2.1-TOPO® (Invitrogen) and sequencing was carried out as described elsewhere using universal M13 Forward (5' GTT TTC CCA GTC ACG AC 3') and M13 Reverse primers (5' CAG GAA ACA GCT ATG AC 3').

RESULTS

Analysis of *SDNP* expression

RNA was extracted from adult male and female worms, and microfilariae separately to examine the *SDNP* gene expression at the different life stages of *S. digitata*. After the reverse transcription of RNA samples, the cDNA was amplified by PCR with gene specific primers of UNF and UNR (Figure 1). RT-PCR analyses of *SDNP* revealed that the gene is expressed in both the adult males and females, and microfilariae.

Determination of *SDNP* copy number

To determine the *SDNP* copy number, a southern blot analysis of genomic DNA of *S. digitata* cleaved with four selected restriction enzymes was carried out by probing with a biotin labelled *SDNP*-derived PCR product (Figure 2). Southern blot analysis revealed a single hybridization band in the lanes cleaved with *Bam*H1 and *Hind*111 restriction enzymes to which *SDNP* does not have any restriction sites. Two hybridization fragments were identified for the *S. digitata* genomic DNA cleaved with *Eco*R1 and *Pvu*1 restriction enzymes to which *SDNP* had unique restriction sites within the gene.

Characterization of *SDNP* genomic region

Four primer combinations were used to PCR amplify the *SDNP* genomic region. The amplification of PCR product was not detected with the primer combination of UNF/UNR for *S. digitata* genomic DNA. However, when primer combinations of UNF/UNR2, UNF2/UNR and UNF2/UNR2 was used, the amplification of ~1.5 kb, ~1.3 kb and ~800 bp fragments, respectively was detected for *S. digitata* genomic DNA (Figures 3 & 4). These results indicated that the approximate size of *SDNP* genomic region is 2.0 kb. pSDC13 DNA containing *SDNP*-cDNA was amplified side-by-side at the same time with the same primer combinations as positive controls which resulted in 621 bp, 392 bp, 334 bp and 100 bp PCR fragments for UNF/UNR, UNF/UNR2, UNF2/UNR and UNF2/UNR2, respectively, indicating the presence of intervening sequences in the genomic region of *SDNP*. This premise was proven

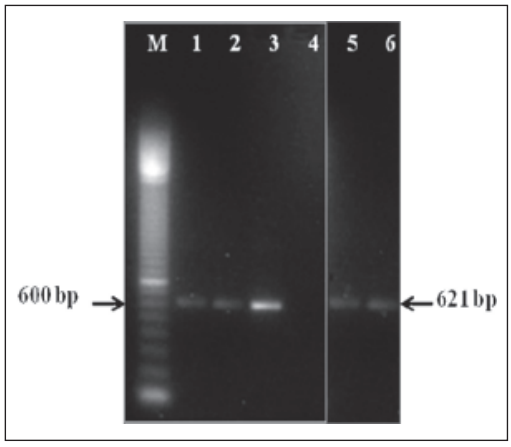


Figure 1. RT-PCR analysis of *SDNP* expression in adult male and female, and microfilariae. Lane (M): 100 bp ladder (Promega, USA); Lane (1): RT-PCR (male); Lane (2): RT-PCR (female); Lane (3): Positive control (pSDC13 DNA amplified with UNF/UNR primers); Lane (4): -Ve control and Lane (5) & (6): RT-PCR (microfilariae)

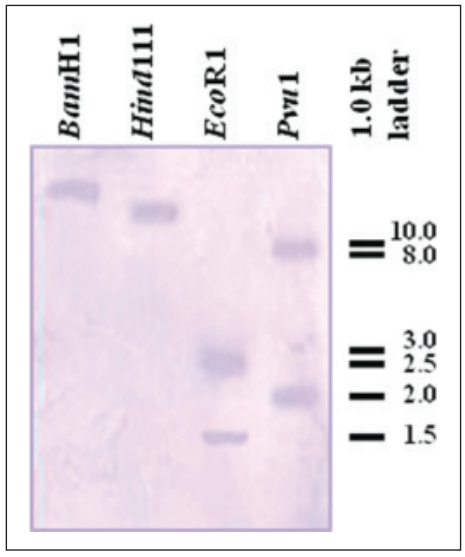


Figure 2. Southern blot analysis of *S. digitata* genomic DNA using biotin labeled probe *SDNP*-ORF. *S. digitata* genomic DNA cleaved with four restriction enzymes of *Bam*H1, *Hind*111, *Eco*R1 and *Pvu*1 and probed with *SDNP*-ORF. 1.0 kb DNA ladder (Amersham biosciences) was used as the size marker

by carrying out Southern blot hybridization analysis using amplified PCR products of *S. digitata* genome and probing with 621 bp fragment that containing almost entire *SDNP*-ORF. Southern blot hybridization indicated that the PCR amplified products from the *S. digitata* genomic DNA are the regions correspond to *SDNP* genomic region (Figure 4).

SDNP nucleotide sequence analysis

The genomic region of *SDNP* was fully sequenced and the size was estimated as 1819 bp from start codon to stop codon. The positions of the exons and introns in the genomic region of *SDNP* were identified by comparing sequence of the latter region with its corresponding coding sequence recovered from cDNA sequencing. This analysis revealed that the coding sequence of *SDNP* was interrupted by three introns at the positions of 163-581, 768-1426 and 1529-1651 (Figure 5). Sizes of the intron and exon regions of *W. bancrofti*, *B. malayi* and *L. loa* were estimated for comparison using published sequences in GenBank (Figure 6). The A+T content of each intron of *SDNP* were compared with the same content of their corresponding introns of *W. bancrofti*, *B. malayi* and *L. loa* (Table 1). Alignment of consensus sequences around the splice donor and acceptor sites of filarial parasitic nematodes is shown in Figure 7 and alternative splice sites predicted by splice site prediction tool in Berkeley Drosophila Genome Project is given in Figure 8.

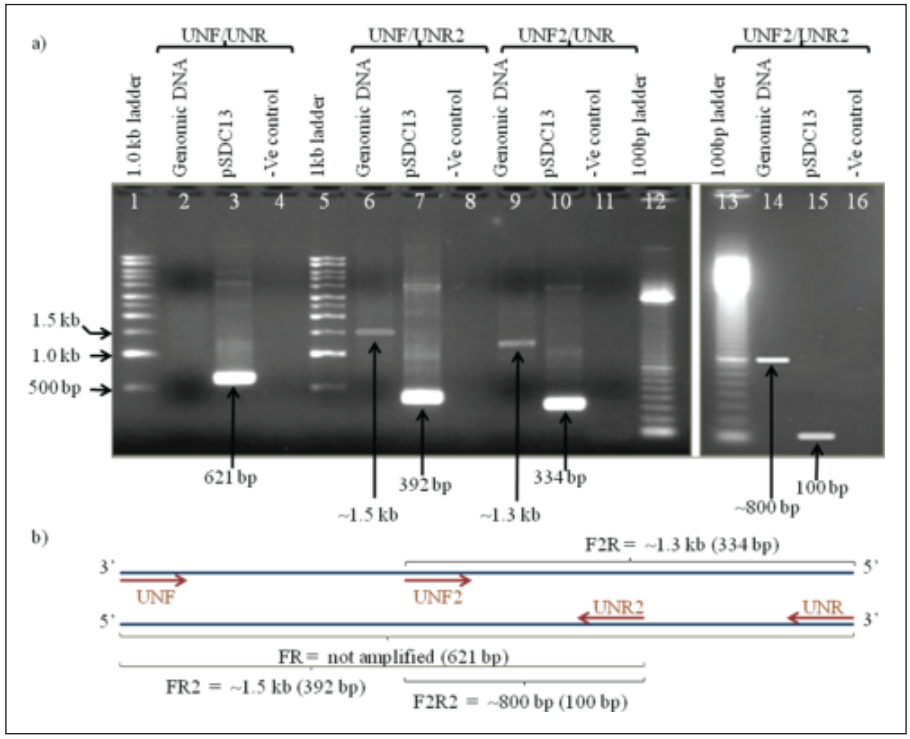


Figure 3. PCR amplification *SDNP*-cDNA (pSDC13) and corresponding genomic region of *SDNP* with different primer combinations for amplicon size comparison. Lane (1) & (5) 1.0 kb ladder (Amersham biosciences); Lane (2), (6), (9) & (14) are *S. digitata* genomic DNA PCR amplified with UNF/UNR, UNF/UNR2, UNF2/UNR & UNF2/UNR2 respectively; Lane (3), (7), (10) & (15) are cDNA (pSDC13 plasmid DNA) amplified with UNF/UNR, UNF/UNR2, UNF2/UNR & UNF2/UNR2 respectively; Lane (12) & (13) 100 bp ladder (Promega); Lane (4), (8), (11) & (16) -Ve controls of each primer combinations (A). Sizes of the PCR products amplified by genomic and cDNA with different primer combinations (B)

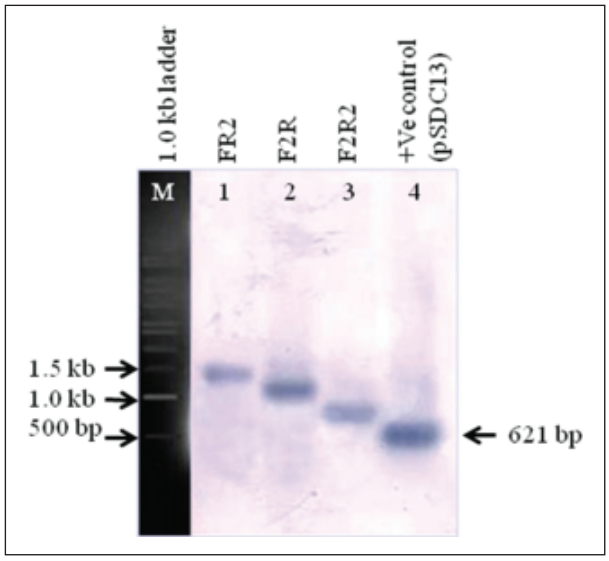


Figure 4. Southern blot analyses of amplified PCR products of *SDNP* genomic region. Lane (M): 1.0 kb ladder (Amersham biosciences); Lane (1) – (3): *S. digitata* genomic DNA amplified with UNF/UNR2, UNF2/UNR and UNF2/UNR2 primer combinations respectively and Lane (4): pSDC13 plasmid DNA amplified with UNF/UNR primers

```
   1  ATGAATGTGAAAACAAGGAAAAGATTTAATGAGGATTCAATTATTGATGATGCTAAGTCATGCGGCACTGTACTACTGAATTCTGAT
        M  N  V  K  T  R  K  R  F  N  E  D  S  I  I  D  D  A  K  S  C  G  T  V  L  L  N  S  D
  88  GCTGTTGCTTCGACTAATTCAAATAGTACAAACGAAAATGAATATGATGTTTGGCTTGTCAGGAAGCCGACGAGGgttgggaaccac
        A  V  A  S  T  N  S  N  S  T  N  E  N  E  Y  D  V  W  L  V  R  K  P  T  R
 175  actttcccagaacccacaacaactggtaccacatcccttccaaaattcacaccacacaacttaaacccacacaatcaattccttcca
 262  ttcaacacaaaacttaacacgcacaattcaccacactcaaccacacttcatccacacaaatcgggcactctgaaacacactaacaaa
 349  acaacgcattctgaccactaaaagcattcacccccggacagctacacacacacaaacaaacactaaactacaagcacacccaactacc
 436  ccatccccggccaccactcaggacagttccacaccaacaaccggcgcaccacaagaactacacccctactcacacaagacccgtggga
 523  cctgccgagaccacaccagcaggacacaaaaccaaacacccccacccccaacgccacagGTTCCATTGAGTGATTTATTGTCAATTA
                                                                  V  P  L  S  D  L  L  S  I
 610  AATTCCCCCATAAAGCAAAAAACCGAATAAGACTTGCAACTCCATCTCGATCGGATGCAGTACCACTTTACTGTCATTTTAGTCAAC
        K  F  P  H  K  A  K  N  R  I  R  L  A  T  P  S  R  S  D  A  V  P  L  Y  C  H  F  S  Q
 697  TTGCTATACCGCATGTTTATATTCCAACTACCGGTATCAGAACTCAGAAAGATGCCATGGCGCTCAAAGCCgttatttattgtttat
         L  A  I  P  H  V  Y  I  P  T  T  G  I  R  T  Q  K  D  A  M  A  L  K  A
 784  tttttctctttttatgtgtttatttttatcttttgattgtaacaggtgagatatctatgagggaatggaagtaacggtacatgtaag
 871  gtaaaagtgagaaaatacagctatgaaacatgaaggaagagcgtgtcctttgttcattgattagaacaagtgaaaatttattatcat
 958  gtgtaaaaataaaaacttgaaatttctacccatttttccaacttcttatcaagcattaaagtaatcattaacattagtattctggta
1045  aagattgtagcaagaaaaaagaaatagctgatgatcaatatattccacaataatatgatatcgtaatgcattatatgccaactctat
1132  atataattcgaaagtgcggcaggatagttgtgtggggtttgtgctgtcgatgtcggtgtgttactgttggatttttatggcggtgt
1219  ttcccatgttcgggggtcgttcccggcacgagaagtttttgtgttttgtctgtgttgtgtgttgttaattgtgtgtccgaaggcctt
1306  tgccccccccatttgtgtggggccacttgggaatttttttctctagtgttttggtttttctttgtacttttttaaatatctctttt
1393  tcaaacgagcatgtttcagcaaatttgatgaaagACAAATTTGATTAAAGGAGTTGTGGTGGTCAGCGAGAAAATTGATTTATTAGA
                                        T  N  L  I  K  G  V  V  V  V  S  E  K  I  D  L  L  D
1480  TAATGCTGTAATTTCTATAAATGATGATACTGCTATTAAAAAAGAATTCgttatattttcagcacatcttcacaacacttatcagcc
        N  A  V  I  S  I  N  D  D  T  A  I  K  K  E  F
1567  aaatttgacacgttatttataattataatagttattaccataaatttgattttaaatcttttcactttcacggtacgacgcctagGG
                                                                                          G
1654  TCCAGCGCTAGGAAATGGAACTTGTGAAATAAATTTTAAAATTAAGCCAATTCGAAAAAAGCCACATTTGCCAACGGATGATGTGAA
         P  A  L  G  N  G  T  C  E  I  N  F  K  I  K  P  I  R  K  K  P  H  L  P  T  D  D  V  K
1741  ACAGCGATTGAAACCGTTTGGAATTATTCCTAAAAAGAAGAAAAAAAGCTACAGACTTCCGAATTTGATTAATTACTGATAATAAAT
         Q  R  L  K  P  F  G  I  I  P  K  K  K  K  K  S  Y  R  L  P  N  L  I  N  Y
1828  AAGATTGTTCACGTTCTATGAATATTTGTGTTGTATGAAAATAAAACCTAAACTAGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
1915  AAA
```

Figure 5. Nucleotide sequence of *SDNP* genomic region. The sequence of exons and the 3' flanking region (obtained from cDNA sequencing) are presented in uppercase letters while the intron sequences are in lowercase letters. The predicted amino acid sequence, in single-letter terminology, is indicated below the nucleotide sequences. The polyadenylation signal AATAAA in the 3' flanking region are boxed. Start and the stop codons are underlined with single lines

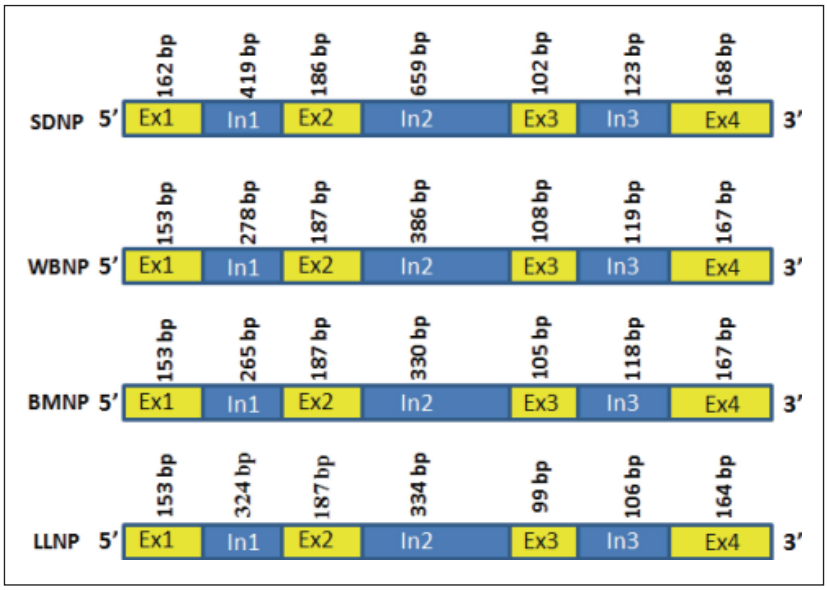


Figure 6. Comparison of genomic structure (introns and exons) of novel parasitic nematode-specific genes. *S. digitata* novel gene (SDNP), *W. bancrofti* novel gene (WBNP, ADBV01005262.1), *B. malayi* novel gene (BMNP, NW_001892970.1) and *L. loa* novel gene (LLNP, NW_003322759.1)

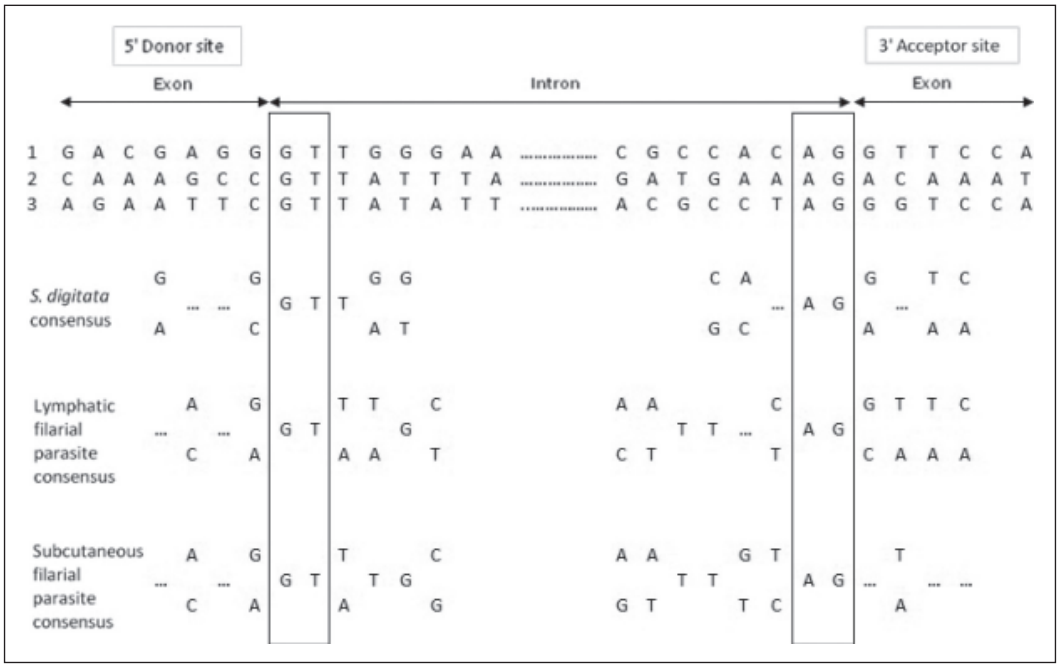
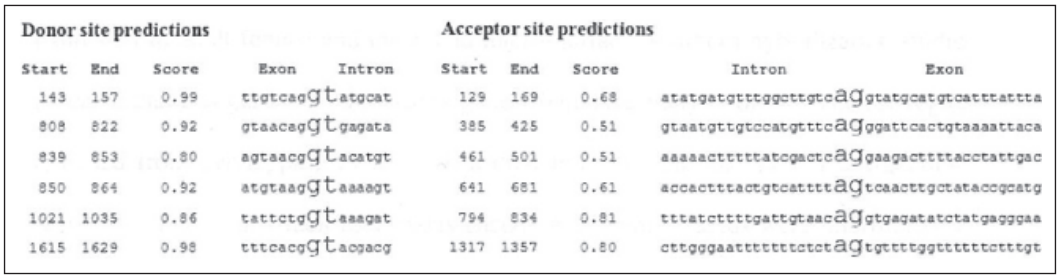


Figure 7. Alignment of the sequences around 5' and 3' splice junctions. Comparison of sequences around the 5' and 3' splice junctions of *SDNP* with the same regions of novel parasitic nematode-specific genes of lymphatic filarial parasites (*W. bancrofti*, and *B. malayi*) and subcutaneous filarial parasite (*L. loa*)

Table 1. Percentage AT nucleotide content in introns, coding region and the region between start and stop codons of novel parasitic nematode-specific genes (NPNSG)

	S. digitata (%)	*W. bancrofti* (%)	*B. malayi* (%)	*L. loa* (%)
Intron 01	49.88	66.19	66.04	66.05
Intron 02	65.71	68.91	68.18	70.06
Intron 03	70.73	70.59	72.88	70.75
Coding region	64.24	60.98	61.93	61.69
NPNSG	61.90	82.19	82.87	82.37



Donor site predictions

Start	End	Score	Exon	Intron
143	157	0.99	ttgtcag	gtatgcat
808	822	0.92	gtaacag	gtgagata
839	853	0.80	agtaacg	gtacatgt
850	864	0.92	atgtaag	gtaaaagt
1021	1035	0.86	tattctg	gtaaagat
1615	1629	0.98	tttcacg	gtacgacg

Acceptor site predictions

Start	End	Score	Intron	Exon
129	169	0.68	atatgatgtttggcttgtc	agtatgcatgtcatttattta
385	425	0.51	gtaatgttgtccatgtttc	aggattcactgtaaaattaca
461	501	0.51	aaaaactttttatcgactc	aggaagacttttacctattgac
641	681	0.61	accactttactgtcattttt	agtcaacttgctataccgcatg
794	834	0.81	tttatcttttgattgtaac	aggtgagatatctatgagggaa
1317	1357	0.80	cttgggaatttttttctct	agtgttttggtttttttctttgt

Figure 8. Alternative splice sites predicted by splice site prediction tool in Berkeley Drosophila Genome Project. High scoring donor and acceptor splice sites were identified with donor and acceptor score cutoff 0.40 (exon/intron boundary shown in larger font)

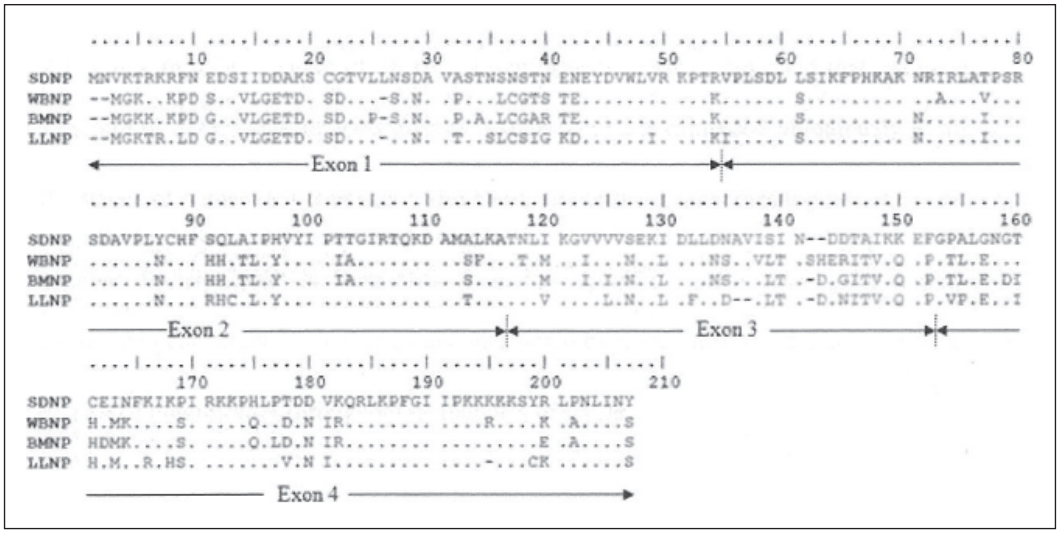


Figure 9. Multiple sequence alignments of amino acid sequences of novel parasitic nematode-specific proteins using Clustal. *W. bancrofti* novel protein (WBNP, EJW79645), *L. loa* novel protein (LLNP, XP_003146179), *B. malayi* novel protein (BMNP, XP_001897047) with *S. digitata* novel protein (SDNP). The sequences were obtained from the protein database at NCBI. Consensus sequences are indicated by dots and the deletions & insertions are indicated by hyphens. Exons (1-4) are shown in different colours

SDNP amino acid sequence analysis

The coding sequence of the *SDNP* encodes a protein of 205 amino acids which showed 66% identity with *W. bancrofti* hypothetical protein (WUBG_09446), 67% identity with *L. loa* hypothetical protein (LOAG_10608)

and 65% identity with *B. malayi* hypothetical protein (Bm1_27980) when similarity search was carriedout over NCBI database and these sequences were multiply aligned and compared (Figure 9). Low degree of sequence identity to *SDNP* was also found for the TAG-267 of *Ceanorhabditis elegans* (22%), *Ceanorhabditis briggsae* (24%) and *Ceanorhabditis remanei* (23%) protein. The PSORT program which predics the sub cellular localization sites of proteins from their amino acid sequences indicated the presence of potential nuclear localization signals in all these proteins.

DISCUSSION

The infection by filarial parasitic nematodes is one of the major causes of morbidity and mortality among humans and their domestic livestock. Despite its medical and economic importance, biology of these nematodes at molecular level is poorly understood. Characterization of parasite nematode-specific genes that their genomes encode, paves the way to understand the biology behind parasitism. Therefore, this study was undertaken to characterize a novel parasitic nematode-specific ORF identified though the analysis of expressed sequence tag of *S. digitata*.

Comparative analysis of the genomic sequences with cDNA sequences showed that the genomic region of *SDNP* containing four exons and the open reading frame (618 bp) was interrupted by three introns (Figure 5). The sizes of exonic regions of *SDNP* are almost similar to the corresponding regions of novel parasitic nematode-specific genes of human filarial parasites *W. bancrofti* (Accession no. ADBV01005262/ AF159281), *B. malayi* (NW_001892970/ XM_001897012) and *L. loa* (NW_003322759/ XM_003146131), despite the variation in intron sizes, for an instance, the second intron of *SDNP* is twice the size of the corresponding region of other three organisms. Further, all these four organisms demonstrated a higher content of A+T in their novel parasitic nematode-specific genes (excluding the 3' and 5' UTR) and it was 61.90%, 82.19%, 82.87% and 82.37% for *S. digitata*, *W. bancrofti*, *B. malayi* and *L. loa*, respectively (Table 1), displaying the characteristic AT-richness observed in nematode genomes (Glockner, 2000). On the other hand, GC content which is an important parameter of genome in the analysis of phylogenetic relationship was 38.10% in the *SDNP* genomic region and 35.76% in the coding region. Alignment of sequence around the 5' (donor) and 3' (acceptor) splice junctions of *SDNP* with novel parasitic nematode-specific genes identified the universal GT-AG splicing rule of the eukaryotes (Hammond & Bianco, 1992). The lymphatic filarial parasites *W. bancrofti* and *B. malayi*, subcutaneous filarial parasite *L. loa* and cattle filarial parasite *S. digitata* were shown to have their own consensus sequences around the splicing junctions (Figure 7). This consensus would be useful to identify exon-intron junctions of other genes of these organisms as well as of other human filarial parasites that may be studied in the future.

Dinucleotide and trinucleotide microsatellites [$(TA)_n$, $(CA)_n$, $(AT)_n$, $(GT)_n$, $(AC)_n$, $(CC)_n$, $(TG)_n$, $(CAA)_n$, $(AAT)_n$, $(CAC)_n$, $(CCA)_n$, $(CAG)_n$, $(ACG)_n$, $(ACA)_n$] were most prominant satallite sequences identified throughout the introns of *S. digiata*, which was not seen in the introns of other patasitic nematode analyzed in this study. Microsatellites are the most popular source of molecular markers for studying population genetic variation in eukaryotes (Castagnone-Sereno *et al.*, 2010). Therefore, these microsatellites might be useful to study the genetic deversity amongst *S. digitata* isolates. The strict conservation of gene structure (exon/intron numbers) amongst the human and cattle filarial nematodes suggests the requirement of the latter for the functional role/s of novel filarial parasite-specific protein in these nematodes.

The codon usage pattern of SDNP revealed that only 55 codons were utilized out of the 61 codons. Certain amino acids

showed a strong preference for a specific codon, AUU was used 63.1% in isoleucine, AAU was used 63.1% in asparagines, AAA was used 87.4% in lysine and GAU was used 68% in aspartic acid. The analysis of the 3' UTR revealed that there are two putative polyadenylation signals at 2 bp and 48 bp downstream from the stop codon respectively (Figure 5) and one located 48 bp has been used as polyadenylation signal. Since, it has been reported that there are two potential putative alternatively spliced forms for *B. malayi* novel parasitic nematode-specific protein (Accession no. XP_001897047 and XP_001897048), the possible sites for alternative splicing was investigated using splice site prediction tool in Berkeley Drosophila Genome Project. As this analysis identified potential splice acceptor and donor sites other than sites identified (Figure 8), the ability of *SDNP* to form alternatively splicing isoforms cannot be excluded. Southern blot analyses of genomic DNA revealed one hybridization band for *Bam*H1 and *Hind*111 and two hybridization bands for *Eco*R1 and *Pvu*1 due to the absence of former restriction sites and the presence of latter restriction sites in *SDNP* indicating the novel parasitic nematode-specific gene of *S. digitata* is a single copy gene.

The gene expression studies carried out using RT-PCR indicated that *SDNP* expressed in both sexes and the microfilariae of *S. digitata* showing its functional roles in these stages. The similarity search over the nucleotide, EST and Parasitic Genomes Database identified homologous sequences from human filarial parasitic nematodes *W. bancrofti*, *B. malayi* and *L. loa*, and African river blindness nematode, *O. volvulus*, pig intestinal parasite, *A. suum* and mouse filarial worm, *L. sigmodontis* (Dassanayake *et al.*, 2009). Investigation on EST databases revealed that homologous counterpart of *SDNP* expressed in all stages in the life cycle of *B. malayi* (adult male, female, molting L_3 larva and infective larva) and embryo and adult female of *A. suum*, egg of *M. hapla*, L_3 larva of *O. volvulus* and unknown stages of *W. bancrofti*. Therefore, with the similarity of amino acid sequences and the presence of potential nuclear localization signal, we can assume that *SDNP* is an orthologue of novel genes of parasitic nematodes and they may perform similar functions in their respective hosts. Therefore, revelation of the biological role of this gene may help to understand the role of homologues in human filarial parasites and other parasitic nematodes.

Furthermore, BLASTP search over the databases at NCBI identified sequences from free living nematodes *C. elegans*, *C. briggsae*, and *C. remanei* to have moderate sequence similarities with *SDNP* in global alignment with insertions and deletions, and also to have a potential C-terminal nuclear localization signal as seen in the parasitic nematode- specific proteins. These proteins are annotated as putative TAG 267 proteins of the above mentioned nematodes [UniProt identifiers: Q23205 (*C. elegans*), E3MLB9 (*C. remanei*), A8XLK2 (*C. briggsae*)]. Therefore, it is tempting to speculate whether the TAG 267 proteins are the true orthologues of novel parasitic nematode-specific proteins in free living nematodes despite that they have undergone significant divergence perhaps since speciation. The TAG 267 proteins have been believed to perform important functional roles in embryonic development ending in birth or egg hatching, growth, hermaphrodite genitalia development, morphogenesis of an epithelium, nematode larval development and positive regulation of growth rate function in Gene Ontology (http://refgene.com/gene/175220).

In conclusion, *SDNP* is a parasitic nematode-specific single copy gene that is expressed ubiquitously in all the stages of nematode's life, having a gene structure of four exons open reading frame interrupted by three introns which is conserved amongst the filarial nematodes and the protein products of such genes often be useful in developing therapeutic agents for filarial infections as its present only in parasitic nematodes.

Acknowledgements. This work was supported by a grant (SIDA/2006/BT/04) awarded by National Science Foundation of Sri Lanka and Research Grant of University of Colombo, Sri Lanka.

REFERENCES

Ahmad, R. & Srivastava, A.K. (2007). Biochemical composition and metabolic pathways of filarial worms *Setaria cervi*: search for new antifilarial agents. *Journal of Helminthology* **81**(3): 261-280.

Castagnone-Sereno, P., Danchin, E.G., Deleury, E., Guillemaud, T., Malausa, T. & Abad, P. (2010). Genome wide survey and analysis of microsatellites in nematodes, with a focus on the plant-parasitic species Meloidogyneincognita. *BMC Genomics* **11**: 598. doi: 10.1186/1471-2164-11-598.

Dassanayake, R.S., Rodrigo, W.W.P., Karunanayake, E.H., Weerasena, O.V.D.S.J. & Chandrasekharan, N.V. (2009). A putative nuclear growth factor-like globular nematode specific protein. *Bioinformation* **3**: 370-374.

Glockner, G. (2000). Large scale sequencing and analysis of AT rich eukaryotic genomes. *Current Genomics* **1**: 289-299.

Hammond, M.P. & Bianco, A.E. (1992). Genes and genomes of parasitic nematodes. *Parasitology Today* **8**(9): 299-305.

John, L., Bright, J.J. & Raj, R.K. (1995). Biological activity and diagnostic use of detergent soluble antigens from *Setaria digitata*. *Journal of biosciences* **20**(1): 69-81.

Kim, N.S., Kim, H.C., Sim, C., Ji, J.R., Kim, N.S. & Park, B.K. (2010). Congenital infection with *Setaria digitata* and *Setaria marshalli* in the thoracic cavity of a Korean calf: a case report. *Veterinarni Medicine* **55**(6): 275-280.

Martin, J., Abubucker, S., Wylie, T., Yin, Y., Wang, Z. & Mitreva. M. (2008). Nematode.net update 2008: improvements enabling more efficient data mining and comparative nematode genomics. *Nucleic Acids*: D571-8. doi: 10.1093/nar/gkn744. Epub.

Muruganathan, A., Karunanayake, E.H. & Tennekoon, K.H. (2010). Cloning and characterisation of alkali myosin light chain gene (mlc-3) of cattle filarial parasite *setaria digitata*. *The IIOAB Journal* **1**(4): 1-10.

Nakai, K. & Horton, P. (1999). PSORT: a program for detecting the sorting signals of proteins and predicting their subcellular localization. *Trends Biochemical Science* **24**: 34-35.

Parkinson, J., Whitton, C., Schmid, R., Thomson, M. & Blaxter, M. (2004). NEMBASE: a resource for parasitic nematode ESTs. *Nucleic Acids Research* **32**: D427-30.

Rao, R.U. (2005). Endosymbiotic Wolbachia of parasitic filarial nematodes as drug targets. *Indian Journal of Medical Research* **122**: 199-204.

Reese, M.G., Eeckman, F.H., Kulp, D. & Haussler, D. (1997). Improved Splice Site Detection in Genie. *Journal of Computational Biology* **4**(3): 311-323.

Rhee, J.K., Choi, E.Y., Park, B.K. & Jang, B.G. (1994). Application of scanning electron microscopy in assessing the prevalence of some Setaria species in Korean cattle. *The Korean Journal of Parasitology* **32**(1): 1-6.

Tung, K.C., Lai, C.H., Ooi, H.K., Yang, C.H. & Wang, J.S. (2003). Cerebrospinal Setariosis with *Setaria marshalli* and *Setaria digitata* infection in Cattle. *Journal of Parasitology* **65**(9): 977-983.

Wijesundera, W.S.S., Chandrasekharan, N.V., Karunanayake, E.H. & Dharmasena, S.P. (1996). Development of a diagnostic DNA probe to detect *Setaria digitata*: The causative parasite of cerebrospinal nematodiasis in goats, sheep and horses. *British Veterinary Journal* **152**(5): 561–571.