Published in final edited form as:

Gene. 2009 December 15; 448(2): 187–191. doi:10.1016/j.gene.2009.06.019.

Marsupial-specific microRNAs evolved from marsupial-specific transposable elements

Eric J. Devor^{a,*}, Andrew S. Peek^{a,1}, William Lanier^b, and Paul B. Samollow^c

^a Molecular Genetics and Bioinformatics, Integrated DNA Technologies, Coralville, IA 52241, USA

^b Interdisciplinary Program in Genetics, University of Iowa, Iowa City, IA 52242, USA

^c Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX 77843-4461, USA

Abstract

Using a direct miRNA cloning strategy we previously identified fourteen marsupial- or species-specific microRNAs in the marsupial species *Monodelphis domestica*. In the present study we examined each of the pre-miRNAs and their flanking sequences and demonstrate that half of these miRNAs evolved from marsupial-specific transposable elements. These findings reinforce the view that transposable elements are a previously unappreciated source of new, lineage-specific microRNAs.

1. Introduction

MicroRNAs (miRNAs) are a class of small (21–24nt) regulatory RNAs whose impact on both comparative and functional genomics is only just being felt. First discovered in 1993 (Lee et al., 1993; Wightman et al., 1993), the number of miRNAs known in animal, plant, and viral genomes has grown into the thousands (Griffiths-Jones et al., 2008) and knowledge regarding their roles in regulating cellular processes, including virtually all aspects of development, differentiation, and, even cell death, continues to expand rapidly (Bartel, 2009; Kim et al., 2009). However, many questions regarding the origin and evolution of miRNAs remain unanswered. For example, while it is becoming clear that miRNAs arise from a variety of sources, including inverted duplications (Voinnet, 2004; Piriyaongsa and Jordan, 2007), pseudogenes (Devor, 2006), and transposable elements (TEs) (Piriyaongsa et al., 2007), the relative importance of these sources in generating new miRNA loci remains unclear. In this paper we focus on data suggesting that the latter of these sources may be more important than previously recognized.

Several studies (e.g., Smalheiser and Torvik, 2005, 2006; Borchert et al., 2006; Piriyaongsa et al., 2007; Piriyaongsa and Jordan, 2008) have shown that miRNAs can and do evolve from TEs and it has been suggested that TEs are an unappreciated source of new, lineage-specific miRNAs (Lanier et al., in preparation). We carried out an extensive *in vitro* and *in silico* survey of the miRNAome of the marsupial mammal *Monodelphis domestica* (Devor and Samollow, 2008; Devor and Peek, 2008). In addition to cloning and mapping 174 miRNAs that are conserved throughout the Mammalia, we identified fourteen miRNAs that are unique to marsupials including three miRNAs that are unique to *M. domestica*. Here, we show that half

*Corresponding author. Department of Obstetrics and Gynecology, University of Iowa Carver College of Medicine, 3234 MERF, Iowa City, IA 52242, USA. Tel.: +1 319 335 8212. eric-devor@uiowa.edu (E.J. Devor).

¹Present address: Roche Molecular Systems, 4300 Hacienda Drive, Pleasanton, CA 94588.

of these lineage-specific miRNAs contain TE signatures and that these signatures are themselves specific to marsupials. Thus, we provide further evidence supporting the view that TEs are a source of lineage-specific miRNAs.

2. Materials and methods

The initial *in vitro* and *in silico* survey of *M. domestica* miRNAs was carried out by direct miRNA cloning from six tissues (brain, heart, lung, liver, kidney, and testes) using the miRCat™ small RNA cloning method (Integrated DNA Technologies) coupled with queries of the completed *M. domestica* genome assembly MonDom5 (Ensembl; Mikkelsen et al., 2007). Marsupial-specific miRNAs were all found by *in vitro* cloning and were validated using MonDom5 to obtain sequences flanking the mature miRNA, the RNA folding package mFOLD (Zuker, 2003) to confirm the presence of a thermodynamically stable pre-miRNA hairpin structure, and miRBase (Griffiths-Jones et al., 2008) to confirm that the pre-miRNA had not been seen in any other species.

The search for TE signatures in these marsupial-specific miRNAs was carried out by submitting the pre-miRNAs, plus 3 kb of upstream and 3 kb of downstream flanking sequence obtained from the MonDom5 genome assembly, to both RepeatMasker (Smit, et al., 1996–2004) and CENSOR (Jurka et al., 2005; Kohany et al., 2006). The TE identifications provided by these two sources were then used to annotate the immediate genomic region around each of the miRNAs. Very short nucleotide runs identified in an miRNA precursor as matching a TE fragment would not be counted as a legitimate TE signature. However, no such short fragments were seen as the shortest miRNA precursor sequence that was masked accounted for more than 40% of one precursor.

3. Results and discussion

Descriptions of the fourteen marsupial-specific miRNAs are provided in Table 1. Twelve of these fourteen miRNAs have been assigned miRBase identification numbers while two await assignment. Eleven of the fourteen miRNAs map to intergenic regions of the *M. domestica* genome while two others, *Mdo-miR-1547* and *Mdo-miR-1548*, map to introns of predicted protein coding genes. The unassigned *Mdo-301* microRNA is located in the 5' UTR of the *TCF12* transcription factor locus. Eleven of the fourteen miRNAs have been identified by us in the genome of the tammar wallaby (*Macropus eugenii*) whereas three, *Mdo-miR-1541*, *Mdo-miR-1544*, and *Mdo-miR-1545*, have so far been found only in the *M. domestica* genome (Table 1). Importantly, seven of the fourteen miRNAs are completely masked (*Mdo-miR-1542-1*, *Mdo-miR-1542-2*, *Mdo-miR-1544*, *Mdo-miR-1545*, and *Mdo-miR-340*) or partially masked (*Mdo-miR-1546* and *Mdo-miR-1547*) by marsupial-specific TE sequence signatures (Table 1). Six of these TE sequence signatures were identified as LINEs and one was identified as a Mariner DNA transposon.

The three microRNAs that appear so far to be unique to *M. domestica* represent extremes of TE annotations. *Mdo-miR-1541* is located in the q-terminal region of chromosome 4 and lies in a sequence extending nearly 2 kb in each direction for which there is no evidence of TE signatures. In contrast, *Mdo-miR-1544* and *Mdo-miR-1545* are members of a cluster of 39 microRNAs spanning a 100 kb region of the X-chromosome that we have previously shown to form two distinct clades descended from a single common ancestor sequence via a series of duplication events (Fig. 1A and Devor and Samollow, 2008). When the entire 103,000 bp region was searched for TE signatures, nearly half (18 of 39) of the precursors were completely or partially masked including the two cloned members that we have designated *Mdo-miR-1544a* and *Mdo-miR-1545a* (Fig. B). All 39 members of this cluster, whether masked or 1 not, are flanked upstream by the same 5' UTR remnants of a marsupial L1 TE and downstream

by another marsupial L1 5' UTR remnant in the opposite orientation. Similar juxtapositions of LINEs and other transposed elements have been shown to produce microRNAs in primate genomes (Smalheiser and Torvik, 2005). The common ancestor of both the *Mdo-miR-1544* and *Mdo-miR-1545* clades thus appears to have evolved from the opposition of two marsupial L1 transposed elements and the entire cluster then diversified via a series of duplications with many of the individual members still retaining sufficient ancestral sequence identity to be masked.

A second *M. domestica* locus in which one microRNA is a duplicate of another is the *Mdo-miR-1542-1/Mdo-miR-1542-2* pair located on chromosome 8. Annotation of the sequence containing these two miRNAs showed a tandem arrangement of six copies of a fragment of the 5' UTR of a marsupial L2 transposed element. The two microRNAs, separated by 124 bp, are found completely within this set of tandem duplications (Fig. 2). Various alignments of the L2 remnant duplications indicate that *miR-1542-2* evolved first from TE fragments and that *miR-1542-1* and its immediate flanking region arose by a subsequent tandem duplication event. The two miRNAs have the same mature sequence but differ at nine of the remaining 68 positions of the precursor resulting in slightly different hairpin structures (Fig. 2). This locus is almost completely conserved in the *M. eugenii* (tammar wallaby) genome. There is a single-base difference between the two species in *miR-1542-1* and no differences in *miR-1542-2*. The immediate flanking and intervening masked sequences are 83% identical.

The 3' halves of two miRNAs, *Mdo-miR-1546*, located on chromosome 3, and *Mdo-miR-1547*, located on chromosome 4, were each found to be masked by our analysis. The last 38 bases (42%) of its 89 base precursor of *Mdo-miR-1546* were masked as a sequence identified as the 3' UTR of a marsupial L2 transposed element. For *Mdo-miR-1547*, the last 49 bases (54%) of its 91 base precursor were masked as a sequence identified as the 3' end of marsupial RTE1 LINE. The similarity value for both of the masked sequences is ~0.75 and the transposed sequences all continue well past the 3' ends of the precursors. Both microRNAs are conserved between the genomes of *M. domestica* and the tammar wallaby with two base differences seen between *Mdo-miR-1546* and *Meu-miR-1546* and between *Mdo-miR-1547* and *Meu-miR-1547*. In none of these four cases does the single-base difference occur in a region of the precursor crucial for target specificity. Two of the substitutions lie near the termini of the stems, one substitution lies within the *miR-1547* loop, and another lies at position 19 of the mature *miR-1546* sequence.

The final masked *M. domestica* microRNA is *Mdo-miR-340*. Initially thought by us to be marsupial-specific, this microRNA is now assigned to the *miR-340* family which is found throughout the eutherian (placental) mammals on the basis of precursor and mature sequence similarities that were not immediately obvious upon initial examination. Annotation of *Mdo-miR-340* indicates that its origin is in the pairing of two nearly identical remnants of a Mariner DNA transposon lying in opposite orientations (Fig. 3A). This juxtaposition, which could have resulted from an inverted duplication, yields a highly stable hairpin ($\Delta G = -41.1$ kcal). The tammar wallaby ortholog shows only 86.6% identity with the *M. domestica* ortholog over the full length of the precursor and a substantially lower hairpin stability ($\Delta G = -31.9$ kcal); nevertheless the mature sequence differs only in position 19 of 22. Interestingly, when we used *miR-340* precursor sequences from human, chimpanzee, rhesus macaque, mouse, rat, cow, and dog, archived in miRBase (Release 12.0), to screen GenBank TRACE Archives for additional marsupial and placental sequences we did not obtain any additional marsupial examples, but we did identify *miR-340* precursor sequences for a further seventeen placental mammal species including a perissodactyl (horse), a cetacean (dolphin), two chiroptera (brown bat and fruit bat), two lagomorphs (rabbit and pika), armadillo, and tenrec. An alignment of *miR-340* precursors of these species with opossum and tammar wallaby is shown in Fig. 3B. Precursor sequences among the placental mammal species show the effects of purifying selection

expected for microRNAs while the two marsupials are not only quite dissimilar to each other but even more so to the placental mammals. This raises the question as to whether or not marsupial *miR-340* actually is orthologous to placental *miR-340*. Three observations bear directly on this question. First, the mature miRNA cloned from *M. domestica* is different from that cloned in the placental mammals. Recently, Wheeler et al. (2009) discussed a phylogenetic miRNA phenomenon they term “seed shift” in which the mature sequence is seen to shift 1–2nt in either the 5′ or 3′ direction thereby changing the seed at positions 2–8. These seed shifts are phylogenetically conserved between two or more taxa. The five base shift we observe in *miR-340* would clearly be an extreme example of this phenomenon. Second, the 3′ stem sequence in placental *miR-340* precursors retains insufficient similarity to the Mariner transposed element seen in the marsupial sequences to be identified as such in any of the twenty-four species in which we have found it. Sequence differences of this magnitude have not been observed in any of the other microRNAs conserved among the Mammalia (Devor and Peek, 2008). Third, the context of *miR-340* in the *M. domestica* genome is different from that in the placental mammals. Among the placental mammals, *miR-340* lies in intron 2 of the ring finger protein gene *RNF130*. In the *M. domestica* genome, *miR-340* maps near the p-terminal end of chromosome 1 whereas *RNF130* maps near the middle of the q-arm more than 300 Mb distant. Moreover, the region in which *Mdo-miR-340* is found is not syntenic with the location of *RNF130* on human chromosome 5 but *M. domestica* *RNF130* is. No other conserved miRNA in the *M. domestica* genome lacks synteny with its ortholog in the human genome (Devor and Samolow, 2008). The most parsimonious explanation for the phylogenetic divergence pattern seen here for *miR-340* would be that all *miR-340*s are descended from a common ancestor and that, following metatherian–eutherian divergence, there was rapid divergence of the mature sequence resulting in changes in the fold back sequence that were subsequently preserved under purifying selection. However, the observations presented here appear to favor a less parsimonious hypothesis. The data presented suggest that the miRNAs identified as *miR-340* in the metatherian and eutherian lineages are in fact different microRNAs that coincidentally evolved from the same sequence of the same TE. Thus, *Mdo-miR-340* is indeed a lineage-specific miRNA.

4. Conclusions

Smalheiser and Torvik (2005, 2006), Borchert et al. (2006), Piriyaongsa et al. (2007), and Lanier et al. (in preparation) have all presented evidence that microRNAs can evolve from sequences of transposed elements. Moreover, both Piriyaongsa et al. (2007) and Lanier et al. (in preparation) suggest that transposed elements are an unappreciated source of lineage-specific microRNAs. We have presented evidence herein that half of the lineage-specific microRNAs that have been discovered in the marsupial species *M. domestica* show evidence of lineage-specific TE origin. This evidence further strengthens the position that TEs are indeed fertile ground for the emergence of new microRNAs.

Acknowledgments

Portions of this paper were presented at the 2nd International Conference and Workshop “Genomic Impact of Eukaryotic Transposable Elements” Asilomar, California February 6–9, 2009. Funding for this work was provided in part by the National Institutes of Health Grant RR014214 to PBS.

Abbreviations

pre-miRNA	microRNA precursor
miRNA	microRNA
TE	transposable element

kb	kilobase
bp	base pair
Mdo	<i>Monodelphis domestica</i>
miR	microRNA locus
ΔG	folding free energy
kcal	kilocalories

References

- Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell* 2009;136:215–233. [PubMed: 19167326]
- Borchert GM, Lanier W, Davidson BL. RNA polymerase III transcribes human microRNAs. *Nat Struct Mol Biol* 2006;13:1097–1101. [PubMed: 17099701]
- Devor EJ. Primate microRNAs *miR-220* and *miR-492* lie within processed pseudogenes. *J Heredity* 2006;97:186–190.
- Devor EJ, Peek AS. miRNA profile of a Triassic common mammalian ancestor and pre-miRNA evolution in the three mammalian reproductive lineages. *Open Genomics J* 2008;1:22–32.
- Devor EJ, Samollow PB. *In vitro* and *in silico* annotation of conserved and nonconserved microRNAs in the genome of the marsupial *Monodelphis domestica*. *J Heredity* 2008;99:66–72.
- Griffiths-Jones S, Saini HK, Dongen SV, Enright AJ. miRBase: tools for microRNA genomics. *Nucleic Acids Res* 2008;36:D154–D158. [PubMed: 17991681]
- Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 2005;110:462–467. [PubMed: 16093699]
- Kim VN, Han J, Siomi MC. Biogenesis of small RNAs in animals. *Nat Rev, Mol Cell Biol* 2009;10:126–139. [PubMed: 19165215]
- Kohany O, Gentles AJ, Hankus L, Jurka J. Annotation, submission, and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics* 2006;7:474. [PubMed: 17064419]
- Lanier W, Tyra HM, Devor EJ, Borchert G, Comeron J, Bhattacharya D. From simple beginnings: the evolution of miRNAs from transposable elements in *Chlamydomonas reinhardtii*. *PLoS Genetics*. in preparation
- Lee RC, Feinbaum RL, Ambros V. The *C. Elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 1993;75:843–854. [PubMed: 8252621]
- Mikkelsen TS, Wakefield MJ, Aken B, et al. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* 2007;447:167–177. [PubMed: 17495919]
- Piriyapongsa J, Jordan IK. A family of human microRNA genes from inverted-repeat transposable elements. *PLoS ONE* 2007;2 (2):e203. [PubMed: 17301878]
- Piriyapongsa J, Jordan IK. Dual coding of siRNAs and miRNAs by plant transposable elements. *RNA* 2008;14:814–821. [PubMed: 18367716]
- Piriyapongsa J, Marino-Ramirez L, Jordan IK. Origin and evolution of human microRNAs from transposable elements. *Genetics* 2007;176:1323–1337. [PubMed: 17435244]
- Smalheiser NR, Torvik VI. Mammalian microRNAs derived from genomic repeats. *Trends Genet* 2005;21:322–326. [PubMed: 15922829]
- Smalheiser NR, Torvik VI. Alu elements within human mRNAs are probable microRNA targets. *Trends Genet* 2006;22:532–536. [PubMed: 16914224]
- Smit AFA, Hubley R, Green P. RepeatMasker Open-3.0. 1996–2004
- Voinnet O. Shaping small RNAs in plants by gene duplication. *Nat Genet* 2004;36:1245–1246. [PubMed: 15565102]
- Wheeler BM, et al. The deep evolution of metazoan microRNAs. *Evol Dev* 2009;11:50–68. [PubMed: 19196333]

- Wightman B, Ha I, Ruvkun G. Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* 1993;75:843–854. [PubMed: 8252621]
- Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 2003;31:3406–3415. [PubMed: 12824337]

Fig. 1.

(A) Neighbor-joining tree for the thirty-nine members of the *Mdo-miR-1544/Mdo-miR-1545* X-chromosome cluster. Separation between the *Mdo-miR-1544* clade and the *Mdo-miR-1545* clade is clearly seen. (B) A clustal W alignment of the precursor sequences of all thirty-nine members of the *Mdo-miR-1544/Mdo-miR-1545* X-chromosome cluster. Nucleotides annotated as transposon in each member of the cluster are shown in lower case. Mature sequences are indicated by bold underline.

Fig. 2.

Genomic context and annotated hairpin sequences of microRNAs *Mdo-miR-1542-2* and *Mdo-miR-1542-1*. (A) Orientation and distance between *Mdo-miR-1542-2* and *Mdo-miR-1542-1* is shown above the line and the annotated transposon sequences are shown below the line. The first five sequences are all from the same fragment of an L2 LINE. The position identifications are, from the left, 3323–3447, 3381–3452, 3384–3448, 3379–3451, and 3389–3444. (B) Precursor hairpins of *Mdo-miR-1542-2* and *Mdo-miR-1542-1* are shown with TE sequences shown in lower case and the mature sequences underlined.

Fig. 3.
 (A) Hairpin structure of the *Mdo-miR-340* microRNA showing the location, orientation, and identity of the Mariner transposon fragments forming the stem-loop. (B) Clustal W alignment of *miR-340* in twenty-six mammalian species. Annotated sequences are presented in lower case and the mature miRNAs are bolded. Also shown in bold are *miR-340** sequences recorded as validated in miRBase.

Table 1

Fourteen marsupial-specific microRNAs.

microRNA	miRBaseID	5' Mature miR Seq 3'	Chromosome coordinates ^a	Genomic location	Wallaby ortholog	TE signature
Mdo-miR-1540	MI0007263	ugaauccaagagcgaugu	Chr.1:580149275-580149333	Intergenic	Yes	No
Mdo-miR-1541	MI0007264	ugguugcucguuuggaugugg	Chr.4:400891413-400891482	Intergenic	No	No
Mdo-miR-1542-1	MI0007265	uaugaucuccaauagccuagc	Chr.8: 87784786-87784857	Intergenic	Yes	L2 LINE
Mdo-miR-1542-2	MI0007266	uaugaucuccaauagccuagc	Chr.8: 87784982-87785058	Intergenic	Yes	L2 LINE
Mdo-miR-1543	MI0007267	uauguccuagucugugcac	Chr.1:330615486-330615566	Intergenic	Yes	No
Mdo-miR-1544a- ^b	MI0007269	acuuccaucccuugcacugu	Chr. X: 14003847-14003918	Intergenic	No	L1 LINE
Mdo-miR-1545a- ^b	MI0007270	ugcaccagggauggaugagcg	Chr. X: 14005426-14005511	Intergenic	No	L1 LINE
Mdo-miR-1546	MI0007271	ucagggaucucagggauggaa	Chr.3:518067372-518067460	Intergenic	Yes	L2 LINE
Mdo-miR-1547	MI0007272	uaucagagucucugggucuccuu	Chr.4:335163695-335163785	Intron	Yes	RTE1
Mdo-miR-1548	MI0007273	ugcauccgcagcgggcuccccu	Chr.2:541547258-541547317	Intron	Yes	No
Mdo-miR-1549	MI0007274	uuccgcccugcagcccgua	Chr.6:60423748-60423810	Intergenic	Yes	No
Mdo-miR-340	MI0007268	aaguaagagauuauuucugu	Chr.1:72641040-72641094	Intergenic	Yes	Mariner
Mdo-204 ^c		guaacagccacgaugguug	Chr. X: 35734446-35734535	Intergenic	Yes	No
Mdo-301 ^c		uccgcuccgcuugugcuggc	Chr.1:164502968-164503023	5' UTR	Yes	No

^aChromosome coordinates in the MonDom5 genome assembly in Ensembl.

^bChromosome coordinates and mature sequences are given for miR-1544a and miR-1545a.

^cUnassigned in miRBase as of Release 13.0.