

Transcriptome of Small Regulatory RNAs in the Development of the Zoonotic Parasite *Trichinella spiralis*

Xiaolei Liu^{1,3}, Yanxia Song^{1,3}, Huijun Lu¹, Bin Tang¹, Xianyu Piao², Nan Hou², Shuai Peng¹, Ning Jiang¹, Jigang Yin¹, Mingyuan Liu^{1*}, Qijun Chen^{1,2*}

1 Key Laboratory of Zoonosis, Ministry of Education, Institute of Zoonosis, Jilin University, Changchun, China, **2** Laboratory of Parasitology, Institute of Pathogen Biology, Chinese Academy of Medical Sciences/Peking Union Medical College, Beijing, China, **3** Biological Therapy Center, The First Affiliated Hospital, Jilin University, Changchun, China

Abstract

Background: *Trichinella spiralis* is a parasite with unique features. It is a multicellular organism but with an intracellular parasitization and development stage. *T. spiralis* is the helminthic pathogen that causes zoonotic trichinellosis and afflicts more than 10 million people worldwide, whereas the parasite's biology, especially the developmental regulation is largely unknown. In other organisms, small non-coding RNAs, such as microRNAs (miRNA) and small interfering RNAs (siRNA) execute post-transcriptional regulation by translational repression or mRNA degradation, and a large number of miRNAs have been identified in diverse species. In *T. spiralis*, the profile of small non-coding RNAs and their function remains poorly understood.

Methodology and Principal Findings: Here, the transcriptional profiles of miRNA and siRNA in three developmental stages of *T. spiralis* in the rat host were investigated, and compared by high-throughput cDNA sequencing technique ("RNA-seq"). 5,443,641 unique sequence tags were obtained. Of these, 21 represented conserved miRNAs related to 13 previously identified metazoan miRNA families and 213 were novel miRNAs so far unique to *T. spiralis*. Some of these miRNAs exhibited stage-specific expression. Expression of miRNAs was confirmed in three stages of the life cycle by qRT-PCR and northern blot analysis. In addition, endogenous siRNAs (endo-siRNAs) were found mainly derived from natural antisense transcripts (NAT) and transposable elements (TE) in the parasite.

Conclusions and Significance: We provide evidence for the presence of miRNAs and endo-siRNAs in *T. spiralis*. The miRNAs accounted for the major proportion of the small regulatory RNA population of *T. spiralis*, while fewer endogenous siRNAs were found. The finding of stage-specific expression patterns of the miRNAs in different developmental stages of *T. spiralis* suggests that miRNAs may play important roles in parasite development. Our data provide a basis for further understanding of the molecular regulation and functional evolution of miRNAs in parasitic nematodes.

Citation: Liu X, Song Y, Lu H, Tang B, Piao X, et al. (2011) Transcriptome of Small Regulatory RNAs in the Development of the Zoonotic Parasite *Trichinella spiralis*. PLoS ONE 6(11): e26448. doi:10.1371/journal.pone.0026448

Editor: Michael Freitag, Oregon State University, United States of America

Received: June 21, 2011; **Accepted:** September 26, 2011; **Published:** November 1, 2011

Copyright: © 2011 Liu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was supported by the National Natural Science Foundation of China (NSFC 30625029, 30825033, 31030064, 30950110328, 81070311), the intramural grant from Institute of Pathogen Biology, CAMS (2008IPB204), and the national science and technology specific projects (2008ZX-10004-011, 2008ZX10401). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: cqj@jlu.edu.cn (QC); liumy@jlu.edu.cn (ML)

† These authors contributed equally to this work.

Introduction

Parasites of the genus *Trichinella* are a group of pathogens with diverse biological and pathological features. Phylogenetic analysis of the mitochondrial DNA of the species identified so far indicated that there are two genetic clades that form unique monophyletic lineages [1]. One clade is represented by *T. pseudospiralis*, which does not capsulate in the muscle cells, while the other clade is represented by *T. spiralis* while does encapsulate in muscle cells, named nurse cells, with a parasite inside the cell, surrounded by a sick layer of collagen. Parasites of *Trichinella* genus are unique intracellular pathogens. Interestingly, their entire life cycle can be completed within an animal, and trichinellosis in human and other mammals was caused through the ingestion of parasite-contaminated meat. This is a typical zoonotic disease that affects more than 10 million people worldwide [2].

After being ingested with the infected muscle tissue, L1 larvae are released and activated in the small intestine, enter the epithelial layer and undergo four times of moultings before maturation into adult worms. Mating is initiated on day 2 after infection and newborn larvae are released by the females into the mucosa as early as 4 days post infection (dpi) [3]. The larvae migrate through the lymphatic and blood vessels, invade striated muscle cells and develop into the infective L1 stage over a period of 2–3 weeks to the next host which complete the life cycle [4]. Thus, unlike other nematodes, *T. spiralis* does not have an embryonic developmental stage in the egg, which differs markedly in biological and molecular characteristics from other nematodes, especially the well-characterized free-living worm *Caenorhabditis elegans*. Intriguingly, compared to other nematodes, *T. spiralis* has a much smaller genome with 64 Mb in nuclear DNA, which

contains $\approx 15,808$ genes [5]. The availability of genome sequence information has made it possible to dissect parasite biology.

Small non-coding RNAs (sncRNAs) are a large group of small endogenous RNAs that have been widely identified in animals [6], plants [7], fungi [8,9] and some viruses [10]. They are generally 21–23 nucleotides in length, which guide various processes involving sequence-specific silencing through chromatin modification, mRNA degradation, and translational repression [11–13]. Based on their origins, structures, associated proteins and biological roles, sncRNAs are divided into three general categories: microRNAs (miRNAs), endogenous small interfering RNAs (endo-siRNAs), and piwi-interacting RNAs (piRNAs) [14]. MiRNA and endo-siRNAs have been discovered in diverse animals and plants and fungi [8] [12], while piRNAs are found only in animals [15]. miRNAs are generated from precursor transcripts by two RNase III-type enzymes, Droscha and Dicer. In animals, single-stranded miRNA is incorporated into the argonaute (Ago) protein complexes (Ago) known as RNA induced silencing complexes (RISC) and binds by partially or completely complementary to the 3' untranslated region (3'UTR) of a target mRNA [6], this results negative control of gene expression by cleavage or inhibition of translation or other regulatory functions [16]. The biological function of miRNAs was first demonstrated in *C. elegans*, where two miRNAs (*Let-7* and *Lin-4*) were shown to be regulators for stage-specific differentiation of the worm [17,18]. Endogenous siRNAs are mainly derived from three sources: transposable elements (TEs), complementary annealed transcripts, also called natural antisense transcripts (NAT) and long 'fold-back' transcripts called hairpin RNAs (hpRNAs) [19]. Unlike miRNAs, endo-siRNA's function requires perfect match with the target mRNA [6,20]. MiRNAs and endogenous siRNAs play important roles in the regulation of fundamental cellular processes, including cell differentiation, stress response, apoptosis, proliferation [12] and metabolism [21,22]. Recent studies further differentiated the functions of endo-siRNAs based on the sources they were generated from [14]. The TE-derived siRNAs are likely functional in the germline cells by repression of transposon activity and consequently keep genome stability, while NAT-derived siRNAs are more functional in somatic cells [14].

In this study, the transcriptional profiles of both miRNA and endo-siRNA in three developmental stages, namely adult (Ad), new born larvae (NBL) and muscle larvae (ML) of *T. spiralis* were systematically investigated and compared by high throughput cDNA sequencing technology ("RNA-seq"). We found that miRNAs were mainly expressed in the adult worm stage and endo-siRNAs were predominantly derived from transposable elements in the genome.

Results and Discussion

Summary of small RNA sequencing

To identify miRNA and endo-siRNA involved in the development of *T. spiralis*, small RNA libraries were generated from three life cycle stages, the new-born larvae, muscular larvae and adult worms. The libraries were directly sequenced using Solexa sequencing technology. A total of about 40 million unfiltered sequence reads (12,368,833 reads from adult worms, 12,867,246 reads from newborn larvae and 14,495,649 reads from muscular larvae) with sizes between 18 and 30 nucleotides were obtained, respectively. After removal of low quality reads and adaptor sequences, the clean reads obtained in Ad, NBL and ML were 11,878,917 (96.0% of total reads), 12,357,960 (96.0% of total reads) and 14,078,375 (97.1% of total reads) respectively. Sequence characterization suggested that the small RNA pools contained miRNAs (20.63%), other non-coding RNAs (tRNA,

tRNA, snoRNA) (0.55%), mRNA-related small RNAs (32.13%), TE-related small RNAs (0.02%) and unknown small RNA transcripts (46.66%) (Fig. 1A, Table S1). The GC content of the small RNAs was around 38.7% and the distribution of the small RNA populations in the libraries generally followed the patterns identified in other organisms, such as schistosomal parasites [23–29]. The libraries generated in this study have significantly high coverage than those in a recent report, which only showed a proportion of miRNAs in one single stage of the parasite [30].

As shown in Fig. 1B, the unique clean reads obtained in Ad, NBL and ML were 1,816,014, 1,691,015 and 1,936,612, respectively (Table S2). The number of small RNAs existed in all three life cycle stages was 393,026 reads, which accounted for 21.6%, 20.3% and 23.2% of the unique clean reads of Ad, NBL and ML stage, respectively (Fig. 1B). The small RNAs commonly expressed in Ad stage and ML stage were 3 times more than that between Ad stage and NBL stage (Fig. 1B), indicating that many genes of Ad stage were activated already in ML and NBL had more stage-specific small RNAs.

Length variation of small RNAs in *T. spiralis*

The length of predicted small RNAs varied from 18 to 30 nt in the three life stages, with most of them between from 18 to 27 nt. Here the size and number of sequences in Ad and ML almost followed the same pattern, whereas the number of sequences between 18 and 22 nt was enriched in NBL (Fig. 2 and Table S3). This implied that, apart from stage-specific activation in NBL, small RNAs may undergo size differentiation during early development of the parasite. A previous study indicated that the length of the non-conserved connector helix in Dicer was the main determinant of product size of the mature small RNAs [31], providing a possible explanation for the stage-specific variation we observed. The mechanism behind the stage-specific small RNA processing remains unknown. Since both piRNA and endo-siRNA were more common in the germline, it is likely that the small RNAs found in NBL were predominantly miRNAs, which may play fundamental roles in stage-specific differentiation. Longer RNAs may be more important for parasite reproduction [32–37].

Identification of miRNAs in *T. spiralis*

To identify candidate miRNAs of *T. spiralis*, the clean unique reads obtained were mapped to the draft *T. spiralis* genome sequences (http://genome.wustl.edu/pub/organism/Invertebrates/Trichinella_spiralis/assembly/Trichinella_spiralis-1.0/) [5] using SOAP. A total of 72,110 (out of 2,099,966) unique small RNA reads that perfectly matched are referred to as miRNA candidates. After screening for secondary structure of the inverted repeats (found with Einverted of Emboss) with RNAfold and evaluation by MirCheck, a total of 240 *T. spiralis* predicted miRNAs were identified.

To further characterize miRNAs in *T. spiralis*, all predicted miRNAs obtained above were compared against a miRNA database, miRBase (Release 15.0). In total, 21 miRNAs were found that had been identified in other species, and belong to 13 different miRNA families (Table 1, Table S4 and S5). Most miRNAs contain a 7 nt region (typically positioned at the 5' side 2–8 nt) known as the "miRNA seed" sequence (Fig. 3). It has been suggested that the seed regions serve to anchor miRNAs to their mRNA targets [14].

The sequencing data showed that, of the 21 conserved miRNAs, *tsp-let-7* and *tsp-miR-87* were found to locate only in the 3' arm of their pre-miRNAs, and *tsp-miR-31* was located only in the 5' arm of the hairpin structures. The remaining 18 miRNAs were derived from both arms of the pre-miRNAs (Table S4 and S5). According to the current model of miRNA maturation, Dicer recognizes the

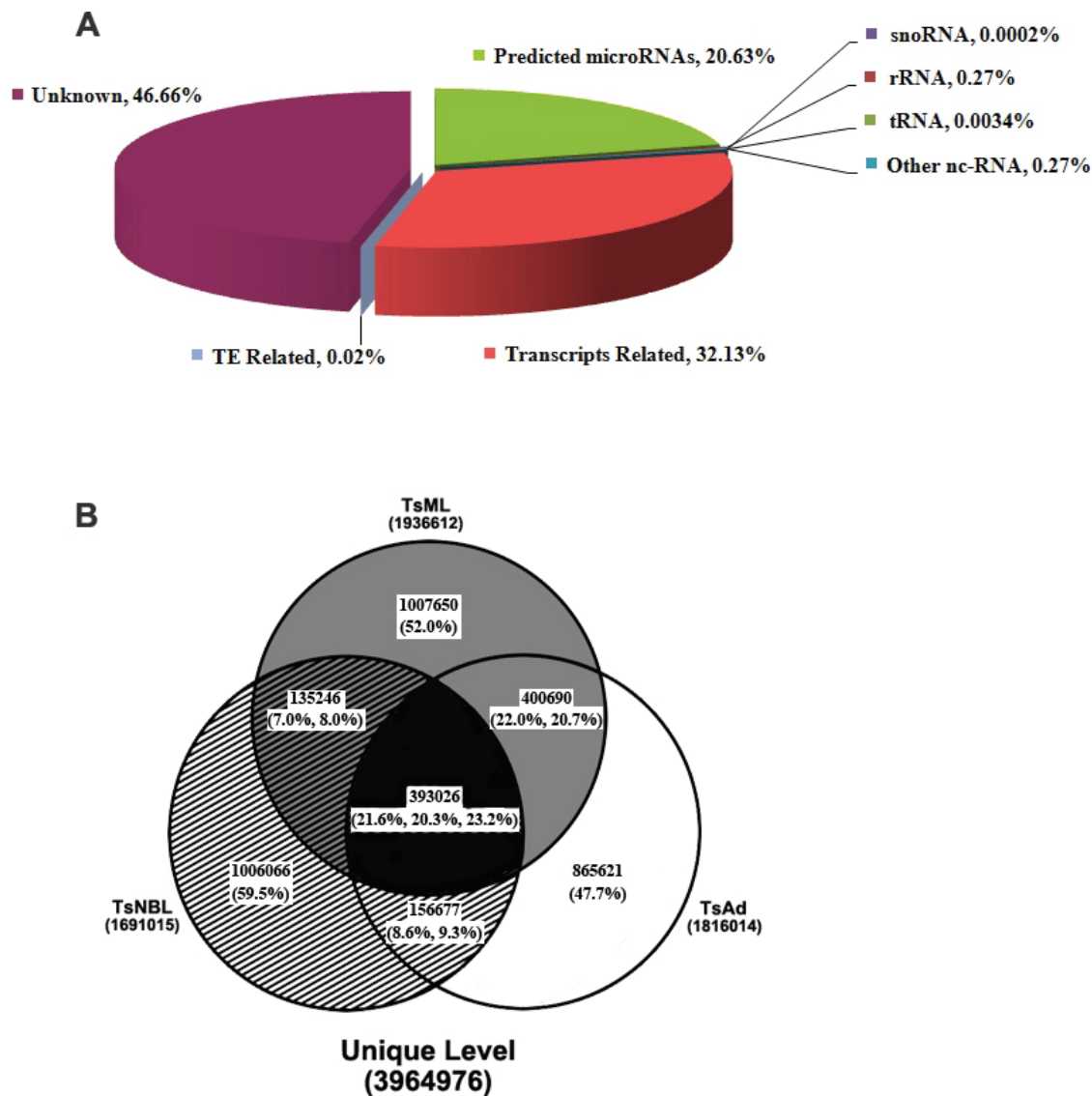


Figure 1. Small non-coding RNA identified in three developmental stages of *T. spiralis*. **A** The composition in percentages of clean reads of small RNAs. **B** Numbers and proportions (bracketed) of unique small RNAs identified in the three stages. doi:10.1371/journal.pone.0026448.g001

double-strand of pre-miRNAs, cleaves away the loop structure and de-associates the duplex to generate antisense and sense strands, *i.e.* the mature miRNA and miRNA* [38]. Although the antisense and sense strands may have different thermostability, which also determined their relative abundance in different tissues, both strands could be functional in miRNA-mediated regulatory pathways [39]. For instance, *tsp*-miR-1-3p was detected at 1,515 TPM (transcripts per million) in the Ad stage, 9,759 in the NBL stage and 3,351 in the ML stage, respectively, whereas its counterpart *tsp*-miR-1-5p was much less abundant (Fig. 4, Table S5). The TPM of *tsp*-miR-100-5p and *tsp*-miR-100-3p, *tsp*-miR-125-5p and *tsp*-miR-125-3p, *tsp*-miR-9-1-5p and *tsp*-miR-9-1-3p, *tsp*-miR-9-2-3p and *tsp*-miR-9-2-5p behaved in similar fashion (Fig. 4, Fig. 5A and Table S5). Thus the anti-sense strands complementary to mRNA targets may play a main regulatory role. However, it cannot be ruled out that both strands generated from the same pre-miRNA execute similar functions. This is because, in some miRNAs, both strands were either similarly expressed or showed stage-specific expression patterns (Fig. 5A and B; Table S5 and S6).

We found 213 novel miRNAs (Table 2, Table S7). Comparison of novel miRNAs expressed in the three developmental stages of *T. spiralis* revealed stage-specific expression patterns (Fig. 5). A majority of the novel miRNAs in NBL stage has a lower expression level than those in Ad stage and ML stage (Fig. 5B, Table S6 and S7). For instance, *tsp*-novel-21, *tsp*-novel-50a-3p, *tsp*-novel-50b-3p were relatively enriched in adult worm, *tsp*-novel-46 were abundant in muscular larvae, whereas *tsp*-novel-108 and *tsp*-novel-83 exhibited a high abundance in all the three stages (Table S7). Of the novel miRNAs, the expression patterns of *tsp*-novel-93-5p and *tsp*-novel-93-3p, *tsp*-novel-50a-3p and *tsp*-novel-50b-3p, *tsp*-novel-101-3p and *tsp*-novel-101a-5p and *tsp*-novel-101b-5p were interesting. *Tsp*-novel-93-5p and *tsp*-novel-93-3p were derived from the same pre-miRNA, but *tsp*-novel-93-5p was mainly expressed in the NBL stage, less in ML and least in Ad. While *tsp*-novel-93-3p was much less expressed than *tsp*-novel-93-5p in all three stages. (Fig. 5B, Table S7). *Tsp*-novel-50a-3p and *tsp*-novel-50b-3p were derived from the 3' arm of two gene copies with a similar sequence. *Tsp*-novel-50a-3p was dominantly expressed in

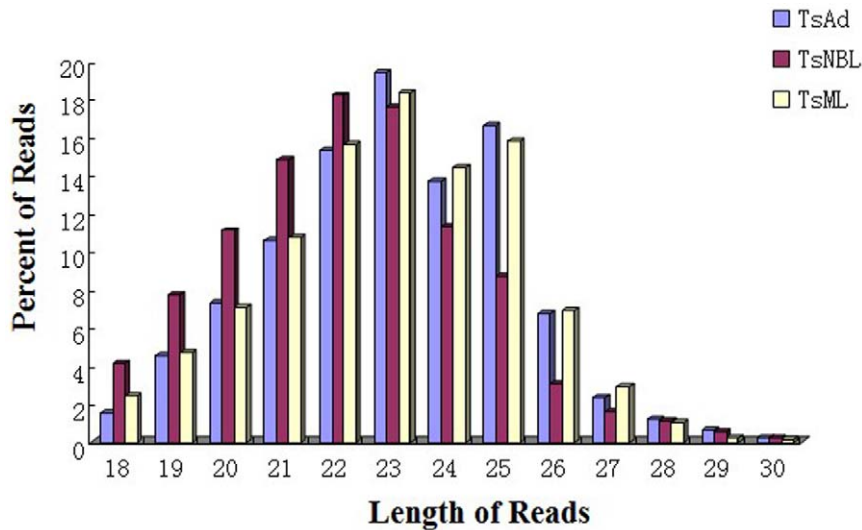


Figure 2. Length distribution of small RNAs in different developmental stages. Length of small RNAs is given on the x-axis in base pairs. And the left Y-axis indicate the percentage of small RNAs after removal of low quality sequences but prior to selection for complete matches to the genome sequence.

doi:10.1371/journal.pone.0026448.g002

Ad, while *tsp*-novel-50b-3p was mainly expressed in NBL and Ad but less in ML. The genes coding for *tsp*-novel-101-3p and *tsp*-novel-101a-5p and *tsp*-novel-101b-5p and their expression patterns were even more complicated. They were encoded by three genes in the genome of *T. spiralis*. The sequences of *tsp*-novel-101-3p derived from the 3 genes were the same, while *tsp*-novel-101b-5p was derived from 1 gene which had a single nucleotide change from *tsp*-novel-101a-5p (Fig. 5B and Table S6 and S7). The later was derived from two gene copies. All these data suggested that there may be multiple layers of control for the stage-specific

expression of miRNA genes, as well as their potential regulatory function in the development of the parasite [40]. Although the mRNA targets of these regulatory miRNAs have not been identified, with the availability of the genome sequence of the parasite, genome-wide association study can be pursued. For instance, miR-1 has been found in many species, from *Drosophila* to human [41–43], suggesting that they are evolutionary conserved, which has been characterized to play essential functions in regulating proliferation and differentiation of muscle cell [41,44]. One homologous to mir-1, *tsp*-miR-1, was observed to express in

Table 1. Conserved (common) miRNAs identified in different developmental stages.

Name	Mature Arm	miR* ^a	Most abundant sequence	Length	Expression (TPM ^b)		
					Ad	NBL	ML
<i>tsp</i> -miR-228	5'	Y	AAUGGCACUGGAUGAAUUCACGG	23	23725	19859	60017
<i>tsp</i> -miR-100	5'	Y	AACCCGUAGAUCCGAACUUGUGU	23	3028	35995	541
<i>tsp</i> -let-7	3'	N	UGAGGUAGUAGGUUGUAUAGUU	22	25573	1067	5554
<i>tsp</i> -miR-1	3'	Y	UGGAAUUGUAAAGAAGUAUGUAG	22	1515	9759	3351
<i>tsp</i> -miR-31	5'	N	AGGCAAGAUGUUGCAUAGCUGA	23	1004	7728	2025
<i>tsp</i> -miR-125	5'	Y	UCCUGAGACCCAAACUUGUGA	22	745	10	414
<i>tsp</i> -miR-252	5'	Y	CUAAGUAGUAGUGCCGAGGUC	22	193	279	290
<i>tsp</i> -miR-9-1	5'	Y	UCUUUGGUUAUCUAGCUGUAUGA	23	239	203	179
<i>tsp</i> -miR-87	3'	N	GUGAGCAAAGUUUCAGGUGUGU	22	78	160	120
<i>tsp</i> -miR-9-2	3'	Y	AUAAGCUAGUUGACCAAAGA	20	47	25	97
<i>tsp</i> -miR-29	3'	Y	UAGCACCAUUGAAUUCAGUG	21	21	8	24
<i>tsp</i> -miR-9-3	3'	Y	UAAAGCUGGAUGACCAAAGU	20	18	8	26
<i>tsp</i> -miR-993	3'	Y	GAAGCUCGUUUUCACAGG	18	5	5	16
<i>tsp</i> -miR-133	5'	Y	ACUGGUUGAGGACGUACCAAUUG	24	1	3	1
<i>tsp</i> -miR-34	5'	Y	UGGACAGUUAUUGCUGGUUGU	23	1	1	2

^aY indicates that sequences from both arms of a pre-miRNA species were found, while N means that only a sequence from one arm was identified.

^bThe abundance value of each miRNA was normalized to "transcripts per million (TPM)". If the value after normalization was less than 1, the normalized value was set as 1.

doi:10.1371/journal.pone.0026448.t001

```

AGGCAAGAUGUUGGCAUAGCUGA
UGGCAAGAUGUUGGCAUAGCUGU
AGGCAAGAUGCUGGCAUAGCUGU
AGGCAAGAUGUUGGCAUAGCUGU
UGGCAAGAUGUUGGCAUAGCUGA
UGGCAAGAUGUUGGCAUAGCUGA
UGGCAAGAUGCUGGCAUAGCUGA
UGGCAAGAUGCUGGCAUAGCUGA
UGGCAAGAUGCUGGCAUAGCUGA
UGGCAAGAUGCUGGCAUAGCUGA
UGGCAAGAUGCUGGCAUAGCUGA
UGGCAAGAUGCUGGCAUAGCUGA
UGGCAAGAUGCUGGCAUAGCUGA
UGGCAAGAUGCUGGCAUAGCUGA
UGGCAAGAUGCUGGCAUAGCUGA
AGGCAAGAUGUUGGCAUAGCUGA
AGGCAAGAUGUUGGCAUAGCUGA
AGGCAAGAUGUUGGCAUAGCUGA
AGGCAAGAUGCUGGCAUAGCUGA
    
```

tsp-miR-31
bfl-miR-31
cfa-miR-31
oan-miR-31
aae-miR-31
cqu-miR-31
dan-miR-31a
der-miR-31a
dgr-miR-31a
dme-miR-31a
dmo-miR-31a
dpe-miR-31a
dps-miR-31a
dpu-miR-31
dse-miR-31a
dsi-miR-31a
dvi-miR-31a
dwi-miR-31a
dya-miR-31a
bma-miR-72
cel-miR-72
sme-miR-31a
sme-miR-31b

Figure 3. Alignment of the *tsp*-miR-31 sequence with homologues from other organisms. The seed sequences, typically positioned at the 5' side 2–8 nucleotides, are shadowed in black. Abbreviations: *B. floridae*, *bfl*; *C. familiaris*, *cfa*; *A. aegypti*, *aae*; *O. anatinus*, *oan*; *C. quinquefasciatus*, *cqu*; *D. grimshawi*, *dgr*; *D. ananassae*, *dan*; *D. erecta*, *der*; *D. melanogaster*, *dme*; *D. mojavensis*, *dmo*; *D. persimilis*, *dpe*; *D. pseudoobscura*, *dps*; *D. sechellia*, *dse*; *D. simulans*, *dsi*; *D. willistoni*, *dwi*; *D. virilis*, *dvi*; *D. yakuba*, *dya*; *B. malayi*, *bma*; *D. pulex*, *dpu*; *C. elegans*, *cel*; *S. mediterranea*, *sme*. doi:10.1371/journal.pone.0026448.g003

three developmental stages of *T. spiralis* and may have similar functions. Further, MiR-100 and *let-7*, the two conserved miRNA in metazoa, play a role in regulation of developmental timing [18,42,45]. Their homologs, *tsp*-miRNA-100 and *tsp*-*let-7*, were found throughout the life cycle of *T. spiralis*. *Tsp*-*let-7* showed very low expression in NBL stage, whereas *tsp*-miRNA-100 was detected in rather high abundance at the same development stage. The abundance of *tsp*-miRNA-100 was almost identical with that of *tsp*-*let-7* in both Ad and ML stage, indicating that miRNA-100 may be more functional in NBL stage.

Characterization of stage-associated miRNAs in *T. spiralis*

Quantitative real-time PCR and northern blots were performed to validate miRNAs identified in *T. spiralis* and their relative expression levels at different developmental stages. Five conserved miRNAs (*tsp*-miR-228, *tsp*-miR-100, *tsp*-*let-7*, *tsp*-miR-1 and *tsp*-miR-31) and five novel miRNAs (*tsp*-novel-108, *tsp*-novel-83, *tsp*-novel-46, *tsp*-novel-86 and *tsp*-novel-21) with relatively higher TPM values identified by sequencing were validated by qRT-PCR and Northern blot. First, the above miRNAs identified in sequencing were all amplified by qRT-PCR (Fig. 6A) and except the *tsp*-*let-7* which was found less expressed in NBL, the qRT-PCR results were all consistent with the TPM values of sequencing results. The reason for *tsp*-*let-7* being found less common in NBL is not known. This may be due to preferential amplification in the tag-generation step before RNA-seq or other unidentified factors. Further, all qPCR amplicons were cloned into the pMD-18T vector (Takara, Dalian, China) and sequenced. The sequencing results demonstrated that the amplicons were identical to the miRNAs sequences (data not shown). Four miRNAs were selected for characterization by northern blots. The results confirmed the qPCR data. However, the probe for *tsp*-miR-100 did not show any signal with total RNA purified from the Ad stage; this was likely due to probe modification or other unidentified factors that prevented annealing of the probes with the targets (Fig. 6B). Since this miRNA was found by both

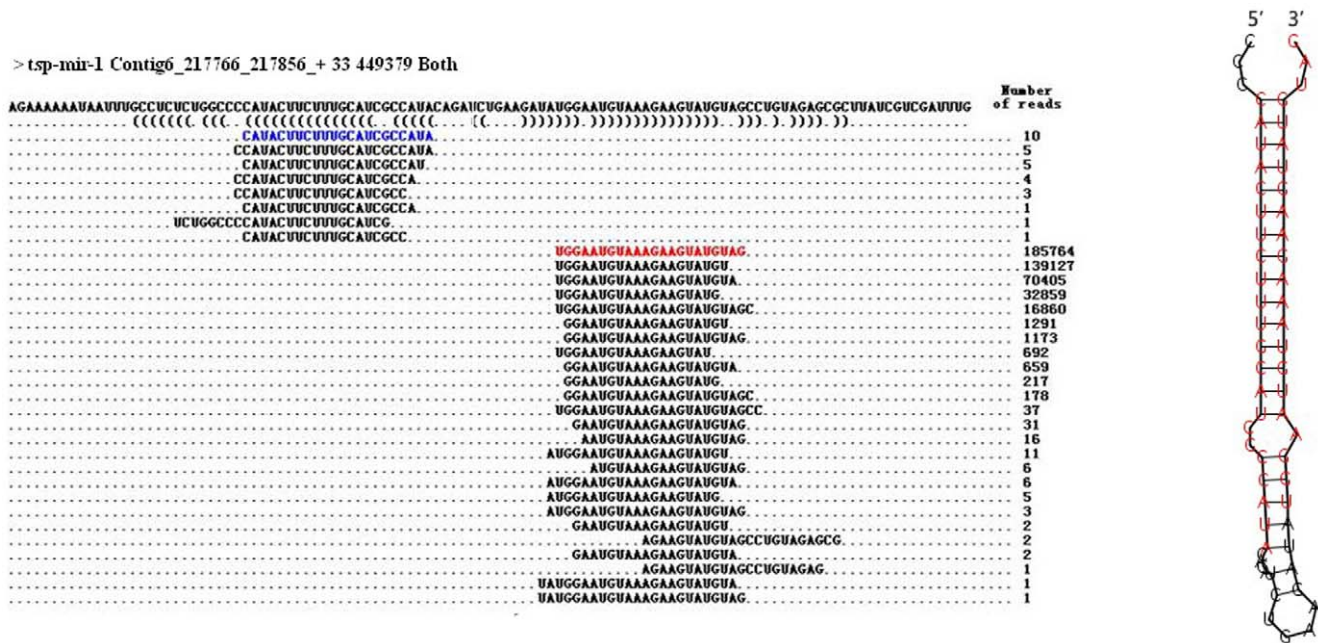


Figure 4. Sequence and predicted secondary stem-loop structure of *tsp*-miR-1 identified in *T. spiralis*. Sequences and the number of sequencing reads matching the *tsp*-miR-1 hairpin were listed. The mature miRNA of *tsp*-miR-1 and the complementary miR* are represented in red and blue respectively. The predicted structure of the pre-miRNA is represented on the right side. doi:10.1371/journal.pone.0026448.g004

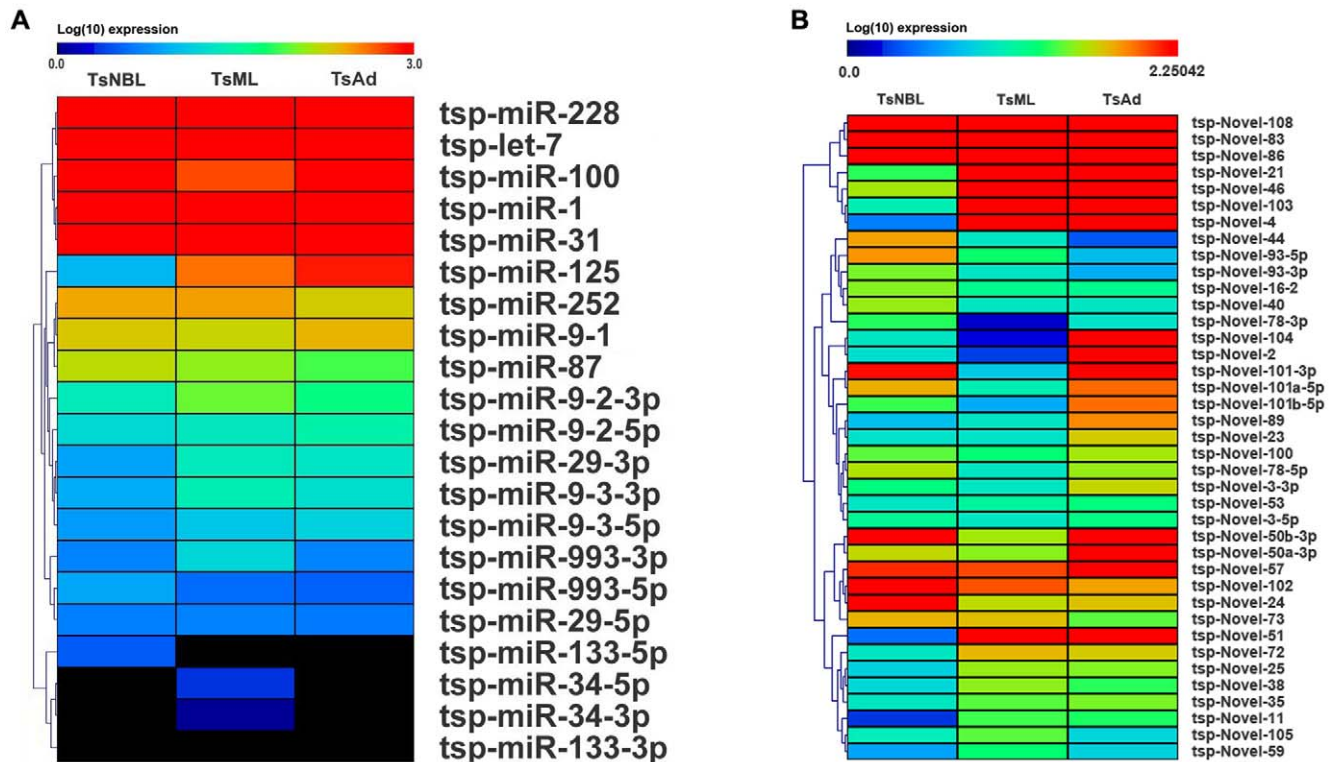


Figure 5. Clustering of expression level of conserved (A) and novel (B) miRNAs. Log₁₀ (expression) is used instead of raw expression. The abundance of each miRNA was normalized to “transcripts per million (TPM)”. If the value after normalization was less than 1, the normalized value was set as 1 to avoid the negative values.
doi:10.1371/journal.pone.0026448.g005

sequencing and qPCR in the Ad stage, we have no doubt that *tsp-mir-100* was indeed expressed at this stage.

Identification of Endogenous siRNA

Endogenous siRNA are extremely diverse and normally not conserved across species. Several types of endogenous siRNAs have been found in *Drosophila melanogaster*, *C. elegans*, *Schistosoma japonicum*, fungi, *Arabidopsis thaliana* and mice [8,9,27,33,36,46–50]. Most of these endogenous siRNAs derive from transposable elements (TE), complementary annealed transcripts (also called natural antisense transcripts, NAT) and long ‘fold-back’ transcripts called hairpin RNAs (hpRNAs) [19]. TE and NAT are the main sources of endogenous siRNAs. Thus, the small RNA transcript reads perfectly matched to TE and NAT were regarded as endogenous siRNAs (endo-siRNAs) [47,48].

TEs are major components of the intergenic regions in the genomes of eukaryotic organisms [47,48]. Based on their structures and modes of integration, TEs are comprised of two main classes [51]. One class includes retrotransposons and retrotransposon-like elements such as Long Interspersed Nuclear Elements (LINE), Long Terminal Repeat Elements (LTR), and Short Interspersed Elements (SINE). The other group includes DNA transposons. We predicted TE structures in the *T. spiralis* genome with RepeatMasker (<http://www.repeatmasker.org/>), and endo-siRNAs matched to TEs were further analyzed. As reported recently [5], the repetitive sequences in the *T. spiralis* genome estimated to be around 18%, which is much lower than in schistosomes and *Drosophila melanogaster* [52,53]. Thus, TE-derived siRNAs accounted for a minor portion of the sncRNAs identified here (Fig. 1A).

The numbers for TE-derived siRNAs identified in the Ad, NBL and ML stages were 2,055, 2,234 and 2908 respectively. No

obvious stage-related variations in endo-siRNA expression were found, suggesting that TE components in the genome might not be very active during embryogenesis or during other development stages. Further, most of these siRNAs were derived from transposons of Long Interspersed Nucleotide Elements (LINE) family and the transposon Charlie 24 accounted for a major portion (Table S8). Sequencing analysis showed that more siRNA are derived from the antisense strand than from the sense strands (data not shown). This suggested that antisense-derived siRNA may execute regulatory functions through hybridization with the mRNA template generated from the sense strand.

Since NAT-derived siRNAs were generated from double-stranded RNAs formed by complementary annealed transcripts (NAT) and long ‘fold-back’ transcripts (called hairpin RNAs, hpRNAs) [19], *T. spiralis* NAT-siRNAs were identified from predicted overlapping genes. The number of NAT-derived siRNAs identified in Ad, NBL and ML was 21,157, 16,243 and 21,251, respectively (Table S9). Thus the number of NAT-derived endo-siRNAs was much higher than that from TE, which suggests that these siRNAs played more regulatory roles in the development and parasitization of *T. spiralis*. Further, all NAT-derived siRNAs were trans-NAT siRNAs and no *cis*-NAT siRNAs were found (Fig. 7 and data not shown), which differs from observations with *D. melanogaster* and mice [33,47–49]. This also suggests that there are fewer coding sequences with internal inverted repeat sequences in the genome of *T. spiralis*.

In summary, we identified and analyzed the expression of miRNAs and endo-siRNAs in three development stages of *T. spiralis* through high through-put RNA sequencing techniques. We found vastly more transcripts of miRNAs than that of endo-siRNAs. A total of 21 conserved miRNAs in 13 metazoan miRNA families and 213

Table 2. Novel miRNAs identified in different developmental stages (only the most abundant are shown here).

MicroRNA Name	Mature Arm	miR* ^a	Most abundant sequence	Length	Expression (TPM ^b)		
					Ad	NBL	ML
<i>tsp</i> -novel-21	3'	Y	UCACCGGGUAAUUAUCACAGC	22	4448	24	348
<i>tsp</i> -novel-108	5'	Y	CUUGGCACUGUAGAUAUCACAGA	24	2705	3722	4804
<i>tsp</i> -novel-83	3'	Y	UUGAGCAAUUUGAUCGUAGC	21	2024	1545	1143
<i>tsp</i> -novel-50a	3'	Y	UCACCGGAUACUAAAACACGUU	22	1022	47	36
<i>tsp</i> -novel-50b	3'	Y	UCACCGGAUACUAAAACACGUGU	23	854	288	40
<i>tsp</i> -novel-103	5'	Y	UUUUUAUGAAGUGGUAAGUAGG	22	821	12	623
<i>tsp</i> -novel-86	3'	Y	UGAGAUCACCGUGAAAGCCUUU	22	696	333	871
<i>tsp</i> -novel-46	3'	Y	UGGACGGCGAAUAGUGGAAG	21	539	40	1789
<i>tsp</i> -novel-4	3'	Y	UGGACGGAUGCUCAGUGGAUGU	22	459	3	860
<i>tsp</i> -novel-104	3'	Y	UCACCGGGCACAAUUGGCGUC	22	270	1	2
<i>tsp</i> -novel-2	3'	Y	CACCCGGAUGCUAAAACACGUA	22	194	9	2
<i>tsp</i> -novel-51	5'	Y	UCGAAUCGCCACAUCGGAAGGC	22	182	3	171
<i>tsp</i> -novel-101	3'	Y	UCACCGGGCACUAAAUCACGUUU	23	178	152	7
<i>tsp</i> -novel-57	3'	Y	UUGAGCAAUCACAGUCGUAG	20	170	134	114
<i>tsp</i> -novel-89	3'	Y	CAUAGGAUUCUAAAACAUGCA	21	80	6	1
<i>tsp</i> -novel-102	5'	Y	ACUGAAAGAGGGAAACGGUUAG	22	69	251	107
<i>tsp</i> -novel-24	3'	Y	UGGCAUACUGGAAACGCGUAGA	23	56	189	46
<i>tsp</i> -novel-72	3'	Y	UAAUGAGCAUGUAGACCUGAGU	22	52	1	60
<i>tsp</i> -novel-100	3'	Y	GACCAAUGCGUUGAUGUAGA	20	40	29	19
<i>tsp</i> -novel-73	5'	Y	UGAAGUUGCACUGGUAUUGGU	22	29	63	58

^aY indicates that the sequences from both strands of a pre-miRNA species were found, while N means that only the sequence from one arm was identified.

^bThe abundance value of each miRNA was normalized to "transcripts per million (TPM)". If the value after normalization was less than 1, the normalized value was set as 1. doi:10.1371/journal.pone.0026448.t002

novel miRNAs were identified in the parasite. Some showed clear stage-specific expression patterns, suggesting a potential regulatory function in the corresponding developmental stages. Endo-siRNAs were mainly derived from natural antisense transcripts with TE-derived siRNAs accounting for only minor proportion of the small RNA population. Thus, the function of endo-siRNAs in *T. spiralis* is likely to regulate gene expression instead of maintaining genome stability. The data of this study provide insight information for further dissection of the parasite's biology.

Materials and Methods

Parasites

Muscle larvae (ML) of *T. spiralis* (strain ISS534) were obtained from rats 35 days post infection (dpi) by digestion of minced skeletal muscle in 1% pepsin, 1% HCl for 3 h at 37°C with agitation as previously described [54]. To purify adult worms and newborn larvae, Wistar rats at 6 weeks of age were orally inoculated with *T. spiralis* (strain ISS534) with a dose of 8000 larvae per rat. At 30 h or 6 dpi, all rats were killed and the entire intestines were removed, opened longitudinally and cut into small pieces (about 0.5–1 cm). The fragmented intestine was put on a layer of gauze which was immersed into 0.9% sodium chloride solution at 37°C and incubated for 3 h. Adult *T. spiralis* worms migrate into the liquid phase, which was harvested by centrifugation. To obtain new-born larvae, adult worms collected at 6 dpi were incubated in Iscove's Modified Dulbecco's Medium (IMDM) in 75-cm² cell culture plates at 37°C, and the newborn larvae were harvested every 12 h. Our study was reviewed and approved by the Ethics Committee of Jilin University (Ethical clearance

application number IZ-2009-III). All animal work was conducted according to Chinese and international guidelines.

RNA isolation

Total RNA of *T. spiralis* (Ad, NBL and ML) was extracted using Trizol reagent (Invitrogen, CA, USA) according to the manufacturer's instructions. RNAs were dissolved in diethylpyrocarbonate (DEPC)-treated water, aliquoted and stored at -80°C. RNA was quantified by measuring the absorbance at 260 nm with a Nanodrop 1000 machine (Thermo Scientific CA, USA).

Construction of small RNA libraries and sequencing

For small RNA library construction and deep sequencing, the 15–30 nt size range base-pair fraction of each RNA sample from the three life cycle stages was first enriched by 15% TBE urea polyacrylamide gel electrophoresis and the Illumina's proprietary adaptors (UCAGAGUUCUACAGUCCGACGAUC and UCGUAUGCCGUCUUCUGCUUGUIdT) were ligated to the 5' and 3' termini of the purified small RNAs which were converted into single-stranded cDNA with Superscript II reverse transcriptase (Invitrogen, CA, USA) and Illumina's small RNA RT-Primer. The cDNAs were PCR-amplified with high fidelity Phusion DNA polymerase (Finnzymes Oy, Finland) in 18 PCR cycles using Illumina's small RNA primer sets. After purification using 6% TBE urea PAGE gels, the PCR products were sequenced by Solexa's sequencing-by-synthesis method (Fig. S1).

Mapping the sequence reads onto the reference genome

The bioinformatic analysis and work flow is shown in Figure S1. After removing the low quality sequence reads and the adapter

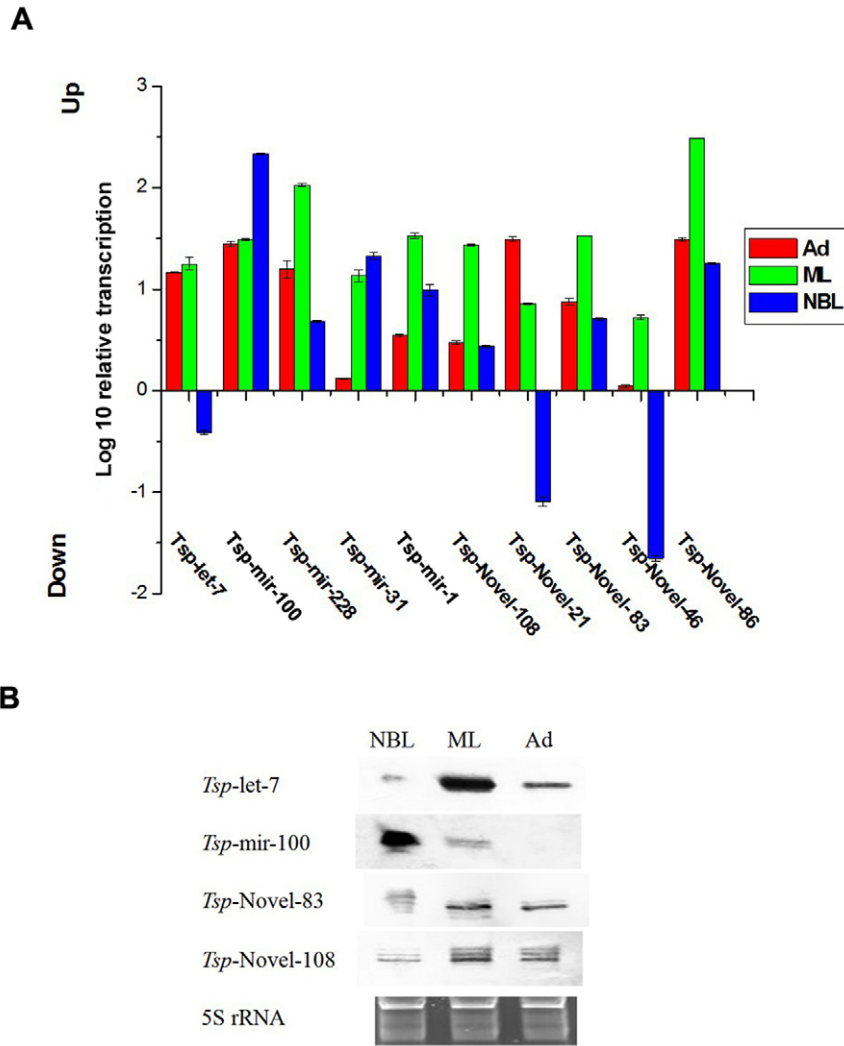


Figure 6. Confirmation of miRNA expression in three developmental stages of *T. spiralis*. **A** The expression level of ten selected miRNAs was in 3 developmental stages of *T. spiralis* measured by qRT-PCR. GAPDH was chosen as an endogenous reference. **B** Identification of the miRNAs by northern-blot. Lanes from left to right: new born larvae (NBL), muscle larvae (ML) and adult (Ad). The 5S rRNA bands were visualized by ethidium bromide staining of polyacrylamide gels and served as loading controls (bottom). doi:10.1371/journal.pone.0026448.g006

sequences according to the criteria of Illumina, all identical sequences were retained with associated count numbers as their expression abundances To determine whether these clean small RNA sequences were candidate miRNAs, the unique reads were mapped onto the *T. spiralis* genome of the Genome Sequencing Center (GSC) at

Washington University St. Louis, (http://genome.wustl.edu/pub/organism/Invertebrates/Trichinella_spiralis/assembly/Trichinella_spiralis-1.0/) with SOAP [55] (<http://soap.genomics.org.cn>).

The perfectly matched reads were searched against the Metazoa mature miRNA of Sanger miRBase with Patscan [56]. Sequence

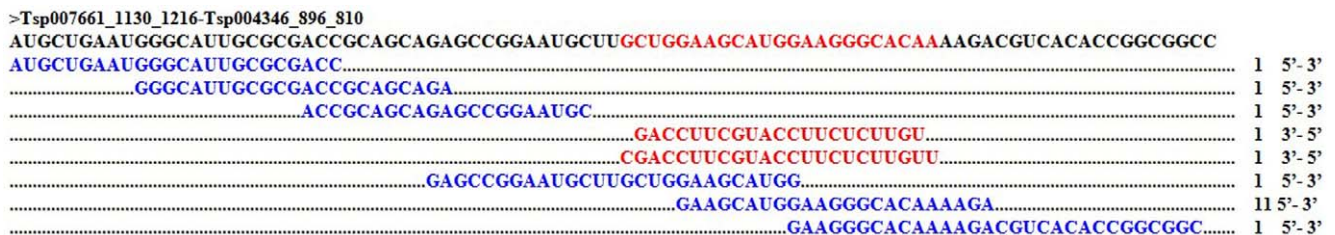


Figure 7. NAT-derived siRNA that matched to the sense and antisense strands. The origin of the genomic sequence (genomic loci) is named on top. siRNAs matched to the sense strand are in blue and those matched to antisense strand are in red. The read numbers and reading direction of the siRNA are listed on the right side. doi:10.1371/journal.pone.0026448.g007

tags of more than 5 reads that matched perfectly or near-perfectly (no more than 1 mismatch and the mismatch not positioned in the seed region) were regarded to be conserved miRNA candidates. For reads that did not match to the miRNA database, we used the software Einverted of Emboss [57] to find the inverted repeats (stem loops or hairpin structure). Each inverted repeat was extended 10 nt on each side, and the secondary structure of the inverted repeat was predicted by RNAfold [58]. Unique reads with a folding free energy of at least 25 kcal/mole ($\Delta G^{\circ}_{\text{folding}} \leq -25$ kcal/mol) were evaluated by MirCheck [59] with modified parameters. Finally, precursors (hairpins) of miRNA that passed MirCheck were inspected manually in order to remove false predictions. The reads passing the inspections were regarded as novel miRNAs.

Similar to credibility interval approaches reported for the analysis of SAGE data [60], we employed IDEG6 [61] to identify miRNAs showing statistically significant differences in relative abundance (as reflected by the total count of individual sequence reads) between the three libraries (corresponding to the three developmental stages of the parasite). We used the general Chi-square method for comparison analysis, which has been commonly applied by others [27,36]. Finally, miRNA with a P value ≤ 0.01 were deemed to be significantly different between the samples of the three developmental stages of the parasite.

The repeated sequences, i.e. potential transposable elements, in the *T. spiralis* genome were predicted by using RepeatMasker (<http://www.repeatmasker.org/>) and the sequences of typical transposons were annotated (data not shown). The sequencing reads that perfectly matched the *T. spiralis* genome were aligned to repeats (TE) using SOAP. The reads that perfectly matched TEs were considered TE-derived siRNAs.

Natural antisense transcripts (NATs) were detected by aligning the *T. spiralis* predicted genes to each other. If a pair of overlapping genes were matched on opposite strands with an E value of $\leq 1e-9$ [62], this pair of overlapping genes was defined as a NAT pair. The reads that perfectly matched the *T. spiralis* genome were aligned to overlapped sequences of NAT pairs with SOAP. The reads that perfectly matched the overlapped regions were considered NAT-derived siRNAs.

miRNAs quantification by real-time PCR

Total RNA purified from the three stages were polyadenylated with *E. coli* poly(A) polymerase (E-PAP) following the manufacturer's protocol of the poly(A)-tailing kit (Ambion, CA, USA). The RNA samples were purified separately from the parasites of the same developmental stages for sequencing purpose. After phenol-chloroform extraction and ethanol precipitation, the polyadenylated products were dissolved in DEPC-treated water and reverse-transcribed with 200 U of SuperScriptTM III Reverse Transcriptase (Invitrogen, CA) and 3' RACE Adapter (5'-GCGAGCACAGAAT-TAATACGACTCACTATAGGT12VN-3') in the FirstChoice RLM-RACE Kit (Ambion) according to the manufacturer's protocol. The 20 μ l RT reaction contained 1 μ g total RNA, 2 μ l 3' RACE Adapter, 0.5 mM dNTP mix (Takara, Dalian, China), 10 U RNase inhibitor and 200 U SuperScriptTM III Reverse Transcriptase.

In quantitative RT-PCR reactions, GAPDH (glyceraldehyde-3-phosphate dehydrogenase) was chosen as an endogenous reference. The forward and reverse primers for GAPDH were 5'-GTGCTGATTACGCTGTTG-3' and 5'-CTAAGCCATTGG-TAGTGC-3'. PCR was done on an Applied Biosystems 7500 system. The following forward primers were designed to confirm the sequencing results of miRNAs that showed differential expression patterns: *tsb*-miR-100 5'-AAC CCG TAG ATC CGA ACT TGT GT-3'; *tsb*-let-7 5'-TGA GGT AGT AGG

TTG TAT AGT T-3'; *tsb*-miR-228 5'-AAT GGC ACT GGA TGA ATT CAC GG-3'; *tsb*-miR-1 5'-TGG AAT GTA AAG AAG TAT GTA G-3'; *tsb*-miR-31 5'-AGG CAA GAT GTT GGC ATA GCT GA-3'; *tsb*-novel-108 5'-CTT GGC ACT GTA AGA ATT CAC AGA-3'; *tsb*-novel-83 5'-TTG AGC AAT TTT GAT CGT AGC-3'; *tsb*-novel-46 5'-TGG ACG GCG AAT TAG TGG AAG-3'; *tsb*-novel-86 5'-TGA GAT CAC CGT GAA AGC CTT T-3'; *tsb*-novel-21 5'-TCA CCG GGT AAT AAT TCA CAG C-3'. The sequence 5'-GCG AGC ACA GAA TTA ATA CGA CT-3' (complementary to the adaptor) was used as a common reverse primer. Relative expression was calculated by the comparative Ct method [63]. ANOVA and Tukey's HSD post-hoc test were used to analyze significant differences among three stage; $P < 0.05$ was considered significant.

Northern blot analysis of miRNA expression

Total RNA of *T. spiralis* (muscle larvae, adult worms and newborn larvae) were separated by electrophoresis on a 12.5% polyacrylamide gel under denaturing (8 M urea) conditions and transferred to Hybond-N⁺ nylon membranes (GE Healthsystems, Uppsala, Sweden). The membranes were crosslinked in a UV crosslinker and baked for 1 h at 80°C. Probes complementary to small RNA sequences were end-labeled with DIG at 5' termini (Takara, Dalian, China). Prehybridization and hybridization were both performed overnight at 53°C in Northernmax Hybridization buffer (Ambion, CA, USA). The blots were washed four times for 30 min in 2×SSC, 0.05% SDS and twice for 15 min in 0.1×SSC, 0.1% SDS at room temperature. The hybridization signal was detected using a DIG Detection Kit (Roche) following manufacturer's instructions. The oligonucleotide probes used for hybridization are as follows:

tsb-mir-100 probes: 5'ACACAA*GTTTC*GGATCT*AC*GG-GTT3'

tsb-let-7 probes: 5'AACTAT*ACA*ACCT*ACT*ACCTCA3'

tsb-novel-108 probes: 5'TCT*GT*GAATTCTT*ACA*GTG-CCAAG3'

tsb-novel-83 probes: 5'GCTAC*GATC*AA*AATT*GCTCA-A3'

(LNA (Locked nucleic acid) substitutions are indicated by a “*”).

Supporting Information

Figure S1 Work-flow of bioinformatics analysis of the small RNA sequences obtained.

(TIF)

Table S1 General information of the small RNA libraries.

(DOC)

Table S2 The common and stage-specific small RNAs in three developmental stages.

(DOC)

Table S3 Length distribution of small RNAs in *T. spiralis*.

(DOC)

Table S4 Conserved miRNAs identified in different developmental stages.

(DOC)

Table S5 Expression levels of conserved miRNAs derived from different arms.

(DOC)

Table S6 Novel miRNAs identified in different developmental stages.

(DOC)

Table S7 The expression of novel miRNAs derived from different arms.

(DOC)

Table S8 siRNAs derived from DNA transposons.

(DOC)

Table S9 siRNAs derived from NAT.

(DOC)

References

- Zarlenga DS, Rosenthal BM, La Rosa G, Pozio E, Hoberg EP (2006) Post-Miocene expansion, colonization, and host switching drove speciation among extant nematodes of the archaic genus *Trichinella*. *Proc Natl Acad Sci U S A* 103: 7354–7359.
- Dupouy-Camet J (2000) Trichinellosis: a worldwide zoonosis. *Vet Parasitol* 93: 191–200.
- Liu MY, Zhu XP, Xu KC, Lu Q, Boireau P (2001) Biological and genetic characteristics of two *Trichinella* isolates in China; comparison with European species. *Parasite* 8: S34–38.
- Wakelin D, Goyal PK (1996) *Trichinella* isolates: parasite variability and host responses. *Int J Parasitol* 26: 471–481.
- Mitreva M, Jasmer DP, Zarlenga DS, Wang Z, Abubucker S, et al. (2011) The draft genome of the parasitic nematode *Trichinella spiralis*. *Nat Genet* 43: 228–235.
- Ambros V (2004) The functions of animal microRNAs. *Nature* 431: 350–355.
- Zhang B, Pan X, Cobb GP, Anderson TA (2006) Plant microRNA: a small regulatory molecule with big impact. *Dev Biol* 289: 3–16.
- Lee HC, Li L, Gu W, Xue Z, Crosthwaite SK, et al. (2010) Diverse pathways generate microRNA-like RNAs and Dicer-independent small interfering RNAs in fungi. *Mol Cell* 38: 803–814.
- Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T (2001) Identification of novel genes coding for small expressed RNAs. *Science* 294: 853–858.
- Cullen BR (2006) Viruses and microRNAs. *Nat Genet* 38 Suppl: S25–30.
- Carrington JC, Ambros V (2003) Role of microRNAs in plant and animal development. *Science* 301: 336–338.
- Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116: 281–297.
- Zhang B, Pan X, Anderson TA (2006) Identification of 188 conserved maize microRNAs and their targets. *FEBS Lett* 580: 3753–3762.
- Carthew RW, Sontheimer EJ (2009) Origins and Mechanisms of miRNAs and siRNAs. *Cell* 136: 642–655.
- Kim VN (2006) Small RNAs just got bigger: Piwi-interacting RNAs (piRNAs) in mammalian testes. *Genes Dev* 20: 1993–1997.
- Vasudevan S, Tong Y, Steitz JA (2007) Switching from repression to activation: microRNAs can up-regulate translation. *Science* 318: 1931–1934.
- Lee RC, Feinbaum RL, Ambros V (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75: 843–854.
- Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, et al. (2000) The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* 403: 901–906.
- Okamura K, Chung WJ, Ruby JG, Guo H, Bartel DP, et al. (2008) The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs. *Nature* 453: 803–806.
- Filipowicz W, Jaskiewicz L, Kolb FA, Pillai RS (2005) Post-transcriptional gene silencing by siRNAs and miRNAs. *Curr Opin Struct Biol* 15: 331–341.
- Poy MN, Eliasson L, Krutzfeldt J, Kuwajima S, Ma X, et al. (2004) A pancreatic islet-specific microRNA regulates insulin secretion. *Nature* 432: 226–230.
- Gauthier BR, Wollheim CB (2006) MicroRNAs: ‘ribo-regulators’ of glucose homeostasis. *Nat Med* 12: 36–38.
- Xu MJ, Liu Q, Nisbet AJ, Cai XQ, Yan C, et al. (2010) Identification and characterization of microRNAs in *Clonorchis sinensis* of human health significance. *BMC Genomics* 11: 521.
- Friedlander MR, Adamidi C, Han T, Lebedeva S, Isenbarger TA, et al. (2009) High-resolution profiling and discovery of planarian small RNAs. *Proc Natl Acad Sci U S A* 106: 11546–11551.
- Poole CB, Davis PJ, Jin J, McReynolds LA (2010) Cloning and bioinformatic identification of small RNAs in the filarial nematode, *Brugia malayi*. *Mol Biochem Parasitol* 169: 87–94.
- Cucher M, Prada L, Mourglia-Ettlin G, Dematteis S, Camicia F, et al. (2011) Identification of *Echinococcus granulosus* microRNAs and their expression in different life cycle stages and parasite genotypes. *Int J Parasitol* 41: 439–448.
- Hao L, Cai P, Jiang N, Wang H, Chen Q (2010) Identification and characterization of microRNAs and endogenous siRNAs in *Schistosoma japonicum*. *BMC Genomics* 11: 55.
- Simoes MC, Lee J, Djikeng A, Cerqueira GC, Zerlotini A, et al. (2011) Identification of *Schistosoma mansoni* microRNAs. *BMC Genomics* 12: 47.
- Chen MX, Ai L, Xu MJ, Zhang RL, Chen SH, et al. (2011) *Angiostrongylus cantonensis*: Identification and characterization of microRNAs in male and female adults. *Exp Parasitol* 128: 116–120.
- Chen MX, Ai L, Xu MJ, Chen SH, Zhang YN, et al. (2011) Identification and characterization of microRNAs in *Trichinella spiralis* by comparison with *Brugia malayi* and *Caenorhabditis elegans*. *Parasitol Res*.
- MacRae IJ, Zhou K, Doudna JA (2007) Structural determinants of RNA recognition and cleavage by Dicer. *Nat Struct Mol Biol* 14: 934–940.
- Saito K, Siomi MC (2010) Small RNA-Mediated Quiescence of Transposable Elements in Animals. *Dev Cell* 19: 687–697.
- Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, et al. (2008) Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* 453: 539–543.
- Wang G, Reinke V (2008) A *C. elegans* Piwi, PRG-1, regulates 21U-RNAs during spermatogenesis. *Curr Biol* 18: 861–867.
- Das PP, Bagijn MP, Goldstein LD, Woolford JR, Lehrbach NJ, et al. (2008) Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the *Caenorhabditis elegans* germline. *Mol Cell* 31: 79–90.
- Ruby JG, Jan C, Player C, Axtell MJ, Lee W, et al. (2006) Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* 127: 1193–1207.
- Batista PJ, Ruby JG, Claycomb JM, Chiang R, Fahlgren N, et al. (2008) PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in *C. elegans*. *Mol Cell* 31: 67–78.
- Lau NC, Lim LP, Weinstein EG, Bartel DP (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 294: 858–862.
- Guo L, Lu Z (2010) The fate of miRNA* strand through evolutionary analysis: implication for degradation as merely carrier strand or potential regulatory molecule? *PLoS One* 5: e11387.
- Lim LP, Lau NC, Weinstein EG, Abdelhakim A, Yekta S, et al. (2003) The microRNAs of *Caenorhabditis elegans*. *Genes Dev* 17: 991–1008.
- Chen JF, Mandel EM, Thomson JM, Wu Q, Callis TE, et al. (2006) The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nat Genet* 38: 228–233.
- Sokol NS, Xu P, Jan YN, Ambros V (2008) *Drosophila* *let-7* microRNA is required for remodeling of the neuromusculature during metamorphosis. *Genes Dev* 22: 1591–1596.
- Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, et al. (2005) Identification of hundreds of conserved and nonconserved human microRNAs. *Nat Genet* 37: 766–770.
- Christodoulou F, Raible F, Tomer R, Simakov O, Trachana K, et al. (2010) Ancient animal microRNAs and the evolution of tissue identity. *Nature* 463: 1084–1088.
- Caygill EE, Johnston LA (2008) Temporal regulation of metamorphic processes in *Drosophila* by the *let-7* and miR-125 heterochronic microRNAs. *Curr Biol* 18: 943–950.
- Okamura K, Balla S, Martin R, Liu N, Lai EC (2008) Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in *Drosophila melanogaster*. *Nat Struct Mol Biol* 15: 998.
- Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, et al. (2008) An endogenous small interfering RNA pathway in *Drosophila*. *Nature* 453: 798–802.
- Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, et al. (2008) Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science* 320: 1077–1081.
- Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, et al. (2008) Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* 453: 534–538.
- Allen E, Xie Z, Gustafson AM, Carrington JC (2005) microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* 121: 207–221.
- Gogvadze E, Buzdin A (2009) Retroelements and their impact on genome evolution and functioning. *Cell Mol Life Sci* 66: 3727–3742.
- Berriman M, Haas BJ, LoVerde PT, Wilson RA, Dillon GP, et al. (2009) The genome of the blood fluke *Schistosoma mansoni*. *Nature* 460: 352–358.
- Biemont C, Vieira C (2006) Genetics: junk DNA as an evolutionary force. *Nature* 443: 521–524.

Acknowledgments

We very much appreciate the effort of scientists at BGI Genomics for their kind assistance with the sequencing of small RNA libraries and Dr. Haipo Sun for assistance with bioinformatics.

Author Contributions

Conceived and designed the experiments: XL ML QC. Performed the experiments: XL YS HL BT XP NH SP NJ. Analyzed the data: XL QC. Contributed reagents/materials/analysis tools: JY. Wrote the paper: XL QC.

54. Despommier DD, Campbell WC, Blair LS (1977) The in vivo and in vitro analysis of immunity to *Trichinella spiralis* in mice and rats. *Parasitology* 74: 109–119.
55. Li R, Li Y, Kristiansen K, Wang J (2008) SOAP: short oligonucleotide alignment program. *Bioinformatics* 24: 713–714.
56. Dsouza M, Larsen N, Overbeck R (1997) Searching for patterns in genomic data. *Trends Genet* 13: 497–498.
57. Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 16: 276–277.
58. Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P (1994) Fast Folding and Comparison of RNA Secondary Structures. *Monatshfte fur Chem Chem Month* 125: 167–188.
59. Jones-Rhoades MW, Bartel DP (2004) Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol Cell* 14: 787–799.
60. Matsubara H, Takeuchi T, Nishikawa E, Yanagisawa K, Hayashita Y, et al. (2007) Apoptosis induction by antisense oligonucleotides against miR-17-5p and miR-20a in lung cancers overexpressing miR-17-92. *Oncogene* 26: 6099–6105.
61. Romualdi C, Bortoluzzi S, D'Alessi F, Danieli GA (2003) IDEG6: a web tool for detection of differentially expressed genes in multiple tag sampling experiments. *Physiol Genomics* 12: 159–162.
62. Li YY, Qin L, Guo ZM, Liu L, Xu H, et al. (2006) In silico discovery of human natural antisense transcripts. *BMC Bioinformatics* 7: 18.
63. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C(T)}$ Method. *Methods* 25: 402–408.