Gene 564 (2015) 125–133

Contents lists available at ScienceDirect

Gene

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

Identification and characterization of microRNAs by deep-sequencing in *Hyalomma anatolicum anatolicum* (Acari: Ixodidae) ticks

Jin Luo [a], Guang-Yuan Liu [a,*], Ze Chen [a], Qiao-Yun Ren [a], Hong Yin [a,b], Jian-Xun Luo [a], Hui Wang [a,b,c]

[a] State Key Laboratory of Veterinary Etiological Biology, Key Laboratory of Veterinary Parasitology of Gansu Province, Key Laboratory of Grazing Animal Diseases MOA, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Science, Lanzhou, Gansu 730046, China
[b] Jiangsu Co-innovation Center for Prevention and Control of Important Animal Infectious Diseases and Zoonoses, Yangzhou, 225009, P. R. China
[c] NERC/Centre for Ecology and Hydrology (CEH) Wallingford, Crowmarsh Gifford, Wallingford, Oxon OX10 8BB, UK

ARTICLE INFO

Article history:
Received 11 October 2014
Accepted 10 January 2015
Available online 13 January 2015

Keywords:
Tick
Hyalomma anatolicum anatolicum
(*H.a. anatolicum*)
MicroRNA (miRNA)
Deep-sequencing

ABSTRACT

Hyalomma anatolicum anatolicum (*H.a. anatolicum*) (Acari: Ixodidae) ticks are globally distributed ectoparasites with veterinary and medical importance. These ticks not only weaken animals by sucking their blood but also transmit different species of parasitic protozoans. Multiple factors influence these parasitic infections including miRNAs, which are non-coding, small regulatory RNA molecules essential for the complex life cycle of parasites. To identify and characterize miRNAs in *H.a. anatolicum*, we developed an integrative approach combining deep sequencing, bioinformatics and real-time PCR analysis. Here we report the use of this approach to identify miRNA expression, family distribution, and nucleotide characteristics, and discovered novel miRNAs in *H.a. anatolicum*. The result showed that miR-1-3p, miR-275-3p, and miR-92a were expressed abundantly. There was a strong bias on miRNA, family members, and nucleotide compositions at certain positions in *H.a. anatolicum* miRNA. Uracil was the dominant nucleotide, particularly at positions 1, 6, 16, and 18, which were located approximately at the beginning, middle, and end of conserved miRNAs. Analysis of the conserved miRNAs indicated that miRNAs in *H.a. anatolicum* were concentrated along three diverse phylogenetic branches of bilaterians, insects and coelomates. Two possible roles for the use of miRNA in *H.a. anatolicum* could be presumed based on its parasitic life cycle: to maintain a large category of miRNA families of different animals, and/or to preserve stringent conserved seed regions with active changes in other places of miRNAs mainly in the middle and the end regions. These might help the parasite to undergo its complex life style in different hosts and adapt more readily to the host changes. The present study represents the first large scale characterization of *H.a. anatolicum* miRNAs, which could further the understanding of the complex biology of this zoonotic parasite, as well as initiate miRNA studies in other related species such as *Haemaphysalis longicornis* and *Rhipicephalus sanguineus* of human and animal health significance.

1. Introduction

Arthropods are a diverse group of organisms that include Chelicerata (ticks, spiders), Myriapoda (centipedes, millipedes), Crustacea (crabs, shrimps), and Insecta (flies, beetles). *H.a. anatolicum* is one of the most widely distributed and economically important tick species that parasitizes livestock including horses, goats, sheep, pigs and some wild animals living in subtropical and tropical regions worldwide (Luo et al., 2003; Guan et al., 2009). This parasite not only weakens animals by sucking their blood, but also transmits different species of the parasitic protozoa from the genus *Theileria*, including *Theileria ovis*, *Theileria annulata* and *Theileria equi* among others (Bartel, 2004; Lewis et al., 2005) and *Babesia* (Luo et al., 2003; Guan et al., 2009). The economic losses caused by ticks are large, although a precise estimate is difficult primarily due to their ability to transmit pathogenic agents directly or indirectly. Control of ticks would greatly reduce the incidence of tick-borne diseases.

MicroRNAs (miRNAs) are small 19–25 nucleotide (nt) long regulatory RNAs that act as post-transcriptional modulators of gene expression in animals and plants (Bartel, 2004). They are estimated to represent 1% of the transcriptome of higher eukaryotes and are predicted to regulate the expression of up to 30% of messenger RNAs (Lewis et al., 2005; Farh et al., 2005). Most miRNAs are encoded in intergenic regions and are transcribed by RNA polymerase II as long primary nuclear miRNAs (pri-miRNAs), which range from hundreds to thousands of nt in length (Lee et al., 2002). One pri-miRNA typically contains a single or several

Abbreviations: *H.a. anatolicum*, *Hyalomma anatolicum anatolicum*; miRNA, MicroRNA; QC, Quality control; *I. scapularis*, *Ixodes scapularis*; sRNA, Small RNA; U, Uracil; G, Guanine; A, Adenine; C, Cytosine.

* Corresponding author at: Department of Parasitology, State Key Laboratory of Veterinary Etiological Biology, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, 1 Xujiaping, Yanchangbu, Lanzhou, Gansu Province 730046, China.
E-mail addresses: liuguangyuan@caas.cn (G.-Y. Liu), huw@ceh.ac.uk (H. Wang).

http://dx.doi.org/10.1016/j.gene.2015.01.019
0378-1119/

miRNA precursors (pre-miRNAs) as stem-loop, hairpin structures flanked by unstructured, single stranded RNA sequences (Paula et al., 2011). Pre-miRNAs are cleaved near their loops by the cytoplasmic RNase III enzyme Dicer to generate a heteroduplex of small RNAs that are then packaged into the RISC complex (Lee et al., 2003). About half of the miRNA genes in *Drosophila melanogaster* are clustered and transcribed from a single polycistronic pri-miRNA (Jaubert et al., 2007).

Recently, deep sequencing technologies have been utilized to profile and discover miRNAs genome-wide. To assist in this process it is normally required to have a reference genome sequence. Currently, there is no reference genome sequence for *Hyalomma anatolicum anatolicum* (*H.a. anatolicum*) and the closest reference genome is the *Ixodes scapularis* genome draft (IscaW1.1) in which 49 miRNAs were identified (Griffiths-Jones et al., 2008). Some miRNAs are highly conserved throughout evolution such as miR-Let-7, present in metazoan lineages such as arthropods and vertebrates (Peterson et al., 2008). Thus, the use of the *I. scapularis* reference genome for which many miRNAs are currently annotated (Paula et al., 2011) may allow not only the identification of highly conserved tick miRNAs, but also aid in the discovery of arthropod-specific miRNAs. Previous studies have shown that miRNAs are continuously being added to metazoan genomes. Once integrated into gene regulatory networks these miRNAs show only rare nucleotide substitutions within the mature miRNA sequence at predictable positions (Yang et al., 1995). This is likely related to the strong purifying selection against changes in secondary structures of pre-miRNAs (Price et al., 2011). Wheeler and colleagues (Wheeler et al., 2009) documented evolutionary stable shifts to the determination of position 1 of the mature sequence that can be displaced towards either the 5′ or 3′ end, a phenomenon called seed shifting, as well as the ability to post-transcriptionally edit the 5′ end of the mature read, changing the identity of the seed sequence and possibly the repertoire of downstream targets.

In light of the probability that miRNA species could also be involved in gene regulation in *H.a. anatolicum*, we investigated the expression profile of miRNAs and identified potential novel miRNAs in *H.a. anatolicum*. Due to similarities in the morphology, life cycle, and modes of transmission among members of the phylum Chelicerata, the miRNA profile of *H.a. anatolicum* will pave way for miRNA studies in other species such as ticks, mites and mosquitoes.

2. Materials and methods

2.1. Tick sampling and RNA preparation

The ticks used in this study were obtained from the tick colony (Xinjiang Uygur Autonomous Region) maintained at the Animal Research Institute (Lanzhou Veterinary Research Institute). For the experiments, unfed-adult female ticks were collected in sterile 5 mL containers and transferred to the laboratory. Prior to experiment, the ticks were rinsed with 0.1% DEPC-treated water and the adult ticks were frozen in liquid nitrogen and subsequently crushed and ground to powder using a sterile mortar and pestle. Total RNA and/or enriched small RNA fraction was isolated from the whole adult ticks using the mirVana microRNA isolation kit as described in the manufacturer's instructions (Ambion, Applied Biosystems). Purity and integrity of the total RNA were examined by standard agarose gel electrophoresis, and the concentration was determined using a BioPhotometer (Eppendorf). The RNA sample was stored at −80 °C until deep sequencing or real time PCR analysis as described below.

2.2. Total RNA isolation, small RNA library preparation, and sequencing

Total RNA isolation and small RNA cloning were performed as described previously (Hafner et al., 2008), and low molecular weight RNA is isolated from the tick samples. Small RNAs (18–30 nt) are purified from the low molecular weight RNA fraction by polyacrylamide gel-based size fractionation and are ligated to a 5′ RNA adapter. To prevent self-ligation of small RNAs and self-ligation of the adapter, the 5′ terminus of the adapter has a hydroxyl group and an excess of adapter over small RNAs is used. Then, a 3′ adapter is ligated to the gel-purified product of the 5′ adapter ligation. The 3′ RNA adapter is modified to prevent circularization and self-ligation; typically, the 3′ hydroxyl is blocked by chemical synthesis of an oligonucleotide containing a 3′non-nucleotidic group, and above the steps were reference with Lu C et al. (Lu et al., 2007). After reverse transcription, a low number of PCR cycles are used to obtain sufficient amount of template for sequencing. The PCR product can be cloned and sequenced with regular PCR cloning vectors. The quality of the small RNA cDNA libraries is usually assessed in a quality control ("QC") step by sequencing about 100 individual clones (Lu et al., 2007).

2.3. Bioinformatics

Here the clean-reads of *H.a. anatolicum* ticks were mapped to the genome of *I. scapularis* ticks (http://www.ncbi.nlm.nih.gov/bioproject/?term=16232) using SOAP (Li et al., 2008) to analyze their expression and distribution on the genome. The program and parameters were SOAP −M 0, −v 0, −p 7, and −r 2 and the −M indicated was a control to mismatch number. The bioinformatics flowchart of *H.a. anatolicum* sRNAs are shown in Supplementary material 1. After masking the adaptor sequences and removing redundant reads and reads smaller than 18 nt, the resulting clean reads were annotated for the small RNA tags with rRNA, scRNA, snoRNA, snRNA and tRNA from GenBank and Rfam. To make every unique small RNA mapped to only one annotation, we follow the following priority rule:rRNAetc (in which GenBank > Rfam) > known miRNA > repeat > exon > intron (Calabrese et al., 2007), and the matched ncRNA tags were removed from the pool of unannotated tag using RepeatMasker (http://www.repeatmasker.org). The remaining reads were searched against the Sanger miRBase (version 18.0) to identify conserved miRNAs. Based on the nomenclature of miRNAs, reads showing high similarity to known miRNAs from other organisms (mismatches ≤ 2) were classified into the same miRNA family (Wheeler et al., 2009; Lu et al., 2007; Li et al., 2008). Family distribution provided a reference for conserved miRNA. The reads that did not match any of the above databases were marked as un-annotated. And all the parameters of software are showed in Supplementary material 3.

2.4. Novel miRNA prediction

These un-annotated short reads were assembled and the resulting contigs were used for the prediction of miRNAs. The characteristic hairpin structure of miRNA precursor can be used to predict novel miRNA. A prediction software Mireap was used to predict novel miRNA by exploring the secondary structure, the Dicer cleavage site, and the minimum free energy of the unannotated small RNA tags which could be mapped to genome. Mireap can be accessed from the following link: http://sourceforge.net/projects/mireap/, and the prediction software developed by BGI — Mireap. Parameters for animal: (1) Minimal miRNA sequence length for 18 nt; (2) maximal miRNA sequence length 26 nt; (3) minimal miRNA reference sequence length 20 nt; (4) maximal miRNA reference sequence length 24 nt; (5) minimal depth of Drosha/Dicer cutting site for 3; (6) maximal copy number of miRNAs on reference are 20 copies; (7) maximal free energy (−18 kcal/mol) allowed for a miRNA precursor; (8) maximal space between miRNA 35 nt; (9) minimal base pairs of miRNA are fourteen; (10) maximal bulge of miRNA is 4; (11) maximal asymmetry of miRNA duplex is 5; and (12) flank sequence length of miRNA precursor are 10 nt. And in order to determine the authenticity of these putative novel miRNAs from ticks, these were again mapped to clean reads from *H.a. anatolicum* ticks using NCBI's BLAST, as recommended.