Identification of microRNA in the protist *Trichomonas vaginalis*

Wei-Chen Lin a,b,1, Sung-Chou Li c,d,1, Wen-Chang Lin c,e, Jyh-Wei Shin f, Song-Nian Hu g, Xiao-Min Yu g,h, Ting-Yun Huang i, Shih-Chieh Chen i, Hua-Chien Chen j, Shu-Jen Chen j, Po-Jung Huang k, Richie Ruei-Chi Gan k, Cheng-Hsun Chiu l, Petrus Tang a,b,i,*

© 2009 Elsevier Inc. All rights reserved.

Abstract

MicroRNAs (miRNAs) are a class of small noncoding RNAs that have important regulatory roles in multicellular organisms. However, miRNA has never been identified experimentally in protist. Direct cloning of 438 expressed miRNA tags by microRNA serial analysis of gene expression from the parasitic protist *Trichomonas vaginalis* identified nine candidate miRNAs. Bioinformatics analysis of the corresponding genomic region revealed that these miRNA candidates contain a classical stem–loop–stem structure of pre-microRNAs. Analysis of the 20 nt long mature tva-miR-001 showed that it is an intergenic miRNA located at the scaffold DS113596. Tva-miR-001 was differentially expressed in the trophozoite, pseudocyst and amoeboid stages. Based on the experimental results of the present study, we provided solid evidence that protist possesses a miRNA regulating network comparable with multicellular organisms for the first time.

Introduction

MicroRNAs (miRNAs) are ~22-nucleotide noncoding RNA species that are found in diverse organisms from viruses, plants to mammals [1–4]. MicroRNAs are encoded by specific genes and play key roles in regulating the translation and degradation of mRNAs by inducing post-transcriptional silencing through base pairing to complementary target sites. The precursor miRNAs are cleaved from the primary transcripts by Drosha in the nucleus to generate a pre-miRNA hairpin structure of approximately 70 nt, which is subsequently processed by the Dicer, a RNAse III enzyme to produce mature miRNA:miRNA* duplexes [5–9]. One strand of the mature miRNA duplexes associates with the RNA induced silencing complex (RISC) to form a miRNA-ribonucleoprotein complex (miRNP) and binds to the target sites of mRNAs, predominately in the untranslated region of the target mRNA for translational repression or mRNA cleavage [7,9,10]. MiRNAs are involved in numerous cellular processes including development, differentiation, proliferation, apoptosis and response to stress. Dysregualtion of miRNA expression also contributes to disease pathology [11–14].

The first two miRNAs, *lin-4* and *let-7*, were identified in the 1990s through genetic experiments in the worm *Caenorhabditis elegans* [15,16]. Since then, a large number of miRNAs have been characterized from many eukaryotic species expect the protists. There has been considerable amount of debate on the existence of miRNA in protist based on the phylogenetic analysis of arguinguata-piwi, dicer-like and RNA-dependent RNA polymerase proteins in eukaryotes [17,18]. Functional miRNA machinery has only been reported in the unicellular green alga *Chlamydomonas reinhardtii* [19] and in the slime mold *Dictyostelium discoideum* [20]. The completion of many protist genome sequencing projects of medical importance in recent years revealed that the presence of putative genes which encodes specific enzymes required for miRNA processing and presentation in these protoist. Proteins containing Dicer like RNAse III domains are identified in the genomes of the pathogenic protists *Trichomonas vaginalis* [21], *Entamoeba histolytica* [22] and *Giardia intestinalis* [23]. In *Trypanosoma brucei*, the causative agent of African sleeping sickness possesses...
that the revolutionary position of T. vaginalis there is a very stringent control of differentially-expressed paralogous families. Large-scale expression sequenced tag analysis showed that miRNA-ribonucleoprotein complex. In plant and animal species, there

Expression of T. vaginalis Argonaute (AGO) genes

Protist genomes suggested the existence of a functional miRNA machinery in T. vaginalis. It will be of great interest to examine whether the different types of miRNAs that we have uncovered are processed by the same or different AGO proteins at different development stages or experimental conditions.

Direct cloning of T. vaginalis miRNA (Tv_miRNA) tags

Small RNA species shorter than 200-nucleotides isolated from T. vaginalis ATCC30236 was separated on SDS PAGE. RNAs shorter than 40-nucleotides were excised from the gel, cloned and sequenced. We sequenced 432 tva-miRNA tags and based on sequence identity we clustered these candidates into 102 groups. The resulting set of the unique sequences was mapped onto the T. vaginalis G3 genome assembly to identify hairpin-like RNA structures encompassing the tva-miRNA tags by using an in-house microRNA prediction pipeline [35]. A summary of the bioinformatics analysis was shown in Table 1. A total of nine tva-miRNA tags (tva-miR-001~009) ranging from 17 to 23 nucleotides long survived from the filter. Interestingly, none of these newly identified tva-miRNA showed significant sequence homology (over 90% identity) with known miRNAs in the miRBase V11. Among them, 5 of the tva-miRNA were located in intergenic region and 4 in the coding regions. The sequences of the identified tva-miRNA and their locations in the genome were shown in Table 2. Tva-miR-003, tva-miR-005, and tva-miR-007 have two copies in the genome while other identified-tva-miRNAs are single copy.

Sequence homology search of predicted miRNAs in related unicellular organisms

We mapped the T. vaginalis mirRAGE tags to 22 protozoan genomes downloaded from different resources and predicted potential hairpin structure surrounding the mapped region. All identified Tva-miRs have at least one homologues miRNA in other protozoan genome. A total of 24 predicted pre-miRNAs were identified in 14 unicellular organisms (Table 3). The sequences and structures of pre-miRNAs are shown in Supplemental Fig. 2. Among all the identified T. vaginalis miRNAs, tva-miR-006 is highly conserved in Cryptosporidium hominis, C. muris, C. parvum, Neospora caninum, Plasmodium knowlesi, Toxoplasma gondii, and Theileria parva. The Plasmodium genome datasets which composed of six malaria genomes contain two-third of the identified tva-miRNA tags. However, we are not able to identify any sequences with significant homology to the identified tva-miRNA tags in the E. histolytica, Leishmania and Trypanosoma genomes.

Results and discussion

Expression of T. vaginalis Argonaute (AGO) genes

Argonaute family protein is the key component of the RISC-like miRNA-ribonucleoprotein complex. In plant and animal species, there are multiple Dicer and Argonaute family members that are genetically and functionally diversified [32,33]. The Trichomonas genome encodes at least two Argonaute proteins which contain Argonaute specific Piwi domains [21]. The expression of Tv_AGO1 (TVAG_453810) and Tv_AGO2 (TVAG_411040) genes were determined by using quantitative real-time PCR. Fig. 1 showed that the expression of Tv_AGO1 gene was significantly higher than Tv_AGO2 gene in the trophozoite stage. Although the yeast Schizosaccharomyces pombe contain an Argonaute homolog (AGO1_SCHO) which is involved only in RNAi processes, all eukaryotic Argonaute family protein studied so far are important regulators in microRNA biogenesis [34]. Therefore, the expression of Tv_AGO1 gene suggested that functional miRNA machinery exists in T. vaginalis. It will be of great interest to examine whether the different types of miRNAs that we have uncovered are processed by the same or different AGO proteins at different development stages or experimental conditions.

Table 1

<table>
<thead>
<tr>
<th>Bioinformatics analysis of the mirRAGE tags</th>
<th>No. of tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>mirRAGE Tags</td>
<td>432</td>
</tr>
<tr>
<td>Unique miRNA groups</td>
<td>102</td>
</tr>
<tr>
<td>miRNA match to genome</td>
<td>22</td>
</tr>
<tr>
<td>miRNA survived Rfam filter</td>
<td>17</td>
</tr>
<tr>
<td>miRNA located on the arm of candidate hairpins</td>
<td>9</td>
</tr>
<tr>
<td>Exonic miRNA</td>
<td>4</td>
</tr>
<tr>
<td>Intergenic miRNA</td>
<td>5</td>
</tr>
</tbody>
</table>
Expression of tva-miR-001 in different stages

The validation of the exact length of the cloned miRNAs is critical for their classification and for future studies such as expression profiling, target prediction and miRNA-target interaction. This step is predominantly relied on Northern hybridization of specific sequences. However, miRNAs that are expressed at low levels or exhibit tissue or developmental-stage specific expression may be difficult to validate using Northern blots. Recent advancements in the sensitivity and specificity of stem–loop RT-PCR technology allowed it to distinguish single nucleotide variation in mature miRNAs [36]. In the present study, we adapted this technology to determine the length of tva-miR-001 by primer walking. Serial primers with only one nucleotide difference were designed to decipher the length of the mature tva-miR-001 by using stem–loop real-time PCR. The amount of RT-PCR products generated by using different combination of primer pairs was shown in Fig. 2. Primer pairs with exact match to the length of the target miRNA usually have a higher affinity and generate more PCR products [42]. The results indicated that the length of tva-miR-20 nt as encompassed by the primer pair tva-miR-001_R1 RT-2 and tva-miR-001_GS_2.

The expression level of tva-miR-001 during cell cycle was also followed by using stem–loop real-time PCR. Fig. 3 showed that the expression level of tva-miR-001 was similar in the trophozoite and the pseudocyst stage. However, the expression level of tva-miR-001 in amoeboid stage was significantly lower than the other stages. Tva-miR-001 is an intergenic miRNA located between the conserved hypothetical proteins TVAG_357920 and TVAG_RC_DS11395_5 within a region of 2562 bps. Tva-miR-001 is located 122 nt upstream from the 5′ end of the conserved hypothetical protein TVAG_357920. Fig. 4 showed that a region of 2000 bps which include partial intergenic region, tva-miR-001 and TVAG_357920. Two putative promoter regions which contain the conserved ATGACA (-35) and TATAAT (-10) sequences were identified upstream of tva-miR-001. It is quite reasonable to postulate that tva-miR-001 and TVAG_357920 are in the same transcriptional unit and probably use the same promoter for gene expression. By using RT-PCR we showed that TVAG_357920 was also expressed in the trophozoite stage (Fig. 5). The coordinated expression of both tva-miR-001 [Fig. 3] and TVAG_357920 in trophozoite supports our hypothesis that they may use the same promoter for gene expression.

In multicellular organisms, three major classes of endogenous small RNAs have been discovered: microRNAs (miRNAs) [5,12], Piwi-associated small RNAs (piRNAs) [37–39], and small interfering RNAs (siRNAs) [40,41]. The identification of a large number of miRNAs in different species prompted us to speculate that these small RNA populations may also exist in protist. The Trichomonas miRNAs that we discovered do have any significant identifiable homology in other organisms although most miRNAs are evolutionarily conserved in related species and some even show conservation between invertebrates and vertebrates [36,42–44]. Since T. vaginalis is the first protist reported up-to-date that may contain functional miRNA machinery, the lack of sequence homology between other organisms does not necessarily reflect the fundamental differences in miRNA silencing pathways. We are using high-throughput next generation sequencing technology to study comparative microRNAome in the related flagellates G. intestinalis, Tritrichomonas foetus and Pentatrichomonas hominis. These flagellates are the most primitive eukaryotes in the phylogenetic tree. The resulting data should provide a more in-depth understanding on the evolution of miRNAs.

Materials and methods

T. vaginalis culture

T. vaginalis ATCC strain 30236 (JH 31A #4) was axenically cultured at 37 °C in yeast extract, iron-serum (YI-S) medium [45].

Table 2

Summary of T. vaginalis miRNA candidates

<table>
<thead>
<tr>
<th>Candidate</th>
<th>miRAGE tags</th>
<th>Position of the tags in the T. vaginalis G3 genome</th>
<th>Length</th>
<th>Sequence (5′–3′)</th>
<th>Genomic location</th>
<th>Start</th>
<th>Stop</th>
<th>Intergenic/exonic</th>
</tr>
</thead>
<tbody>
<tr>
<td>tva-miR-001</td>
<td>TIGR_Assembly_92719</td>
<td>Intergenic</td>
<td>23</td>
<td>TTTACTCCATGGATAAATT</td>
<td>47,027</td>
<td>47,005</td>
<td></td>
<td>Intergenic</td>
</tr>
<tr>
<td>tva-miR-002</td>
<td>TIGR_Assembly_88546</td>
<td>Intergenic</td>
<td>17</td>
<td>ATTTCTAATCAATCC</td>
<td>45,896</td>
<td>45,912</td>
<td></td>
<td>Intergenic</td>
</tr>
<tr>
<td>tva-miR-003</td>
<td>TIGR_Assembly_97090</td>
<td>Exonic</td>
<td>23</td>
<td>CGACTTAAAGAAGAAACAGGAT</td>
<td>10,892</td>
<td>10,869</td>
<td></td>
<td>Exonic</td>
</tr>
<tr>
<td>tva-miR-004</td>
<td>TIGR_Assembly_96185</td>
<td>Intergenic</td>
<td>18</td>
<td>GCATGGATAATCGGAGTATC</td>
<td>20,335</td>
<td>20,319</td>
<td></td>
<td>Intergenic</td>
</tr>
<tr>
<td>tva-miR-005</td>
<td>TIGR_Assembly_92899</td>
<td>Exonic</td>
<td>23</td>
<td>TGGATAAGTGTTTTTATCCTGGG</td>
<td>15,420</td>
<td>15,403</td>
<td></td>
<td>Exonic</td>
</tr>
<tr>
<td>tva-miR-006</td>
<td>TIGR_Assembly_88558</td>
<td>Intergenic</td>
<td>18</td>
<td>TGACTAAGGCTTTTATTCTCCGG</td>
<td>3618</td>
<td>3602</td>
<td></td>
<td>Intergenic</td>
</tr>
<tr>
<td>tva-miR-007</td>
<td>TIGR_Assembly_84258</td>
<td>Intergenic</td>
<td>18</td>
<td>TTATTTCTCATCTGATGAT</td>
<td>94,718</td>
<td>94,701</td>
<td></td>
<td>Intergenic</td>
</tr>
<tr>
<td>tva-miR-008</td>
<td>TIGR_Assembly_93905</td>
<td>Intergenic</td>
<td>23</td>
<td>GGACCGACGTTGCATGGTT</td>
<td>33,980</td>
<td>33,963</td>
<td></td>
<td>Intergenic</td>
</tr>
<tr>
<td>tva-miR-009</td>
<td>TIGR_Assembly_86319</td>
<td>Exonic</td>
<td>17</td>
<td>CCGAAAGATCTGGCTGGG</td>
<td>94,347</td>
<td>94,329</td>
<td></td>
<td>Exonic</td>
</tr>
</tbody>
</table>

Table 3

T. vaginalis miRNA tags with sequence homology in related unicellular organisms

<table>
<thead>
<tr>
<th>tva-miR-001</th>
<th>Ch</th>
<th>Cm</th>
<th>Cp</th>
<th>Gl</th>
<th>Nc</th>
<th>Pb</th>
<th>Pc</th>
<th>P′</th>
<th>Pg</th>
<th>Pk</th>
<th>Py</th>
<th>Tg</th>
<th>Tp</th>
<th>Dd</th>
</tr>
</thead>
<tbody>
<tr>
<td>tva-miR-002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tva-miR-003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tva-miR-004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tva-miR-005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tva-miR-006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tva-miR-007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tva-miR-008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tva-miR-009</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

tva-miR-001 tags with sequence homology in the genomes of related unicellular organisms and located on the stem region of predicted pre-miRNAs are indicated by black boxes.

Ch: Cryptosporidium hominis; Cm: C. muris; Cp: C. parvum; Gl: Giardia lamblia; Nc: Neospora caninum; Pb: Plasmodium berghei; Pc: P. chabaudi; P′: P. falciparum; Pg: P. gallinaceum; Pk: P. knowlesi; Py: P. yoelii; Tg: Toxoplasma gondii; Tp: Theileria parva; Dd: Dicyostelium discoideum.
supplemented with 10% horse serum. Growth of the parasite was monitored by using trypan blue exclusion hemocytometer counts. Late-logarithmic phase parasites with more than 95% viable cells were harvested for RNA isolation.

Cloning of microRNA in T. vaginalis

Total RNAs were isolated by using a RNAspin mini isolation kit as described by the manufacturer (GE Healthcare). The small RNAs (shorter than 40 nucleotides) were isolated by using the miRNA isolation kit and purified by FlashPAGE™ Fractionator (Ambion). Isolated small RNAs were dephosphorylated with calf intestinal alkaline phosphatase (NEB, Beverly, MA) and extracted by phenol/chloroform. The aqueous phase which contain small RNAs were transferred to a new tube and precipitated by adding glycogen to a final concentration of 2 μg/ml and 3 volumes of absolute ethanol. The small RNAs were ligated to the miRAGE 3′ adaptor: 5′pUUUCTATCATGGACTGidT 3′ (p:5′ phosphate; idT: 3′-inverted deoxythymidine) with T4 RNA ligase (NEB, Beverly, MA). RNA/DNA hybrid was separated on 15% polyacrylamide PAGE gel with TBE/Urea and stained with ethidium bromide. The 38- to 45-base RNA/DNA bands were excised from the gel and gel-extracted by incubating the gel slices in 0.3 M NaCl at 4 °C under constant agitation. The supernatant was ethanol-precipitated, ligated to the miRAGE 5′ adaptor (5′TGGGAATTCCTCACTrArArA 3′) and separated on 15% polyacrylamide PAGE gel.

Fig. 2. Stem–loop real-time PCR of tva-miR-001. The length of the mature tva-miR-001 was determined by using primer walking. The primer pairs used for stem loop real-time PCR assay were shown in the top panel. The related fold of stem loop real-time PCR products of different primer combinations was shown in the bottom panel with reference to the primer pairs GS_1 and RT_1.

---

tva-miR-001 GS_1: CGCCGGACATTCTATGGGA

tva-miR-001 GS_2: CGCCGGCAATTCTATGGAT

tva-miR-001 RT_1: CAUUCUAUGGAAUAUAUAAGCTCACTG TGGCG G

TTTTTATAAGTGTCACAC ACGGC A

TTA TG

tva-miR-001 RT_2: CAUUCUAUGGAAUAUAUAAGCTCACTG TGGCG G

TTTTTATAAGTGTCACAC ACGGC A

TTA TG

tva-miR001 RT_3: CAUUCUAUGGAAUAUAUAAGCTCACTG TGGCG G

TTTTTATAAGTGTCACAC ACGGC A

TTA TG
After gel extraction of RNA/DNA products ranging from 56 to 63 bases, reverse transcription of the ligation products was performed by using PCR 3′ primer (5′-TACAGTCCATGGATAGAAA-3′). This is followed by PCR using the reverse PCR 3′ primer and the forward PCR 5′ primer (5′-CATGGGAATTCCTCACTAAA-3′). The PCR product was purified by phenol/chloroform extraction, ethanol precipitated, and digested with EcoRI and NcoI. The restriction digest was concatamerised with T4 DNA ligase and the ligation products were recovered from the gel. After additional phenol/chloroform extraction and ethanol precipitation, the ligation products were ligated to a EcoRI/NcoI-linearized pGEM-T Easy cloning vector (Promega) by using T4 DNA ligase and transformed the ligation mixture into TOP10 competent cells (Invitrogen). Added SOC and plated out on LB Amp agar plates containing IPTG and X-gal. The inserts in positive colonies were amplified by PCR using the M13-forward and M13-reverse primers. Purified PCR products were subjected to nucleotide sequencing. After removing the adaptor sequences, miRAGE tags were collapsed to a non-redundant set and matched to the T. vaginalis G3 genome. Genome matches were clustered if neighboring matches fell within either 100 nucleotides of each other. The increased sizes of the matches are used for precursor hairpin structure analysis. If the miRAGE tags were located on one arm of a predicted hairpin and the region of the hairpin corresponding to that sequence contained >16 base pairs [46], the candidate locus was examined manually for characteristics of known miRNAs, using criteria described previously [35].

The microRNA prediction pipeline

Potential hairpin structures in the genomic scaffolds were identified by using a microRNA prediction pipeline as described previously [35]. Briefly, the pipeline use Smalooop to identify putative hairpins from the genomic sequence. Each identified 5′- and 3′- hairpin pairs was further screened by using various filters based on sequences homology conservation and sequence structure features. Any candidates for which the GC content, core minimum free energy (mfe), hairpin mfe and ch_ratio values falling within the selected reference range values were regarded as a positive candidate miRNA. We used the same prediction pipeline to screen the genome of 22 related unicellular organisms.

Quantitative real-time PCR and stem–loop real-time PCR

Quantitative real-time PCR was used to study the gene expression of the T. vaginalis Argonaute genes (TVAG_453810, TVAG_411040).
Reverse-transcription PCR was used to study the gene expression of *TVAG_357920*, a conserved hypothetical protein downstream of tva-miR-001.

The expression of miRNA in *T. vaginalis* was validated by using a modified stem–loop real-time PCR (SL-RTPCR) method [47]. All the gene specific primers used in the present study were listed in Table 4. Reverse transcription was carried out in a reaction mixture containing total RNA, 50 nM stem–loop RT primer, 0.25 mM dNTPs, 0.75 U/μl ThermoScript™ III reverse transcriptase, 0.2 U/μl RNase out, and 0.05 M DTT (ThermoScript™ III RT-PCR System, Invitrogen). The RT reaction mixture was incubated at 16 °C for 30 min, 50 °C for 30 min and then stopped by 85 °C for 5 min. RT-PCR was performed using a 20 μl PCR mixture contains 1 μl reverse transcription product, master mix, 0.5 M real-time forward and reverse primers. The reactions were incubated in a 96-well plate at 95 °C for 10 min, followed by 40 cycles of 95 °C for 30 s, 50 °C for 1 min, and 72 °C for 30 s. 60S RNA was used as an internal control for normalization in all experimental groups. The expression level of miRNAs was determined by five independent SL-RTPCR reactions and analyzed according to the 2−ΔΔCt method [48].

### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2009.01.004.

### References
