Journal of Animal and Veterinary Advances 11 (18): 3469-3473, 2012
ISSN: 1680-5593
© Medwell Journals, 2012

Characterization of MicroRNAs in *Paragonimus westermani* by Solexa Deep Sequencing and Bioinformatics Analysis

L. Ai, M.X. Chen, S.H. Chen, Y.N. Zhang, H. Li, Y.C. Cai, Y. Lu,
L.G. Tian, X.N. Zhou and J.X. Chen
National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention,
Laboratory of Parasite and Vector Biology, Ministry of Health,
WHO Collaborating Center for Malaria, Schistosomiasis and Filariasis,
200025 Shanghai, P.R. China

Abstract: MicroRNAs (miRNAs) are small non-coding RNA which generate from large hairpin precursors and play important role on post-transcriptional regulators of target genes. Although, many individual miRNAs have recently been extensively studied, there has been very little research on miRNA transcriptomes in trematode. By the method of using high throughput Solexa sequencing technology, 11.02 million clean reads have been obtained of *Paragonimus westermani*. Among the clean reads, 1.14 million ones (10.37%) were perfectly mapped onto the *S. japonicum* genome which included 20,322 (0.69%) unique sequences. A number of 9,492,613 (86.17%) reads had no match with public databases and marked as un-annotated RNAs. Nucleotide bias analysis found that the known miRNAs showed high bias and the guanine was the dominant nucleotide, particularly at the 2nd and 23rd positions which were almost at the beginning and end of conserved miRNAs. Three novel miRNA corresponding to 18 precursors were predicted. To the knowledge, this is the first report of miRNA profiles in *P. westermani* which will contribute to better understanding of the complex biology of this zoonotic parasite. The reported data of *P. westermani* miRNAs should provide valuable references for miRNA studies of closed related zoonotic trematode.

Key words: *Paragonimus westermani*, microRNAs (miRNAs), Solexa deep sequencing, novel, China

INTRODUCTION

Paragonimus westermani which parasitize in the lung of human with various kinds of mammal animals as definitive hosts. This trematode causes paragonimiasis in many Asian countries such as Japan, Korea, China, etc. where culturally people eat raw or undercooked freshwater crab or crayfishes infected with infective metacercariae (Mukae *et al*., 2001; Liu *et al*., 2008; Kim *et al*., 2009). Currently, paragonimiasis has been recognized as a serious food-borne disease with the symptom as lung cancer or tuberculosis which always led to misdiagnosis (Lane *et al*., 2009; Sohn *et al*., 2009; Ikehara *et al*., 2010).

Although, the research of classification, pathogenicity, immunologic mechanism, etc. on *P. westermani* have been carried out for many years (Zhao *et al*., 2007; Liu *et al*., 2008; Devi *et al*., 2010), increasing knowledge of microRNAs (miRNAs) is able to reveal riboregulators throughout different animal phyla (Grimson *et al*., 2008), uncover developmental genetic switches (Flynt and Lai, 2008) and explore novel biomarkers for disease (Zhang, 2008; Mo, 2012).

The microRNAs (miRNA) have been found in many organisms such as viruses, plants and mammals (Grant-Downton *et al*., 2009; Dolken *et al*., 2010; Su *et al*., 2011) which are non-coding small RNA molecules with the length of 18-24 nucleotides and play essential roles in gene expressions of living organisms (Mead and Tu, 2008; Reddy *et al*., 2009). In the miRNA research of parasites, miRNAs have been identified in model organism Caenorhabditis elegans as well as some zoonotic parasites such as *Schistosoma japonicum*, *Clonorchis sinensis*, *Toxoplasma gondii*, *Trichinella spiralis* and *Angiostrongylus cantonensis* (Wang *et al*., 2010; Xu *et al*., 2010; Warf *et al*., 2011; Chen *et al*., 2011a, b). However, there have been no reports of miRNAs in *P. westermani*.

Therefore, the aim of the present study was to investigate the expression profile of miRNAs in adult

Corresponding Author: J.X. Chen, National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention, 200025 Shanghai, The People's Republic of China

P. westermani and to predict potential novel miRNAs by Solexa deep sequencing method and bioinformatics analysis.

MATERIALS AND METHODS

Total RNA preparation and sequencing: The adult *P. westermani* were collected from lung of the artificial infected dog in the lab and the infected metacercaria from freshwater crab in Zhejiang province which were used to feed the dog. After harvesting, the parasites were transferred immediately to sterile physiological saline (37°C) in a sterile beaker and washed 5 times to remove contaminants from the host. One adult was cut into pieces and transferred to RNase-free 1.5 mL screw-top cryotube containing RNAlater (Sigma). Then, the sample was kept at 4°C for overnight and stored at -70°C until use. The taxonomic identity of the trematode was confirmed as *P. westermani* by Polymerase Chain Reaction (PCR) amplification and subsequent sequence analysis of the mitochondrial cytochrome c oxidase subunit 1 gene (*cox1*) (Park *et al*., 2003).

Trizol (Invitrogen) was used to prepare total RNA in terms of the protocol provided by the manufacturer. RNA concentration and purity were evaluated photometrically by measuring the absorbance at 260 and 280 nm, through NanoDrop ND-1000 spectrophotometer (Nanodrop Technologies, Wilmington, DE) and Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA). After obtaining pure RNA, it stored at -70°C (Lau *et al*., 2001; Xu *et al*., 2010). About 20 μg total RNA was separated by electrophoresis on a Novex 15% urea gel (Invitrogen Co., Ltd.) with RNA in the region of 18-30 nt purified. Next, a Reverse Transcription PCR (RT-PCR) kit (Invitrogen) was used for reverse transcription of the small RNAs to form complementary DNA (cDNA). Finally, the cDNA was then sequenced with a Solexa sequencer at Huada Genomics Institute Co., Ltd. China.

High-throughput sequencing and computational analysis: Data analysis refered as earlier (Xu *et al*., 2010; Chen *et al*., 2011a, b). Firstly, adaptors, reads smaller than 18 nt and low quality sequences were trimmed from the raw data. Next, the reads were searched in GenBank Rfam Database (Version 10.1) to remove non-coding RNAs including rRNA, tRNA, snRNA, snoRNA and other ncRNAs. RepeatMasker (http://www.repeatmasker.org) was used to eliminate repetitive sequences in the clean reads. Reads of 18-30 nt were mapped onto the genome of *S. japonicum* (http://lifecenter.sgst.cn/schistosoma/cn) using Short Oligo nucleotide Analysis Package Software (SOAP, http://soap.genomics.org.cn) (Bartel, 2004; Rhoades *et al*., 2002). Further bioinformatics analysis including unique reads screening, candidate precursors, inverted repeats, family distribution of conserved miRNA and the nucleotide bias were performed as described by Xue *et al*. (2008).

RESULTS AND DISCUSSION

High throughput sequencing technologies provide more convenient assay to detect known and novel miRNAs which can also open doors to directly show differences in expression levels. Given that miRNAs play essential roles in gene expression, research of miRNAs in parasites can prob mechanisms that the parasites respond to environmental and developmental signals through miRNA-mediated gene expression (Liu *et al*., 2010). Therefore, the present study examined the expression profiles of miRNAs in *P. westermani* by Solexa deep sequencing combined with bioinformatics analysis.

There were 12.99 million reads yielded with Solexa sequencing. After filtering low quality tags including 5' and 3' adaptors and adaptor-adaptor ligation, a total of 12.87 million reads with high quality were obtained. Length distribution analysis showed that most reads distributed among 20-23 nt. The highest percentage was 16.03% of 22 nt reads followed by 15.08% of 23 nt reads (Fig. 1). After removing reads smaller than 18 nt (12.72%), a total of 11.02 million clean reads being obtained. Among the clean reads, 1.14 million ones (10.37%) were perfectly mapped onto the *S. japonicum* genome which included 20,322 (0.69%) unique sequences. However, it was found that reads matched to the reference genome of *S. japonicum* was very low. However, it can reach as high as 70.5% or higher in some other species (Xue *et al*., 2008). The most possible reason for the phenomenon might be that the reference genome was *S. japonicum*

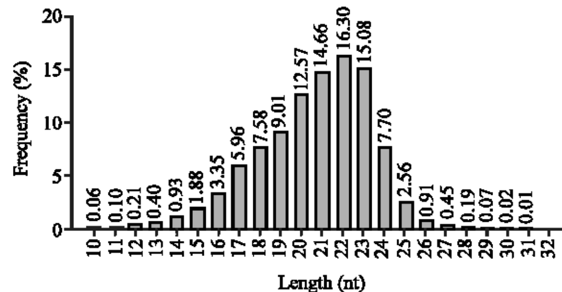

Fig. 1: Length distribution analysis of small RNAs of *Paragonimus westermani*. Analysis of the 12.87 million reads with high qualities after filtering low quality tags, 5' and 3' adaptor and contamination formed by adaptor-adaptor ligation

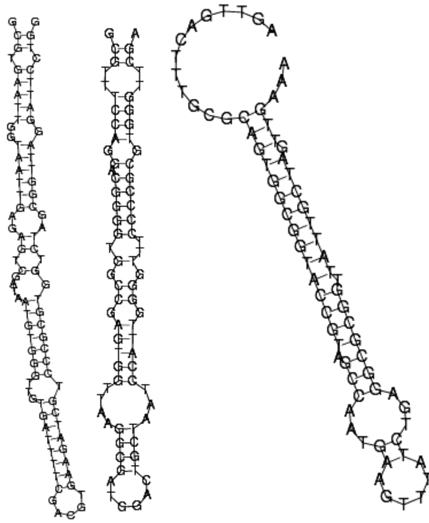

Fig. 2: Predicted structures of the novel miRNA precursors of *Paragonimus westermani* evaluated by RNA-Fold Software

Table 1: Detailed classification of reads of *Paragonimus westermani*

	P. westermani	
Classification	Total sRNAs (%)	Unique sRNAs (%)
Total	11015886 (100.00)	2949024 (100.00)
miRNA	568896 (5.16)	219 (0.01)
rRNA	761689 (6.91)	36143 (1.23)
snRNA	1532 (0.01)	690 (0.02)
snoRNA	2170 (0.02)	314 (0.01)
tRNA	165276 (1.50)	7368 (0.25)
Unann	9492613 (86.17)	2902039 (98.41)

genome rather than *P. westermani* because the *P. westermani* genome is not available currently. For the same reason, there were only three novel miRNAs predicted in adult *P. westermani* (Fig. 2). When the genome of *P. westermani* or other closely related trematodes become available, more miRNAs will be found in *P. westermani.*

Among the clean reads, 930,667 (8.44%) were found as rRNA, tRNA, snRNA and snoRNA when searched against Rfam Database (Table 1). Two types of repeats of LINE/RTE:0 and LINE/RTE:1 with 1668 reads (474 unique ones). The percentage of known miRNA was 568,898 total reads corresponding to 221 unique ones focusing on 18 conserved miRNAs. In addition to the known miRNAs, rRNA and repeats mentioned above, a number of 9,492,613 (86.17%) reads had no match with public databases and marked as un-annotated RNAs. Among them, 2,902,039 were unique reads (Fig. 3).

Some kinds of miRNAs were expressed with high predominance. The miR-71 had the most abundant reads with a percentage of 68.80% (391,382 reads). The miR-71 occupied most portions of the total reads which indicated

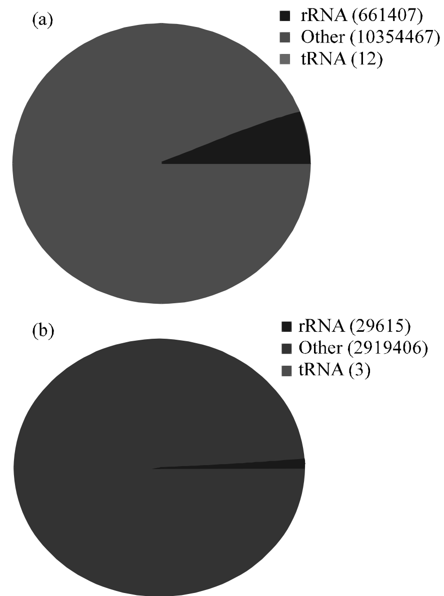

Fig. 3: Coverage of small RNAs in *Paragonimus westermani* by Solexa deep sequencing. a) Total reads; b) Unique reads

Table 2: Expressed miRNAs in *Paragonimus westermani*

miR name	Express
sja-miR-71	391382
sja-bantam	67557
sja-miR-71b-5p	38382
sja-miR-2a-3p	24809
sja-miR-7-5p	18178
sja-let-7	12321
sja-miR-277	7284
sja-miR-61	3667
sja-miR-10-5p	2719
sja-miR-219-5p	1325
sja-miR-2b-3p	625
sja-miR-124-3p	261
sja-miR-2e-3p	242
sja-miR-2c-3p	83
sja-miR-2d-3p	50
sja-miR-36-3p	6
sja-miR-307	2
sja-miR-3488	2

that the miR-71 was essential for the survival of the trematode. This phenomenon has also been shown in *S. japonicum*, *Fasciola* sp. as well as Taenia cestode (Xue *et al*., 2008; Ai *et al*., 2012; Xu *et al*., 2012). In addition to the miR-71, some other miRNAs had particularly high copies in *P. westermani* such as bantam, miR-2a-3p, miR-7-5p. bantam with a percentage of 11.88% with 67,557 reads which was also with high copy in *S. japonicum* (Xue *et al*., 2008). However, let-7 family was the biggest miRNA family among the conserved miRNAs in *C. elegans* and *C. sinensis* (Reinhart *et al*., 2000; Xu *et al*., 2010). The let-7 or its family members were only 2.17% with 12,321 reads identified in *P. westermani* in the present study (Table 2).

Table 3: Positions and rate of sRNA in *Paragonimus westermani*

Position on sRNA	A (%)	U (%)	C (%)	G (%)	A + U (%)	C + G (%)
1	1.190	95.750	0.002	3.047	96.950	0.000
2	2.170	0.000	4.320	95.520	2.170	3.050
3	94.460	0.015	1.200	4.320	97.480	5.520
4	89.880	0.350	6.670	3.090	90.230	9.770
5	93.310	2.390	1.200	3.100	95.700	4.300
6	1.140	4.250	0.070	94.540	5.390	94.610
7	92.100	0.350	5.220	2.340	92.450	7.550
8	2.040	1.200	93.700	3.060	3.240	96.760
9	6.660	20.560	0.350	72.430	27.220	72.780
10	73.870	16.220	0.066	9.840	90.090	9.906
11	1.200	77.940	3.060	17.800	79.140	20.860
12	16.590	6.720	1.200	75.490	23.310	76.690
13	0.002	7.420	2.440	90.140	7.422	92.580
14	4.320	92.930	2.350	0.400	97.250	2.750
15	85.850	4.340	0.350	9.456	90.190	9.810
16	10.240	3.060	1.140	85.560	13.300	86.700
17	0.017	94.830	0.900	4.253	94.850	5.150
18	0.000	4.590	2.040	93.370	4.590	95.410
19	90.120	5.920	0.900	3.060	96.040	3.960
20	4.320	5.560	1.540	88.580	9.880	90.120
21	88.530	5.680	0.000	5.740	94.260	5.740
22	0.990	84.460	13.260	1.290	85.450	14.550
23	1.050	0.000	0.000	98.940	1.050	98.940
Average	33.060	23.240	6.170	37.620	56.420	39.460

Analysis of the 221 unique reads of known miRNAs showed that G was the dominant nucleotide in mature miRNAs with the average percentage of 37.62% and showed high percentage at the 23rd nucleotide of 98.94%. The highest percentage of (G+C) was found at the 23rd positions with a percentage of 98.94%. The U was most often used as the first nucleotide in miRNAs of *P. westermani* with an average percentage of 95.75%. (A+U) was found in most reads with a percentage of 56.42% in average (Table 3).

CONCLUSION

The present study represents the miRNAs characterization of *P. westermani* which will be useful to help understanding the complex biology of this trematode and to explore effective control methods for paragonimosis.

ACKNOWLEDGEMENTS

This research is supported in part by the Shanghai Postdoctoral Sustentation Fund (Grant No. 12R21416500), China Postdoctoral Science Foundation (Grant No. 2012M520353), National S&T Major Program (Grant No. 2008ZX10004-011,2009ZX10004-302,2009ZX10004-201), and National Key Technology R and D Program (Grant No. 2008BAI56B03, the Open Funds of the State Key Laboratory of Veterinary Etiological Biology, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences (Grant Nos. SKLVEB2009KFKT008, SKLVEB2009KFKT014,SKLVEB2011KFKT004,SKLVEB-2010KFKT009, SKLVEB2010KFKT010 and SKLVEB2011-KFKT010).

REFERENCES

Ai, L., M.J. Xu, M.X. Chen, Y.N. Zhang and S.H. Chen *et al*., 2012. Characterization of microRNAs in *Taenia saginata* of zoonotic significance by Solexa deep sequencing and bioinformatics analysis. Parasitol. Res., 110: 2373-2378.

Bartel, D.P., 2004. MicroRNAs: Genomics, biogenesis, mechanism and function. Cell, 116: 281-297.

Chen, M.X., L. Ai, M.J. Xu, R.L. Zhang and S.H. Chen *et al*., 2011a. *Angiostrongylus cantonensis*: Identification and characterization of microRNAs in male and female adults. Exp. Parasitol., 128: 116-120.

Chen, M.X., L. Ai, M.J. Xu, S.H. Chen and Y.N. Zhang *et al*., 2011b. Identification and characterization of microRNAs in *Trichinella spiralis* by comparison with *Brugia malayi* and *Caenorhabditis elegans*. Parasitol. Res., 109: 553-558.

Devi, K.R., K. Narain, T. Agatsuma, D. Blair and M. Nagataki *et al*., 2010. Morphological and molecular characterization of *Paragonimus westermani* in northeastern India. Acta Tropica, 116: 31-38.

Dolken, L., A. Krmpotic, S. Kothe, L. Tuddenham and M. Tanguy *et al*., 2010. Cytomegalovirus microRNAs facilitate persistent virus infection in salivary glands. PLoS Pathog, 10.1371/journal.ppat.1001150.

Flynt, A.S. and E.C. Lai, 2008. Biological principles of microRNA-mediated regulation: Shared themes amid diversity. Nat. Rev. Genet., 9: 831-842.

Grant-Downton, R., G. Le Trionnaire, R. Schmid, J. Rodriguez-Enriquez and S. Hafidh *et al*., 2009. MicroRNA and tasiRNA diversity in mature pollen of *Arabidopsis thaliana*. BMC Genomics, 10.1186/1471-2164-10-643.

Grimson, A., M. Srivastava, B. Fahey, B.J. Woodcroft and H.R. Chiang *et al*., 2008. Origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature, 455: 1193-1197.

Ikehara, M., Y. Komase, A. Morita, H. Yamaguchi and T. Yamamoto, 2010. Paragonimiasis in a person whose symptoms were shown 22 years after emigrating to Japan from Laos. J. Infect. Chemother., 16: 49-52.

Kim, E.M., J.L. Kim, S.I. Choi, S.H. Lee and S.T. Hong, 2009. Infection status of freshwater crabs and crayfish with metacercariae of *Paragonimus westermani* in Korea. Korean J. Parasitol., 47: 425-426.

Lane, M.A., M.C. Barsanti, C.A. Santos, M. Yeung, S.J. Lubner and G.J. Weil, 2009. Human paragonimiasis in North America following ingestion of raw crayfish. Clin. Infect. Dis., 49: e55-e61.

Lau, N.C., L.P. Lim, E.G. Weinstein and D.P. Bartel, 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. Science, 294: 858-862.

Liu, Q., F. Wei, W. Liu, S. Yang and X. Zhang, 2008. Paragonimiasis: An important food-borne zoonosis in China. Trends Parasitol., 24: 318-323.

Liu, Q., W. Tuo, H. Gao and X.Q. Zhu, 2010. MicroRNAs of parasites: Current status and future perspectives. Parasitol. Res., 107: 501-507.

Mead, E.A. and Z. Tu, 2008. Cloning, characterization and expression of microRNAs from the Asian malaria mosquito, *Anopheles stephensi*. BMC Genomics, 10.1186/1471-2164-9-244.

Mo, Y.Y., 2012. MicroRNA regulatory networks and human disease. Cell. Mol. Life Sci., 69: 3529-3531.

Park, G.M., K.I. Im and T.S. Yong, 2003. Phylogenetic relationship of ribosomal ITS2 and mitochondrial COI among diploid and triploid *Paragonimus westermani* isolates. Korean J. Parasitol., 41: 47-55.

Reddy, A.M., Y. Zheng, G. Jagadeeswaran, S.L. Macmil and W.B. Graham *et al*., 2009. Cloning, characterization and expression analysis of porcine microRNAs. BMC Genomics, 10.1186/1471-2164-10-65.

Reinhart, B.J., F.J. Slack, M. Basson, A.E. Pasquinelli and G. Ruvkun *et al*., 2000. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature, 403: 901-916.

Rhoades, M.W., B.J. Reinhart, C.B. Burge, B. Bartel and D.P. Bartel, 2002. Prediction of plant microRNA targets. Cell., 110: 513-520.

Sohn, B.S., Y.J. Bae, Y.S. Cho, H.B. Moon and T.B. Kim, 2009. Three cases of paragonimiasis in a family. Korean J. Parasitol., 47: 281-285.

Su, Z., J. Xia and Z. Zhao, 2011. Functional complementation between transcriptional methylation regulation and post-transcriptional microRNA regulation in the human genome. BMC Genomics, 10.1186/1471-2164-12-S5-S15.

Wang, Z., X. Xue, J. Sun, R. Luo and X. Xu *et al*., 2010. An in-depth description of the small non-coding RNA population of *Schistosoma japonicum* schistosomulum. PLoS Negl. Trop. Dis.

Warf, M.B., W.E. Johnson and B.L. Bass, 2011. Improved annotation of C. elegans microRNAs by deep sequencing reveals structures associated with processing by Drosha and Dicer. RNA, 17: 563-577.

Xu, M.J., L. Ai, J.H. Fu, J.N. Alasdair and Q.Y. Liu *et al*., 2012. Comparative characterization of MicroRNAs from the liver flukes *Fasciola gigantica* and *F. hepatica*. PloS.

Xu, M.J., Q. Liu, A.J. Nisbet, X.Q. Cai and C. Yan *et al*., 2010. Identification and characterization of microRNAs in Clonorchis sinensis of human health significance. BMC Genomics, 10.1186/1471-2164-11-521.

Xue, X., J. Sun, Q. Zhang, Z. Wang, Y. Huang and W. Pan, 2008. Identification and characterization of novel microRNAs from *Schistosoma japonicum*. PloS, 10.1371/journal.pone.0004034.

Zhang, C., 2008. MicroRNomics: A newly emerging approach for disease biology. Physiol. Genomics, 33: 139-147.

Zhao, Q.P., S.U. Moon, B.K. Na, S.H. Kim and S.H. Cho *et al*., 2007. *Paragonimus westermani*: Biochemical and immunological characterizations of paramyosin. Exp. Parasitol., 115: 9-18.