International Journal for Parasitology 45 (2015) 253–267

Contents lists available at ScienceDirect

International Journal for Parasitology

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

High-throughput characterization of *Echinococcus* spp. metacestode miRNomes

Marcela Cucher [a,*,1], Natalia Macchiaroli [a,1], Laura Kamenetzky [a], Lucas Maldonado [a], Klaus Brehm [b], Mara Cecilia Rosenzvit [a,*]

[a] *Instituto de Investigaciones en Microbiología y Parasitología Médica (IMPaM), Facultad de Medicina, Universidad de Buenos Aires (UBA)-Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET), Buenos Aires, Argentina*
[b] *Institute of Hygiene and Microbiology, University of Würzburg, Würzburg, Germany*

ARTICLE INFO

Article history:
Received 28 August 2014
Received in revised form 4 December 2014
Accepted 8 December 2014
Available online 4 February 2015

Keywords:
sRNA-seq
Echinococcus
microRNA
Small RNA
Parasite
Platyhelminth
Echinococcosis

ABSTRACT

Echinococcosis is a worldwide zoonosis of great public health concern, considered a neglected disease by the World Health Organisation. The cestode parasites *Echinococcus granulosus* sensu lato (s. l.) and *Echinococcus multilocularis* are the main aetiological agents. In the intermediate host, these parasites display particular developmental traits that lead to different patterns of disease progression. In an attempt to understand the causes of these differences, we focused on the analysis of microRNAs (miRNAs), small non-coding regulatory RNAs with major roles in development of animals and plants. In this work, we analysed the small RNA expression pattern of the metacestode, the stage of sanitary relevance, and provide a detailed description of *Echinococcus* miRNAs. Using high-throughput small RNA sequencing, we believe that we have carried out the first experimental identification of miRNAs in *E. multilocularis* and have expanded the *Echinococcus* miRNA catalogue to 38 miRNA genes, including one miRNA only present in *E. granulosus* s. l. Our findings show that although both species share the top five highest expressed miRNAs, 13 are differentially expressed, which could be related to developmental differences. We also provide evidence that uridylation is the main miRNA processing mechanism in *Echinococcus* spp. These results provide detailed information on *Echinococcus* miRNAs, which is the first step in understanding their role in parasite biology and disease establishment and/or progression, and their future potential use as drug or diagnostic targets.

1. Introduction

Echinococcus spp. cestode parasites are the causative agents of echinococcosis, the most relevant species being *Echinococcus granulosus* sensu lato (s. l.) and *Echinococcus multilocularis*. Echinococcosis is a worldwide zoonosis of great public health concern and is considered a neglected disease by the World Health Organisation (WHO). *Echinococcus* spp. require two mammalian hosts to complete their life cycles: a definitive host (carnivores) and an intermediate host (mostly ungulates in the case of *E. granulosus* s. l., and wild rodents in the case of *E. multilocularis*). Humans act as accidental intermediate hosts. Intermediate hosts contract the disease by ingesting infective eggs with oncospheres which migrate to the target organ, generally the liver, and develop to the next larval stage, the metacestode. The metacestode is lined with an inner germinal layer that produces immature worms (protoscoleces) which, after ingestion by the definitive host, develop into adult worms in the gut and reproduce sexually, releasing infective eggs with the host faeces into the environment.

Echinococcus granulosus s. l. and *E. multilocularis* metacestodes, the stage of sanitary relevance, show remarkable developmental differences. While the *E. granulosus* s. l. metacestode is unilocular and can only bud endogenously, *E. multilocularis* grows in a more aggressive manner as it can also bud exogenously, infiltrating and colonising surrounding tissues. *Echinococcus multilocularis* can also colonise distant foci due to the metastatic nature of its germinative cells which spread via lymph nodes and blood vessels (Mehlhorn et al., 1983), and which were recently shown to be the only proliferating cells, i.e. those can serve exclusively as a source for metastases (Koziol et al., 2014).

* Corresponding authors at: IMPaM, Paraguay 2155, Piso 13, CP 1121 Buenos Aires, Argentina. Tel.: +54 11 5950 9500x2192 (M. Cucher), +54 11 5950 9500x2194 (M.C. Rosenzvit).
E-mail addresses: marcecucher@gmail.com (M. Cucher), mrosenzvit@fmed.uba.ar (M.C. Rosenzvit).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.ijpara.2014.12.003
0020-7519/

The genomes of *E. multilocularis* and *E. granulosus* sensu stricto (s. s.) (G1 genotype) were recently sequenced and assembled (Tsai et al., 2013; Zheng et al., 2013), and a first, preliminary, analysis suggests that approximately 10–14% of *Echinococcus* spp. genomes are protein encoding regions (Tsai et al., 2013; Zheng et al., 2013). This highlights the fact that there is still a vast proportion of the genome that needs to be explored including non-coding RNAs such as small RNAs (sRNAs). Within this class of small regulatory RNAs, microRNAs (miRNAs) have been identified in many different organisms ranging from viruses to higher eukaryotes and their relevance as master regulators of gene expression is now broadly accepted. miRNAs are ~22 nucleotides (nt) long RNAs involved in the control of nearly all cellular pathways from development to oncogenesis in animals and plants (Ameres and Zamore, 2013). They exert their role by negatively regulating their target genes by mRNA cleavage or, more commonly in metazoans, by translational repression, mRNA destabilisation or a combination of both (Bartel, 2009). Animal miRNAs are processed in the nucleus from long primary RNA transcripts (pri-miRNAs) into ~70 nt long stem loop intermediates, known as miRNA precursors (pre-miRNAs), from which mature miRNAs are processed by a Dicer enzyme in the cytoplasm (Bartel, 2004). The mature miRNA is assembled into an effector miRNA Induced Silencing Complex (miRISC) which then associates with partially complementary sequences commonly located in the 3′ untranslated region (UTR) of target mRNAs. Generally, the fate of the complementary strand of the mature miRNA, also known as star strand or miRNA*, is degradation (Hutvágner et al., 2001; Lau et al., 2001). However, reports on the detection of miRNAs* were first available for *Drosophila melanogaster* (Aravin et al., 2003; Schwarz et al., 2003). These results were confirmed and the ratio miRNA:miRNA* could be more robustly quantified when the sequencing scale dramatically increased by using high-throughput sRNA sequencing (sRNA-seq) and abundant miRNAs* sequences were detected (Ruby et al., 2007; Stark et al., 2007). Then, the regulatory activity of some miRNAs* was experimentally validated (Okamura et al., 2008). Usually, miRNAs exist as an heterogeneous group of multiple sequence isoforms, also known as isomiRs, that may differ in their length or base composition due to different mechanisms such as the addition of non-templated nucleotides (tailing) (Ameres and Zamore, 2013). Two main characteristics of miRNAs set them apart from other classes of sRNAs: the fact that many of them are phylogenetically conserved, especially in the seed region (nts 1–7 or 2–8) of the mature sequence and the short hairpin structure (pre-miRNA) generated during their biogenesis (Bartel, 2004). These two features allowed the development of computational algorithms, which aid in the identification of miRNAs from high-throughput sequencing data, such as miRDeep (Friedländer et al., 2008).

In a previous report, we experimentally confirmed that miRNAs are present in *E. granulosus* s. l. using a conventional cloning methodology and Northern blot (Cucher et al., 2011). Furthermore, we determined by in silico homology comparisons that this class of sRNAs is also present in *E. multilocularis*. Here we further performed the characterization of the miRNomes of both species' metacestodes using a high-throughput sequencing approach.

2. Materials and methods

2.1. Parasite material

Two fertile hydatid cysts were obtained from naturally infected swine livers provided by abattoirs from Buenos Aires province, Argentina. After aseptic aspiration of hydatid fluid with a syringe, cyst walls were carefully recovered with forceps and extensively washed in PBS to remove host cells and protoscoleces. One fraction of protoscoleces from each cyst was used to determine viability by an eosin exclusion test. Samples showing more than 90% viability were frozen in liquid nitrogen and stored at −80 °C until RNA extraction. *Echinococcus granulosus* s. l. species/genotype was determined by sequencing a fragment of the mitochondrial cytochrome c oxidase subunit 1 (Cox1), as described in Cucher et al. (2011). The resulting species/genotype of both cysts was *Echinococcus canadensis* (G7 genotype).

Echinococcus multilocularis (isolate G8065) metacestodes were obtained from 10 experimentally infected female CF1 mice (6–8 weeks old) as described in Spiliotis and Brehm, 2009. Mice were housed in the animal facilities of Instituto de Investigaciones en Microbiología y Parasitología Médica (IMPaM), Facultad de Medicina, Universidad de Buenos Aires (UBA)-Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET), Buenos Aires, Argentina in a temperature-controlled light cycle room with food and water ad libitum according to respective Institutional Guidelines. After 4 months of infection, mice were euthanised by cervical dislocation, parasite material was recovered (Spiliotis and Brehm, 2009), immediately homogenised in TriPure (Roche, USA) and stored at −80 °C until use. Animal experiments were approved by the Institutional Animal Care and Use Committee of School of Medicine, Universidad de Buenos Aires and carried out according to the National Law of Animal Protection N° 14346, government regulations of the National Administration of Drugs, Food and Medical Technology (Regulation N° 6344/96), the National Food Safety and Quality Service (SENASA, resolution No.RS617/2002, Argentina) and the National Institutes of Health (NIH), USA, Guide for the Care and Use of Laboratory Animals.

2.2. RNA isolation

For *E. canadensis* (G7) sRNA library construction, an RNA preparation enriched for RNA of <200 nt, was obtained from cyst walls using the mirVana miRNA Isolation Kit (Ambion, USA). For *E. multilocularis*, RNA was purified with TriPure (Roche) following the manufacturer's instructions. The aqueous phase obtained after the organic phase separation step was enriched in <200 nt RNAs with a mirVana miRNA Isolation Kit. An additional centrifugation step at 12,000*g* for 10 min at 4 °C was performed after homogenisation with either TriPure or miRVana Lysis Buffer to remove the insoluble material of the laminated layer. RNA was then precipitated overnight at −20 °C with 0.1 vol. of 3 M sodium acetate (pH 5.2), 2.5 vol. of 100% ethanol and glycogen. RNA was centrifuged at 14,000*g* for 30 min at 4 °C, air dried at room temperature and resuspended in nuclease-free water. RNA concentration and integrity were determined using a Qubit Fluorometer (Invitrogen, USA) and an *Agilent* 2100 *Bioanalyzer*, respectively.

2.3. Library construction and sRNA-seq

For each sRNA library construction, 100–200 ng of RNA enriched in RNAs <200 nt were used as starting material. For each sample type, *E. multilocularis* and *E. canadensis* (G7), two libraries were constructed from independent samples (biological replicates). For *E. multilocularis* libraries, RNA isolated from metacestodes recovered from five different mice were used for each library. For *E. canadensis* (G7), RNA isolated from individual cyst walls was used for each library. Libraries were constructed with the TruSeq Small RNA Sample Preparation Kit (Illumina, USA) according to the manufacturer's instructions. Library size selection was performed in order to recover sRNAs of ~22 nt long. sRNA libraries and sequencing experiments were performed at Macrogen, Korea.

2.4. Source of genome assemblies and annotations

The high quality *E. multilocularis* genome assembly version 4 (Tsai et al., 2013) was obtained from the Wellcome Trust Sanger Institute, UK (ftp://ftp.sanger.ac.uk/pub/pathogens/Echinococcus). The *E. granulosus* s. s. (G1) complete genome assembly was obtained from FTP sites of the Chinese National Human Genome Center at Shanghai (http://chgc.sh.cn/Eg) (Zheng et al., 2013). *Echinococcus* spp. genome annotation (coding sequence (CDS), tRNA, rRNA) was obtained from the GeneDB website (http://www.genedb.org). Additional rRNA sequences from flatworms (Mallatt et al., 2012) and flatworm repetitive DNA elements were downloaded from the NCBI website (http://www.ncbi.nlm.nih.gov). *Echinococcus granulosus* s. s. (G1) long non-coding RNAs (lncRNAs) (Parkinson et al., 2012) were retrieved from the PartiGeneDB website http://www.compsysbio.org/partigene/. *Echinococcus* spp. mature and hairpin (pre-miRNAs) sequences, as well as metazoan mature miRNAs, were obtained from miRBase 20 (http://www.mirbase.org/). All annotated sequences, together with novel miRNA precursor sequences identified in this study, were used to construct an in-house database for sRNA library data classification.

2.5. sRNA library data pre-processing

sRNA library data were pre-processed using FASTX-Toolkit http://hannonlab.cshl.edu/fastx_toolkit/ before mapping to the corresponding reference genome. After adapter trimming, low quality reads and reads <18 nt were removed to obtain clean reads. Identical clean reads were then collapsed into unique sequences with associated read counts.

2.6. Classification of sRNA library data

To classify all sRNA library sequences as miRNAs, rRNA, tRNA, CDS/sense CDS/antisense, lncRNAs and repeats, the processed reads were first mapped to the corresponding reference genome with Bowtie (version 0.12.7) (Langmead et al., 2009) with the option-v 2 that reports read mappings with up to two mismatches. All mapped reads were then analysed by BLASTN (*e*-value 0.01) against our in-house database (see Section 2.5). Reads with no matches were grouped as "Unknown".

2.7. miRNA identification

The miRDeep2 software package (Friedländer et al., 2012) was used for miRNA prediction. Unique sequences were mapped to the *E. multilocularis* reference genome with the read aligner Bowtie (mapper module) using default parameters (alignments with 0 mismatches in the first 18 nt of a read sequence, $\leqslant$2 mismatches after nt 18 and reads with $\leqslant$5 loci in the reference genome). For miRNA prediction with the core algorithm of miRDeep2, all metazoan mature miRNAs and hairpins including previously reported *Echinococcus* spp. sequences (retrieved form miRBase release 20) were used. The initial miRDeep2 output list of candidate miRNAs of each library was manually curated to generate a final high confidence set of miRNAs retaining only those with (i) miRDeep2 score $\geqslant$4, (ii) mature reads in both libraries per species, (iii) presence of star strand and (iv) read counts $\geqslant$100 (for non-conserved miRNAs). For further curation, the candidate novel precursor sequences were then analysed using BLASTN (*e*-value 0.01) against sets of rRNAs, tRNAs, CDS, lncRNAs and repeats. Predictions that matched with these categories were removed. Since the *E. canadensis* (G7) genome is not yet available, the precursor sequences for this parasite were predicted using the *E. granulosus* s. s. (G1) complete genome assembly (Zheng et al., 2013). The reported mature sequence for each miRNA corresponds to the sequence with the higher number of read counts which may not be the one reported by the software.

2.8. Annotation of miRNAs

To identify conserved miRNAs in *Echinococcus* spp., mature miRNA sequences were compared with previously reported miRNAs (miRBase 20) applying a 70% nt identity cut-off and a seed match criterion: identical nts at positions 1–7 or 2–8 (Winter et al., 2012). Those miRNAs that did not meet these requirements were considered novel candidate miRNAs.

To identify miRNA families within *Echinococcus* spp., all-against-all pairwise sequence alignments were computed using BLAST and all sequences sharing the seed region (nt 1–7 or nt 2–8) were considered to belong to the same family.

2.9. Expression analysis

For analysis of miRNA abundance, read counts of each individual miRNA in a sample were normalised to the total number of mature miRNA read counts in that sample according to Friedländer et al. (2009). Differential expression analysis of miRNAs between *E. multilocularis* and *E. canadensis* (G7) was performed with DESeq software (Anders and Huber, 2010). miRNAs expressed in both species that showed $-1 \geqslant \log_2$ fold change $\geqslant 1$ and *P* value adjusted <0.05 were considered differentially expressed.

Arm usage was determined by analysing product ratios of the 5′ (5p) and 3′ (3p) arms. When the minor product of a pre-miRNA showed $\geqslant$30% of reads with respect to the major product, miRNAs produced from both arms were considered mature miRNAs (Okamura et al., 2008). For the arm usage conservation analysis, data available in miRBase (release 20) was used for comparative purposes.

2.10. Editing and post-transcriptional modifications analysis

miRDeep2 outputs were manually inspected to determine the presence of editing and/or post-transcriptional modifications in mature miRNA sequences. For this analysis, only those sequences with $\geqslant$150 read counts (Fernandez-Valverde et al., 2010) and representing $\geqslant$2.5% of the total reads for the corresponding mature miRNA in each library were considered. When more than one sequence with the same type of modification for a given miRNA reached these criteria, the read count numbers of the modified sequences were added. A miRNA was considered to be under the effect of editing or post-transcriptional modifications when the modified sequence/s read count number was $\geqslant$10% of the total read counts of the corresponding miRNA (Farazi et al., 2012).

2.11. Genomic context and chromosomal location

The GeneDB *E. multilocularis* gene database was used to determine the upstream and downstream genomic context of *E. multilocularis* miRNAs. For chromosomal location analysis, *E. multilocularis* chromosomes were downloaded from ftp://ngs.sanger.ac.uk/scratch/project/pathogens/Echinococcus/multilocularis_Genome_v4/, precursor miRNAs were mapped and their positions (including chromosome and strand) were registered. When the nearest gene was annotated under the terms "hypothetical protein", "expressed protein" or "expressed conserved protein", the following up- or downstream gene was registered. Synteny analysis was plotted in accordance with Krzywinski et al. (2009).

3. Results

3.1. General sRNA sequencing results

In order to characterise the miRNomes of *Echinococcus* spp. metacestodes, we sequenced samples from the two most relevant species of the genus, *E. multilocularis* and *E. granulosus* s. l. In the case of the *E. granulosus* s. l. complex, the samples belonged to the species *E. canadensis*, G7 genotype.

An initial overview of the sequencing results showed a good performance of the sRNA-seq experiment according to the percentage of mapped reads to both *Echinococcus* genomes and the high reproducibility between biological replicates, taking into account the fraction corresponding to miRNAs (Pearson's correlation coefficient ~1 for both species) (Table 1). With respect to library depth, similar numbers of mapped reads were obtained for both species datasets (~13,000,000 and ~12,000,000 read counts for *E. multilocularis* and *E. canadensis*, respectively).

Interestingly, a remarkably higher percentage of mapped reads corresponding to miRNAs was observed in *E. canadensis* (G7), with mean values of 69.0% ± 4.2 and 13.9% ± 4.0 for *E. canadensis* and *E. multilocularis*, respectively (Fig. 1). This may be a consequence of the higher percentage of mapping reads sorted as "rRNA" and "Others" in *E. multilocularis*. A different proportion of rRNA in different libraries has been reported for irradiated planarians compared with wild type (Friedländer et al., 2009), as well as in *Caenorhabditis elegans* embryo preparations compared with larva and adult preparations (Kato et al., 2009). In both reports, the authors speculate that this could be due to rRNA degradation. Another possibility is that the fraction of miRNAs is indeed larger in *E. canadensis* (G7) and displaces the fraction of rRNA fragments to be sequenced. It is interesting to note that different contents of reads corresponding to rRNA, similar to those reported here, were also found for two species of parasitic nematodes, *Brugia pahangi* and *Haemonchus contortus* (Winter et al., 2012).

The composition of the category "Others" is depicted in Supplementary Table S1. As can be observed, most sequences could not be annotated and were therefore considered as "Unknown".

Finally, the coverage of the different datasets was analysed according to Friedländer et al. (2009). The coverage was calculated as the overlap of mapped reads with known *Echinococcus* miRNAs. All of the previously reported miRNAs for *Echinococcus* spp. were detected, being the median miRNA count value by conventional cloning and sequencing equal to two (1–36) (Cucher et al., 2011). Taking into account the same miRNAs, the median miRNA count by sRNA-seq for *E. multilocularis* was 2,731 (38-642,066) and for *E. canadensis* (G7) was 8,458 (21-4,539,475). Finally, the 5p and 3p product sequences of all *Echinococcus* pre-miRNAs were determined (Table 2, Supplementary Table S2).

3.2. Echinococcus spp. miRNA catalogue

Here we provided experimental evidence of *E. multilocularis* miRNA expression and expanded the *Echinococcus* spp. miRNA repertoire, with the addition of 12 conserved and three candidate miRNA genes with no orthologues in other organisms according to miRBase version 20.0 (Table 2, Supplementary Tables S2 and S3).

We considered that both arms of the same hairpin produced two mature miRNAs when the number of read counts of the minor product represented ⩾30% of the read counts from the major product originating from the opposite arm. By doing this, we observed that six pre-miRNAs showed expression from both arms, one in *E. multilocularis* and five in *E. canadensis* (G7) (Supplementary Tables S4 and S5).

Hence, the set of mature miRNAs expressed by *E. multilocularis* in the intermediate host is composed of 33 conserved mature sequences (grouped in 28 miRNA families), three genus-specific sequences, two of which are new miRNAs and one miRNA* sequence regarded as a mature miRNA. The *E. canadensis* (G7) repertoire is composed of 32 conserved mature sequences (grouped in 27 miRNA families), as well as four genus-specific miRNAs, three of which are new candidate miRNAs, a species-specific miRNA and five abundant miRNA* sequences annotated as mature miRNAs according to the criterion described above. Therefore, the total number of mature miRNAs is 37 for *E. multilocularis* and 42 for *E. canadensis* (G7), while the number of pre-miRNAs is 37 and 38 for *E. multilocularis* and *E. canadensis* (G7), respectively (Supplementary Tables S2 and S3). For *E. canadensis* (G7), the repertoire reported here represents the miRNAs expressed in a natural infection.

Regarding the features of *Echinococcus* miRNAs, the average length is 22 nt and the first 5p nucleotide is mainly U, in 76% of the cases in *E. multilocularis* and 68% in *E. canadensis* (G7) (Table 2), as already described for this class of sRNAs in flies and worms (Lau et al., 2001; Stark et al., 2007).

In a previous report we could not identify a valid hairpin sequence for miR-96 in *E. granulosus* s. s. (G1) (Cucher et al., 2011), probably due to the fact that only a draft version of the genome was available. Using the recently published high quality *E. granulosus* s. s. (G1) genome (Zheng et al., 2013), pre-mir-96 could be determined (Supplementary Fig. S1).

With respect to miR-31-5p and miR-31-3p, which were only detected in *E. multilocularis* datasets, and miR-new-3-3p and miR-4990-5p, which were detected only in *E. canadensis* (G7) datasets, it was possible to determine that the *E. granulosus* s. s. (G1) genome codes for the precursor sequence of miR-31 with 100% identity to the precursor sequence encoded by *E. multilocularis* (Supplementary Fig. S1). Also, there is a region in the *E. multilocularis* genome that codes for a hairpin sequence 98% identical to mir-new-3 (Table 2). The polymorphisms are located on the mature miRNA sequence downstream of the seed region (Supplementary Table S2 and Fig. S1). Finally, with the current high quality version of the *E. multilocularis* genome (Tsai et al., 2013), it was possible to assert that there is no region that upon transcription can fold into a valid hairpin containing miR-4990 on one of the arms of the stem, which contrasts to *E. granulosus* s. s. Interestingly, the most similar region in the *E. multilocularis* genome shows a 71% identity to pre-mir-4990 and the sequence corresponding to the mature miRNA has only one polymorphism in the seed region.

Mature miRNA sequences between *E. multilocularis* and *E. canadensis* (G7) share on average 99.1% identity, being completely identical in 85.7% (30/35) of cases, taking into account only those miRNAs shared by both species (Table 2). Regarding the pre-miRNAs sequences, the identity is on average 98.4%. This value is higher than the pre-miRNA identity calculated for *Schistosoma mansoni* and *Schistosoma japonicum* which is 84.9% (miRBase version 20.0).

miR-1992 was the miRNA with the lowest identity between *E. multilocularis* and *E. canadensis* (G7) mature sequences (86.4%, Table 2). It also showed the peculiarity that the seed region differs between both organisms. This may be the result of "seed shifting", which is one mechanism of evolving new miRNA roles and occurs when the mature sequence of a given miRNA is moved one or several nucleotides relative to its original position (de Wit et al., 2009). However, since seed shifting is observed in highly diverged species (de Wit et al., 2009), a higher number of read counts should be obtained for emu-miR-1992 to assert its seed region and confirm or refute this hypothesis.

Table 1
General results of small RNA-seq experiments from whole parasite homogenates of *Echinococcus multilocularis* and *Echinococcus canadensis* (G7 genotype) metacestodes.

	E. multilocularis		*E. canadensis* (G7 genotype)	
	Sample 1 ($N = 5$)	Sample 2 ($N = 5$)	Sample 1 ($N = 1$)	Sample 2 ($N = 1$)
Raw reads	40,485,054	38,353,801	17,854,364	19,958,589
Trimmed reads	29,422,027	25,461,944	16,860,263	17,483,714
Clean reads ($\geqslant$18 nt)[a]	24,703,158	20,396,074	16,431,381	16,364,826
Mapped reads[b]	13,793,487	13,160,068	11,836,512	12,596,993
Percentage of mapped reads[c]	55.8%	64.5%	72.0%	77.0%
Unique reads	589,137	684,133	320,130	282,429
Pearson's correlation coefficient (microRNAs fraction)	0.998		0.993	

[a] No base calling ambiguities in the sequences.
[b] Results obtained allowing up to two mismatches to the reference genome (*E. multilocularis* or *E. granulosus* s.s., G1 genotype).
[c] Relative to clean reads.

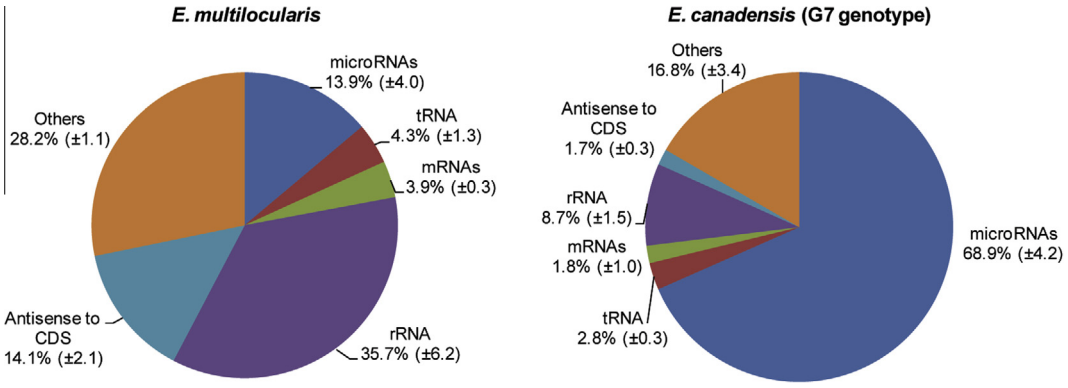


Fig. 1. Classification of the RNA species obtained in the datasets from the metacestode stage of *Echinococcus multilocularis* and *Echinococcus canadensis* (G7 genotype). Results are shown as average percentages (± S.D.) of biological replicates. CDS, coding sequences. Others, sequence groups with no formal annotation (unknown), repetitive sequences (sense/antisense), long non-coding RNAs (sense/antisense), sequences antisense to tRNAs and rRNAs (see Supplementary Table S1).

3.3. miRNA expression profile of Echinococcus spp. metacestodes

When analysing the expression profiles, we observed that both species show the same top five most expressed miRNAs, which are miR-10, let-7, bantam, miR-71 and miR-9 (Fig. 2A). Let-7 and miR-10 together account for ~55% and ~70% of total miRNA expression in *E. multilocularis* and *E. canadensis* (G7) datasets, respectively, while bantam accounts for ~18% in *E. multilocularis* samples. Regarding the miRNAs organised in clusters, which most likely share the promoter region (Fig. 2B), miR-71 from cluster miR-71/2b/2c shows a fourfold higher expression compared with miR-2b and miR-2c in both datasets; and miR-4989 from cluster miR-277/4989 shows a sixfold or twofold increase compared with miR-277 in *E. multilocularis* and *E. canadensis* (G7) datasets, respectively (Fig. 2A).

We analysed whether the mature product preferentially originated from one arm of the hairpin. In this way, we observed that 70% of the conserved miRNAs showed preferential expression from the 3p arm in both species. In the case of the three new candidate miRNAs, the mature sequences also belonged to the 3p arm (Supplementary Tables S4 and S5). This is in agreement with observations made on nematodes, fruit fly and plants but not in vertebrates where the predominant product comes from the 5p arm (de Wit et al., 2009). Curiously, according to data available in miRBase (version 20.0), the mature homologous sequences for miR-1, miR-61, miR-133 and miR-277 in *S. japonicum* are produced from the 5p arm, which differs in *Echinococcus* spp. In *Schmidtea mediterranea* the homologous sequence to miR-2162-3p is also located on the 3p arm of the precursor but in contrast to *Echinococcus* spp. the predominant arm product is produced from the 5p arm (Friedländer et al., 2009). We also observed that six pre-miRNAs showed expression from both arms: pre-mir-31 for *E. multilocularis* (Supplementary Table S4) and pre-mir-87, pre-mir-124b, pre-mir-125, pre-mir-153 and pre-mir-1992 for *E. canadensis* (G7) (Supplementary Table S5).

A comparative expression analysis was performed for those miRNAs expressed by both species. This analysis was performed using the software DESeq, and the requirement for a miRNA to be considered up- or down-regulated was that it showed a statistically significant expression change $\geqslant$twofold ($-1 \geqslant \log_2 \geqslant 1$). As shown in Fig. 3A, most miRNAs (27/40) did not display any significant difference between *E. multilocularis* and *E. canadensis* (G7) datasets. Five miRNAs were up-regulated in *E. canadensis* (G7) (miR-10-5p, miR-87-5p, miR-277-3p, miR-new-1-3p, miR-new-2-3p) and eight miRNAs were up-regulated in *E. multilocularis* (bantam-3p, miR-2a-3p, miR-2c-3p, miR-36a-3p, miR-36b-3p, miR-124a-3p, miR-124b-3p, miR-125-5p).

As mentioned previously, four miRNAs showed a species-specific expression profile: miR-31-5p and miR-31-3p, which were only detected in *E. multilocularis* datasets, and miR-new-3-3p and miR-4990, which were only in *E. canadensis* (G7) (Table 2, Fig. 3B). We classified the former three miRNAs as expression-specific due to the fact that both reference genomes code for valid hairpin sequences for those (Supplementary Fig. S1), in contrast to egr-miR-4990.

3.4. IsomiRs

Generally, miRNAs are present as isoforms or isomiRs which may differ in length due to imprecise precursor processing, terminal trimming or the addition of non-templated nts (tailing) (Ameres and Zamore, 2013). Tailing normally comprises the addi-

Table 2
Catalog of mature microRNAs (miRNAs) of *Echinococcus multilocularis* and *Echinococcus canadensis* (G7 genotype) in the metacestode stage.

microRNA	miRNA mature sequence[a]		Normalised read counts mean (%) (±S.D.)		Percentage of identity (mature miRNAs)	Percentage of identity (precursor miRNAs)	miRNA length (nt)		Identification method[c]
	E. multilocularis	*E. canadensis* (G7 genotype)	*E. multilocularis*	*E. canadensis*			*E. multilocularis*	*E. canadensis*	
bantam-3p	UGAGAUCGCGAUUACAGCUGAU	UGAGAUCGCGAUUACAGCUGAU	17.784 (±1.477)	3.368 (±0.154)	100.0	99.0	22	22	RNA-seq
let-7-5p	UGAGGUAGUGUUUCGAAUGUCU	UGAGGUAGUGUUUCGAAUGUC[2]	24.979 (±0.686)	23.108 (±2.020)	95.5	100.0	22	21	Sanger, RNA-seq
miR-1-3p	UGGAAUGUUGUGAAGUAUGU	UGGAAUGUUGUGAAGUAUGU	0.014 (±0.002)	0.013 (±0.004)	100.0	100.0	20	20	Sanger, RNA-seq
miR-2a-3p	AAUCACAGCCCUGCUUGGAACC	AAUCACAGCCCUGCUUGGAACC	0.752 (±0.078)	0.310 (±0.122)	100.0	100.0	22	22	Sanger, RNA-seq
miR-2b-3p	UAUCACAGCCCUGCUUGGGAC	UAUCACAGCCCUGCUUGGGAC[2]	1.955 (±0.399)	0.943 (±0.191)	100.0	98.0	21	21	Sanger, RNA-seq
miR-2c-3p	UCACAGCCAAUAUUGAUGAAC	UCACAGCCAAUAUUGAUGAAC	2.022 (±0.083)	0.577 (±0.261)	100.0	100.0	21	21	Sanger, RNA-seq
miR-7a-5p	UGGAAGACUGGUGAUAUGUUGU	UGGAAGACUGGUGAUAUGUUGUA	0.003 (±0.000)	0.001 (±0.001)	95.7	100.0	22	23	Sanger, RNA-seq
miR-8-3p	UAAUACUGUUCGGUUAGGACGCCA	UAAUACUGUUCGGUUAGGACGCCA	0.047 (±0.010)	0.042 (±0.006)	100.0	92.5	24	24	Sanger, RNA-seq
miR-9-5p	UCUUUGGUUAUCUAGCUGUGU[b]	UCUUUGGUUAUCUAGCUGUGU[b]	4.298 (±0.149)	3.266 (±0.088)	100.0	100.0	21	21	Sanger, RNA-seq
miR-10-5p	CACCCUGUAGACCCGAGUUUGA	CACCCUGUAGACCCGAGUUUGA	29.466 (±0.448)	51.874 (±5.700)	100.0	97.0	22	22	Sanger, RNA-seq
miR-31-5p	UGGCAAGAUACUGGCGAAGCUGA	ND	0.001 (±0.000)	NA	NA	100.0	23	NA	RNA-seq
miR-31-3p	AGCUUCGUCUGGUCUUGCUGCA	ND	0.001 (±0.000)	NA	NA	100.0	22	NA	RNA-seq
miR-36a-3p	UCACCGGGUAGACAUUCCUUGC	UCACCGGGUAGACAUUCCUUGC	0.013 (±0.002)	0.000 (±0.000)	100.0	99.0	22	22	RNA-seq
miR-36b-3p	UCACCGGGUAGUUAUUACGCCU	UCACCGGGUAGUUAUUACGCCU	0.020 (±0.002)	0.002 (±0.000)	100.0	99.0	22	22	RNA-seq
miR-61-3p	UGACUAGAAAGAGCACUCACAUC	UGACUAGAAAGAGCACUCACAUC	2.006 (±0.039)	3.238 (±1.325)	100.0	97.0	23	23	RNA-seq
miR-71-5p	UGAAAGACGAUGGUAGUGAGAU	UGAAAGACGAUGGUAGUGAGAU	8.748 (±0.265)	4.555 (±0.913)	100.0	100.0	22	22	Sanger, RNA-seq
miR-87-5p	–	CCACCUGUCAUUUUGCUCGAACC	NA	0.085 (±0.013)	NA	100.0	NA	23	Sanger, RNA-seq
miR-87-3p	GUGAGCAAAGUUUCAGGUGUGC	GUGAGCAAAGUUUCAGGUGUGC	0.436 (±0.058)	0.184 (±0.023)	100.0	100.0	22	22	Sanger, RNA-seq
miR-96-5p	AUUGGCACUUUUGGAAUUGU	AUUGGCACUUUUGGAAUUGUC	0.033 (±0.012)	0.025 (±0.004)	95.2	94.9	20	21	Sanger, RNA-seq
miR-124a-3p	UAAGGCACGCGGUGAAUGCC	UAAGGCACGCGGUGAAUGCCA[b]	0.009 (±0.003)	0.003 (±0.002)	95.2	96.4	20	21	Sanger, RNA-seq
miR-124b-5p	–	GUAUUCUACGCGAUGUCUUGGUA	NA	0.010 (±0.007)	NA	98.0	NA	23	Sanger, RNA-seq
miR-124b-3p	UAAGGCACGCGGUGAAUACC	UAAGGCACGCGGUGAAUACC	0.034 (±0.012)	0.006 (±0.002)	100.0	98.0	20	20	Sanger, RNA-seq
miR-125-5p	UCCCUGAGACCCUAGAGUUGUC	UCCCUGAGACCCUAGAGUUGUC	0.155 (±0.031)	0.045 (±0.027)	100.0	97.0	22	22	Sanger, RNA-seq
miR-125-3p	–	CAACUCUAAUGUCCCGGGUUAU	NA	0.021 (0.018)	NA	97.0	NA	22	Sanger, RNA-seq
miR-133-3p	UUGGUCCCCAUUAACCAGCCGCCb	UUGGUCCCCAUUAACCAGCCGCC[b]	0.008 (±0.004)	0.005 (±0.003)	100.0	100.0	23	23	RNA-seq
miR-153-5p	–	AUGCUUACGAGACGUGCACUC	NA	0.002	NA	99.0	NA	21	RNA-seq

				(±0.001)					
miR-153-3p	UUGCAUAGUCUCAUAAGUGCCAb	UUGCAUAGUCUCAUAAGUGCCA[b]	0.007 (±0.003)	0.003 (±0.002)	100.0	99.0	22	22	Sanger, RNA-seq
miR-184-3p (former miR-4988)[d]	GGGACGGAAGUCUGAAAGGUUU	GGGACGGAAGUCUGAAAGGUUU	1.140 (±0.083)	1.543 (±0.218)	100.0	98.0	22	22	Sanger, RNA-seq
miR-190-5p	AGAUAUGUUUGGGUUACUUGGUG	AGAUAUGUUUGGGUUACUUGGUG	0.533 (±0.031)	0.305 (±0.147)	100.0	95.4	23	23	Sanger, RNA-seq
miR-219-5p	UGAUUGUCCAUUCGCAUUUCUUG	UGAUUGUCCAUUCGCAUUUCUUG	0.177 (±0.007)	0.272 (±0.033)	100.0	99.0	23	23	Sanger, RNA-seq
miR-277-3p	UAAAUGCAUUUUCUGGCCCGUA	UAAAUGCAUUUUCUGGCCCGUA	0.412 (±0.028)	1.084 (±0.053)	100.0	100.0	22	22	Sanger, RNA-seq
miR-281-3p	UGUCAUGGAGUUGCUCUCU	UGUCAUGGAGUUGCUCUCU	0.163 (±0.019)	0.069 (±0.018)	100.0	98.0	19	19	RNA-seq
miR-307-3p	UCACAACCUACUUGAUUGAGGGG	UCACAACCUACUUGAUUGAGGGG	0.442 (±0.110)	0.647 (±0.068)	100.0	95.9	23	23	RNA-seq
miR-745-3p	UGCUGCCUGGUAAGAGCUGUGA	UGCUGCCUGGUAAGAGCUGUGA	0.531 (±0.004)	0.598 (±0.045)	100.0	96.6	22	22	Sanger, RNA-seq
miR-1992-5p	–	UUUCAUUGGUCAAUUGCUAA	NA	0.000 (±0.000)	NA	97.1	NA	20	RNA-seq
miR-1992-3p	UCAGCAGUUGUACCAUUGAAAU	AGCAGUUGUACCAUUGAAAUG	0.001 (±0.000)	0.000 (±0.000)	86.4	97.1	22	21	RNA-seq
miR-2162-3p	UAUUAUGCAACUUUUCACUCC	UAUUAUGCAACUUUUCACUCC	0.546 (±0.068)	0.503 (±0.060)	100.0	100.0	21	21	RNA-seq
miR-3479a-3p	UAUUGCACGUUCUUUCGCCAUC	UAUUGCACGUUCUUUCGCCAUC	0.356 (±0.003)	0.270 (±0.089)	100.0	100.0	22	22	RNA-seq
miR-3479b-3p	GAUUGCACUACCCAUCGCCCAC[b]	GAUUGCACUACCCAUCGCCCAC[b]	0.253 (±0.017)	0.270 (±0.054)	100.0	96.0	22	22	RNA-seq
miR-4989-3p	AAAAUGCACCAACUAUCUGAGA	AAAAUGCACCAACUAUCUGAGA	2.563 (±0.036)	2.310 (±1.246)	100.0	100.0	22	22	Sanger, RNA-seq
miR-4990-5p	ND	UGUCUCCUCACGGGUUUAAACCC	NA	0.004 (±0.003)	NA	NA	NA	23	Sanger, RNA-seq
miR-new-1-3p	UAAUUCGAGUCAACAGGGUCGUU	UAAUUCGAGUCAACAGGGUCGUU	0.007 (±0.000)	0.023 (±0.007)	100.0	100.0	23	23	RNA-seq
miR-new-2-3p	UAAAUGCAAAAUAUCUGGUUAUG	UAAAUGCAAAAUAUCUGGUUAUG	0.087 (±0.034)	0.397 (±0.116)	100.0	100.0	23	23	RNA-seq
miR-new-3-3p	ND	GCAGGUGACUCCAAAACUUUUG[b]	NA	0.027 (±0.021)	NA	98.0	NA	22	RNA-seq
				Average	99.1	98.4	21.9	21.9	

5p, 5′; 3p, 3′; ND, not detected; NA, not applicable.

[a] The canonical (100% identical to the reference genome) most frequent read from both biological replicates is reported. When the minor product of a precursor (pre)-miRNA showed a read count number ⩾30% with respect to the major product, it was also considered as mature miRNA.

[b] The most frequent read corresponded to an isomiR (miRNA isoform).

[c] Sanger: corresponds to miRNAs detected by conventional cloning and sequencing in Cucher et al. (2011). RNA-seq: corresponds to miRNAs detected in this work by high throughput RNA sequencing.

[d] Only miR-4988-5p (not miR-4988-3p) was previously identified by conventional cloning (Cucher et al., 2011).

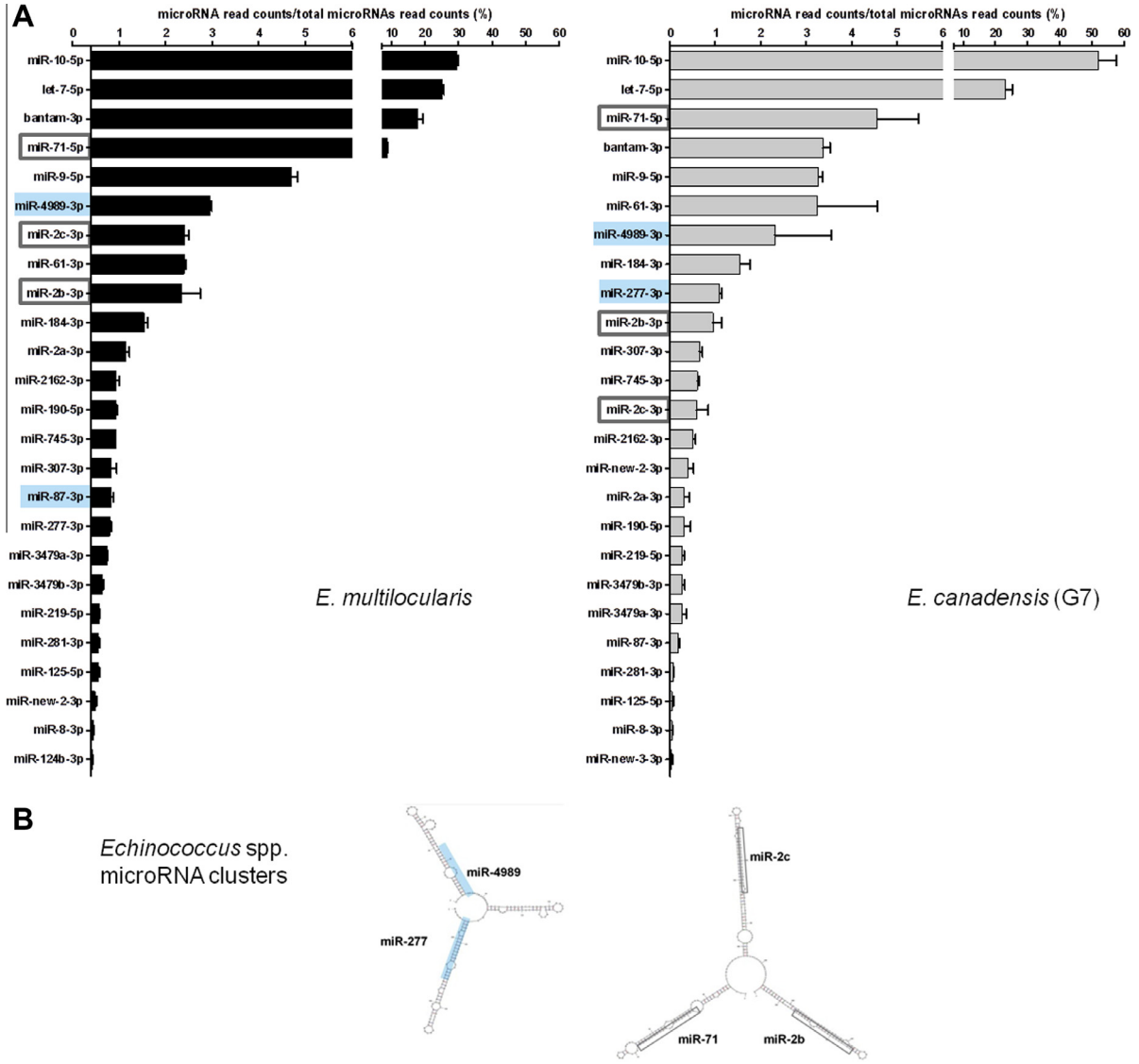


Fig. 2. *Echinococcus* spp. microRNA expression profiles. (A) Top 25 most abundant microRNAs in *Echinococcus multilocularis* and *Echinococcus canadensis* (G7 genotype) datasets. The shaded and framed microRNAs are members of different microRNA clusters. The read counts of individual microRNAs were normalised to the total number of mature microRNA read counts within each library. Results are shown as average percentages (± S.D.) of biological replicates. (B) *Echinococcus* spp. microRNA clusters.

tion of adenosine or uridine at the 3p end which can be identified because the added nts do not map to the genome (Kim et al., 2010). We searched for the presence of tailing in our datasets and considered a miRNA to be tailed when the isomiR read counts were ⩾10% with respect to the total number of reads for that particular miRNA (Farazi et al., 2012). In this way, we exclusively observed miRNA processing by addition of a single 3p terminal U in 12 miRNAs in *E. multilocularis* and 18 miRNAs in *E. canadensis* (G7) (Fig. 4). Poly-uridylated reads were also detected in some cases but the number of read counts was lower than the cut-off value (150 counts), so those were not taken into consideration.

The isomiRs identified in *E. multilocularis* and *E. canadensis* (G7) corresponded to the same miRNAs, have the same sequences and were detected in each biological replicate of each species (Supplementary Table S6), even though for six isomiRs in *E. multilocularis*, data did not fulfill the read counts and/or percentages of read cut-off values. Interestingly, in six cases (miR-2b-3p, miR-9-5p, miR-124b-5p, miR-745-3p, miR-3479b-3p and miR-new-3-3p) in *E. canadensis* (G7) and two cases (miR-9-5p and miR-3479b-3p) in *E. multilocularis*, the isomer-U was more abundant than the unmodified form. Regarding the top five most expressed miRNAs, three in *E. canadensis* (miR-9, miR-10, miR-71) and two in *E. multilocularis* (miR-9 and miR-71) showed ⩾10% uridylation.

Finally, the different miRNAs within one cluster showed different levels of uridylation. In cluster miR-71/2b/2c, only miR-2b and miR-71 are uridylated, with miR-2b displaying a higher degree of uridylation. In cluster miR-277/4989, the only uridylated member is miR-277. Taking into account these results and the expression level of each miRNA (Fig. 2), it can be observed that generally the less expressed miRNAs from each cluster show a higher degree of uridylation (except for miR-2c).

miRNA primary transcripts have been described to undergo RNA editing by adenosine deaminases that modify adenosine into inosine (Kim et al., 2010), which is read as guanosine in sequencing assays since it pairs to cytosine. According to our results, there is no evidence of editing on *Echinococcus* spp. miRNAs, at least in the mature sequences of the stage analysed (the metacestode).

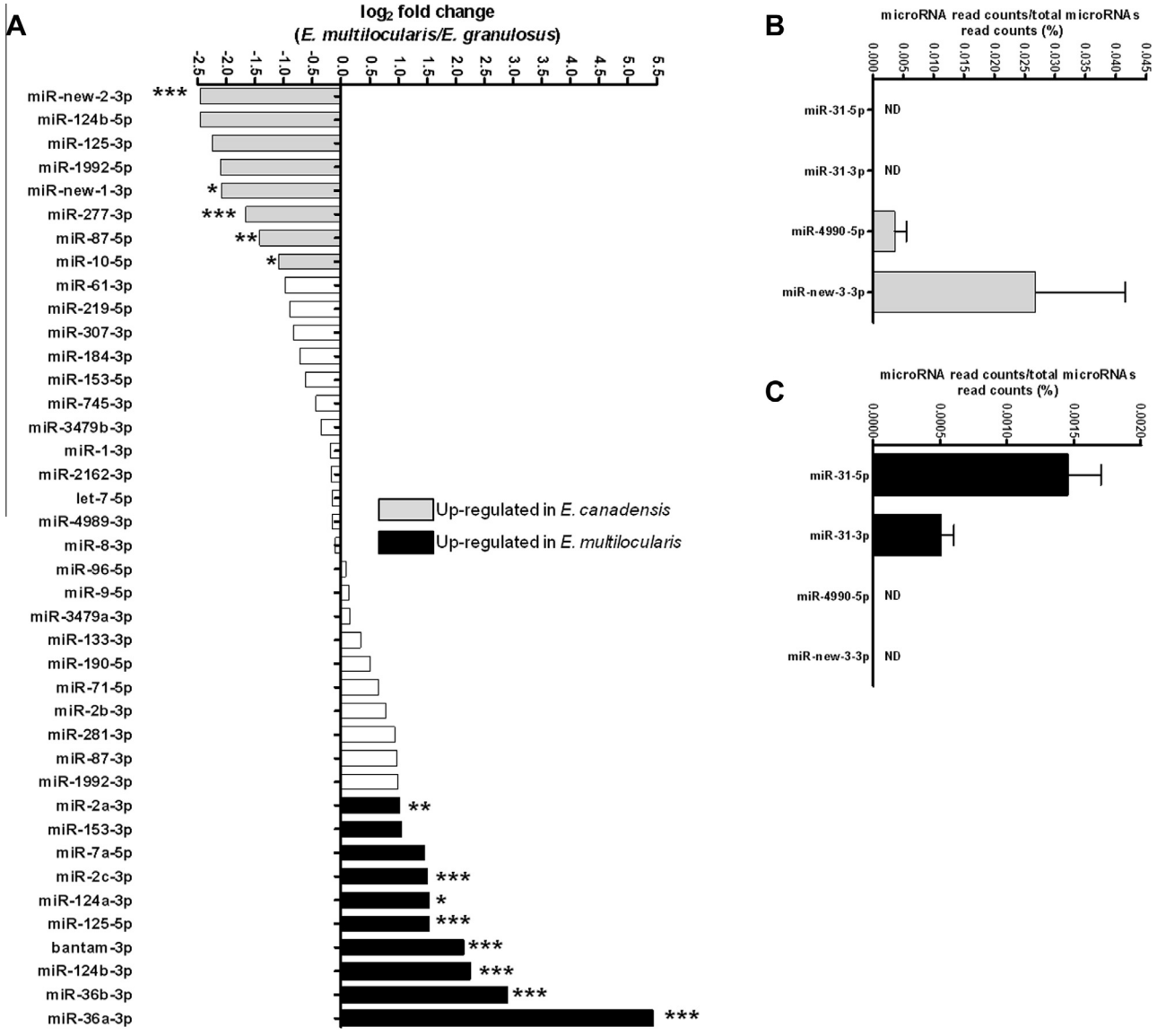


Fig. 3. Differential expression of microRNAs between *Echinococcus multilocularis* and *Echinococcus canadensis* (G7 genotype) datasets. (A) Fold change expression analysis performed with the DESeq algorithm. Differences were regarded as significant for fold changes $\geqslant 2$ ($-1 \geqslant \log_2 \geqslant 1$) and $^{*}P < 0.05$, $^{**}P < 0.01$ or $^{***}P < 0.001$. Normalised expression levels of microRNAs detected only in (B) *E. canadensis* or (C) *E. multilocularis* datasets. ND, not detected. The read counts of each microRNA were normalised to the total number of mature microRNA read counts within each library. Results are shown as average percentages (± S.D.) of biological replicates.

3.5. Chromosomal distribution and genomic context of E. multilocularis pre-miRNAs

All pre-miRNAs could be mapped to the *E. multilocularis* genome and located on the nine chromosomes (Fig. 5A) except for pre-mir-4990, which is *E. granulosus* s. l.-specific, and pre-mir-new-3 which maps to a still unassembled contig. Fifty percent (17/34) of *E. multilocularis* conserved miRNAs present synteny with *S. mansoni* miRNAs according to the chromosomal distribution reported for this related parasite (de Souza Gomes et al., 2011; Marco et al., 2013) (Fig. 5B). The comparative analysis of synteny with *E. granulosus* s. l. could not be performed since the *E. granulosus* s. s. (G1) genome is not yet assembled into chromosomes.

Nearly 50% of *Echinococcus* pre-miRNAs described to date are located on the two largest chromosomes, 1 and 2 (Tsai et al., 2013), while no conserved pre-miRNAs are located on chromosomes 7 or 8. Only one of the novel pre-miRNAs (new-1) is located on chromosome 8 (Fig. 5A).

Let-7 and miR-1992 are located on chromosome 9, the smallest chromosome. This chromosome showed transient trisomy in protoscoleces and metacestodes of two different isolates of *E. multilocularis* (Tsai et al., 2013), which is in general agreement with karyotype plasticity observed in tapeworms (Spakulova et al., 2011). This raises the question whether these miRNAs are transcriptionally active during this event and what impact their over-expression may have on parasite development. let-7 is not clustered with mir-125 as reported for many bilaterians (Campo-Paysaa et al., 2011) except for *C. elegans* (Sokol, 2012). These pre-miRNAs are even located on different chromosomes in *Echinococcus*. The same situation has been observed in *S. mansoni* (de Souza Gomes et al., 2011).

Regarding the genomic location of *E. multilocularis* pre-miRNAs, 81.1% (30/37) are located in intergenic regions and 18.9% (7/37) are located in introns of protein-coding genes according to current GeneDB annotation (Release February 2014). Three of those are located in introns of sequences with no functional annotation (Supplementary Table S7).

The intronic pre-miRNAs that are located in genes with defined annotation are mir-96, mir-190, mir-3479b and mir-4988 (now renamed mir-184, see below). In agreement with previous reports, mir-190 is located within the *talin* gene as observed in many bila-

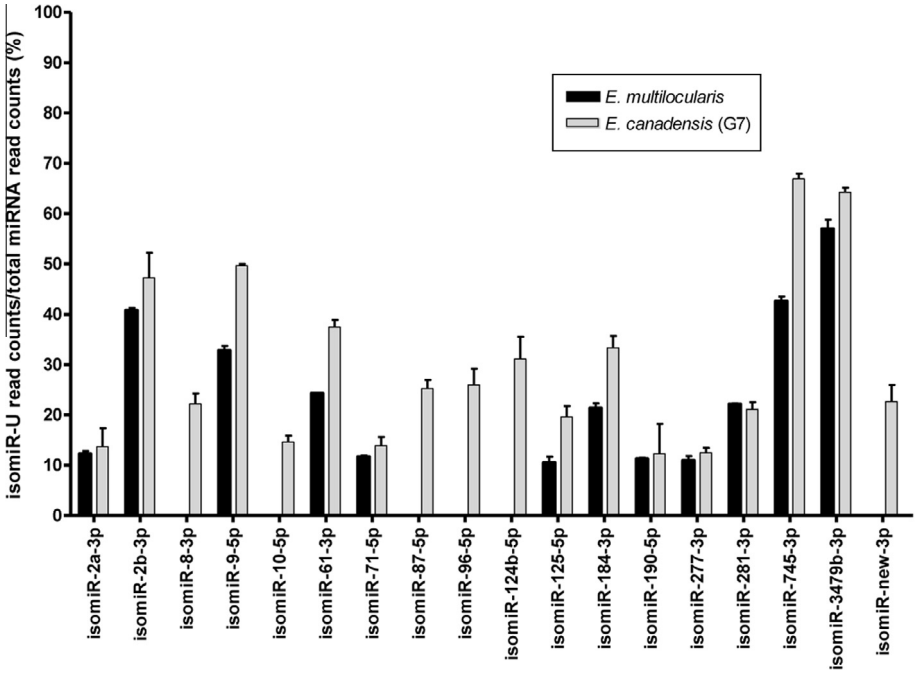


Fig. 4. Subset of microRNAs that showed mono-uridylated isoform/s (IsomiR-U). Only those isoforms with ⩾150 read counts representing ⩾2.5% of total read counts for the corresponding microRNAs in each library were taken into consideration. When more than one isoform was detected, the percentages were added. Results are shown as average percentages (± S.D.) of biological replicates.

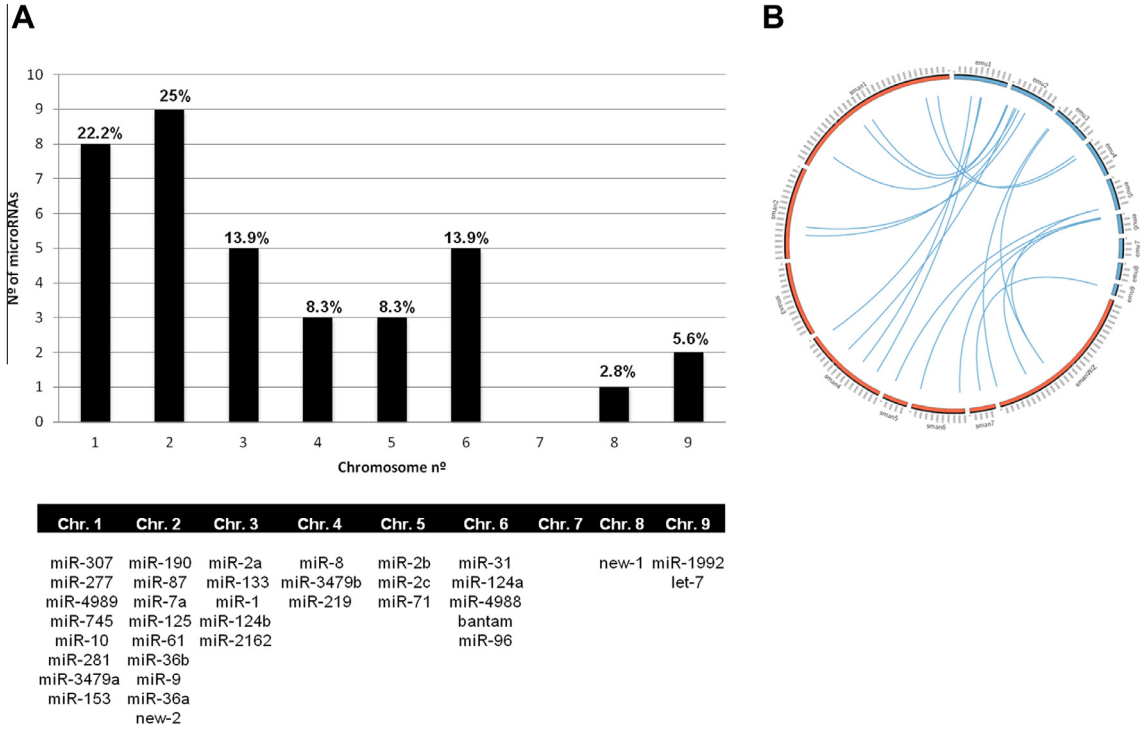


Fig. 5. Chromosomal distribution of *Echinococcus* microRNAs and synteny analysis with *Schistosoma mansoni*. (A) Chromosomal location and distribution of *Echinococcus multilocularis* microRNAs. The microRNAs are placed according to their relative position on each chromosome (Chr.). Percentages indicate the proportions of microRNAs on each chromosome with respect to the total number (Nº) of microRNAs. (B) Synteny analysis of conserved microRNAs between *E. multilocularis* (emu; blue (black)) and the trematode parasite *S. mansoni* (sman; orange (grey)). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

terians (Campo-Paysaa et al., 2011) including *S. mansoni* (de Souza Gomes et al., 2011). With respect to pre-mir-3479b, in *S. mansoni* this miRNA homologue is also intronic. All of those are in sense orientation with their host gene except for mir-4988 (mir-184). The host genes of mir-96, mir-190, mir-3479b and mir-4988 (mir-184) show a low but detectable level of expression in the metacestode stage according to previously published FPKM data (Tsai et al., 2013).

One of the most highly conserved miRNAs, mir-10, is located downstream of *HoxB4a* as reported for bilaterians (Mansfield and

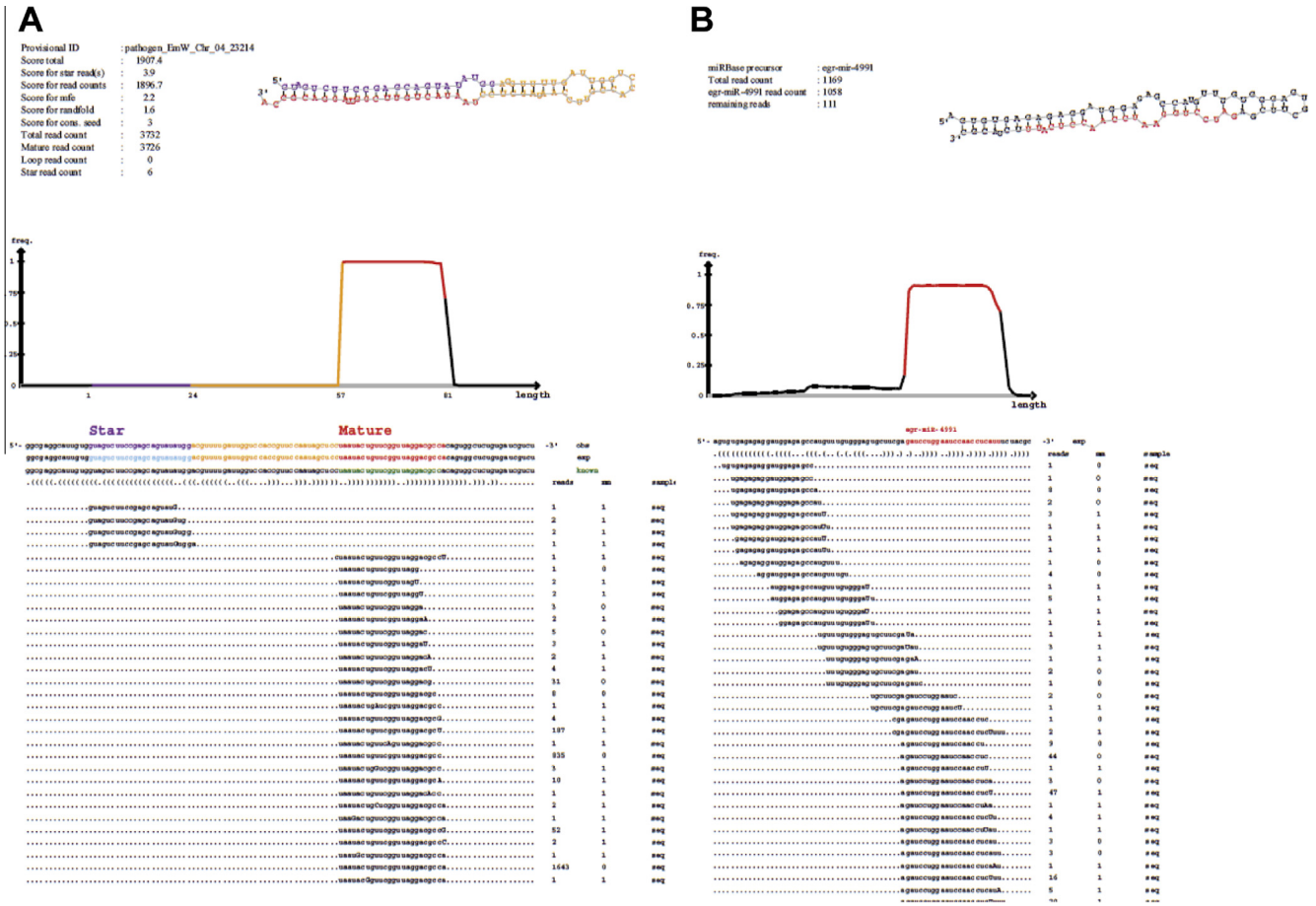


Fig. 6. Read data from a small RNA library from *Echinococcus canadensis* (G7 genotype). (A) Representative miRDeep2 output of a bona fide microRNA (egr-miR-1). Note that the mature and star reads can be clearly identified. In this case, no loop reads were sequenced. (B) miRDeep2 output not compatible with microRNA (miRNA) biogenesis products (former egr-miR-4991). Note that reads overlap throughout the entire putative pre-miRNA sequence. Freq., frequency.

McGlinn, 2012). In *Echinococcus* this miRNA is in antisense orientation, as was reported for the Hox gene embedded miRNAs mir-iab4 in *D. melanogaster* and mir-196 in vertebrates (Mansfield and McGlinn, 2012).

According to previous reports, platyhelminth miRNA clusters size up to 500 bp (Wang et al., 2010; Cucher et al., 2011; Sasidharan et al., 2013). *Echinococcus* spp. miRNA clusters are mir-71/2b/2c and mir-277/4989 (Cucher et al., 2011). It has been proposed that *Echinococcus* mir-1 and mir-133 form another cluster (Jin et al., 2013) as observed in many other organisms (Campo-Paysaa et al., 2011). Due to the fact that mir-1 and mir-133 are located approximately 12 kb from each other, further assays should be performed to determine their co-transcription as a polycistronic unit. Furthermore, according to current gene annotation (GeneDB release February 2014) these miRNAs are not located on one of the *mib-1* introns, as generally observed in metazoans (Campo-Paysaa et al., 2011), but in an intergenic region between the genes E3 ubiquitin protein ligase mib 1 and mind bomb. These miRNAs are located in antisense orientation to their neighbouring genes as reported for the mir-1/133 cluster (Campo-Paysaa et al., 2011).

3.6. Re-annotation of Echinococcus miRNAs

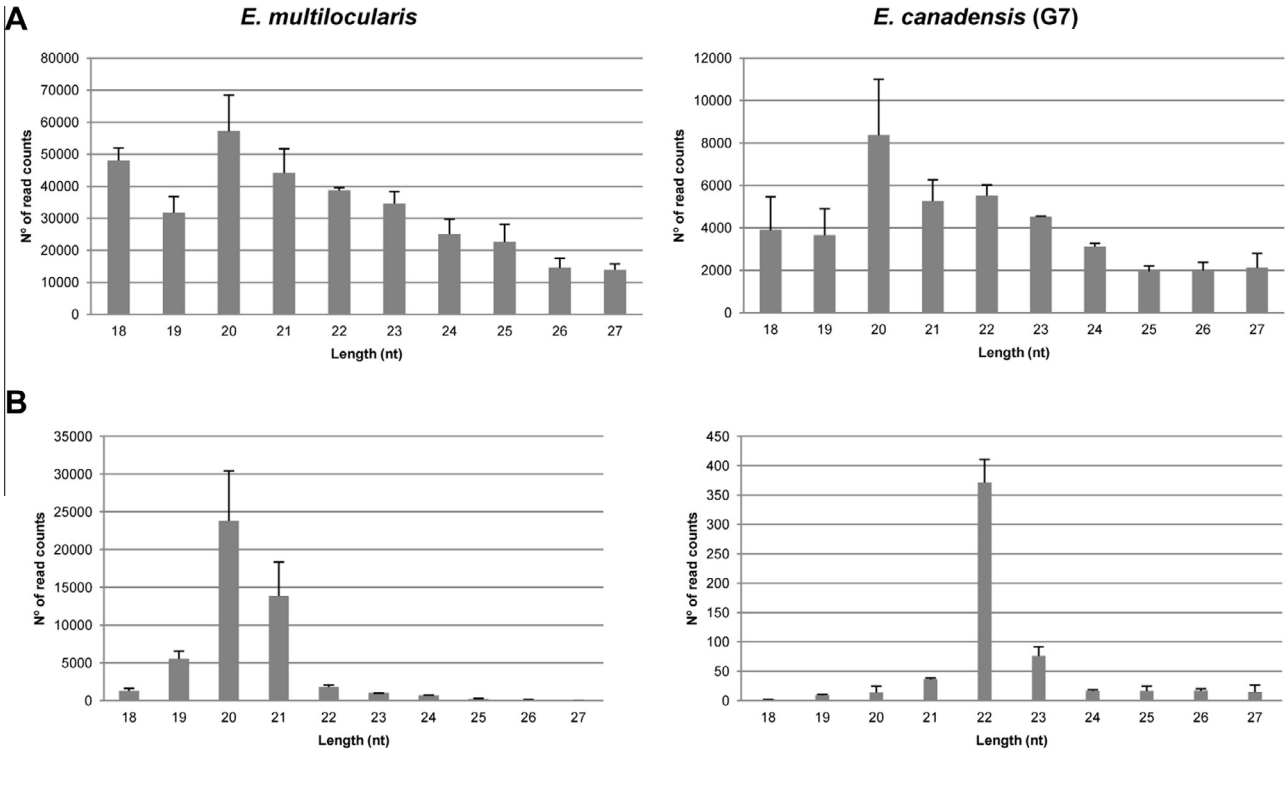
The data obtained in this report allowed the confirmation of miRNA sequences previously identified experimentally (Cucher et al., 2011) or predicted in silico (Cucher et al., 2011; Jin et al., 2013) and the determination of the products from both arms of all *Echinococcus* pre-miRNAs.

In this way, we confirmed that the mature product of mir-4988 corresponds to the 3p arm and not the 5p arm as we concluded previously (Cucher et al., 2011) and that it is not an *Echinococcus*-specific miRNA since miR-4988-3p is homologous to miR-184 according to the criteria we used to classify the sequences (see Section 2.8). Thus, we propose to rename miR-4988 as miR-184.

Also, due to the high depth of the libraries, it could be determined that miR-4991 is not a real miRNA according to its miRDeep2 pattern which is not compatible with miRNA biogenesis products (Fig. 6). Even though it was cloned and sequenced from *E. canadensis* (G7) protoscoleces (Cucher et al., 2011), and from *E. multilocularis* and *E. canadensis* (G7) metacestode samples (this work), its length is within the expected range and the flanking genome sequence can fold into a stem-loop structure. This sequence maps to a region in *E. multilocularis* genome (version 4) annotated as "hypothetical protein" (70/73 positive nucleotides). As already reported, the pattern of reads that map to a predicted hairpin locus provides the necessary information to discriminate between a bona fide miRNA and other transcribed fragments (Kozomara and Griffiths-Jones, 2014). Taking this result into consideration we excluded miR-4991 from the *Echinococcus* miRNA repertoire.

3.7. Other sRNAs

With respect to other sRNAs of approximately 22 nt long, reads antisense to CDS were detected for both species. To analyse the presence of endogenous siRNAs (endo-siRNAs), we only considered those sequences that were 100% identical to the corresponding genome in coding regions and ~22 nt long. As shown in Fig. 7A,



Dataset	Sequence	Length (nt)	Nº read counts		Gene model	Product
			Sample 1	Sample 2		
E. multilocularis	GAAGGAUGUCGUUAGUCUAC	20	17231	25540	EmuJ_000048500	non capsid protein NS 1
E. canadensis (G7 genotype)	GUUUGCUGGUUGUCGUGGUUCC	22	356	319	EgG_000323300.1	Histone H2B

Fig. 7. Analysis of antisense reads to coding sequences for *Echinococcus* spp. (A) Length distribution of antisense reads to CDS, 100% identical to the corresponding reference genome (*Echinococcus multilocularis* or *Echinococcus granulosus* sensu stricto). (B) Length distribution of antisense reads to gene models EmuJ_000048500 or EgG_000323300.1 compatible with small interfering RNAs (siRNAs) in *E. multilocularis* or *E. canadensis* datasets. (C) Detail of antisense reads compatible with siRNAs in *E. multilocularis* or *E. canadensis* datasets. Results are shown as averages (± S.D.) of biological replicates. nt, nucleotide; nº, number.

there is no evidence of a high predominance of this type of small RNA among the CDS-antisense reads population, i.e. there is no peak at ~22 nt. Nevertheless, the siRNA pathway machinery is present and functional in *Echinococcus* (Mizukami et al., 2010; Spiliotis et al., 2010). The length-based identification of siRNAs may have been hindered by the fact that the libraries correspond to whole-parasite homogenates and endo-siRNAs may be more abundant in certain cell types than others. Hence, we changed the approach and analysed the presence of siRNAs by first studying the composition of the 20–22 nt long peaks from each species dataset to determine whether there were predominant sequences in each peak which were antisense to a single gene model. As this occurred for both species, we proceeded to analyse the individual length distribution of the antisense reads against those gene models. By doing this, we identified a subset of candidate endo-siRNAs targeting histone H2B transcripts (gene model EgG_323300) in *E. canadensis* (G7). As depicted in Fig. 7B, the length distribution of the candidate siRNAs peaks, as expected, at 22 nt. For *E. multilocularis*, we found antisense reads targeting the transcript annotated as non-capsid protein NS1 (gene model EmuJ_000048500), however the length distribution of the reads shows a peak at 20 nt (Fig. 7B). Overall, the great difference in read count numbers corresponding to antisense sequences between both species and the low density of candidate endo-siRNAs are remarkable.

4. Discussion

We report, to our knowledge for the first time, experimental evidence on *E. multilocularis* miRNA expression. We also report the expansion of the *Echinococcus* miRNA collection by the identification of 15 additional pre-miRNAs. The current repertoire of *Echinococcus* spp. miRNAs is composed of 38 genes, grouped in 32 families.

We experimentally detected expression of the in silico predicted *Echinococcus* miRNAs bantam, miR-31, miR-61, miR-133, miR-281 and miR-2162 (Jin et al., 2013) and miR-36, miR-184 (former miR-4988), miR-281 and miR-1992, which had been assumed to be lost in *Echinococcus* (Fromm et al., 2013). This highlights the importance of experimental confirmation of in silico data used to predict loss or gain of miRNA genes.

One of the newly identified miRNAs, miR-31, has been already described in other organisms and is encoded by both reference genomes (*E. multilocularis* and *E. granulosus* s. s.) but its expression could only be detected in *E. multilocularis* datasets. However, the species-specific expression of this miRNA awaits confirmation by other techniques as it was only detected with a low read count number (mean = 25). We also identified three novel candidate miRNAs, with all of those present in both parasites genomes, but only two showed expression in both species datasets. One of these

miRNAs (miR-new-2) is a new member of the family miR-277/4989 as it shares the same seed region. The third one (miR-new-3) was only detected in *E. canadensis* (G7) with a moderate read count number (mean = 2146).

The sRNA-seq data analysis did not yield a remarkably higher number of miRNAs for *Echinococcus* spp. compared with the conventional cloning and sequencing methodology used previously (Cucher et al., 2011). This could be due to the lack of information on sRNAs expressed by oncospheres and adult worms that could enlarge the repertoire of *Echinococcus* miRNAs. Another possibility is that many *Echinococcus* spp. miRNAs evolved recently, i.e. those have no homologues in other organisms, and according to the stringent pipeline of miRNA discovery used in this work, were discarded. Finally, the low number of miRNAs would be in agreement with the hypothesis of a loss of conserved miRNAs in platyhelminths (Fromm et al., 2013).

Regarding the miRNA expression profiles of *E. multilocularis* and *E. canadensis* (G7), we observed that both species showed the same top five most highly expressed miRNAs (bantam, let-7, miR-9, miR-10 and miR-71) during infection of the intermediate host. In the *E. multilocularis* datasets, miR-10, let-7 and bantam represent ~70% of total miRNA expression and in *E. canadensis* (G7) this value is reached only by miR-10 and let-7 expression.

miR-71, one of the top five most expressed miRNAs in *Echinococcus* metacestodes (this report; Macchiaroli et al., 2015) and protoscoleces (Macchiaroli et al., 2015), is also one of the most highly expressed miRNAs in different life-cycle stages of *Schistosoma* spp. (Huang et al., 2009; Hao et al., 2010; Wang et al., 2010; Cai et al., 2011). This suggests that miR-71 may play stage-independent roles relevant to parasite homeostasis or that its range of targets is so wide that it regulates multiple genes in different cell types. In line with the latter, it has been shown that miR-71 plays many different roles in *C. elegans*, such as regulation of diapause survival and developmental recovery functions (Zhang et al., 2011), resistance to heat and oxidative stress, promotion of longevity (Boulias and Horvitz, 2012) and differentiation of olfactory neurons (Hsieh et al., 2012).

With respect to the miRNAs differentially expressed between *E. multilocularis* and *E. canadensis* (G7) samples, five miRNAs were significantly up-regulated in *E. canadensis* (G7) and eight were significantly up-regulated in *E. multilocularis*. From these, miR-36a and miR-36b showed the most remarkable up-regulation in *E. multilocularis* with approximately 40-fold and sevenfold, respectively. This miRNA family has been shown to be enriched in planarian neoblasts, the cells responsible for the extraordinary regenerative capacity of this platyhelminth (Friedländer et al., 2009; Lu et al., 2009), and would be in accordance with the higher regenerative capacity of *E. multilocularis* with respect to *E. granulosus* s. l. Bantam, miR-2a and miR-2c were also significantly up-regulated in *E. multilocularis* samples. The roles of these miRNAs have been experimentally validated in *D. melanogaster* where they were probed to down-regulate the expression of pro-apoptotic genes (Brennecke et al., 2003; Stark et al., 2003), in agreement with the highly proliferative nature of *E. multilocularis*.

miR-4989 has been previously proposed as a cestode-specific miRNA (Jin et al., 2013) and showed a similar moderate expression level in both species datasets, being among the top 10 most highly expressed miRNAs. It would be interesting to discover its role in *Echinococcus* biology and whether this miRNA is relevant in the determination of the parasitic nature of these and other cestodes.

Interestingly, 73% (27/37) and 78.6% (33/42) of miRNAs in *E. multilocularis* and *E. canadensis* (G7), respectively, showed low expression levels (⩽1%) (Table 2). Since whole-organism homogenates were used for sRNA-seq, low expression miRNAs may be cell-specific miRNAs whose expression levels were underrepresented among abundant miRNAs with widespread expression patterns.

Echinococcus canadensis (G7) presents a wider catalogue of expressed miRNAs due to production of mature miRNAs from both arms of some precursors. This is the case for mir-87, mir-124b, mir-125, mir-153 and mir-1992. Only one pre-miRNA (mir-31) showed this expression pattern in *E. multilocularis* and was exclusively detected in this species dataset. However, data on mir-31 and mir-1992 expression should be considered with caution due to the low number of reads. Further experiments such as in situ hybridization assays will help to clarify whether the different arm products are expressed in different cell types.

In plants and animals, non-template nt additions in sRNAs are commonly observed. The added nts are mostly uridine or adenosine which can apparently induce the degradation or stabilization of the modified sRNA (Kim et al., 2010). Although there is increasing evidence that uridylation tags sRNAs for degradation, its role on metazoan miRNA biogenesis, turnover and/or function needs further research (Scott and Norbury, 2013). Here, we observed that only certain miRNAs were 3p modified by the presence of a single non-template added uridine. Furthermore, the modified miRNAs coincide in *E. multilocularis* and *E. canadensis* (G7) datasets, suggesting that non-templated nt addition is not the result of sequencing "noise".

Among the different mono-uridylated miRNAs, some were produced from the 5p arm and others from the 3p arm of the corresponding precursors. The modified products coming from the 5p arms are miRNAs mono-uridylated in their mature form, however, the modified products coming from the 3p arm could have been modified either at the precursor or mature miRNAs (Kim et al., 2010). This suggests that there could be at least two different steps or mechanisms that act on the mono-uridylation of *Echincoccus* miRNAs.

It can be speculated that only the mature miRNAs are modified as the normal fate of star strands is degradation. This was the case for those miRNA:miRNAs* pairs with a number of miRNA* read counts higher than the cut-off value selected for this analysis.

Two of the modified miRNAs were miR-87-5p and miR-125-5p, which according to our criterion originate from precursors that produce mature miRNAs from both arms in the *E. canadensis* (G7) datasets. The most abundant products for each pre-miRNA corresponded to miR-87-3p and miR-125-5p. This implies that the 3p arm products are not under this type of regulatory mechanism or, in the case of miR-125-3p, that it simply corresponds to a star sequence with a low rate of degradation. It should be taken into account that these results disguise the fact that not all of the miRNAs are expressed in all of the different cell types, so it could happen that the same miRNA is tagged for degradation in certain cell types and not others, i.e. a mix of canonical miRNAs and isomiR-U are detected in whole parasite samples.

It is worth noting that the members of *Echinococcus* miRNA clusters undergo different levels of individual processing. *Echinococcus* miRNA clusters are small in size and not compatible with individual gene transcription (clusters size ~250 bp). Hence, it would be expected to detect a similar level of expression for the different miRNAs within one cluster if no post-transcriptional mechanisms are affecting their stability. However, miR-4989 from cluster miR-277/4989 shows a higher level of expression and no evidence of uridylation, in contrast to its partner, miR-277. This was also observed among the members of cluster miR-71/2b/2c, where miR-2b is uridylated in a higher proportion than miR-71, which shows a higher level of expression. No uridylation was detected for miR-2c, suggesting that its lower expression level is due to a different mechanism of post-transcriptional regulation (Fig. 4). Post-transcriptional processing of miRNAs also has practical importance since it can influence the annealing of stem-loop reverse transcription primers (Chen et al., 2005) or PCR reverse-specific primers which are in both cases designed to anneal to the 3p end nts of miRNAs. Finally, no editing (A to I) was detected

on mature sequences. However, we cannot rule out the possibility that the pri- or pre-miRNAs are undergoing editing.

In contrast to vertebrate miRNAs which are mainly intronic (Godnic et al., 2013), most *Echinococcus* miRNA loci are located in intergenic regions as already reported for worms, flies (Griffiths-Jones et al., 2008), planaria (Friedländer et al., 2009) and the helminth parasites *Schistosoma* spp. (Cai et al., 2011; de Souza Gomes et al., 2011), *H. contortus* (Winter et al., 2012), *Ascaris suum* (Wang et al., 2011) and *Brugia malayi* (Poole et al., 2010). In *Echinococcus*, all intragenic miRNAs are located on introns of protein-coding genes, according to current gene annotation (GeneDB release February 2014). Intronic miRNAs may depend on their host gene promoter region for their expression or can have independent promoters. In the latter case, it cannot be ruled out that both promoters may influence the transcription of intronic miRNAs (Isik et al., 2010). In either case, intergenic miRNAs have the advantage of being transcriptionally independent of a host gene which might allow their expression in a wider range of cell types, tissues or developmental stages, enlarging their spectrum of targets. In the case of platyhelminths, especially parasites, this would be one way to maximise the action of the low number of miRNAs they have.

As reported for *S. mansoni* (de Souza Gomes et al., 2011), *Echinococcus* miRNAs identified to date are single-locus genes. The opposite situation is observed in the free-living platyhelminth, *S. mediterranea* (Palakodeti et al., 2006) and the parasitic nematodes *B. malayi*, *H. contortus* and *A. suum* (Poole et al., 2010; Wang et al., 2011; Winter et al., 2012) where miRNAs with more than one locus, i.e. there is more than one precursor that originates the exact same mature miRNA, have been reported. This is also in agreement with the miRNA-loss hypothesis in platyhelminths, mostly in the parasitic ones (Fromm et al., 2013), according to which not only fewer miRNAs have been retained across evolution, but also their copy number is reduced to a minimum.

According to the size of the sequenced libraries (~22-nt long), the other expected main class of sRNAs to be detected was endo-siRNAs. An initial analysis showed that ~10% of the *E. multilocularis* datasets and ~2% in the *E. canadensis* (G7) datasets corresponded to reads antisense to CDS (Fig. 1). Since this overall scenario was obtained by allowing two mismatches to the reference genomes, to assess the presence of reads compatible with endo-siRNAs we retained only those antisense sequences which showed 100% identity to the corresponding genome. By doing this, the percentage of reads antisense to CDS decreased to ~3% and ~0.4% for *E. multilocularis* and *E. canadensis* (G7), respectively, and no characteristic ~22 nt long peak was observed in the overall antisense population. There are many possible reasons for this: (i) siRNAs may not be a main regulatory mechanism, at least in the analysed stage (metacestode), (ii) siRNAs might not have been detected with the library construction protocol used since they might have different 5p ends as reported for *A. suum* where 5p polyphosphate siRNAs were detected (Wang et al., 2011) or (iii) siRNAs in *Echinococcus* spp. might not mainly target coding sequences as has been reported in the murine male-germ line, where endo-siRNAs target mRNAs almost exclusively at the 3p UTR (Song et al., 2011).

In this work we provide detailed information on *Echinococcus* spp. miRNAs, including the first experimental report on *E. multilocularis* miRNA expression as well as a high-throughput characterization of the miRNA expression profiles of the metacestode stage of *E. multilocularis* and *E. canadensis* (G7). The results reported here show that both species profiles display some differentially expressed miRNAs, including a few with a species-specific pattern. Also, we performed a systematic analysis on the genomic context of each identified miRNA and on its post-transcriptional modifications. To our knowledge this is the first report describing mono-uridylation of platyhelminth miRNAs. This information provides a valuable characterization on miRNA expression in the cestode parasites of the genus *Echinococcus*, aetiological agents of echinococcosis, which is the first step in understanding their role in parasite biology and, eventually, in disease establishment and/or progression and their future potential use as drug or diagnostic targets.

Acknowledgements

This work was supported by Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT), Argentina, projects PICT 2010N° 2252 and PICT-CABBIO 2012N° 3044. M.C. was a recipient of a Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET), Argentina, postdoctoral fellowship.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ijpara.2014.12.003.

References

Ameres, S.L., Zamore, P.D., 2013. Diversifying microRNA sequence and function. Nat. Rev. Mol. Cell Biol. 14, 475–488. http://dx.doi.org/10.1038/nrm3611.

Anders, S., Huber, W., 2010. DESeq: differential expression analysis for sequence count data. Genome Biol. 11, R106. http://dx.doi.org/10.1186/gb-2010-11-10-r106.

Aravin, A.A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J., Tuschl, T., 2003. The small RNA profile during *Drosophila melanogaster* development. Dev. Cell. http://dx.doi.org/10.1016/S1534-5807(03)00228-4.

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

Bartel, D.P., 2009. MicroRNAs: target recognition and regulatory functions. Cell 136, 215–233. http://dx.doi.org/10.1016/j.cell.2009.01.002.

Boulias, K., Horvitz, H.R., 2012. The *C. elegans* MicroRNA mir-71 acts in neurons to promote germline-mediated longevity through regulation of DAF-16/FOXO. Cell Metab. 15, 439–450. http://dx.doi.org/10.1016/j.cmet.2012.02.014.

Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B., Cohen, S.M., 2003. Bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in *Drosophila*. Cell 113, 25–36.

Cai, P., Hou, N., Piao, X., Liu, S., Liu, H., Yang, F., Wang, J., Jin, Q., Wang, H., Chen, Q., 2011. Profiles of small non-coding RNAs in *Schistosoma japonicum* during development. PLoS Negl. Trop. Dis. 5, e1256. http://dx.doi.org/10.1371/journal.pntd.0001256.

Campo-Paysaa, F., Semon, M., Cameron, R.A., Peterson, K.J., Schubert, M., 2011. MicroRNA complements in deuterostomes: origin and evolution of microRNAs. Evol. Dev. 13, 15–27. http://dx.doi.org/10.1111/j.1525-142X.2010.00452.x.

Chen, C., Ridzon, D.A., Broomer, A.J., Zhou, Z., Lee, D.H., Nguyen, J.T., Barbisin, M., Xu, N.L., Mahuvakar, V.R., Andersen, M.R., Lao, K.Q., Livak, K.J., Guegler, K.J., 2005. Real-time quantification of microRNAs by stem-loop RT-PCR. Nucleic Acids Res. 33, e179. http://dx.doi.org/10.1093/nar/gni178.

Cucher, M., Prada, L., Mourglia-Ettlin, G., Dematteis, S., Camicia, F., Asurmendi, S., Rosenzvit, M., 2011. Identification of *Echinococcus granulosus* microRNAs and their expression in different life cycle stages and parasite genotypes. Int. J. Parasitol. 41, 439–448.

De Souza Gomes, M., Muniyappa, M.K., Carvalho, S.G., Guerra-Sa, R., Spillane, C., 2011. Genome-wide identification of novel microRNAs and their target genes in the human parasite *Schistosoma mansoni*. Genomics 98, 96–111. http://dx.doi.org/10.1016/j.ygeno.2011.05.007.

De Wit, E., Linsen, S.E.V., Cuppen, E., Berezikov, E., 2009. Repertoire and evolution of miRNA genes in four divergent nematode species. Genome Res. 19, 2064–2074. http://dx.doi.org/10.1101/gr.093781.109.

Farazi, T.A., Brown, M., Morozov, P., ten Hoeve, J.J., Ben-Dov, I.Z., Hovestadt, V., Hafner, M., Renwick, N., Mihailovic, A., Wessels, L.F., Tuschl, T.A., 2012. Bioinformatic analysis of barcoded cDNA libraries for small RNA profiling by next-generation sequencing. Methods 58, 171–187. http://dx.doi.org/10.1016/j.ymeth.2012.07.020.

Fernandez-Valverde, S.L., Taft, R.J., Mattick, J.S., 2010. Dynamic isomiR regulation in *Drosophila* development. RNA 16, 1881–1888. http://dx.doi.org/10.1261/rna.2379610.

Friedländer, M.R., Chen, W., Adamidi, C., Maaskola, J., Einspanier, R., Knespel, S., Rajewsky, N., 2008. Discovering microRNAs from deep sequencing data using miRDeep. Nat. Biotechnol. 26, 407–415. http://dx.doi.org/10.1038/nbt1394.

Friedländer, M.R., Adamidi, C., Han, T., Lebedeva, S., Isenbarger, T.A., Hirst, M., Marra, M., Nusbaum, C., Lee, W.L., Jenkin, J.C., Sánchez Alvarado, A., Kim, J.K., Rajewsky, N., 2009. High-resolution profiling and discovery of planarian small RNAs. Proc. Natl. Acad. Sci. U.S.A. 106, 11546–11551.

Friedländer, M.R., MacKowiak, S.D., Li, N., Chen, W., Rajewsky, N., 2012. MiRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res. 40, 37–52. http://dx.doi.org/10.1093/nar/gkr688.

Fromm, B., Worren, M.M., Hahn, C., Hovig, E., Bachmann, L., 2013. Substantial loss of conserved and gain of novel MicroRNA families in flatworms. Mol. Biol. Evol. 30, 2619–2628. http://dx.doi.org/10.1093/molbev/mst155.

Godnic, I., Zorc, M., Jevsinek Skok, D., Calin, G.A., Horvat, S., Dovc, P., Kovac, M., Kunej, T., 2013. Genome-wide and species-wide in silico screening for intragenic microRNAs in human, mouse and chicken. PLoS One 8, e65165. http://dx.doi.org/10.1371/journal.pone.0065165.

Griffiths-Jones, S., Saini, H.K., Van Dongen, S., Enright, A.J., 2008. MiRBase: tools for microRNA genomics. Nucleic Acids Res. 36, D154–D158. http://dx.doi.org/10.1093/nar/gkm952.

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.

Hsieh, Y.W., Chang, C., Chuang, C.F., 2012. The MICRORNA mir-71 inhibits calcium signaling by targeting the TIR-1/Sarm1 adaptor protein to control stochastic L/R neuronal asymmetry in *C. elegans*. PLoS Genet. 8, e1002864. http://dx.doi.org/10.1371/journal.pgen.1002864.

Huang, J., Hao, P., Chen, H., Hu, W., Yan, Q., Liu, F., Han, Z.G., 2009. Genome-wide identification of *Schistosoma japonicum* microRNAs using a deep-sequencing approach. PLoS One 4, e8206. http://dx.doi.org/10.1371/journal.pone.0008206.

Hutvágner, G., McLachlan, J., Pasquinelli, A.E., Bálint, E., Tuschl, T., Zamore, P.D., 2001. A cellular function for the RNA-interference enzyme dicer in the maturation of the let-7 small temporal RNA. Science 293, 834–838. http://dx.doi.org/10.1126/science.1062961.

Isik, M., Korswagen, H.C., Berezikov, E., 2010. Expression patterns of intronic microRNAs in *Caenorhabditis elegans*. Silence 1, 5. http://dx.doi.org/10.1186/1758-907X-1-5.

Jin, X., Lu, L., Su, H., Lou, Z., Wang, F., Zheng, Y., Xu, G.T., 2013. Comparative analysis of known miRNAs across platyhelminths. FEBS J. 280, 3944–3951. http://dx.doi.org/10.1111/febs.12395.

Kato, M., De Lencastre, A., Pincus, Z., Slack, F.J., 2009. Dynamic expression of small non-coding RNAs, including novel microRNAs and piRNAs/21U-RNAs, during *C. elegans* development. Genome Biol. 10, R54. http://dx.doi.org/10.1186/gb-2009-10-5-r54.

Kim, Y.K., Heo, I., Kim, V.N., 2010. Modifications of Small RNAs and their associated proteins. Cell 143, 703–709. http://dx.doi.org/10.1016/j.cell.2010.11.018.

Koziol, U., Rauschendorfer, T., Zanon Rodríguez, L., Krohne, G., Brehm, K., 2014. The unique stem cell system of the immortal larva of the human parasite *Echinococcus multilocularis*. Evodevo 5, 10. http://dx.doi.org/10.1186/2041-9139-5-10.

Kozomara, A., Griffiths-Jones, S., 2014. MiRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 42, D68–D73. http://dx.doi.org/10.1093/nar/gkt1181.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., Marra, M.A., 2009. Circos: an information aesthetic for comparative genomics. Genome Res. 19 (9), 1639–1645. http://dx.doi.org/10.1101/gr.092759.109.

Langmead, B., Trapnell, C., Pop, M., Salzberg, S., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25. http://dx.doi.org/10.1186/gb-2009-10-3-r25.

Lau, N.C., Lim, L.P., Weinstein, E.G., Bartel, D.P., 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. Science 294, 858–862. http://dx.doi.org/10.1126/science.1065062.

Lu, Y., Smielewska, M., Palakodeti, D., Lovci, M.T., Aigner, S., Yeo, G.W., Graveley, B.R., 2009. Deep sequencing identifies new and regulated microRNAs in *Schmidtea mediterranea*. RNA 1483–1491. http://dx.doi.org/10.1261/rna.1702009.

Macchiaroli, N., Cucher, M., Zarowiecki, M., Maldonado, L., Kamenetzky, L., Rosenzvit M.C., 2015. microRNA profiling in the zoonotic parasite *Echinococcus canadensis* using a highthroughput approach. Parasit. Vectors, 8, 83. http://dx.doi.org/10.1186/s13071-015-0686-8.

Mallatt, J., Craig, C.W., Yoder, M.J., 2012. Nearly complete rRNA genes from 371 Animalia: updated structure-based alignment and detailed phylogenetic analysis. Mol. Phylogenet. Evol. 64, 603–617. http://dx.doi.org/10.1016/j.ympev.2012.05.016.

Mansfield, J.H., McGlinn, E., 2012. Evolution, expression, and developmental function of Hox-embedded miRNAs. Curr. Top. Dev. Biol. 99, 31–57. http://dx.doi.org/10.1016/B978-0-12-387038-4.00002-1 [].

Marco, A., Kozomara, A., Hui, J.H.L., Emery, A.M., Rollinson, D., Griffiths-Jones, S., Ronshaugen, M., 2013. Sex-biased expression of microRNAs in *Schistosoma mansoni*. PLoS Negl. Trop. Dis. 7, e2402. http://dx.doi.org/10.1371/journal.pntd.0002402.

Mehlhorn, H., Eckert, J., Thompson, R.C., 1983. Proliferation and metastases formation of larval *Echinococcus multilocularis*. II. Ultrastructural investigations. Z. Parasitenkd. 69, 749–763.

Mizukami, C., Spiliotis, M., Gottstein, B., Yagi, K., Katakura, K., Oku, Y., 2010. Gene silencing in *Echinococcus multilocularis* protoscoleces using RNA interference. Parasitol. Int. 59, 647–652. http://dx.doi.org/10.1016/j.parint.2010.08.010.

Okamura, K., Phillips, M.D., Tyler, D.M., Duan, H., Chou, Y., Lai, E.C., 2008. The regulatory activity of microRNA* species has substantial influence on microRNA and 3′ UTR evolution. Nat. Struct. Mol. Biol. 15, 354–363. http://dx.doi.org/10.1038/nsmb.1409.

Palakodeti, D., Smielewska, M., Graveley, B.R., 2006. MicroRNAs from the Planarian *Schmidtea mediterranea*: a model system for stem cell biology. RNA 12, 1640–1649. http://dx.doi.org/10.1261/rna.117206.to.

Parkinson, J., Wasmuth, J.D., Salinas, G., Bizarro, C.V., Sanford, C., Berriman, M., Ferreira, H.B., Zaha, A., Blaxter, M.L., Maizels, R.M., Fernández, C., 2012. A transcriptomic analysis of *Echinococcus granulosus* larval stages: implications for parasite biology and host adaptation. PLoS Negl. Trop. Dis. 6, e1897. http://dx.doi.org/10.1371/journal.pntd.0001897.

Poole, C.B., Davis, P.J., Jin, J., McReynolds, L.A., 2010. Cloning and bioinformatic identification of small RNAs in the filarial nematode, *Brugia malayi*. Mol. Biochem. Parasitol. 169, 87–94. http://dx.doi.org/10.1016/j.molbiopara.2009.10.004.

Ruby, J.G., Stark, A., Johnston, W.K., Kellis, M., Bartel, D.P., Lai, E.C., 2007. Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. Genome Res. 17, 1850–1864. http://dx.doi.org/10.1101/gr.6597907.

Sasidharan, V., Lu, Y.-C., Bansal, D., Dasari, P., Poduval, D., Seshasayee, A., Resch, A.M., Graveley, B.R., Palakodeti, D., 2013. Identification of neoblast- and regeneration-specific miRNAs in the planarian *Schmidtea mediterranea*. RNA 19, 1394–1404. http://dx.doi.org/10.1261/rna.038653.113.

Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., Zamore, P.D., 2003. Asymmetry in the assembly of the RNAi enzyme complex. Cell 115, 199–208. http://dx.doi.org/10.1016/S0092-8674(03)00759-1.

Scott, D.D., Norbury, C.J., 2013. RNA decay via 3′ uridylation. Biochim. Biophys. Acta 1829, 654–665. http://dx.doi.org/10.1016/j.bbagrm.2013.01.009.

Sokol, N.S., 2012. Small temporal RNAs in animal development. Curr. Opin. Genet. Dev. 22, 368–373. http://dx.doi.org/10.1016/j.gde.2012.04.001.

Song, R., Hennig, G.W., Wu, Q., Jose, C., Zheng, H., Yan, W., 2011. Male germ cells express abundant endogenous siRNAs. Proc. Natl. Acad. Sci. U.S.A. 108, 13159–13164. http://dx.doi.org/10.1073/pnas.1108567108.

Spakulova, M., Orosova, M., Mackiewicz, J.S., 2011. Cytogenetics and chromosomes of tapeworms (Platyhelminthes, Cestoda). Adv. Parasitol. 74, 177–230. http://dx.doi.org/10.1016/B978-0-12-385897-9.00003-3.

Spiliotis, M., Brehm, K., 2009. Axenic in vitro cultivation of *Echinococcus multilocularis* metacestode vesicles and the generation of primary cell cultures. Methods Mol. Biol. 470, 245–262. http://dx.doi.org/10.1007/978-1-59745-204-5_17.

Spiliotis, M., Mizukami, C., Oku, Y., Kiss, F., Brehm, K., Gottstein, B., 2010. *Echinococcus multilocularis* primary cells: improved isolation, small-scale cultivation and RNA interference. Mol. Biochem. Parasitol. 174, 83–87. http://dx.doi.org/10.1016/j.molbiopara.2010.07.001.

Stark, A., Brennecke, J., Russell, R.B., Cohen, S.M., 2003. Identification of *Drosophila* microRNA targets. PLoS Biol. 1, E60. http://dx.doi.org/10.1371/journal.pbio.0000060.

Stark, A., Kheradpour, P., Parts, L., Brennecke, J., Hodges, E., Hannon, G.J., Kellis, M., 2007. Systematic discovery and characterization of fly microRNAs using 12 *Drosophila* genomes. Genome Res. 17, 1865–1879. http://dx.doi.org/10.1101/gr.6593807.

Tsai, I.J., Zarowiecki, M., Holroyd, N., Garciarrubio, A., Sanchez-Flores, A., Brooks, K.L., Tracey, A., Bobes, R.J., Fragoso, G., Sciutto, E., Aslett, M., Beasley, H., Bennett, H.M., Cai, J., Camicia, F., Clark, R., Cucher, M., De Silva, N., Day, T.A., Deplazes, P., Estrada, K., Fernández, C., Holland, P.W.H., Hou, J., Hu, S., Huckvale, T., Hung, S.S., Kamenetzky, L., Keane, J.A., Kiss, F., Koziol, U., Lambert, O., Liu, K., Luo, X., Luo, Y., Macchiaroli, N., Nichol, S., Paps, J., Parkinson, J., Pouchkina-Stantcheva, N., Riddiford, N., Rosenzvit, M., Salinas, G., Wasmuth, J.D., Zamanian, M., Zheng, Y., Cai, X., Soberón, X., Olson, P.D., Laclette, J.P., Brehm, K., Berriman, M., 2013. The genomes of four tapeworm species reveal adaptations to parasitism. Nature 496, 57–63. http://dx.doi.org/10.1038/nature12031.

Wang, Z., Xue, X., Sun, J., Luo, R., Xu, X., Jiang, Y., Zhang, Q., Pan, W., 2010. An "in-depth" description of the small non-coding RNA population of *Schistosoma japonicum* schistosomulum. PLoS Negl. Trop. Dis. 4, e596. http://dx.doi.org/10.1371/journal.pntd.0000596.

Wang, J., Czech, B., Crunk, A., Wallace, A., Mitreva, M., Hannon, G.J., Davis, R.E., 2011. Deep small RNA sequencing from the nematode *Ascaris* reveals conservation, functional diversification, and novel developmental profiles. Genome Res. 21, 1462–1477. http://dx.doi.org/10.1101/gr.121426.111.

Winter, A.D., Weir, W., Hunt, M., Berriman, M., Gilleard, J.S., Devaney, E., Britton, C., 2012. Diversity in parasitic nematode genomes: the microRNAs of *Brugia pahangi* and *Haemonchus contortus* are largely novel. BMC Genomics 13, 4. http://dx.doi.org/10.1186/1471-2164-13-4.

Zhang, X., Zabinsky, R., Teng, Y., Cui, M., Han, M., 2011. MicroRNAs play critical roles in the survival and recovery of *Caenorhabditis elegans* from starvation-induced L1 diapause. Proc. Natl. Acad. Sci. 108, 17997–18002. http://dx.doi.org/10.1073/pnas.1105982108.

Zheng, H., Zhang, W., Zhang, L., Zhang, Z., Li, J., Lu, G., Zhu, Y., Wang, Y., Huang, Y., Liu, J., Kang, H., Chen, J., Wang, L., Chen, A., Yu, S., Gao, Z., Jin, L., Gu, W., Wang, Z., Zhao, L., Shi, B., Wen, H., Lin, R., Jones, M.K., Brejova, B., Vinar, T., Zhao, G., McManus, D.P., Chen, Z., Zhou, Y., Wang, S., 2013. The genome of the hydatid tapeworm *Echinococcus granulosus*. Nat. Genet. 45, 1168–1175. http://dx.doi.org/10.1038/ng.2757.