

High Throughput Genome-Wide Survey of Small RNAs from the Parasitic Protists *Giardia intestinalis* and *Trichomonas vaginalis*

Xiaowei (Sylvia) Chen,* Lesley J. Collins,*† Patrick J. Biggs,* and David Penny*†

*Allan Wilson Centre for Molecular Ecology and Evolution, Massey University, Palmerston North, New Zealand; and †Institute of Molecular Biosciences, Massey University, Palmerston North, New Zealand

RNA interference (RNAi) is a set of mechanisms which regulate gene expression in eukaryotes. Key elements of RNAi are small sense and antisense RNAs from 19 to 26 nt generated from double-stranded RNAs. MicroRNAs (miRNAs) are a major type of RNAi-associated small RNAs and are found in most eukaryotes studied to date. To investigate whether small RNAs associated with RNAi appear to be present in all eukaryotic lineages, and therefore present in the ancestral eukaryote, we studied two deep-branching protozoan parasites, *Giardia intestinalis* and *Trichomonas vaginalis*. Little is known about endogenous small RNAs involved in RNAi of these organisms. Using Illumina Solexa sequencing and genome-wide analysis of small RNAs from these distantly related deep-branching eukaryotes, we identified 10 strong miRNA candidates from *Giardia* and 11 from *Trichomonas*. We also found evidence of *Giardia* short-interfering RNAs potentially involved in the expression of variant-specific surface proteins. In addition, eight new small nucleolar RNAs from *Trichomonas* are identified. Our results indicate that miRNAs are likely to be general in ancestral eukaryotes and therefore are likely to be a universal feature of eukaryotes.

Introduction

Since its discovery in 1998 (Fire et al. 1998), RNA interference (RNAi) has been found in animals (Collins and Cheng 2006), plants (Gazzani et al. 2004), and some protists (Ullu et al. 2004). It is implicated in a wide range of gene-silencing mechanisms including downregulating mRNA levels (Sen and Roy 2007), heterochromatin assembly and maintenance (Grewal and Elgin 2007), DNA elimination (Collins and Cheng 2006), promoter silencing, developmental control (Morris et al. 2004), upregulation of transcription during the cell cycle (Vasudevan et al. 2007), and transposon silencing (Hartig et al. 2007). Key elements that guide all the above processes are small RNAs with size ranges of 19–26 nt.

Four major types of small RNAs associated with RNAi have been extensively studied: short-interfering RNAs (siRNAs), repeat-associated short-interfering RNAs, microRNAs (miRNAs) (Meister and Tuschl 2004), and piwi RNAs (Lau et al. 2006). These RNAs are processed from complementary or near-complementary double-stranded RNAs (dsRNAs) precursors into 21- to 26-nt RNAs by Dicer or Drosha RNase III family endonucleases in the cytoplasm (Bernstein et al. 2001; Lee et al. 2003). Dicer homologues have been found in most eukaryotes including deep-branching unicellular parasites such as *Giardia intestinalis* and *Trichomonas vaginalis* (Finn et al. 2006) (hereafter referred to as *Giardia* and *Trichomonas*, respectively). After cleavage by Dicer, the short dsRNAs are then incorporated into ribonucleoprotein particles which assemble the RNA-induced silencing complex (RISC) (Hammond et al. 2001). The assembly of RISC also requires energy-driven unwinding of the siRNA or miRNA duplexes and conformational changes of preassembled ribonucleoprotein particles. The single-stranded siRNA or miRNA guides the RISC complex to the target mRNA and is strongly bound to the Argonaute (Ago) protein which then cleaves the target mRNA (Song et al. 2004). In some organisms

such as *Neurospora crassa* (Forrest et al. 2004), *Caenorhabditis elegans* (Smardon et al. 2000), *Schizosaccharomyces pombe* (Martienssen et al. 2005), and plants (Gazzani et al. 2004), an RNA-dependent RNA polymerase (RdRp) is also essential for dsRNA-triggered gene silencing. The RdRp is likely to use the siRNA as primers and convert the target RNAs into dsRNAs and a second wave of gene silencing is initiated.

Several protozoan parasites have been studied in searching for evidence of RNAi, including *Trypanosoma* (Ullu et al. 2004), *Plasmodium* (Malhotra et al. 2002), and *Giardia* (Ullu et al. 2005; Macrae et al. 2006; Prucca et al. 2008; Saraiya and Wang 2008). The presence of RNAi has been suggested in the deep-branching eukaryote *Giardia* (Macrae et al. 2006; Prucca et al. 2008; Saraiya and Wang 2008). Detailed biochemical and structural studies have been carried out for the *Giardia* Dicer protein homologue, showing that recombinant *Giardia* Dicer could cleave dsRNAs into 25- to 26-nt short fragments in vitro (Macrae et al. 2006; MacRae et al. 2007). The *Giardia* genome contains protein homologues of Ago and RdRp (Morrison et al. 2007). Recent studies also showed that Dicer, Ago, and RdRp are all available for RNAi regulation of *Giardia* variant-specific surface protein (VSP) expression (Prucca et al. 2008) as well as an miRNA derived from a snoRNA (Saraiya and Wang 2008). Results from these studies also support the idea that RNAi mechanism is likely to have occurred in the last common ancestor of eukaryotes (Collins and Penny 2009).

Giardia and *Trichomonas* are both single cellular anaerobic eukaryotes belonging to the group of Excavates (Keeling et al. 2005). They both have gone through reductive evolution which resulted in either mitochondria in *Giardia* (Tovar et al. 2003) or hydrogenosomes in *Trichomonas* (Dyall et al. 2004). Mitochondria and hydrogenosomes appear to be two reduced forms of mitochondria (Embley et al. 2003; Mentel and Martin 2008). Despite the similarly reduced cellular components, *Giardia* and *Trichomonas* are separated by long evolutionary distance within Excavates (Hampel et al. 2009), making them comparable yet distant models for our study. Previous studies on *Giardia* non-coding RNAs (ncRNAs; Collins et al. 2004; Chen et al. 2007, 2008) showed that sequences of ncRNAs from

Key words: ancestral eukaryotes, miRNA, protists, RNA evolution.

E-mail: sylvia.x.chen@gmail.com.

Genome Biol. Evol. Vol. 2009:165–175.

doi:10.1093/gbe/evp017

Advance Access publication July 6, 2009

deep-branching eukaryotes can be highly divergent from other well-studied eukaryotes. Therefore, by using traditional methods, it is hard to identify functional ncRNAs.

In this study, we used high throughput Solexa-sequencing technology (Illumina) to search for previously unidentified small RNAs (including miRNAs and siRNAs) from two protozoan parasites *Giardia* and *Trichomonas*. Despite extensive biochemical studies on the RNAi mechanism of *Giardia*, little is known about the endogenous small RNA (20–60 nt) population in either *Giardia* or *Trichomonas*. Previous studies on ncRNAs from these two organisms showed presence of eukaryotic-specific RNAs such as snoRNAs (Yang et al. 2005; Chen et al. 2007), spliceosomal snRNAs (Chen et al. 2008; Simoes-Barbosa et al. 2008), and RNase P (Marquez et al. 2005). A number of antisense RNAs were also found in *Giardia* (Ullu et al. 2005). Therefore, the presence of other basic small RNAs such as miRNAs and siRNAs is expected. In addition, there have been many previously uncharacterized noncoding RNAs identified in *Giardia* (Chen et al. 2007), indicating the likely presence of new classes of ncRNAs in deep-branching eukaryotes. Large-scale RNA analysis has not previously been done for *Trichomonas*, another member of the Metamonada subclade of Excavates (Hampl et al. 2009). Comparing the ncRNA contents of *Trichomonas* with those of *Giardia* could lead to a better understanding of the evolution of RNA processing in eukaryotes. Using Illumina Solexa sequencing on small RNAs from *Giardia* and *Trichomonas*, we identified 10 strong miRNA candidates from *Giardia* and 11 from *Trichomonas* as well as a number of putative miRNA candidates from both organisms. We also found evidence supporting the presence of siRNA in *Giardia*. In addition, eight new snoRNAs from *Trichomonas* are identified. Our results strongly support RNAi-related small RNAs as a general feature of eukaryotes.

Materials and Methods

Total RNA Preparation and Sequencing

Giardia intestinalis (WB strain) trophozoites were collected from TY1-S-33 growth media at a concentration of 1.4×10^7 cells/ml by centrifugation (10 min, 2,500 rpm, 4 °C). Total RNA was prepared using Trizol (Invitrogen) according to the protocol provided by the manufacturer. The pure RNA was resuspended in distilled water.

Trichomonas vaginalis was grown in Trichomonas broth (Fort Richard) at 37 °C for 3–4 days and harvested by centrifugation (10,000 rpm, 15 min at room temperature). Growth media was removed and cells were resuspended in equal volumes of 2× LETS buffer (200 mM LiCl, 20 mM EDTA, 20 mM Tris pH 7.8, and 2% SDS). An equal volume of phenol:chloroform (5:1, pH 5) was added to the suspension, and the mixture was vortexed for 10 s. Phases were separated by centrifugation at 14,000 rpm for 5 min at room temperature, and the upper phase was further extracted twice with phenol:chloroform, then once with chloroform. Finally, total RNA was precipitated by adding LiCl to a final concentration of 0.2 M and 3 volumes of 100% EtOH, then incubated at –80 °C for 1 h.

For Solexa sequencing, 10 µg of total RNAs were separated on a 15% denaturing acrylamide 8 M urea gel and RNAs ranging from 10 to 200 nt were cut out from the gel and prepared according to Illumina's small RNA preparation protocol. Eight and 12 pmol (in each lane) of *Giardia* and *Trichomonas* cDNA were used for sequencing on an Illumina Genome Analyzer for 35 cycles.

Analysis of Solexa Short-Read Sequences

Initial Solexa data required computational filtering and trimming to remove an expected portion of adaptor sequences (due to the short length of some RNAs). These filtered sequences were mapped to previously identified and computationally predicted ncRNAs from *Giardia* and *Trichomonas* using seqmap-1.0.8 (source code from <http://biogibbs.stanford.edu/~jiangh/SeqMap/>). Sequence assembly was performed using Velvet version 0.7.20 (source code from <http://www.ebi.ac.uk/~zerbino/velvet/>). Solexa short reads with lowest length of 17 nt were used for assembly.

Analysis of miRNA Candidates and snoRNAs

Prediction of possible miRNA precursors were initiated by searching for hairpin loops in the genomes of *Giardia* and *Trichomonas* using smalooop (source code from <http://arep.med.harvard.edu/miRNA/pgmlicense.html>). The output sequences were filtered based on structural criteria using RNAfold from the Vienna package (<http://www.tbi.univie.ac.at/~ivo/RNA/>) and custom C programs according to the following criteria: location in non-protein-coding regions; complementary to 3' untranslated regions (UTRs); having an RNAfold-determined minimum free energy ≤ -32.5 kcal/mol for *Giardia* and ≤ -20 kcal/mol for *Trichomonas* (because of the latter's high A/T content); and no multiloops. Subsequently, the filtered hairpin sequences were mapped with Solexa output sequences using Seqmap (Jiang and Wong 2008). Finally, putative candidates were evaluated by their complementary binding to 3' UTRs, as a primary feature of many miRNAs (Pasquinelli et al. 2005). The filtered sequences were checked for expression by comparing with our Solexa sequencing results using seqmap and then checked against 3' UTR sequences using Blast. The 3' UTR sequences of 50 nt were extracted from the genomes by a custom C program. Prediction of snoRNAs was done using snoscan-0.9 (source code from <http://lowelab.ucsc.edu/snoRNAdb/code/>). The custom C programs are available upon request from the corresponding author.

Reverse Transcriptase–Polymerase Chain Reaction

All the reverse transcriptase–polymerase chain reactions were done using Invitrogen ThermoScript First strand cDNA synthesis kit and subsequent PCRs were done using Roche Taq polymerase. The primers used are listed in supplementary table S2 (Supplementary Material online).

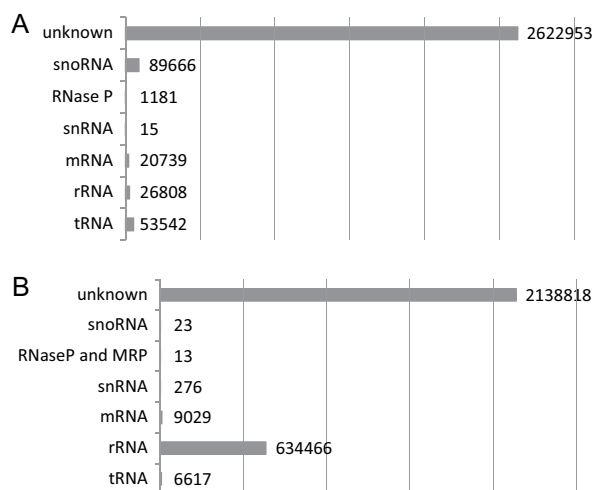


FIG. 1.—Coverage of RNA species by Solexa sequencing. The filtered reads were mapped against known RNA sequences in *Giardia* and *Trichomonas*, and the numbers of reads were counted for the corresponding transcripts. The majority of reads in both organisms are mapped to unknown transcripts, which potentially contain novel ncRNAs. (A) Reads count for *Giardia*. (B) Reads count for *Trichomonas*.

Results

High Throughput Sequencing of *Giardia* and *Trichomonas* Small RNAs

Cultured *G. intestinalis* and *T. vaginalis* were each harvested for total RNA extraction at exponential growth phase, and small RNAs (10 to 200 nt) were purified by size fractionation. cDNA was synthesized following Illumina's small RNA preparation protocol and were then sequenced using an Illumina Genome Analyzer (also known as Solexa sequencing). The initial output of 36-nt sequences was filtered for adaptor sequences, and the resulting output contained 2,761,362 sequences for *Giardia* and 2,789,242 sequences for *Trichomonas*. All sequences from the Solexa sequencing were uploaded onto a MySQL database (v. 5.0.45) and also visualized with GBrowse (v.1.69).

To evaluate the RNA coverage of the sequencing, the filtered sequences were mapped (see Materials and Methods) to previously known ncRNAs from each organism. Allowing the maximum of 2-nt mismatches, 30 tRNAs (Morrison et al. 2007), 3 rRNAs (Morrison et al. 2007), 51 snoRNAs (Yang et al. 2005; Chen et al. 2007), 1 RNase P (Marquez et al. 2005), 4 spliceosomal snRNAs (Collins et al. 2004; Chen et al. 2008), and 21 previously found but uncharacterized ncRNAs (Chen et al. 2007) were recovered in *Giardia*. In *Trichomonas*, 165 tRNAs (Aurrecochea et al. 2008), 3 rRNAs (Aurrecochea et al. 2008), 5 spliceosomal snRNAs (Simoes-Barbosa et al. 2008), 1 RNase P and 1 RNase MRP (Piccinelli et al. 2005), and 8 new snoRNAs were recovered. In total, 188,425 unique sequences from *Giardia* and 648,707 unique sequences from *Trichomonas* mapped to ncRNAs indicated above or to transcripts of protein-coding genes. The remainder corresponds to unknown transcripts. The coverage of various RNA species in both organisms is shown in figure 1.

Identification of miRNA Candidates

In order to effectively represent Solexa output, all filtered sequences which did not map to any known transcripts were assembled into contigs using the de novo short sequence assembler Velvet (Zerbino and Birney 2008). The resulting contigs were checked by Blasting them against the respective genomes. Two strategies were used for further analysis.

Strategy-1: Identifying New miRNAs by Sequence Similarity

The first way to identify miRNA candidates was based on sequence similarity where we compared Solexa output sequences with known mature miRNAs. Initially, sequences of all published miRNAs with annotation were downloaded from the miRBase (Release 12.0 <http://microrna.sanger.ac.uk/>) and Blasted (Altschul et al. 1990) against de novo assembled Solexa sequences. From the Blast-checked Solexa de novo contigs, candidates of miRNAs were identified. Six Solexa de novo contigs from *Giardia* contained sequences with a high degree of similarity to previously known miRNAs. The candidate sequences and alignments with known miRNAs are shown in figure 2A. Although the miRNA candidates align well with the known miRNAs, the corresponding pre-mRNA sequences do not seem to share distinct sequence similarities. These candidate sequences all exist as single copy in the genome. A structural study of dsRNA cleavage by *Giardia* Dicer has previously shown that *Giardia* Dicer protein tends to cleave dsRNAs into small RNAs of 25 or 26 nt (MacRae et al. 2007). This is consistent with a recent study on wild-type *Giardia* RNA (Saraiya and Wang 2008). Therefore, based on alignments, putative length, and sequence similarities among candidates, we can predict mature miRNA candidates of *Giardia*, and we found Solexa de novo contigs that fit our prediction.

Two of the *Giardia* miRNA candidates Gim5 and Gim6 are located on the antisense strands of annotated genes: GL50803_11290 Kinase (CMGC CDK) and GL50803_11912 hypothetical protein, respectively. The other *Giardia* miRNA candidates have antisense matches to predicted open reading frames that do not contain annotated genes. Potential miRNA targets were also searched in the 3' UTR regions of annotated genes. The UTR regions of *Giardia* are typically short with <20 nt at 5' end and <50 nt at 3' end (Elmendorf et al. 2001). Therefore, sequences 50-nt 3' to all annotated genes were extracted to represent all possible 3' UTRs. Blast results showed partial complementary matching of all six *Giardia* miRNA candidates to 3' UTRs, and at least two of them (Gim1 and Gim5) showed extensive matches. Examples of potential *Giardia* miRNA-target binding are shown in figure 3A, together with the Solexa contigs where the examples of miRNAs were found.

To date, there have not been any studies on RNAi in *Trichomonas*. Seven *Trichomonas* de novo contigs from our Solexa sequences had a high degree of sequence similarity with mature miRNAs known from other species. The

A		B	
Gim1 gga-miR-1791	AUCAACGUGACUGAUGCUGGCUCU -GCGAUGUGACUGAUGCAGGCUG- * * * * *	Tvm1 ath-miR862-5p	-GUAUAGGUGCGAGCUUGUGAAU UCCAUAAGGUGCGAGCAUGUGC-- * * * * *
Gim2 dan-miR-311a	AUCUCGCACAUUACC GGCCUCU -UAUUGCACUUAUACC GGCCUGA- * * * * *	Tvm2 gma-miR1534	CAUUUUGGGUAAAUGGUCAAU UAUUUUGGGUAAAUGUCAU- * * * * *
Gim3 gga-miR-202	GUGCAGAGGCAUGGAGCACGGGAA-- ---AGAGGCAUAGAGCAUGGGAAA * * * * *	Tvm3 dre-miR-725	CAAUUCAGUCAUUUUUCUGU--- ---UUCAGUCAUUUUUCUAGUAGU * * * * *
Gim4 oan-miR-1336	GUGGUCUGCAUCUGGACCUUUCACU UCUUUCUGCAUCUGAACCUUUUC- * * * * *	Tvm4 mmu-miR-743b-3p	CAACAGACAUAUUGCUGAAUAG- -GAAAGACAUCAGUCUGAAUAGA * * * * *
Gim5 cel-miR-34	-GGCCGUGUGGUUAGGUGGUUGUUG AGGCAGUGUGGUUAGCUGGUUG-- * * * * *	Tvm5 osa-miR1852	UAAUUGGAUCAGAAUGCAGU- --AUUUGGAUCAGAAUGCAGU * * * * *
Gim6 gga-miR-1673	GUGGUGAGUAGAAGUCAGAUUAUA GGUGUGAGUGGAAGUCAGAGGU-- * * * * *	Tvm6 mml-miR-891	UCAUCCUUAUCCUCAGUCAUUGA UGCAACUUAUCCUCAGUCAUUGA * * * * *
		fru-miR-152 tni-miR-152 Tvm7	UCAGUGCAUAACAGAACUUUGU UCAGUGCAUAACAGAACUUUGU AUUUGGCAUAUAGAACUUUGC * * * * *

FIG. 2.—Predicted miRNA candidates and their alignments with published miRNAs: (A) *Giardia* miRNA candidates and alignments; (B) *Trichomonas* miRNA candidates and alignments. Six *Giardia* miRNA candidates and seven *Trichomonas* miRNA candidates all show extensive sequence homology with known miRNAs from a number of organisms. The full names of known miRNAs are gga-miR-1791 MIMAT0007705 *Gallus gallus* miR-1791, dan-miR-311a MIMAT0008471 *Drosophila ananassae* miR-311a, gga-miR-202 MIMAT0003355 *Gallus gallus* miR-202, oan-miR-1336 MIMAT0006829 *Ornithorhynchus anatinus* miR-1336, cel-miR-34 MIMAT0000005 *Caenorhabditis elegans* miR-34, gga-miR-1673 MIMAT0007557 *Gallus gallus* miR-1673, fru-miR-152 MIMAT0003103 *Fugu rubripes* miR-152, tni-miR-152 MIMAT0003104 *Tetraodon nigroviridis* miR-152, DRE-miR-725 MIMAT0003753 *Danio rerio* miR-725, ATH-miR862-5p MIMAT0004307 *Arabidopsis thaliana* miR862-5p, mmu-miR-743b-3p MIMAT0004840 *Mus musculus* miR-743b-3p, gma-miR1534 MIMAT0007397 *Glycine max* miR1534, mml-miR-891 MIMAT0006530 *Macaca mulatta* miR-891, and osa-miR1852 MIMAT0007772 *Oryza sativa* miR1852.

candidate sequences and alignments with known miRNAs are shown in figure 2B. The studied 3' UTRs in *Trichomonas* are also relatively short (Davis-Hayman et al. 2000; Leon-Sicaireos Cdel et al. 2003; Leon-Sicaireos et al. 2004). Consequently, 50-nt 3' to all annotated genes were extracted to represent possible 3' UTRs in *Trichomonas* in this study. Four of seven miRNA candidates of *Trichomonas* (Tvm2, Tvm3, Tvm6, and Tvm7) showed extensive matches to 3' UTRs; therefore, they were predicted to potentially target these UTR sequences. In addition, four *Trichomonas* candidates (Tvm2, Tvm3, Tvm4, and Tvm7) also have full-length antisense matches to annotated genes. Examples of *Trichomonas* potential miRNA-target binding are shown in figure 3B as well as the de novo contigs where the examples of miRNAs were found.

As a negative control, two randomized databases were generated with the size equivalent to the *de novo* assembled Solexa contigs of *Giardia* or *Trichomonas*, using a Markov chain (Lowe and Eddy 1999) based on dinucleotide frequencies. Known miRNA sequences were Blasted against the two databases. Results showed an average of 11.3-nt match to the *Giardia* database and 11.4 nt for the *Trichomonas* database, indicating the homology search results presented in figure 2 are significantly positive.

Strategy-2: Searching for New miRNAs by Definition

Another way to identify additional candidates was through extracting putative miRNA-containing genomic regions. Previous studies have shown that the sequences of ncRNAs in *Giardia* are highly diverged from other eukary-

otic model organisms (Chen et al. 2007, 2008), but at least some ncRNAs (e.g., spliceosomal snRNAs and RNase P) in *Trichomonas* have a high degree of sequence similarity to ncRNAs from other organisms (Simoes-Barbosa et al. 2008). Therefore, we expect that the majority of miRNAs in *Giardia* and at least some miRNAs in *Trichomonas* do not share sequence similarity with currently published miRNAs because of the large evolutionary distance (Keeling et al. 2005). In order to look for other possible miRNA candidates, additional analyses based on structural and sequence criteria were carried out as the second strategy to isolate genomic regions possibly containing miRNA precursors. These regions would then be confirmed by Solexa short-reads coverage.

Both genomes were obtained from EuPathDB (<http://eupathdb.org/eupathdb/>). However, due to its large size (~180 Mb) the *Trichomonas* genome was first masked to exclude protein-coding and repeat regions. In general, most miRNA precursors adopt a conserved single hairpin structure, which is recognized by Dicer and Dicer-like proteins in the cytoplasm (Lee et al. 2002; Bartel 2004). A number of computational tools have been developed and used to conduct genome-wide miRNA predictions (Lim et al. 2003; Doran and Strauss 2007; Huang et al. 2007). Here, we used the algorithm smalloop (Grad et al. 2003) to look for hairpins with a length of less than 95 bases. Resulting hairpins were then filtered as described in the Experimental Procedures. To determine the threshold of miRNA target prediction based on complementary binding, we used a simulated control database with the size of *Giardia* genome. The control test showed that the average length of a random complementary binding is about 11 bp.

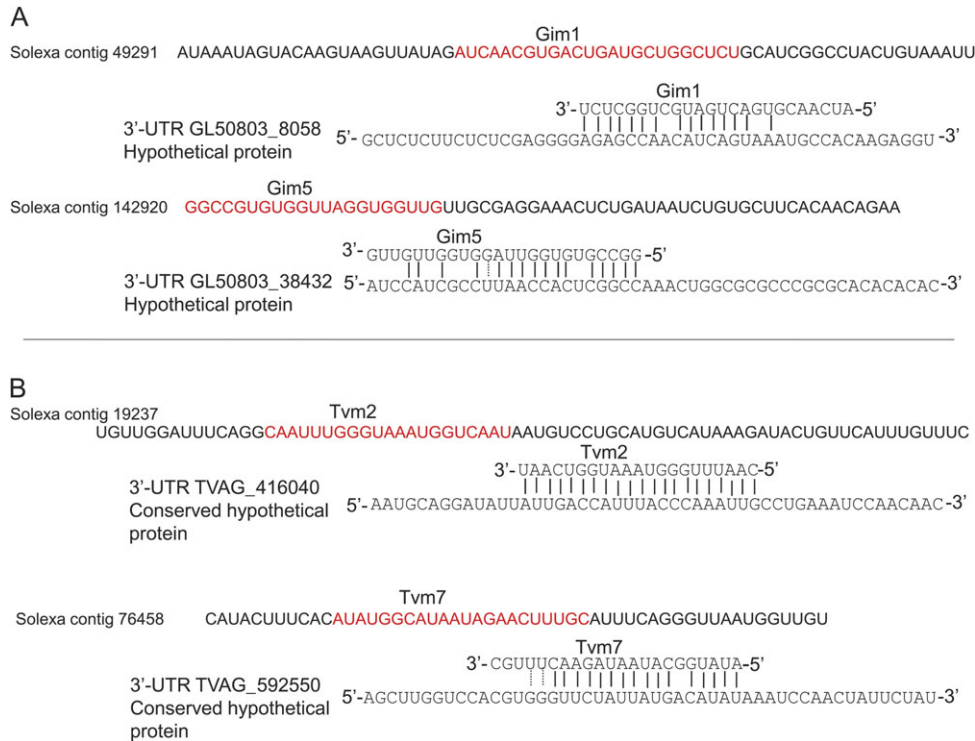


FIG. 3.—Possible miRNA-3' UTR target binding for *Giardia* and *Trichomonas* miRNA candidates resulting from the homology search: (A) Examples of *Giardia* miRNA candidates; (B) Examples of *Trichomonas* miRNA candidates. Examples of *Giardia* and *Trichomonas* miRNA candidates show extensive base pairing with predicted 3' UTR targets. In general, *Trichomonas* candidates have longer base pairing compared with *Giardia* candidates, and all predicted target sites are within 50-nt 3' to the stop codon. The candidate sequences are marked red on corresponding Solexa contigs.

Therefore, to effectively avoid false positives, only hairpins with extensive complementary bindings (over 15 nt) to 3'UTRs were considered as strong candidates. To justify the results of this approach, all the candidates were also run through the existing miRNA-target prediction software miRanda (Enright et al. 2003). To include a negative control, we used a shuffled UTR search with miRanda. The output of miRanda contains numerous false-positive predictions with average total scores of 107 for *Giardia* and 126 for *Trichomonas*. All the positive hits from miRanda are justified with infinite z-scores. The strong candidates from our own approach are also found by miRanda, with the *Giardia* candidates' scores above 140 and *Trichomonas* candidates' scores above 160, indicating positive results. Information for all miRNA candidates found is listed in table 1. Examples of predicted precursor structures and miRNA targets binding from additional computational analysis are shown in figure 4A (*Giardia*) and figure 4B (*Trichomonas*).

Most of the miRNA candidates we identified in *Giardia* and *Trichomonas* have predicted targets within the 3' UTR regions of annotated genes and some have potential antisense targets to mRNAs. The majority of *Trichomonas* miRNA candidates have many identical or nearly identical copies scattered in different genomic contigs. It is possible that some *Trichomonas* ncRNAs are highly duplicated in the same way as observed for *Trichomonas* tRNAs. However the possibility of pseudogenes cannot be excluded because the genome is currently very frag-

mented, with 17,290 scaffolds and 38,210 repeated genes (Aurrecochea et al. 2008), and many genes have not yet been characterized. Hence, we cannot determine at this stage how many copies of each ncRNA could be present.

It is necessary to mention that although we present only a small number of sequences as miRNA candidates from *Giardia* and *Trichomonas*, a larger number of other Solexa de novo contigs also have potential to be miRNA candidates (data not shown). These sequences are not presented here as candidates, either due to shorter complementary binding (between 9 and 12 bp) to predicted targets or due to the atypical folding of predicted precursors (e.g., low folding energy for many candidates in *Trichomonas*). It is possible that novel miRNAs exist in the two parasites, and further work is needed to characterize these transcripts.

Candidates of siRNAs in *Giardia*

In our previous study (Chen et al. 2007), we characterized from *Giardia* an unusual long-tandem repeated RNA named Girep-1. Continuing studies have revealed four other similar RNAs (7–10 copies in tandem), named Girep-2 to -5. The expression of RNAs from Girep-1 to -5 on both sense and antisense strands were confirmed (fig. 5A). All the Girep RNAs are non-protein coding, with an exception of the antisense strand of Girep-1 being a hypothetical mRNA transcript (GL50803_227577). The Gireps are all direct repeats located at different positions

Table 1
miRNA Candidates and Predicted Targets from *Giardia* and *Trichomonas*

Mir Candidate	Genomic Coordinates	Predicted Mature Sequence	Possible Targets
Gim1	CH991782: 610570–610593	AUCAACGUGACUGAUGCUGGCUCU	3' UTR GL50803_8508 hypothetical protein
Gim2	CH991767: 1378309–1378332	AUCUCGCACAUUAACCGGCCUCCU	Not found
Gim3	CH991769: 132337–132314	GUGCAGAGGCAUGGAGCACGGGAA	3' UTR GL50803_23634 hypothetical protein; 3' UTR GL50803_9079 ERP2
Gim4	CH991779: 949975–949998	GUGGUCUGCAUCUGGACCUUCACU	3' UTR GL50803_14602 hypothetical protein
Gim5	CH991779: 908521–908498	GGCCGUGUGGUUAGGUGGUUGUUG	3' UTR GL50803_48432 hypothetical protein; 3' UTR GL50803_10843 Thymus-specific serine protease precursor
Gim6	CH991782: 669487–669463	GUGGUGAGUAGAAGUCAGAUUAUAA	3' UTR GL50803_15063 Long chain fatty acid CoA ligase 5; 3' UTR GL50803_114815 Tenascin precursor
Gim7	CH991769: 687079–687104; 687267–687242	GCGGUCGCUUGGGUCCAGCGGGUTC	3' UTR GL50803_88901 hypothetical protein
Gim8	CH991813: 431–68 (5 copies in tandem)	GGUCGGUUAGCUCAGUCGGUAGAGCG	3' UTR GL50803_20250 hypothetical protein
Gim9	CH991776: 417414–417439	GUAGGAUGCCCCAGAGACUGCCGAG	3' UTR GL50803_13412 acidic ribosomal protein P0
Gim10	CH991782: 268174–268149	AAACUCUCCGCACAGGGGCGCGCCUG	3' UTR GL50803_94658 hypothetical protein
Tvm1	DS114515: 7733–7712	GUAUAGGUCGAGCUUGUGAAU	Not found
Tvm2	DS177933: 152–132 (first of 50 copies)	CAUUUUGGGUAAAUGGUCAAU	3' UTR TVAG_416040 conserved hypothetical protein; mRNA TVAG_225980 conserved hypothetical protein
Tvm3	DS176142: 693–673 (first of seven copies)	CAUUCAGUCAUUCUUUCUGU	3' UTR TVAG_493570 conserved hypothetical protein; mRNA TVAG_450550 conserved hypothetical protein
Tvm4	DS160029: 395–416 (first of 21 copies)	CAACAGACAUAAUGCUGAAUAG	mRNA TVAG_389700 conserved hypothetical protein
Tvm5	DS113666: 32359–32338	UAAUAUGGAAUCAGAAUGCAGU	Not found
Tvm6	DS177803: 248–269 (first of many copies)	UCAUCCUUACCUCAGUCAUUGA	3' UTR TVAG_186870 conserved hypothetical protein
Tvm7	DS177310: 460–481 (first of 49 copies)	AUAUGGCAUAAUAGAACUUUGC	3' UTR TVAG_592550 conserved hypothetical protein; mRNA TVAG_089350 conserved hypothetical protein
Tvm8	DS174663: 730–750 (first of many copies)	UUGAAAAUAGAUGGUUCGC	3' UTR TVAG_080530 conserved hypothetical protein; mRNA TVAG_567710 conserved hypothetical protein
Tvm9	DS162040: 804–784 (first of two copies)	UUUCAAUUGGACAAUUGAAU	3' UTR TVAG_518140 conserved hypothetical protein; mRNA TVAG_258870 conserved hypothetical protein
Tvm10	DS176757: 229–209 (first of many copies)	CUGUUUGGAAGUUGUAUCCA	3' UTR TVAG_029730 conserved hypothetical protein
Tvm11	DS177474: 213–232 (first of many copies)	UGCGACAAGCUUGCCCAUGG	3' UTR TVAG_140140 conserved hypothetical protein; mRNA TVAG_123455 conserved hypothetical protein

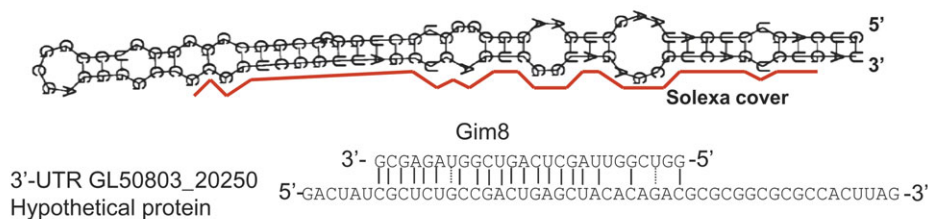
in the genome. Multiple alignments of all Gireps revealed considerable homology among the five sets of sequences and also the presence of shared motifs. The shared motifs and tandem-repeated pattern suggests that these RNAs belong to one group. All five Gireps show a high degree of sequence similarity with a number of VSP genes. The patterns of sequence match are variable but all involve the repeating units of Gireps being partially aligned to repeating units of VSP genes (fig. 5B).

VSP gene expression is crucial for the surface antigenic variation of *Giardia* trophozoites (Nash et al. 1988). The sequences and structures of VSP proteins are highly similar; however, in a single trophozoite, only one VSP is expressed at a time out of a total of 150–200 genes (Nash et al. 2001). Both RNAi (Ullu et al. 2004) and epigenetic mechanisms (Kulakova et al. 2006) have been suggested for the regulation of VSP gene expression. A recent study has identified a snoRNA-derived miRNA

that has the potential to regulate VSP expression (Saraiya and Wang 2008), and another study has demonstrated that Dicer, Argonaut, and RdRp could be involved in RNAi regulation of VSP expression (Prucca et al. 2008). Based on previous studies on *Giardia* RNAi (Ullu et al. 2005; Macrae et al. 2006; MacRae et al. 2007; Prucca et al. 2008; Saraiya and Wang 2008), it is likely that small RNA regulation is involved in VSP expression. However, the overall mechanism is still uncertain. Thus, our findings of potential siRNAs show that the Girep family of RNAs have strong potential to be involved in regulation of VSP expression.

Analysis of our Solexa sequencing results reveal unequal frequencies of matching reads on the sense and antisense strands of Gireps, as shown in table 2. (i.e., the numbers of Solexa short-reads matching to the plus and minus strands of Gireps are uneven.) In addition, Blast showed that all the sense and antisense transcripts of Gireps have sequence matches to *Giardia* mRNAs including many

A Precursor structure of Gim8



B Precursor structure of Tvm9

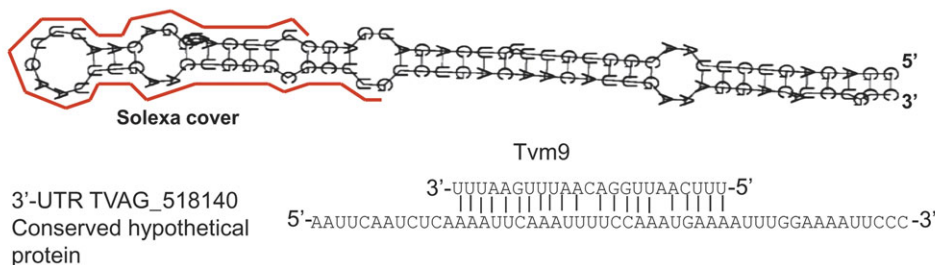


FIG. 4.—Examples of additional miRNA candidates from genomic sequence analysis: (A) Examples of *Giardia* miRNA candidates and predicted target binding. (B) Examples of *Trichomonas* miRNA candidates and predicted target binding. Regions on the predicted precursor sequences marked by red are covered by Solexa de novo contigs. The example of *Trichomonas* miRNA (B) shows a feature of many candidates that the mature miRNA may be in the loop region of the precursor hairpin. However, more evidence is needed to show if this feature is general in *Trichomonas*. The *Giardia* example (A) shows a typical stem location of the mature miRNA.

VSP genes. In total, there are 18 mRNAs that have regions with high degree of sequence similarity to Gireps. Each Girep sequence is similar to more than one VSP gene (supplementary table S1, Supplementary Material online). Searching the *Giardia* genome revealed additional non-repeated sequences that are similar to the Gireps. Also comparing Gireps with the latest *Giardia* EST database (Morrison et al. 2007) has revealed many matches. This observation suggests that a large portion of the total VSP genes are covered by expressed homologous noncoding sequences. A recent study on *Giardia* RNAi showed that similar mRNAs of different VSPs could be cleaved by Dicer to produce short RNAs (Prucca et al. 2008). This leads to the possibility that Girep RNAs can act as matching RNAs to bind VSPs and result in the production of siRNAs, which in turn silence other homologous VSPs. It has been previously shown that both sense and antisense siRNAs can downregulate gene expression in other organisms (Schwarz et al. 2003; Lin et al. 2005; Clark et al. 2008). Therefore, the sense and antisense transcripts of Gireps are likely to function in a similar way.

Solexa sequencing enables the detection of transcription of both sense and antisense strands. Mapping Solexa output sequences to VSP genes revealed bidirectional transcription, consistent with results from a previous study (Prucca et al. 2008). However, the numbers of hits for each strand are highly unequal (table 2). Although 1,633 Solexa sequences map to the plus strands of 147 VSPs, only 133 sequences map to the minus strands of 69 VSPs. Among all the VSPs, 44 have Solexa hits on both strands. This phenomenon is consistent with results from other species (e.g., human) that both strands can be transcribed (Werner

et al. 2009). With both strands indicating expression in our results, we cannot determine if sense, antisense, or regulation from both strands is involved in *Giardia* Girep–VSP regulation. We do, however, suggest that such bidirectional expression may be more common among eukaryotes than originally thought.

New ncRNAs Identified from *Trichomonas*

We identified new noncoding RNAs from *Trichomonas*, including eight C/D box snoRNAs, and we also confirmed the expression of *Trichomonas* MRP which has only been predicted computationally (Piccinelli et al. 2005). In this study, C/D box snoRNAs were initially identified by snoscan-0.9b (Schattner et al. 2005) with rRNAs used as potential targets. Solexa sequencing confirmed expression of these snoRNAs. Analysis of these *Trichomonas* C/D box snoRNAs indicated that the identified C/D box snoRNAs can adopt either of the two common structures shown in figure 6 depending on the length of the RNA. Longer snoRNAs appear to have a D' box, and the antisense recognition regions to rRNAs are more likely to locate towards the 3' ends of snoRNAs, whereas shorter snoRNAs tend to have the antisense recognition regions towards the 5' ends of snoRNAs.

The conserved structures of *Trichomonas* C/D box snoRNAs are relatively reduced compared with model eukaryotes by lacking C' boxes and terminal stems. This is similar to snoRNAs previously identified in *Giardia* (Yang et al. 2005; Chen et al. 2007), but the functional sequence motif C box is slightly longer and more conserved than in *Giardia*. Together with the fact that ncRNAs in *Trichomonas*, such as snRNAs (Simoes-Barbosa et al. 2008), are more

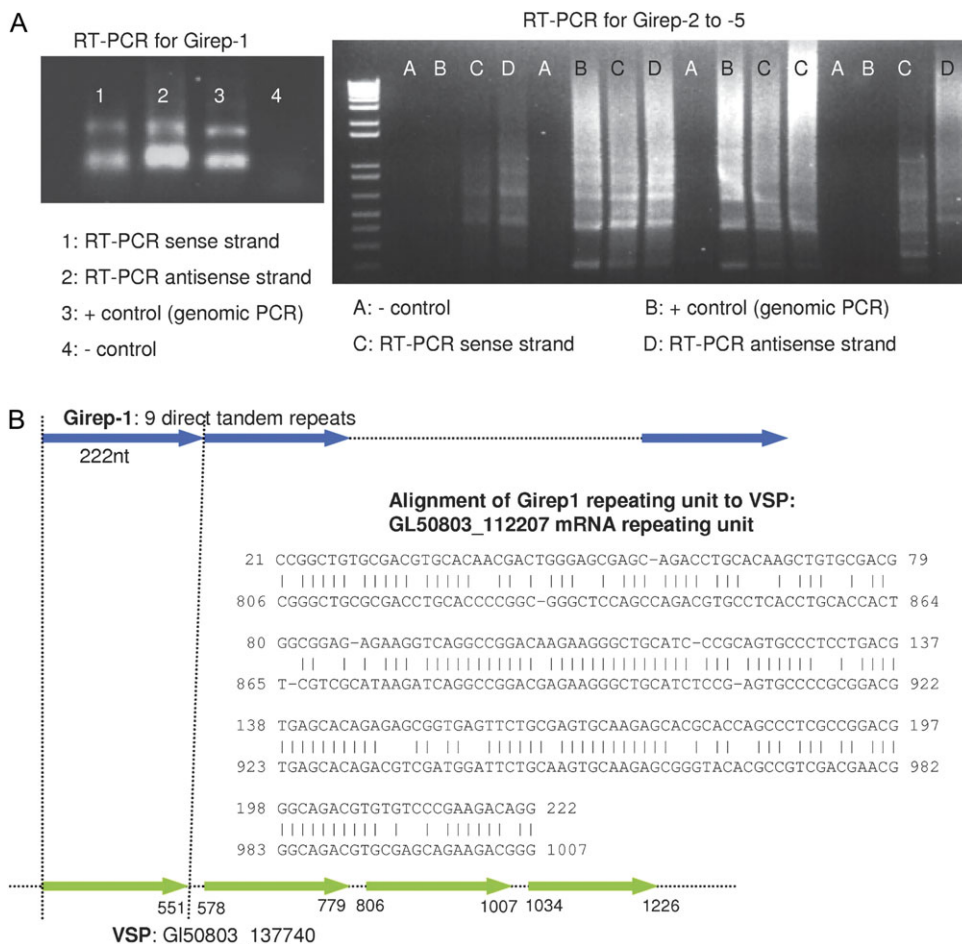


FIG. 5.—Expression of Girep RNAs and alignment with VSP mRNA sequence. (A) Expression of sense and antisense strands of Girep RNAs. From the figure, it is clear that at least one of each of the Girep sequences are transcribed at both sense and antisense strands, indicated by the reverse transcriptase–polymerase chain reaction (RT-PCR). The products of RT-PCR and positive control PCR all have multiple bands, indicating the tandem repeating pattern of Girep sequences. (B) General pattern of matching between a Girep sequence and a VSP mRNA. This figure shows the sequence alignment between the 222-nt repeating unit of Girep-1 and the repeating unit of *Giardia* VSP gene (GL50803_137740). The two sequences are highly similar at sequence level, indicating a strong relation between them.

similar to those of higher eukaryotes, it is possible that the general RNA-processing in *Trichomonas* represents an evolutionarily less reduced state than *Giardia*.

Discussion

Evolution of RNAi Inferred from Studies of Parasitic Protists

The evolutionary relationships of the deepest lineages among eukaryotes is yet uncertain (Keeling et al. 2005;

Hampel et al. 2009). Our strategy has been to look for common features in all deep eukaryotic lineages in order to infer the features of the last common ancestor of all living eukaryotes, thereafter ancestral eukaryote (Collins and Penny 2005). Results from various studies (Malhotra et al. 2002; Ullu et al. 2004, 2005; Macrae et al. 2006; MacRae et al. 2007; Prucca et al. 2008; Saraiya and Wang 2008) have led to the idea that the RNAi mechanism is likely to have occurred in the ancestral eukaryote (Collins and Penny 2009). *Giardia* and *Trichomonas* have both gone

Table 2
Gireps and Solexa Short-Read Coverage

Name	No. of Repeating Units	Length of Repeating Unit	Location	Solexa + Strand Hits	Solexa – Strand Hits
Girep-1	9	222	CH991779:330469–332457	7	0
Girep-2	8	222	CH991779:395054–396828	14	1
Girep-3	7	228	CH991782:209411–210987	44	5
Girep-4	8	228	CH991804:1–1810	2	46
Girep-5	10	225	CH991763:274924–277375	13	2

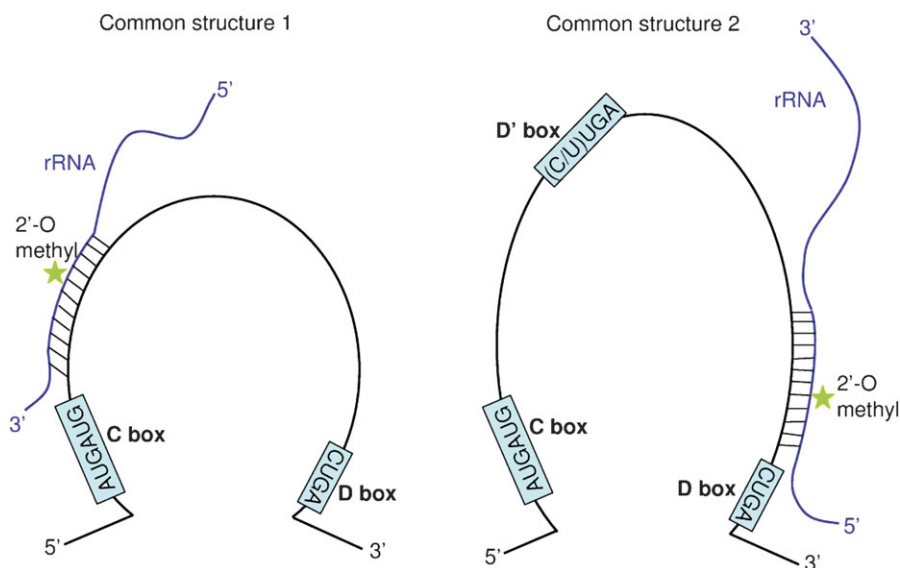


FIG. 6.—Common structures of *Trichomonas* C/D box snoRNAs. *Trichomonas* C/D box snoRNAs adopt either of the above common structures. Shorter snoRNAs tend to have the left-hand side form and longer ones the right-hand side form. Shorter sequences usually have ribosomal recognition sites close to their 5' ends, whereas longer sequences generally have less conserved D' boxes and ribosomal recognition regions close to their 3' ends.

through reductive evolution (Tovar et al. 2003; Dyall et al. 2004), yet they are distantly related, making them comparable models for inferring properties of ancestral eukaryote.

Recent studies have begun to explore the mechanism of *Giardia* RNAi. However, despite the extensive biochemical studies on the protein components (Macrae et al. 2006; MacRae et al. 2007; Prucca et al. 2008), little is known about the endogenous RNAs that may be involved in RNAi and other types of small-RNA regulated gene expression. To date only one report has characterized a single miRNA in *Giardia* (Saraiya and Wang 2008). Our approach has revealed the existence of many miRNA candidates in both *Giardia* and *Trichomonas*, indicating that, like other well-studied model eukaryotes, Excavates also possess small RNAs functioning in RNAi and other regulatory pathways. Therefore, despite some differences among individuals, it appears that all major lineages of eukaryotes share common general features of RNAi.

There are some differences between *Giardia* and *Trichomonas* miRNAs. *Giardia* miRNA candidates generally have less extensive base pairing to 3' UTR targets than *Trichomonas*. However, shorter complementary binding can result in effective RNAi (Lin et al. 2005), and therefore, it is possible that *Giardia* miRNAs do not require full complementarity to their targets, and a single miRNA may target a number of different UTR regions. Compared with *Giardia*, *Trichomonas* miRNA candidates appear more typical in their target recognition. A previous study on *Trichomonas* spliceosomal snRNAs (Simoes-Barbosa et al. 2008) revealed high degree of similarity to human snRNAs. This is consistent with our observation that *Trichomonas* miRNAs found in this study have a more typical eukaryotic target-recognition feature.

Both *Giardia* and *Trichomonas* appear to have reduced protein components in their RNAi pathways. The *Giardia* Dicer protein lacks the dsRNA-binding domain

and DEAD-box helicase domain (compared with the human Dicer) but still can fully function to cleave synthetic dsRNAs in vitro and in vivo (Macrae et al. 2006; MacRae et al. 2007; Prucca et al. 2008). The Ago protein of *Giardia* has also been shown to be functional despite lacking a PAZ domain (Prucca et al. 2008). Biochemical studies have yet to be done on *Trichomonas* RNAi proteins. The predicted Dicer protein homologue in *Trichomonas* lacks a PAZ domain compared with *Giardia*, but it has a typical Ago protein homologue. From the presence of protein homologues and miRNA candidates, we can suggest that *Trichomonas* is likely to have a typical RNAi pathway.

In addition to miRNA candidates, it is highly likely that the Girep RNAs identified from *Giardia* can act as sense- or antisense-matching RNAs to VSP mRNAs and produce siRNAs in a pathway involving Dicer. Based on the evidence of VSP gene regulation by homologous VSP mRNAs in a recent study (Prucca et al. 2008), Girep RNAs may well be a class of *Giardia* endogenous RNAs that participate in VSP gene regulation. However, this suggestion requires further experimental verification. Together with evidence from another study that snoRNA-derived miRNAs may also function in VSP gene regulation

(Saraiya and Wang 2008), we may conclude at this stage that VSP genes may be regulated by a combination of small RNAs derived from different sources.

In this study, we used high throughput sequencing technology (Solexa sequencing, Illumina) to look for previously unidentified small RNAs from the genomes of two distantly related Excavate parasites *Giardia* and *Trichomonas*. Comparing the identified small RNA contents in the two parasites has led to a better understanding of the evolution of RNA processing throughout the Excavates group and in relation with other eukaryotic lineages. As well as confirming previously identified RNAs in both organisms, including many lowly expressed transcripts, and in addition

to the new ncRNAs reported here, the remaining Solexa output sequences are yet to be characterized. Current work is underway to analyze Solexa sequences for additional characteristics based on genome location (e.g., 5' UTRs, etc.), sequence, and structural similarities. It is certain that next-generation RNA sequencing to a high coverage can uncover novel ncRNAs that cannot be characterized with traditional methods, thus providing valuable information on the genome-wide picture of RNA processing.

Funding

This work was supported by the Allan Wilson Centre of Molecular Ecology and Evolution; and the Institute of Molecular Biosciences.

Acknowledgments

We thank Lorraine Berry and Maurice Collins of the Allan Wilson Centre Genome Service for sample preparation and sequencing. The *Giardia* culture was kindly provided by Errol Kwan, Protozoa Research Unit, Hopkirk Institute, Massey University; and the *Trichomonas* samples were collected by Lynn Rogers at Medlab Central, Palmerston North. Tim White helped with simulated databases.

Literature Cited

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215:403–410.
- Aurrecoechea C, et al. 2008. GiardiaDB and TrichDB: integrated genomic resources for the eukaryotic protist pathogens *Giardia lamblia* and *Trichomonas vaginalis*. *Nucleic Acids Res.* 37:D526–D530.
- Bartel DP. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell.* 116:281–297.
- Bernstein E, Caudy AA, Hammond SM, Hannon GJ. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature.* 409:363–366.
- Chen XS, Rozhdetsvensky TS, Collins LJ, Schmitz J, Penny D. 2007. Combined experimental and computational approach to identify non-protein-coding RNAs in the deep-branching eukaryote *Giardia intestinalis*. *Nucleic Acids Res.* 35:4619–4628.
- Chen XS, White WT, Collins LJ, Penny D. 2008. Computational identification of four spliceosomal snRNAs from the deep-branching eukaryote *Giardia intestinalis*. *PLoS ONE.* 3:e3106.
- Clark PR, Pober JS, Kluger MS. 2008. Knockdown of TNFR1 by the sense strand of an ICAM-1 siRNA: dissection of an off-target effect. *Nucleic Acids Res.* 36:1081–1097.
- Collins L, Penny D. 2005. Complex spliceosomal organization ancestral to extant eukaryotes. *Mol Biol Evol.* 22:1053–1066.
- Collins LJ, Macke TJ, Penny D. 2004. Searching for ncRNAs in eukaryotic genomes: maximizing biological input with RNAmotif. *J Integr Bioinformatics.* Available at: <http://journal.imbio.de/>.
- Collins LJ, Penny D. 2009. The RNA infrastructure: dark matter of the eukaryotic cell? *Trends Genet.* 25:120–128.
- Collins RE, Cheng X. 2006. Structural and biochemical advances in mammalian RNAi. *J Cell Biochem.* 99:1251–1266.
- Davis-Hayman SR, Shah PH, Finley RW, Lushbaugh WB, Meade JC. 2000. *Trichomonas vaginalis*: analysis of a heat-inducible member of the cytosolic heat-shock-protein 70 multigene family. *Parasitol Res.* 86:608–612.
- Doran J, Strauss WM. 2007. Bio-informatic trends for the determination of miRNA-target interactions in mammals. *DNA Cell Biol.* 26:353–360.
- Dyall SD, Yan W, Delgadillo-Correa MG, Lunceford A, Loo JA, Clarke CF, Johnson PJ. 2004. Non-mitochondrial complex I proteins in a hydrogenosomal oxidoreductase complex. *Nature.* 431:1103–1107.
- Elmendorf HG, Singer SM, Nash TE. 2001. The abundance of sterile transcripts in *Giardia lamblia*. *Nucleic Acids Res.* 29:4674–4683.
- Embley TM, van der Giezen M, Horner DS, Dyal PL, Foster P. 2003. Mitochondria and hydrogenosomes are two forms of the same fundamental organelle. *Philos Trans R Soc Lond B Biol Sci.* 358:191–201 discussion 201–192.
- Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. 2003. MicroRNA targets in *Drosophila*. *Genome Biol.* 5:R1.
- Finn RD, et al. 2006. Pfam: clans, web tools and services. *Nucleic Acids Res.* 34:D247–D251.
- Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature.* 391:806–811.
- Forrest EC, Cogoni C, Macino G. 2004. The RNA-dependent RNA polymerase, QDE-1, is a rate-limiting factor in post-transcriptional gene silencing in *Neurospora crassa*. *Nucleic Acids Res.* 32:2123–2128.
- Gazzani S, Lawrenson T, Woodward C, Headon D, Sablowski R. 2004. A link between mRNA turnover and RNA interference in *Arabidopsis*. *Science.* 306:1046–1048.
- Grad Y, Aach J, Hayes GD, Reinhart BJ, Church GM, Ruvkun G, Kim J. 2003. Computational and experimental identification of *C. elegans* microRNAs. *Mol Cell.* 11:1253–1263.
- Grewal SI, Elgin SC. 2007. Transcription and RNA interference in the formation of heterochromatin. *Nature.* 447:399–406.
- Hammond SM, Boettcher S, Caudy AA, Kobayashi R, Hannon GJ. 2001. Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science.* 293:1146–1150.
- Hampel V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AG, Roger AJ. 2009. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic “supergroups.” *Proc Natl Acad Sci USA.* 106:3859–3864.
- Hartig JV, Tomari Y, Forstemann K. 2007. piRNAs—the ancient hunters of genome invaders. *Genes Dev.* 21:1707–1713.
- Huang TH, Fan B, Rothschild MF, Hu ZL, Li K, Zhao SH. 2007. MiRFinder: an improved approach and software implementation for genome-wide fast microRNA precursor scans. *BMC Bioinformatics.* 8:341.
- Jiang H, Wong WH. 2008. SeqMap: mapping massive amount of oligonucleotides to the genome. *Bioinformatics.* 24:2395–2396.
- Keeling PJ, Burger G, Durnford DG, Lang BF, Lee RW, Pearlman RE, Roger AJ, Gray MW. 2005. The tree of eukaryotes. *Trends Ecol Evol.* 20:670–676.
- Kulakova L, Singer SM, Conrad J, Nash TE. 2006. Epigenetic mechanisms are involved in the control of *Giardia lamblia* antigenic variation. *Mol Microbiol.* 61:1533–1542.
- Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, Kingston RE. 2006. Characterization of the piRNA complex from rat testes. *Science.* 313:363–367.
- Lee Y, Jeon K, Lee JT, Kim S, Kim VN. 2002. MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J.* 21:4663–4670.

- Lee Y, et al. 2003. The nuclear RNase III Droscha initiates microRNA processing. *Nature*. 425:415–419.
- Leon-Sicairos Cdel R, Perez-Martinez I, Alvarez-Sanchez ME, Lopez-Villasenor I, Arroyo R. 2003. Two *Trichomonas vaginalis* loci encoding for distinct cysteine proteinases show a genomic linkage with putative inositol hexakisphosphate kinase (IP6K2) or an ABC transporter gene. *J Eukaryot Microbiol*. 50(Suppl):702–705.
- Leon-Sicairos CR, Leon-Felix J, Arroyo R. 2004. tvcp12: a novel *Trichomonas vaginalis* cathepsin L-like cysteine proteinase-encoding gene. *Microbiology*. 150:1131–1138.
- Lim LP, Lau NC, Weinstein EG, Abdelhakim A, Yekta S, Rhoades MW, Burge CB, Bartel DP. 2003. The microRNAs of *Caenorhabditis elegans*. *Genes Dev*. 17:991–1008.
- Lin X, Ruan X, Anderson MG, McDowell JA, Kroeger PE, Fesik SW, Shen Y. 2005. siRNA-mediated off-target gene silencing triggered by a 7 nt complementation. *Nucleic Acids Res*. 33:4527–4535.
- Lowe TM, Eddy SR. 1999. A computational screen for methylation guide snoRNAs in yeast. *Science*. 283:1168–1171.
- MacRae IJ, Zhou K, Doudna JA. 2007. Structural determinants of RNA recognition and cleavage by Dicer. *Nat Struct Mol Biol*. 14:934–940.
- Macrae IJ, Zhou K, Li F, Repic A, Brooks AN, Cande WZ, Adams PD, Doudna JA. 2006. Structural basis for double-stranded RNA processing by Dicer. *Science*. 311:195–198.
- Malhotra P, Dasaradhi PV, Kumar A, Mohammed A, Agrawal N, Bhatnagar RK, Chauhan VS. 2002. Double-stranded RNA-mediated gene silencing of cysteine proteases (falcipain-1 and -2) of *Plasmodium falciparum*. *Mol Microbiol*. 45:1245–1254.
- Marquez SM, Harris JK, Kelley ST, Brown JW, Dawson SC, Roberts EC, Pace NR. 2005. Structural implications of novel diversity in eucaryal RNase P RNA. *RNA*. 11:739–751.
- Martienssen RA, Zaratiegui M, Goto DB. 2005. RNA interference and heterochromatin in the fission yeast *Schizosaccharomyces pombe*. *Trends Genet*. 21:450–456.
- Meister G, Tuschl T. 2004. Mechanisms of gene silencing by double-stranded RNA. *Nature*. 431:343–349.
- Mentel M, Martin W. 2008. Energy metabolism among eukaryotic anaerobes in light of Proterozoic ocean chemistry. *Philos Trans R Soc Lond B Biol Sci*. 363:2717–2729.
- Morris KV, Chan SW, Jacobsen SE, Looney DJ. 2004. Small interfering RNA-induced transcriptional gene silencing in human cells. *Science*. 305:1289–1292.
- Morrison HG, et al. 2007. Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. *Science*. 317:1921–1926.
- Nash TE, Aggarwal A, Adam RD, Conrad JT, Merritt JW Jr. 1988. Antigenic variation in *Giardia lamblia*. *J Immunol*. 141:636–641.
- Nash TE, Lujan HT, Mowatt MR, Conrad JT. 2001. Variant-specific surface protein switching in *Giardia lamblia*. *Infect Immun*. 69:1922–1923.
- Pasquinelli AE, Hunter S, Bracht J. 2005. MicroRNAs: a developing story. *Curr Opin Genet Dev*. 15:200–205.
- Piccinelli P, Rosenblad MA, Samuelsson T. 2005. Identification and analysis of ribonuclease P and MRP RNA in a broad range of eukaryotes. *Nucleic Acids Res*. 33:4485–4495.
- Prucca CG, Slavin I, Quiroga R, Elias EV, Rivero FD, Saura A, Carranza PG, Lujan HD. 2008. Antigenic variation in *Giardia lamblia* is regulated by RNA interference. *Nature*. 456:750–754.
- Saraiya AA, Wang CC. 2008. snoRNA, a novel precursor of microRNA in *Giardia lamblia*. *PLoS Pathog*. 4:e1000224.
- Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res*. 33:W686–W689.
- Schwarz DS, Hutvagner G, Du T, Xu Z, Aronin N, Zamore PD. 2003. Asymmetry in the assembly of the RNAi enzyme complex. *Cell*. 115:199–208.
- Sen CK, Roy S. 2007. miRNA: licensed to kill the messenger. *DNA Cell Biol*. 26:193–194.
- Simoes-Barbosa A, Meloni D, Wohlschlegel JA, Konarska MM, Johnson PJ. 2008. Spliceosomal snRNAs in the unicellular eukaryote *Trichomonas vaginalis* are structurally conserved but lack a 5'-cap structure. *RNA*. 14:1617–1631.
- Smardon A, Spoerke JM, Stacey SC, Klein ME, Mackin N, Maine EM. 2000. EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in *C. elegans*. *Curr Biol*. 10:169–178.
- Song JJ, Smith SK, Hannon GJ, Joshua-Tor L. 2004. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science*. 305:1434–1437.
- Tovar J, Leon-Avila G, Sanchez LB, Sutak R, Tachezy J, van der Giezen M, Hernandez M, Muller M, Lucocq JM. 2003. Mitochondrial remnant organelles of *Giardia* function in iron-sulphur protein maturation. *Nature*. 426:172–176.
- Ullu E, Lujan HD, Tschudi C. 2005. Small sense and antisense RNAs derived from a telomeric retroposon family in *Giardia intestinalis*. *Eukaryot Cell*. 4:1155–1157.
- Ullu E, Tschudi C, Chakraborty T. 2004. RNA interference in protozoan parasites. *Cell Microbiol*. 6:509–519.
- Vasudevan S, Tong Y, Steitz JA. 2007. Switching from repression to activation: microRNAs can up-regulate translation. *Science*. 318:1931–1934.
- Werner A, Carlile M, Swan D. 2009. What do natural antisense transcripts regulate? *RNA Biol*. 6:43–48.
- Yang CY, Zhou H, Luo J, Qu LH. 2005. Identification of 20 snoRNA-like RNAs from the primitive eukaryote, *Giardia lamblia*. *Biochem Biophys Res Commun*. 328:1224–1231.
- Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*. 18:821–829.

James Lake, Associate Editor

Accepted June 30, 2009