

Identification and characterization of conserved miRNAs with its targets mRNA in *Trichinella Spiralis*

Dyavegowda Padmashree & Narayanaswamy Ramachandraswamy*

Department of Biochemistry, Central College Campus, Bangalore University, Bangalore, Karnataka, India; Narayanaswamy Ramachandraswamy - E-mail: drnrswamy@bub.ernet.in; Corresponding author; Telephone number: +91 080 22961335

Received June 16, 2016; Revised July 10, 2016; Accepted July 11, 2016; Published August 15, 2016

Abstract:

microRNAs (Small regulatory non-coding RNAs) have an important role in gene regulation and evolutionarily conserved molecules. *Trichinella spiralis* infect majority of species. Therefore, it is of interest to identify conserved miRNAs and their targets using sequences from EST, GSS and full length nucleotides obtained from NCBI against previously reported worm miRNAs. We identify 11 novel miRNAs in *T. spiralis* by using bioinformatics-homology based search. In addition, we predicted target mRNA genes form complementary base pair in seed region of miRNAs. Further, gene annotation using Uniprot shows that these target genes of miRNAs are involved in various metabolism, enzymatic activity and constituents of membrane components.

Keywords: miRNA, computational, *T. spiralis*, prediction, annotation

Background:

MicroRNAs (miRNAs) are endogenous small, single stranded, non-coding RNA (~19-25 nucleotides) molecules that regulate gene expression at post-transcriptional level either by mRNA cleavage or repressing translation [1, 2]. Larger number of miRNAs have been identified and characterized in various organisms such as animals, plants, viruses and in parasites [3-5]. miRNA sequences are highly conserved across the species [6, 7]. Therefore, it is of interest to identify and characterize conserved miRNA in *T. spiralis* by using computational analysis.

Nematode of the genus *Trichinella* are zoonotic parasites which are distributed Worldwide [8, 9]. Different species in the genus *Trichinella* have been identified. *Trichinella spiralis* (*T. spiralis*) is most prevalent and pathogenic disease causing nematode in human. It causes food-borne diseases called trichinellosis which infect majority of species involving human [10]. Human trichinellosis has been documented in 55 countries, hence considered as an emerging or re-emerging diseases [11-13]. Infection occurs by consumption of raw or undercooked meat contaminated with infective muscle larvae of the parasite [14, 15]. After ingestion muscle larvae, releases from their capsules and gets activated into the intestinal infective larvae and then ISSN 0973-2063 (online) 0973-8894 (print)

penetrate into the host intestinal epithelium where they molt and mature to adult worm [16]. The female worm reproduce the new born larvae, they are passed into tissue, enter the lymphatic system and travel through the bloodstream into the muscle fibre, once the worms reach the muscular tissue it gets induced into encapsulated nurse cell [17, 18] which can survive in host for years. They then wait for a new host infection, thereby the new generation begins. However no miRNA has been reported in miRBase for this genus which provided new opportunity for predicting conserved novel miRNAs and their putative target genes. Identification of miRNAs and the function of target genes in *T. spiralis* may provide the idea to know the mechanism of infection and their metabolic processes.

Methodology:

Collection of sequences and reference miRNAs set

T. spiralis EST (expressed sequence tag), GSS (genome survey sequence) and nucleotide sequences were retrieved from NCBI (<http://www.ncbi.nlm.nih.gov/>), used for miRNA prediction. A total of 25,368 sequences were deposited in EST database (dbEST) (<http://www.ncbi.nlm.nih.gov/dbEST/>), 1 GSS and 35,787 nucleotides respectively. To identify the potential conserved miRNA in *T. spiralis*, previously known mature miRNA

sequences belonging to the phylum Platyhelminthes, Nemertea, Annelida and Nematoda were retrieved from miRBase database Release 21 (<http://www.mirbase.org/cgi-bin/browse.pl>) [19]. These mature miRNA sequences were used as reference miRNAs set to search the conserved miRNAs in *T. spiralis*.

Computational tools

The freely available BioEdit software [20] is used for local alignment tool BlastN search, used to predict conserved miRNA in *T. spiralis* using EST, GSS and nucleotides. The sequences were retrieved from NCBI and saved in FASTA format. Commercial

available Geneious software version 7.1.5 were used for pairwise and multiple alignments with default parameter, in addition, it contain RNAfold program [21] to predict secondary structure of pre-miRNAs.

Nomenclature of miRNAs

The predicted miRNA in *T. spiralis* were named according with miRBase [22]. The mature miRNA sequence are designed with 'miR', and precursor hairpin are labelled as 'mir' with the prefix in three letter 'tsp' for *Trichinella spiralis*, first letter from genus and first two letter from species.

tsp-miRNAs	Length	Mature miRNA sequences	(G+C) %	MFE (kcal/mol)	AMFE	MFEI
9	120	UCUUUGGUUAUCUAGCUGUAUGA	34.8	-45.7	-38.9	1.0
34	71	UGGCAGUGUAAUUAGCUGGUUG	45.1	-32.7	-46.0	1.0
72	120	AGGCAAGAUGUUGGCAUAGCUGA	51.7	-45.9	-38.2	0.7
79	120	AUAAAGCUAGGUUACCAAAGCUA	44.2	-43.6	-36.3	0.8
86	120	UAAGUGAAUGCUUUGCCACAGACU	44.2	-43.6	-36.3	0.8
100	84	AACCCGUAGAUCGGAACUUGUGUU	54.8	-29.9	-35.6	0.6
124	123	UAAGGCACGCGGUGAAUGCCAA	51.2	-47.2	-38.3	0.7
125	91	UCACAAGUUUGGUCUCAGGGA	48.4	-28.3	-31.1	0.7
153	108	CAUCACUUUUGUGACUAUGCAA	44.4	-35.2	-32.6	0.7
252	120	CUAAGUAGUAGUGCCGCAGGUCA	49.2	-46.5	-38.8	0.8
263	82	GCCGUGAAUUCAUCCAGUGCCAUU	45.1	-29.7	-36.21	0.8

Table 1: Predicted characterized features of precursor miRNA in *T. spiralis* are given.

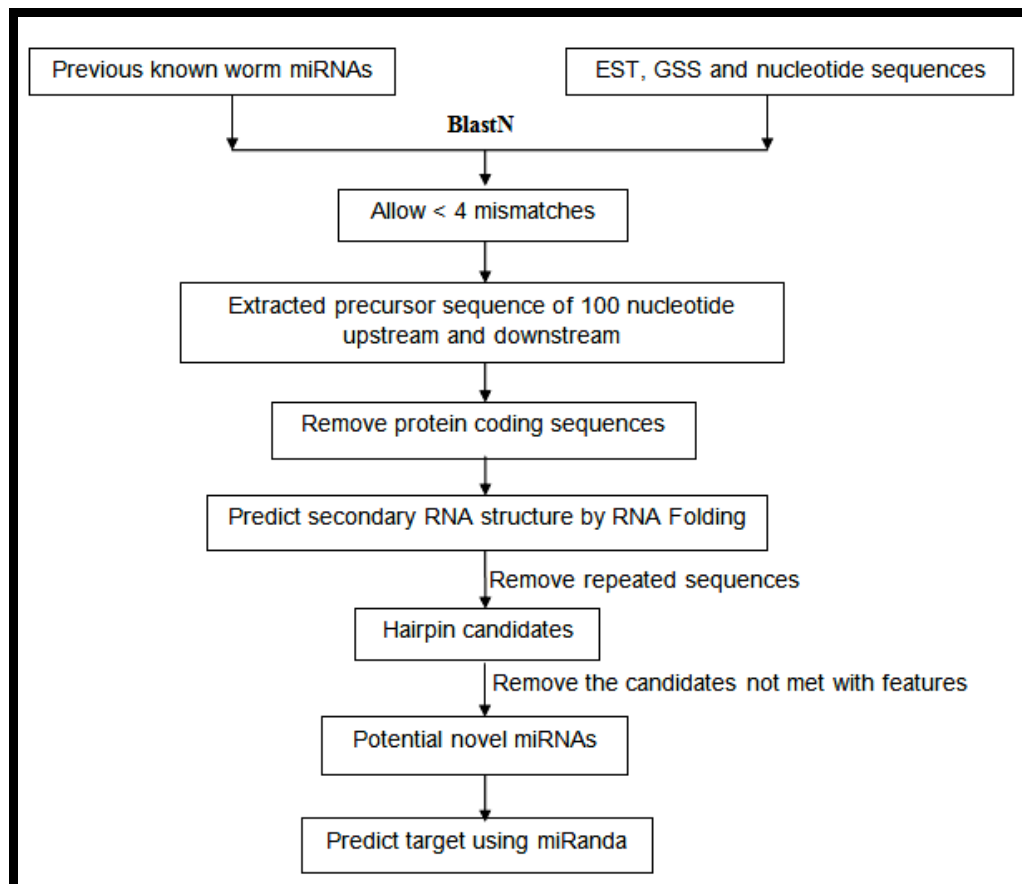


Figure 1: Flowchart for identification and characterization of miRNA in *Trichinella spiralis*.

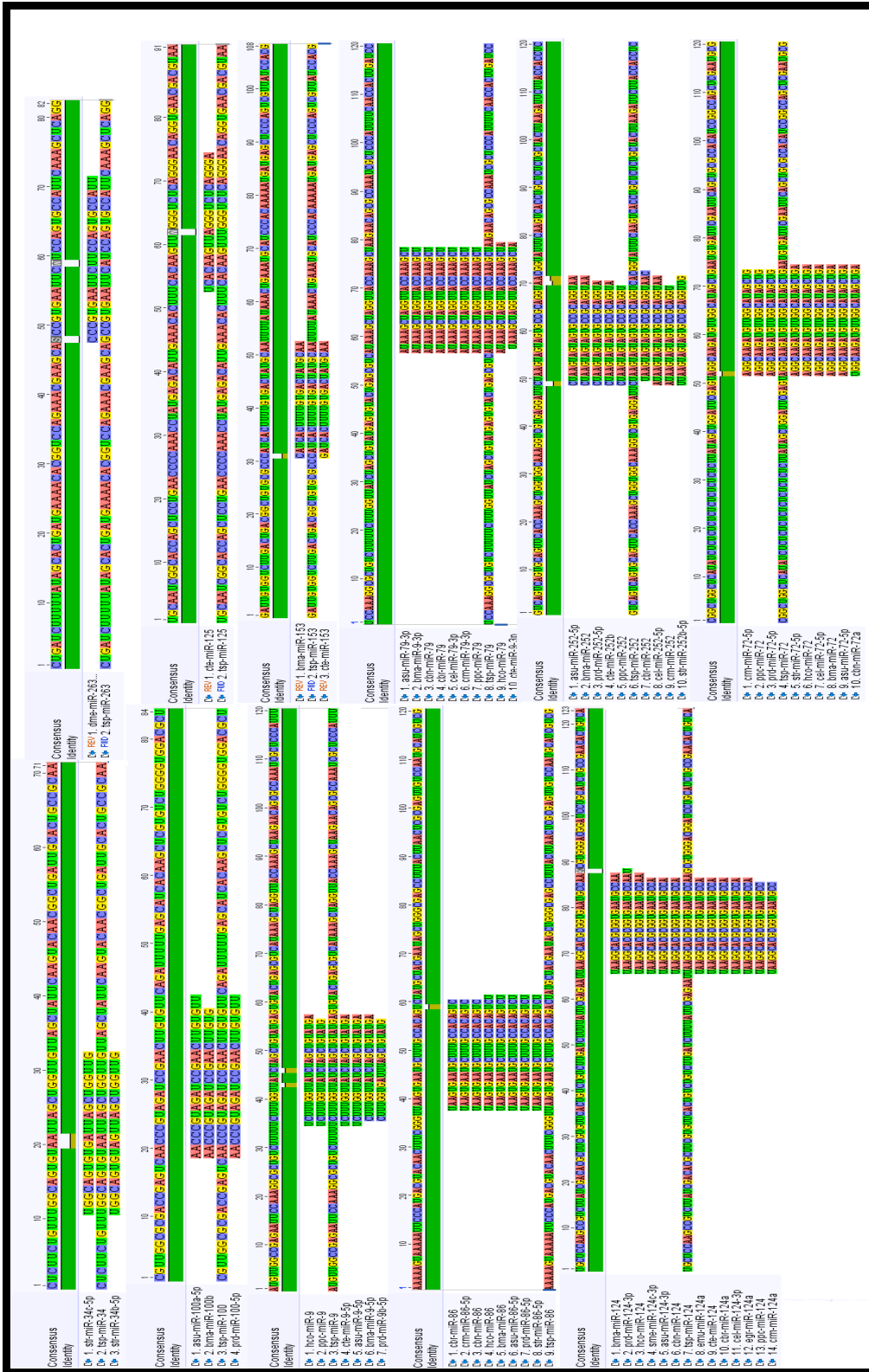


Figure 2: Sequence alignment of mature miRNA of *T. spiralis* and mature miRNA of worm database. The sequences are more than 90% conserved. cte-Capitella teleta, prd-*Panagrellus redivivus*, ppc-*Pristionchus pacificus*, bma-*Brugia malayi*, asu-*Ascaris suum*, cel-*Caenorhabditis elegans*, cbr-*Caenorhabditis briggsae*, emu-*Echinococcus multilocularis* str-*Strongyloides ratti*, crm-*Caenorhabditis remanei*, hco-*Haemonchus contortus*, sme-*Schmidtea mediterranea*, emu-*Echinococcus multilocularis* str-*Strongyloides ratti*, egr-*Echinococcus granulosus*, cbr-*Caenorhabditis breneri*, dme-*Drosophila melanogaster* and lsp-*Trichinella spiralis*.

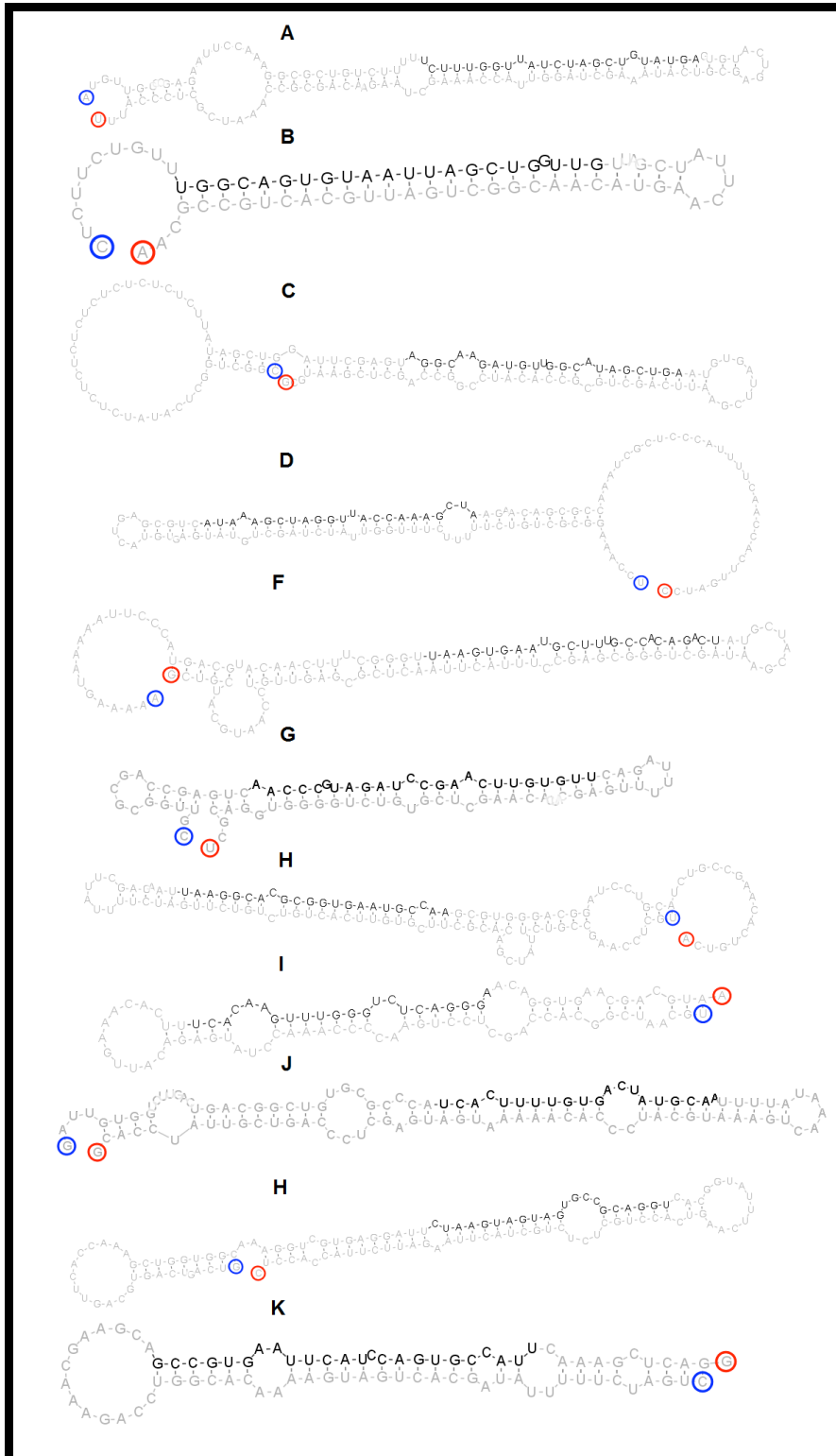


Figure 3: Predicted hairpin secondary structures of newly identified miRNAs in *T. spiralis*. The mature miRNA sequences are highlight in dark black colour. miR-9; B. miR-34; C. miR-72; D. miR-79; E. miR-86; F. miR-100; G. miR-124; H. miR-125; I. miR-153; J. miR-252 & K. miR-263

ISSN 0973-2063 (online) 0973-8894 (print)

Identification and prediction of precursor microRNAs

Procedure for identifying potential conserved novel miRNAs is summarized in **Figure 1**. The following steps were considered: local nucleotide BLAST were run with the reference miRNAs set sequences searched against *T. spiralis* sequences at low e-value threshold using BioEdit software. The results of BLASTN were further analyzed by applying two conditions: 1) at least 19-24 nucleotides length were assumed between the predicted mature miRNAs, more than 90% identity between the sequences. 2) Mismatches between the sequences with 0-4 nucleotide are allowed with reference set of worm mature miRNAs.

The precursor sequences were searched at 50 nucleotides upstream and downstream from the location of mature miRNAs. The following conditions are considered to predict miRNA sequences according to Zhang [23]. 1) The minimum length of pre-miRNA to be 60 nucleotides. 2) The precursor miRNA sequences can fold into an appropriate hairpin secondary structure, the mature miRNA sequence site may be in any one arm of hairpin structure. 3) Not more than 6 mismatches between predicted mature miRNA and the opposite miRNA* sequence in the hairpin structure. 4) No breakage or loop in mature miRNA and miRNA* sequences. 5) Predicted secondary structures had retained minimum free energy (MFE) (≤ -18 kcal/mol). 6) 30-70% of G+C content.

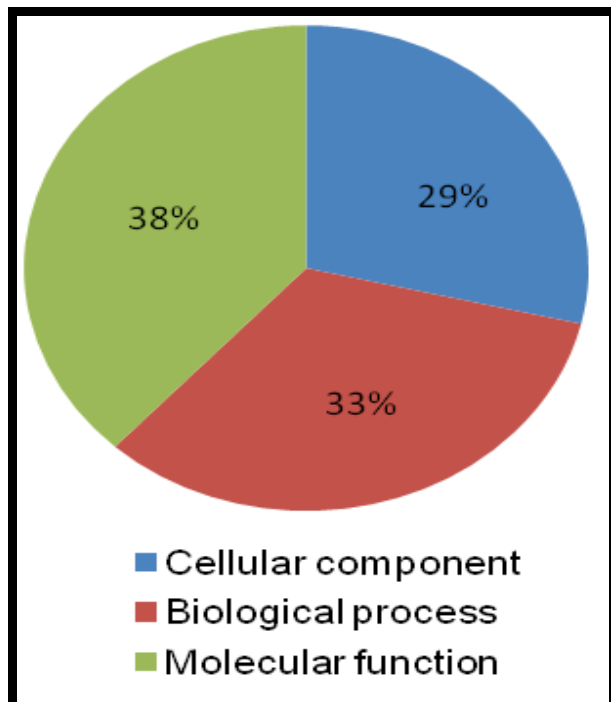


Figure 4: Pie chart showing Functional categories (cellular component, biological process and molecular function) of predicted target genes of novel miRNAs in *T. spiralis*.

miRNAs target prediction and annotation

Totally 16,589 gene sequences are available in NCBI for *T. spiralis*. To search the targets we prefer miRanda software version 3.3a

[24], released 2010 available at <http://www.microrna.org/miranda>. The program used thermodynamics and dynamic programming alignments, along with statistical parameters were followed in these studies with the MFE threshold -20kcal/mol, Score Threshold 120, Scaling Parameter 4 for complementary nucleotide match score. Outputs were subsequently screened out to reduce false positive results. In addition, function of predicted target genes are allowed to understand regulatory network in the cellular, biological and molecular processes using gene annotation Uniprot database (www.uniprot.org).

Results and Discussion:

In silico prediction of miRNAs in *T. spiralis*

Different computational approaches were used to identify miRNA in both animal and plants [25-27]. In the present study, we identify the conserved miRNAs in *T. spiralis* using EST, GSS and nucleotides (DNA and RNA sequences) by comparative genomic approach. The collected sequences from these sources were retrieved and saved in FASTA format. In BioEdit software we created local blast for nucleotide database as reference miRNAs set which contains all known worm mature miRNA sequences retrieved from miRBase, then BlastN program were performed against the *T. spiralis* sequences with default parameter and low e-value threshold. The resulted output were filtered by applying two conditions 1) at least 18-24 nucleotides length were assumed between the predicted mature miRNAs, more than 90% identity between the sequences. 2) Mismatches between the sequences with 0-4 nucleotides are allowed with reference sets of worm mature miRNAs. **Figure 2** shows the high degree of sequence similarity between the predicted mature miRNA sequences and previously known miRNAs. Gap in alignment is due to insertion or deletion of the nucleotides. Further, the precursor sequences were searched at 50 nucleotides upstream and downstream from the location of mature miRNA. Since miRNA are non-coding nucleotides, we performed BLASTX 2.2.30 search against non-redundant (NR) protein databases in order to discard the protein coding miRNA candidates. The remaining precursor miRNA candidates were analysed to predict secondary structure by using RNAfold. **Figure 3** shows precursor miRNA and mature sequences in stem loop structures. All the mature miRNA were found in the stem portion of hairpin structures containing less than 6 mismatches in the opposite arm without break or loop inside the mature sequences. Mature sequence resides either in 3' arm or in 5' arm of the secondary structures. In our studies we considered from 60-120 nucleotides and 19 to 24 nucleotides of precursor miRNAs and mature miRNA sequences respectively. Within the genome, repeated sequences were removed, finally we identified 11 potential novel miRNAs namely *tsp-miRNA-9*, 34, 72, 79, 86, 100, 124, 125, 153, 252 and 263 respectively.

Characterization of *T. spiralis* miRNA

In our study, we could able to identify some of the features of precursor miRNA, such as MFE, have a predominant role to determine the secondary structure of nucleic acids. Lower the MFE, higher the stability of the structure [28]. Adjusted minimal

free energy (AMFE), is calculated by (MEF/Length of RNA sequence)*100, minimal free energy index (MFEI) calculated by AMFE/ (G+C) %. All these features are calculated for each miRNA candidate. MFEI is an important criteria to distinguish miRNA from other types of RNA (coding and non-coding) [23]. In our studies MFE range from -28.3 to -47.2 kcal/mol respectively, GC content (30-70%) and MFEI ranges from 0.6 to 1.0 **Table 1**.

miRNA Target genes and their function

Generally, miRNAs regulate gene expression through translational repression or mRNA cleavage. To understand the target genes of predicted miRNA candidates in *T. spiralis*, we used miRanda software with the above mentioned criteria. miRNA target sites are most commonly found in the 3'UTR of mRNA, which search for the sequence for complementary base site recognised 2-8 nucleotides (seed region) from the 5' end of the mature miRNA. After filtering certain criteria 843 targets were obtained for 11 novel miRNA candidates.

Further the functional information of each miRNA target genes were searched using Uniprot database. Gene annotations were functionally categorized into three parts: cellular component, biological process and molecular functions. In our analysis around 38% are contributed for molecular function, 33% and 29% responsible in biological process and cellular component respectively **Figure 4**. The target genes in the biological category are highly involved in various metabolism and signal transduction, in the cellular process mainly the component of the membranes followed by nucleus and in the molecular process highest percentage in enzymatic activity, nucleic acid and ion binding. The predicted miRNA candidate target genes were found in various regulatory roles during growth and developments. Our studies help us to understand important regulatory network and roles of miRNAs in *T. spiralis*.

Conclusion:

T. spiralis is a food-borne zoonotic nematode causative agent of trichinellosis in humans. Our study is the first attempt to identify and characterize conserved novel miRNAs with their target mRNA in *T. spiralis*. In this report we identify 11 novel miRNA candidates which were evolutionarily conserved with previously known miRNAs registered in miRBase. Analysis of miRNA targets within the genome of *T. spiralis* helps to understand the regulatory function of each miRNA candidate. In addition, gene annotation analysis shows that most of the target transcripts are involved in various metabolism, enzymatic activity and constituents of membrane components. Overall this helps to understand conserved miRNAs in evolution, genome organization and gene regulatory networks in *T. spiralis*. The

predicted result from a computational approach would be the next step in the future to guide experimental methods to design, which would shed more light into the pathogenesis-related control, in better understanding of prevention of infection.

References:

- [1] Bartel DP, *Cell*. 2004 **116**:281. [PMID: 14744438]
- [2] Pillai RS *et al. Trends in cell biology*. 2007 **17**:118. [PMID: 17197185]
- [3] Zhang B *et al. J Cell Physiol*. 2007 **210**:279. [PMID: 17096367]
- [4] Grundhoff A *et al. RNA*. 2006 **12**:733. [PMID: 16540699]
- [5] Xue X *et al. PLoS ONE* 2008 **3**:e4034. [PMID: 19107204]
- [6] Wheeler BM *et al. Evol Dev*. 2009 **11**:50. [PMID: 19196333]
- [7] Weber MJ, *FEBS J*. 2005 **272**:59. [PMID: 15634332]
- [8] Dupouy-Camet J, *Vet Parasitol*. 2009 **159**:194. [PMID: 19054620]
- [9] Dupouy-Camet J, *Vet Parasitol*. 2000 **93**:191. [PMID: 11099837]
- [10] Pozio E, *Vet Parasitol*. 2007 **149**:3. [PMID: 17689195]
- [11] Murrell KD *et al. Emerg Infect Dis*. 2011 **17**:2194. [PMID: 22172230]
- [12] Cui J *et al. Vet Parasitol*. 2013a **194**:133. [PMID: 23422779]
- [13] Dupouy-Camet J, *Vet Parasitol* 2009 **159**:194. [PMID: 19054620]
- [14] Campbell WC *et al. Ed Istituto Superior di Sanita Press Rome, Italy*. 1993.
- [15] Krol-Pakulska E *et al. Przegl Epidemiol*. 2008; **62**:553. [PMID: 19108518]
- [16] Gagliardo LF *et al. Infect Immun*. 2002; **70**:1853. [PMID: 11895947]
- [17] Polvere RI *et al. Exp Parasitol*. 1997; **86**:191. [PMID: 9225769]
- [18] Despommier DD, *Parasitol Today* 1998; **14**:318. [PMID: 17040798]
- [19] Griffiths-Jones S *et al. Nucleic Acids Res*. 2008 **36**: D154. [PMID: 17991681]
- [20] Hall TA, *Nucleic Acids Symposium Series*. 1999 **41**: 95.
- [21] Hofacker IL, *Nucleic Acids Res*. 2003 **31**: 3429. [PMID: 12824340]
- [22] Griffiths-Jones S *et al. Nucleic Acids Res*. 2006 **34**: D140. [PMID: 16381832]
- [23] Zhang BH *et al. Cell Mol Life Sci*. 2006 **63**: 246. [PMID: 16395542]
- [24] Enright AJ *et al. Genome Biol*. 2003 **5**:R1. [PMID: 14709173]
- [25] Lai EC *et al. Genome Biol*. 2003 **4**: R42. [PMID: 12844358]
- [26] Jones-Rhoades MW & Bartel DP, *Mol Cell*. 2004 **14**: 787 [PMID: 15200956]
- [27] Brown JR & Sanseau P, *Drug Discov Today*. 2005 **10**: 595 [PMID: 15837603]
- [28] Bonnet E *et al. Bioinformatics*. 2004 **20**: 2911. [PMID: 15217813]

Edited by P Kanguane

Citation: Padmashree & Ramachandraswamy, *Bioinformatics* 2016 **12**(5) 279-284.

License statement: This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License.