Gene 590 (2016) 317–323

Contents lists available at ScienceDirect

Gene

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

Research paper

Sequence data mining in search of hookworm (*Necator americanus*) microRNAs

Abhijeet P. Kulkarni [a,*], Smriti P.K. Mittal [b]

[a] Bioinformatics Centre, Savitribai Phule Pune University, Pune 411007, Maharashtra, India
[b] Department of Zoology, Savitribai Phule Pune University, Pune 411007, Maharashtra, India

ARTICLE INFO

Article history:
Received 10 February 2016
Accepted 29 May 2016
Available online 31 May 2016

Keywords:
microRNAs
ESTs
Computational prediction
Hookworm

ABSTRACT

The new world hookworm, *Necator americanus* is a soil-transmitted nematode responsible for Necatoriasis (a type of helminthiasis) in hosts such as humans, dogs, and cats. *N. americanus* genome and transcriptome has been sequenced and a draft assembly analysis has been published highlighting protein coding genes and possible drug target proteins. Hookworm microRNA identification, annotations and their public release is yet to be attempted. The same is evident from lack of hookworm miRNA information in related popular public nucleotide sequence repositories such as miRBase, GenBank, WormBase etc. Therefore, in the present study we addressed these issues using EST and assembled transcript sequence information of hookworm. Using computational approaches, we identified three miRNAs precursor sequences and their mature forms. We also identified their potential targets from hookworm ESTs and transcripts, and from human transcriptome. Overall, the results indicate presence of nematode specific miRNA homologs in *N. americanus* and shades light on their putative targets in worm itself and the human host.

1. Introduction

Helminthiasis (worm infection) is one of the most important infections worldwide that affects poor and deprived people. The worm species responsible for Helminthiasis in human host predominantly include roundworms (*Ascaris lumbricoides*), the whipworm (*Trichuris trichiura*) and the hookworms (the new world hookworm: *Necator americanus* and the old world hookworm: *Ancylostoma duodenale*) (Hawdon, 2014). Among these, the hookworm infection occurs through skin exposure to soil contaminated with worm larvae that penetrates the skin (Bethony et al., 2006). It has been estimated that hookworms infect around 576–740 million people, worldwide (World Health Organization [WHO], 1996). Further, most of the infected human hosts are reported to remain asymptomatic (de Silva et al., 2003; Stoltzfus et al., 1997) and the infection may persist for many years in the host and impair the physical and intellectual development.

The new world hookworm, *N. americanus* female produces 5000–10,000 eggs per day (Brooker and Bundy, 2009). After deposition on to soil, the eggs develop into larvae under favourable conditions. The stage three, L3 larvae of *N. americanus* are capable of rapid penetration into skin generally through hand and legs of the host. The larvae then migrate through dermis, enter in the bloodstream and move to the lungs within a period of 10 days (Liu et al., 1999). The lung infection is usually asymptomatic or host may develop mild cough, sore throat, and fever in some cases. Through the ciliary action of the respiratory tract, the larvae are carried to the glottis, which are swallowed and carried to the small intestine and mature further. The worms attach themselves to the mucosal layer of the small intestine which may rupture the arterioles and venules. Similarly, they also secrete hyaluronidase and neutrophil inhibitory factor which help in their intestinal adherence and protect them from host immune system, respectively (Liu et al., 1999). The worms attend the sexual maturity within 3–5 weeks and the female start producing eggs which appear in the faeces of infected host (Liu et al., 1999). Thus, intestinal blood loss is the major clinical indicator of hookworm infections (Hotez et al., 2005). Therefore, in symptomatic phase, the disease develops iron deficiency anaemia, protein malnutrition, growth and mental retardation with lethargy (Hotez et al., 2005). Further, persistent nature of worm infections indicates evolution of an adaptive molecular mechanism in worms to escape or impair host immune system (Quinnell et al., 2004). However, little is known about these mechanisms and innate immune response to metazoans in general and hookworms in particular (Finlay et al., 2014). Overall, in poor communities the infections affects physical, intellectual and economical development of infected individuals.

Abbreviations: miRNA, microRNA; EST, expressed sequence tag; WHO, World Health Organization; NCBI-nr database, National Center for Biotechnology Information Non-Redundant protein database; miRNA*, complementary counter part of miRNA (opposite miRNA); MFEI, minimum folding energy index; MFE, minimum folding energy; UTR, untranslated region; GO analysis, gene ontology analysis; NM, number of mismatch; LM, length of mature miRNA; LP, length of precursor.

* Corresponding author.
E-mail address: abhijeet@bioinfo.net.in (A.P. Kulkarni).

http://dx.doi.org/10.1016/j.gene.2016.05.039
0378-1119/

With advancement in the DNA sequencing technologies and development of next-generation sequencing methods, several sequencing efforts focusing on health care, improved diagnosis and precise disease treatments have been commenced. Sequencing projects are also undertaken to address important questions on biology of infectious agents (Lv et al., 2015). In the same research lines, *N. americanus* genome and transcriptome has also been sequenced and a draft assembly analysis has been published highlighting protein coding genes and possible drug target proteins (Tang et al., 2014).

microRNAs (miRNAs), an important class of non-coding RNA, are small 21–25 bp in length and are known to regulate gene expression program of cells post-transcriptionally (Ha and Kim, 2014). High-throughput miRNA sequencing is also an important research tool used for better understanding of miRNA mediated regulation of gene expression in cells and organisms (Ha and Kim, 2014). In case of infectious agents, understanding post-transcriptional regulation of gene expression and role of miRNA in the process certainly helps to better understand disease biology and holds a great therapeutic promise. Therefore, apart from experimental methods, homology based computational prediction of miRNAs in organisms of varied interest is regularly been carried out (Wang et al., 2004; Zhang et al., 2006). Being members of nematoda group where miRNAs were discovered for the first time (Lee et al., 1993; Reinhart et al., 2000), the enteric parasite worms also contain miRNA genes and have well established post-transcriptional gene expression regulation programs (Zheng et al., 2013). Further, it has been reported that the worms use miRNAs as weapon by secreting them out and targeting host mRNAs which then cause translation inhibition/down regulation of targets (Britton et al., 2014). Therefore, identification of miRNAs in these enteric parasites and worms in particular is of great therapeutic importance.

N. americanus miRNAs identification, annotations and their public release has not been attempted as yet. The same is evident from lack of hookworm miRNA information in related popular public nucleotide sequence repositories such as miRBase, GenBank, WormBase etc. Therefore, using homology based computational approach, we tried to identify nematode specific miRNA homologs in *N. americanus* expressed sequence tags (ESTs) and the assembled transcript sequences. We then identified putative targets of these miRNAs in worm itself and in human transcriptome. Thus, in the present manuscript, we present the identified miRNAs, their putative targets in hookworm and human and discuss their possible role in worm biology and disease progression.

2. Materials and methods

2.1. miRNA reference set

In order to search potential miRNAs in *N. americanus*, mature miRNA sequences of various nematode species (*Ascaris suum*, *Brugia malayi*, *Caenorhabditis brenneri*, *Caenorhabditis elegans*, *Caenorabhditis remanei*, *Haemonchus contortus*, *Pristionchus pacificus*, *Panagrellus redivivus* and *Strongyloides ratti*) were curated from miRBase Database (Release 21: June 2014). During sequence extraction, only high confidence value miRNA (as indicated by miRBase database) were curated. Although some of these miRNAs were initially identified by computational approaches, a majority of them have subsequently been validated by experimental methods. Thus, a total of 2473 mature miRNA sequences were curated including redundancies across species. In order to remove the redundancy and to retain unique miRNAs, the sequences were clustered at 100% identity using uclustv1.2.22q clustering tool (Edgar, 2010). Thus a non-redundant miRNA reference set containing 2211 mature miRNAs was prepared.

2.2. N. americanus ESTs and assembled transcript sequences (test set)

N. americanus ESTs were downloaded from EST database (dbEST) available at National Center for Biotechnology Information (NCBI). A total of 6694 EST sequences were curated. Similarly, the assembled *N. americanus* transcriptome containing 9693 transcript sequences was downloaded from nematode.net FTP server available at http://nematode.net/NN3_frontpage.cgi?navbar_selection=home&subnav_selection=transcript_assembly_ftp. In order to remove redundancy, the sequence information from both the sources was pooled and subjected to clustering at 100% identity using uclustv1.2.22q clustering tool (Edgar, 2010). Thus a non-redundant test set containing 15918 sequences was prepared.

2.3. Software and servers used

For homology searches during miRNA predictions and target predictions, BLASTn was performed locally using standalone blast package v2.2.29+-1.x86_64.rpm downloaded from (ftp://ftp.ncbi.nih.gov/blast/executables/LATEST/). MirEval (http://tagc.univ-mrs.fr/mireval) was used to predict miRNA precursor (Ritchie et al., 2008). In order to remove protein-coding sequences and to retain non-protein encoding sequences, the predicted precursor sequences were subjected to BLASTx analysis against National Center for Biotechnology Information Non-Redundant (NCBI-nr) protein database (ftp://ftp.ncbi.nlm.nih.gov/blast/db/) using standalone blast package. MFOLD v3.6 was used online at (http://unafold.rna.albany.edu/?q=mfold) to analyze RNA secondary structure (Zuker, 2003).

2.4. Prediction of miRNAs

Procedure used for searching potential miRNAs in *N. americanus* is shown in Fig. 1. In brief, the non-redundant mature miRNA sequences were subjected to BLASTn analysis against *N. americanus* test sequence database. The BLASTn parameters used during analysis were as follows: The output format selected was 0 that indicates actual alignments and related parameters; Expect value was set to 1000 to increase chance of finding hit against potential sequences; low complexity sequence filter was chosen; the number of alignments were set to 250, word size match between query and database sequence was 7. Test set sequences were considered as potential miRNA candidates only if they fit in the following rules: (1) at least 18 nucleotide alignment exists between the query and subject sequence and (2) the alignment does not have more than 3 mismatches. The test sequences following these rules were considered as potential miRNAs and used for miRNA precursor prediction using MirEval software. The predicted precursors were then subjected to BLASTx analysis against NCBI-nr database for removing the protein-coding sequences.

2.5. Prediction of RNA secondary structure

The retained non-protein encoding precursor sequences (the potential miRNA sequences) were subjected to secondary structure prediction using MFOLD v3.6 (Zuker, 2003). Following rules were then applied to confirm the candidates as miRNA homologs: (1) the sequence could fold into hairpin secondary structure with an appropriate stem and terminal loop (2) the potential miRNA sequence resides completely within stem of hairpin structure (3) the number of mismatches between predicted miRNA and its opposite miRNA (i.e. miRNA*) does not exceed 6. (4) Large loop or break should not be present in the miRNA and miRNA* sequences, and (5) the predicted secondary structure should have higher MFEI and negative MFE (Wang et al., 2011).

The MFEI was calculated using the following equation:

$$MFEI = [(MFE/\text{length of the RNA sequence}) \times 100]/(G+C)\%MFE$$

denotes the negative folding free energies (ΔG).

Thus, the precursor sequences following these criteria and their mature counterparts were considered as potential miRNA of *N. americanus*. These sequences along with their nematode specific homologs were