miRNA Profile of a Triassic Common Mammalian Ancestor and Pre-miRNA Evolution in the Three Mammalian Reproductive Lineages

Eric J. Devor[*,1] and Andrew S. Peek[2]

[1]Molecular Genetics and [2]Bioinformatics, Integrated DNA Technologies, 1710 Commercial Park, Coralville, Iowa 52241, USA

Abstract: MicroRNAs (miRNAs) are major factors in the regulation of gene expression. Recent evolutionary studies of miRNAs indicate that important biological innovations, such as the advent of bilateral symmetry and placental reproduction, are accompanied by bursts of miRNA creation which are subsequently conserved *via* purifying selection. The emergence of eutherian (placental) mammals followed by as much as fifty million years the appearance of the first true mammals in the late Triassic, some 230 million years ago. We have utilized microRNA inventories of eutherian, metatherian (marsupial), monotreme (platypus), and chicken genomes to assemble a minimal microRNA profile of the last common ancestor of all mammals consisting of 171 miRNAs. This profile suggests that the rise of placental reproduction launched a more than three-fold expansion of microRNAs. In addition to expansion of the microRNA repertoire, the conserved microRNAs from five mammalian and one avian genome show evidence for conforming to a canonical phylogenetic history as well as dramatic deviations from the assumptions of molecular clock-like rates and the equality of substitution rates among lineages. We also show that many of these basal mammalian miRNAs are highly expressed in eutherian placenta thus creating an opportunity to gain insight on how microRNAs acquire new targets and new functions.

INTRODUCTION

Molecular paleontology, or molecular phylogeny, is the use of comparative genomics on extant species coupled with knowledge of the fossil record to infer molecular characteristics of their extinct ancestors as well as to detail evolutionary events that have occurred between extant species and their last common ancestors. If both sources of information are reliable, there are useful, important, and even unique insights to be had. Nowhere has this been more true than in evolutionary developmental biology, or evo devo for short, in which organism-level development and the genetic factors responsible for it are being traced over hundreds of millions of years [1, 2].

One aspect of development that has emerged over the last few years to assume a crucial position is the role of small, non-coding RNAs, in particular the microRNAs (miRNAs), and this has not escaped the notice of those equipped to venture into the realm of molecular paleontology. Sempere *et al*. [3] presented an evolutionary history of 292 non-paralogous Metazoan microRNAs (miRNAs). They observe a large number of microRNAs arising among certain taxa such as bony fishes and eutherian (placental) mammals which is consistent with other observations that major developmental events, like the advent of bilateral symmetry, the emergence of the vertebrates, and the placental reproduction strategy among mammals, were each accompanied by expansions in the microRNA repertoire of the relevant taxa [4-6]. However, none of these studies was able to take advantage of the non-eutherian mammalian genomes that have recently become available.

We recently reported on an *in silico* and *in vitro* miRNA survey of the marsupial species *Monodelphis domestica* [7]. Here we have updated that survey and have used the updated data to carry out an *in silico* miRNA survey of another marsupial genome, the tammar wallaby (*Macropus eugenii*), and the genome of the platypus (*Ornithorhynchus anatinus*), one of only two monotreme species still extant. These miRNA data, coupled with the most recent miRNA database releases for the human (*Homo sapiens*), mouse (*Mus musculus*), and chicken (*Gallus gallus*) genomes [8], provide us with the opportunity to employ the principles of comparative genomics to simultaneously address two aspects of miRNA evolution. The first of these is an initial attempt to create what could be called a minimal miRNA profile of the last common ancestor of the Mammalia that the paleontologic record suggests lived in the Upper Triassic, some 230 million years ago. This profile consists of 171 miRNAs that are shared among placental, marsupial, and monotreme genomes. The second aspect we have addressed, based upon a subset of the comparative data that includes the chicken miRNAome, is a detailed look at phylogenetic patterns and evolutionary rates not only for the complete miRNA precursor sequences shared among the six species, but also for partitions of those precursors that we have designated by their function within the pre-miRNA hairpin. Results of these analyses indicate that, while overall canonical phylogenetic relationships among the six species are recovered using complete pre-miRNA sequences, partitioned pre-miRNA sequences that comprise non-expressed fold back regions opposite mature miRNAs and hairpin sequences that are neither expressed nor in the loop, deviate slightly from canonical. We also find that substitution rates vary considerably between species within hairpin partitions and that only the fold back region displays molecular clock-like behavior.

*Address correspondence to this author at the Integrated DNA Technologies, 1710 Commercial Park, Coralville, Iowa 52241, USA; Tel: 319-626-8450; E-mail: rdevor@idtdna.com

MATERIALS AND METHODS

An Updated *Monodelphis domestica* miRNA Screen

The *in silico* strategy followed in the original *M. domestica* miRNA screen [7] was also followed here using the most current *M. domestica* genome assembly (MonDom5) and release 10.0 of miRBase. Briefly, pre-miRNA sequences for chicken, mouse, and human were screened in MonDom5 using BLAST. The two criteria used for accepting a hit in the *M. domestica* genome were a minimum 90% length match and a minimum 90% sequence identity over the pre-miRNA sequence. All accepted miRNA matches were then validated by a search of miRBase with the *M. domestica* pre-miRNA sequence to establish identity and a comparison with the miRNA map location in the human genome *via M. domestica/H. sapiens* syntenic maps compiled at the Broad Institute (M. Kamal, personal communication) to establish synteny. We also validated the mature miRNA sequences using a 95% sequence match minimum.

M. domestica-Based *M. Eugenii* and *O. anatinus* miRNA Screens

Once the updated *M. domestica* pre-miRNA file was completed and validated, the entire pre-miRNA sequence file was screened against the current whole genome sequences of the tammar wallaby (*M. eugenii*) in GenBank Trace Archive and the platypus (*O. anatinus*) genome assembly in Ensembl (Oana-5.0). The same criteria used for accepting matches in the opossum genome were applied to the wallaby and platypus genomes when using the opossum pre-miRNA sequences as queries. Validation consisted of a search of candidate pre-miRNAs in miRBase only for these species since the current state of the two genomes, particularly the wallaby genome, does not permit employing the syntenic chromosome mapping criterion as was done for *M. domestica*.

Sequence Alignments, Partition Assignment, Phylogenetic Hypotheses, Rate Tests

We were able to assemble a subset of 94 pre-miRNAs that are shared among the genomes of chicken, platypus, wallaby, opossum, mouse, and human. These pre-microRNA sequences were aligned among genomes using Clustal W [9]. Once aligned, we partitioned the pre-miRNAs based on a scheme that reflects the function of the subsequences within the hairpin structure. Partition 1 consists of only mature miRNA sequences. Partition 2 consists only of that part of the hairpin that is clearly a loop sequence as shown in miRBase for the human miRNA. Partition 3 consists of the fold back part of the hairpin directly opposite the mature miRNA sequence. Finally, Partition 4 consists of any remaining sequence in the hairpin. We made a distinction based, again, on the human miRNA in miRBase between a fold back (Partition 3) and a "star" sequence. Since star sequences lie opposite the mature miRNA and are expressed, at least as far as has been reported in mouse and human miRNAs, and appear to play a role in both miRNA and target-site evolution [10], we placed them in Partition 1. Again, for clarity, we reserve the term "fold back" for Partition 3 sequences to distinguish them from the expressed Partition 1 "star" sequences.

Phylogenetic reconstructions and hypotheses were constructed and tested using the PHYLIP package [11], specifically the programs dnaml and dnamlk. Rates among sites were treated as gamma distributed with an alpha = 0.25 and 4 rate categories, where these parameters were determined empirically from the present data. There are 945 binary rooted trees for 6 taxa and maximum likelihood topologies resulted from 10-fold jumbling the input order of the 6 taxa with global rearrangements and optimization. Bootstrap values are the result from 100 bootstrap replicates from the original dataset, with identical parameter conditions for phylogeny construction. Relative rate tests were performed using RRTree [12] under a Kimura 2-parameter substitution model with probabilities for rejecting the null hypothesis of rate equality calculated by the method of Li and Bousquet [13]. Branch length substitution base rate estimates were confirmed using the routine r8s [14] employing a penalized likelihood method.

RESULTS

An Updated *M. domestica* miRNA Map

Screening MonDom5 with pre-miRNA sequences for chicken, mouse, and human from miRBase 10.0 identified 171 conserved miRNAs in the opossum genome including a few corrections of the previously reported data [7]. These microRNAs now encompass 107 single transcripts, 20 tandem loci, and six polycistronic loci. The complete, updated opossum miRNA file, with chromosome coordinates, is shown in miRBase style in Supplemental File **1** and a corrected and updated chromosome map indicating relative positions is shown in Supplemental File **2**. In addition to adding 50 miRNAs to the opossum miRNA map that were not previously reported [7], there is one observation that merits specific comment. In miRBase, there are two *miR-147* loci in human and chicken and one locus in the other species listed, including mouse and rat. When precursors for the *miR-147s* were screened in MonDom5, four loci were identified. These have been designated in Supplemental File **1** as *mdo-miR-147b1-4*. We aligned the *mdo-miR-147b1-4* pre-miRNAs plus a kilobase (kb) of upstream and downstream flanking sequence from MonDom 5 and observed sequence identities extending well beyond the pre-miRNAs. This is indicative of a larger duplication event but we were unable to detect any unequivocal tell-tale duplication identifiers such as flanking repeats.

Inclusion of *M. eugenii* and *O. anatinus* Genomes

Once validated, all of the opossum pre-miRNA sequences were used to screen the genome of the tammar wallaby in Trace Archives in GenBank, and of the platypus in Ensembl. Of the 171 sequences screened, 164 (95.9%) were found in at least one of the two genomes (Table **1**). Overall, 136 (79.5%) were found in the wallaby and 147 (86.0%) were found in the platypus with 119 (69.6%) found in both genomes. It must be noted again that, since the current status of neither the wallaby nor the platypus genomes is as complete as the opossum genome, the comparative data presented here should not be considered complete for either species. It should also be noted that the seven miRNAs mapped in the opossum genome but not found in either the wallaby or platypus genomes are present in the chicken genome as well as in both mouse and human genomes. Thus, it is likely that these seven miRNAs will be found in other

Table 1.   Conservation of Opossum MicroRNAs

microRNA	Chicken	Platypus	Wallaby
mdo-let-7a-1		X	X
mdo-let-7a-2	X	X	X
mdo-let-7a-3	X	X	X
mdo-let-7b	X	X	X
mdo-let-7d	X	X	X
mdo-let-7f-1	X	X	X
mdo-let-7f-2	X	X	X
mdo-let-7g	X	X	X
mdo-let-7i	X	X	X
mdo-mir-1-1	X	X	X
mdo-mir-1-2	X	X	
mdo-mir-7-1	X	X	X
mdo-mir-9-1	X		X
mdo-mir-9-2	X	X	X
mdo-mir-9-3	X		
mdo-mir-10a	X	X	X
mdo-mir-10b	X	X	X
mdo-mir-15a	X	X	X
mdo-mir-15b	X	X	
mdo-mir-16-1	X	X	X
mdo-mir-16-2	X	X	
mdo-mir-17	X	X	X
mdo-mir-18	X	X	X
mdo-mir-18b	X	X	X
mdo-mir-19a	X	X	X
mdo-mir-19b-1	X	X	X
mdo-mir-19b-2		X	X
mdo-mir-20	X	X	X
mdo-mir-20b	X	X	X
mdo-mir-21	X	X	X
mdo-mir-22	X	X	X
mdo-mir-23a		X	X
mdo-mir-23b	X	X	X
mdo-mir-24-1	X	X	X
mdo-mir-24-2		X	X
mdo-mir-25		X	
mdo-mir-26a-2	X	X	X
mdo-mir-26b		X	X
mdo-mir-27a		X	X
mdo-mir-27b	X	X	X
mdo-mir-29a	X	X	X
mdo-mir-29b-1	X	X	X
mdo-mir-29b-2	X	X	
mdo-mir-29c	X	X	X

(Table 1) contd…..

microRNA	Chicken	Platypus	Wallaby
mdo-mir-30a	X	X	
mdo-mir-30b	X	X	X
mdo-mir-30c-1	X	X	X
mdo-mir-30c-2	X	X	X
mdo-mir-30d	X	X	X
mdo-mir-30e	X	X	X
mdo-mir-31	X	X	X
mdo-mir-32	X	X	X
mdo-mir-34a	X		X
mdo-mir-34b	X	X	X
mdo-mir-34c	X	X	
mdo-mir-92-1	X	X	X
mdo-mir-92a-2		X	X
mdo-mir-92b	X		
mdo-mir-93	X		
mdo-mir-96	X	X	X
mdo-mir-100	X	X	X
mdo-mir-101-1	X	X	X
mdo-mir-101-2	X	X	X
mdo-mir-103-1	X	X	X
mdo-mir-103-2	X	X	X
mdo-mir-106a	X	X	X
mdo-mir-107	X	X	
mdo-mir-122a-2	X	X	X
mdo-mir-124a-1		X	X
mdo-mir-124a-2		X	X
mdo-mir-124a-3	X	X	X
mdo-mir-125b-1		X	X
mdo-mir-125b-2	X	X	X
mdo-mir-126	X	X	X
mdo-mir-128a	X	X	X
mdo-mir-128b		X	X
mdo-mir-129-1		X	X
mdo-mir-129-2	X	X	X
mdo-mir-130a	X	X	X
mdo-mir-130b			X
mdo-mir-132	X		X
mdo-mir-133a-1	X	X	X
mdo-mir-133a-2(c)	X	X	
mdo-mir-133b	X	X	X
mdo-mir-135a-2	X	X	X
mdo-mir-135b	X	X	X
mdo-mir-137	X	X	X
mdo-mir-138-1	X	X	X
mdo-mir-138-2	X	X	X

(Table 1) contd…..

microRNA	Chicken	Platypus	Wallaby
mdo-mir-140	X	X	X
mdo-mir-141	X	X	X
mdo-mir-142		X	X
mdo-mir-143	X	X	X
mdo-mir-144		X	X
mdo-mir-145	X	X	X
mdo-mir-146a	X	X	X
mdo-mir-147b-1	X	X	
mdo-mir-147b-2			
mdo-mir-147b-3			
mdo-mir-147b-4			
mdo-mir-148a		X	X
mdo-mir-148b			X
mdo-mir-152	X	X	X
mdo-mir-153-1	X	X	X
mdo-mir-153-2	X	X	X
mdo-mir-155	X	X	X
mdo-mir-181a-1	X	X	X
mdo-mir-181a-2	X		X
mdo-mir-181b-1	X	X	X
mdo-mir-181b-2	X	X	X
mdo-mir-181c		X	X
mdo-mir-182		X	X
mdo-mir-183		X	X
mdo-mir-184	X	X	X
mdo-mir-186			X
mdo-mir-187	X	X	X
mdo-mir-190	X	X	X
mdo-mir-190b	X	X	X
mdo-mir-191		X	X
mdo-miR-192			X
mdo-mir-193a	X	X	
mdo-mir-193b		X	X
mdo-mir-194-2	X	X	X
mdo-mir-196a-2	X	X	X
mdo-mir-196b	X		X
mdo-mir-199a-2	X	X	X
mdo-mir-199b		X	
mdo-mir-200a	X	X	X
mdo-mir-200b	X	X	X
mdo-mir-200c	X		
mdo-mir-203	X		X
mdo-mir-204	X	X	X
mdo-mir-205a	X	X	X

(Table 1) contd…..

microRNA	Chicken	Platypus	Wallaby
mdo-mir-206	X		X
mdo-mir-208			X
mdo-mir-212		X	
mdo-mir-214	X		X
mdo-mir-215	X		X
mdo-mir-216	X	X	X
mdo-mir-217	X	X	X
mdo-mir-218-1	X		X
mdo-mir-218-2	X		X
mdo-mir-219-2	X		X
mdo-mir-220b		X	X
mdo-mir-221	X	X	X
mdo-mir-222a	X	X	X
mdo-mir-223	X		X
mdo-mir-301a	X	X	X
mdo-mir-302a	X	X	X
mdo-mir-302b	X	X	X
mdo-mir-302c	X	X	X
mdo-mir-302d	X	X	X
mdo-mir-338		X	
mdo-mir-363		X	
mdo-mir-365-2	X	X	X
mdo-mir-367	X	X	X
mdo-mir-375	X	X	X
mdo-mir-383	X		X
mdo-mir-425		X	
mdo-mir-429	X	X	X
mdo-mir-449	X		
mdo-mir-451	X	X	
mdo-mir-455	X		
mdo-mir-490	X		X
mdo-mir-499	X		
mdo-mir-551a		X	
mdo-mir-551b			X
mdo-mir-555		X	
mdo-mir-599			X
mdo-mir-632		X	
mdo-mir-768			X
mdo-mir-801		X	X
mdo-mir-873		X	X
mdo-mir-875			X

BLAST hits for 174 *Monodelphis domestica* precursor microRNAs (pre-miRNAs) in the chicken, platypus, and tammar wallaby genomes. Acceptance of a pre-miRNA hit consists of a minimum 90% identity of the precursor for both sequence composition and length. Note that, with the exception of mdo-miR-147b2-4, all *M. domestica* precursors were similarly validated in miRBase [8] in the mouse and human genomes.

non-eutherian genomes as those sequence annotations continue to improve. It is also likely that a few additional conserved miRNAs will be found but we are confident that this number will not increase the total more than a few percent.

Pre-miRNA sequences from the wallaby and/or platypus genomes, as identified by BLAST, were aligned with the opossum sequences using Clustal W with default parameters [9] in order to assess sequence variation among the three non-eutherian species. As a first pass comparison of these aligned sequences, pre-miRNA sequences were separated into "left arm" and "right arm" precursors based upon the position of the primary mature miRNA sequence relative to the loop of the stem-loop miRNA hairpin, following Berezikov *et al.* [15]. Thus, for this exercise, we did not distinguish between fold back and star designations. Taking the first base of the mature sequence as position zero, all sequence differences between the opossum precursor and the wallaby and platypus precursors were tallied by position without regard to the specifics of the nucleotide differences. The same procedure was then carried out for all 171 opossum precursors compared to their equivalent human and mouse precursors. These position-specific precursor nucleotide changes are shown in Fig. (1). As can be seen, the pattern of position-

by-position sequence differences shows that there is a greater amount sequence variation between opossum and platypus than between opossum and wallaby. There are 135 positions in both left (Fig. **1A**) and right armed (Fig. **1B**) microRNAs where the number of differences between opossum (mdo) and platypus (oan) are greater than between opossum and wallaby (meu), whereas only 15 positions where the reverse is true (H_O: difference(mdo - oan) = difference(meu - oan), sign-test, $P = 1.1 \times 10^{-25}$). It is also seen that the overall pattern of variation between opossum and platypus is quite similar to that observed between opossum and combined mouse and human for the same miRNA precursors (Fig. **1A** and **1B** insets). Two important specific features of these positional sequence difference maps are, that the tendency toward invariant "seed" regions (positions 1-8 in the mature sequences) and the aptly-named "camel" pattern of variation distributed along the length of the precursor are both maintained throughout the Mammalia [15-17].

Phylogenetic Analyses

Differences among the microRNAs of the monotreme, marsupial and eutherian mammals suggested that the rates of
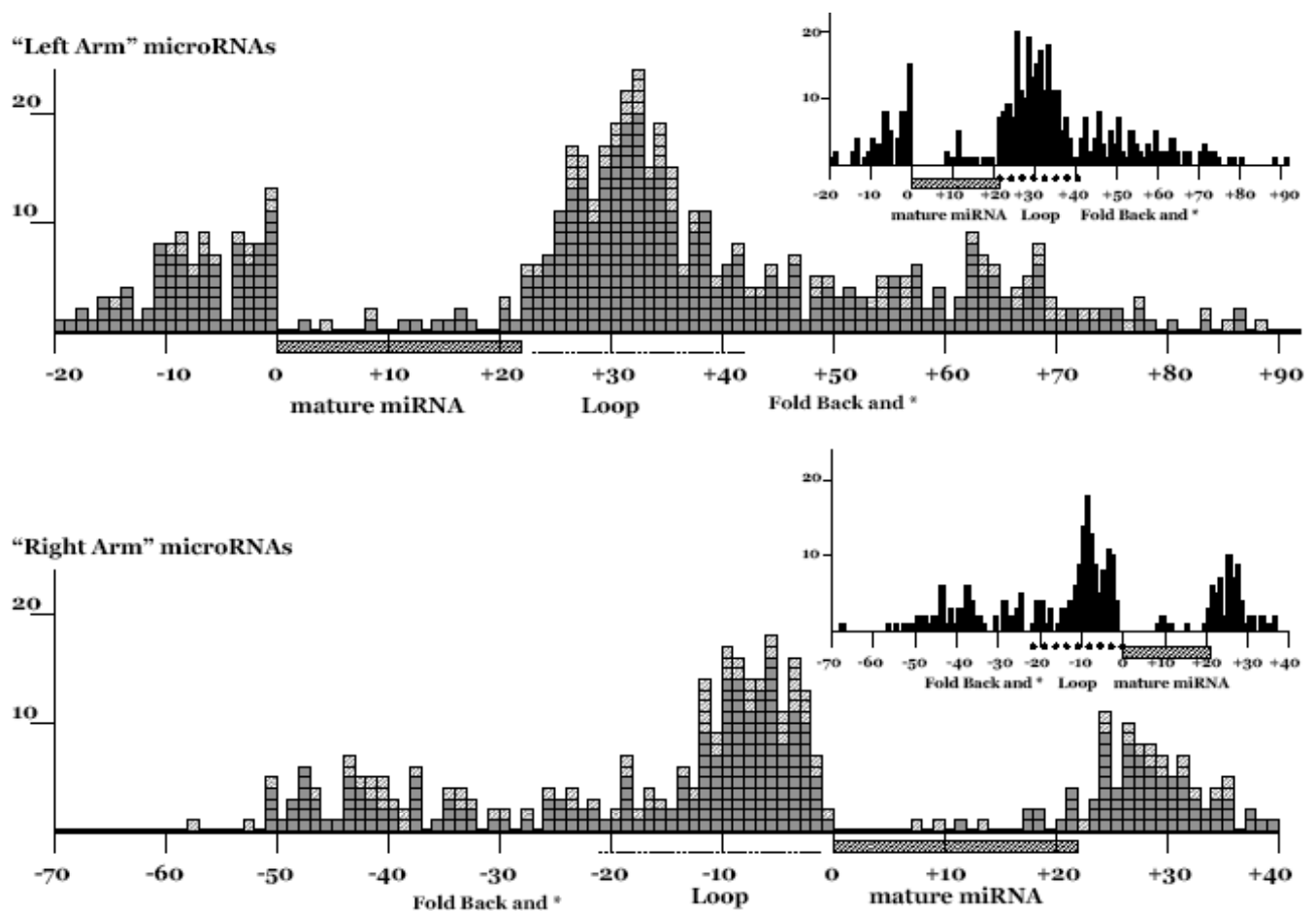
Fig. (1). Position-by-position maps of opossum pre-miRNA base substitutions versus wallaby (cross hatched squares) and platypus (grey squares) orthologs. Pre-miRNAs were classified as "left arm" or "right arm" based upon whether the mature sequence preceded the loop (left) or followed the loop (right) (see [10]). Base substitutions were scored without regard to the transition or transversion specifics of the substitution and indels were scored as single substitutions at the position(s). pre-miRNA partitions are mature sequence, loop, combined fold back and star, and other hairpin sequence. The insets are position-by-position maps of opossum pre-miRNA base substitutions versus mouse and human orthologs for the same miRNAs.
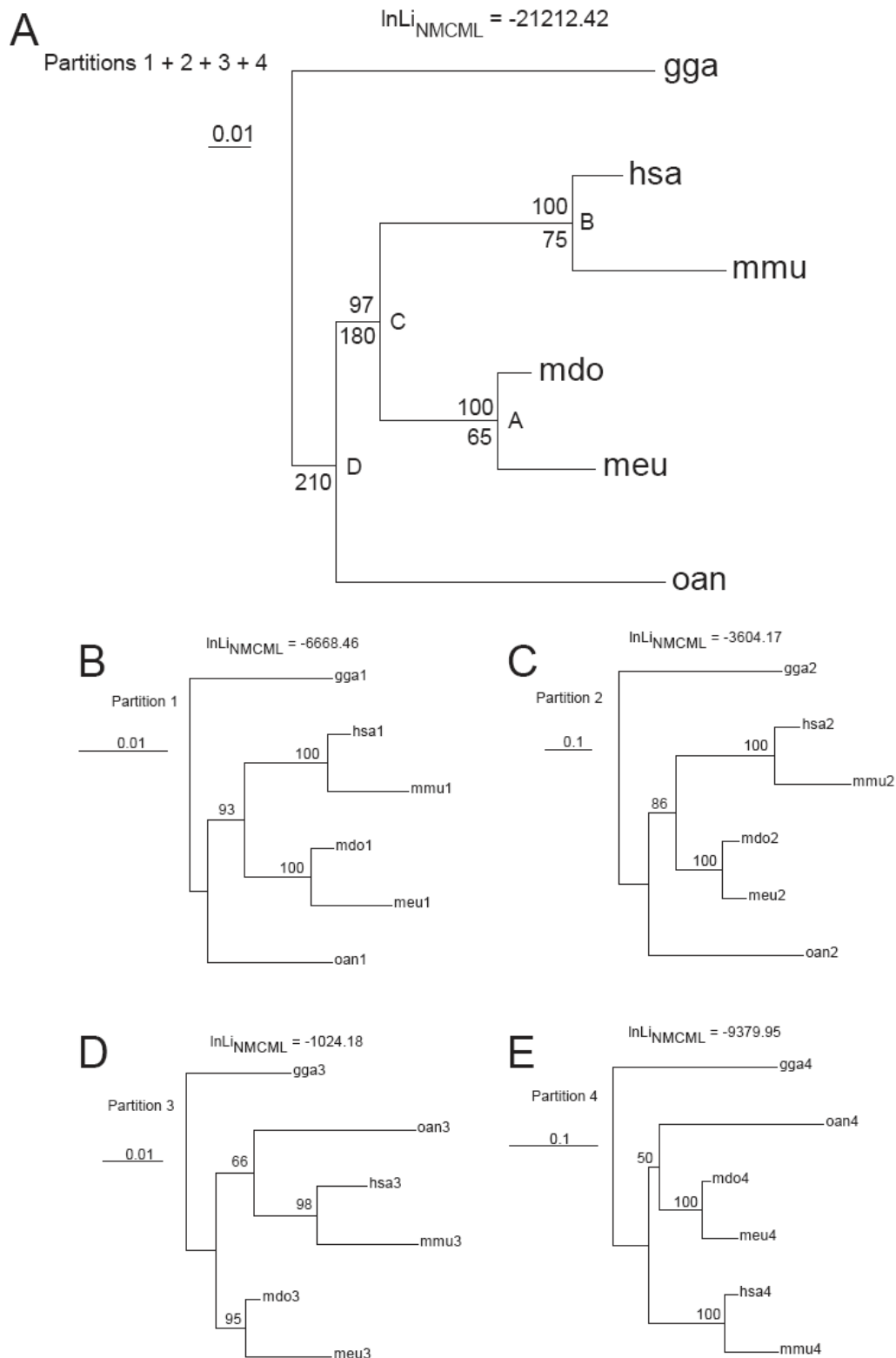
Fig. (2). Phylogenetic trees constructed from a subset of 94 pre-miRNAs shared among the six species shown (hsa = human, mmu = mouse, mdo = opossum, meu = wallaby, oan = platypus, and gga = chicken). (**A**) The unpartitioned pre-miRNA tree. 100 bootstrap support values are shown above the line at each node while below the lines are the estimated dates, in millions of years, of the last common ancestor based upon fossil data. These latter values are used for the calculations presented in Supplemental File **7** and the letters identifying each node refer to the calculations presented in Supplemental Files **6** and **7**. B. The partition 1, mature miRNAs and star sequence, tree. (**B**) The partition 2, loop, tree. (**C**) The partition 3, fold back sequence, tree. (**D**) The partition 4, other hairpin sequence, tree. 100 bootstrap support values for each branch are shown for each partition tree. Log-likelihood values (lnLi) are for each tree under the non-molecular clock model and are those used in Tables **2** and **3**.

Table 2. **Tests of Topological Hypotheses for the Pre-miRNA Subset (1+2+3+4) and for Each Partition: Mature miRNA and "Star" Sequences (1), Loop Sequences (2), Foldback Sequences (3), and Remaining Hairpin Sequences (4)**

Data Partition	Base Positions	lnLi$_{NMCCT}$[1]	lnLi$_{NMCML}$[2]	2(lnLi$_{NMCML}$ − lnLi$_{NMCCT}$)	P	Sig
1+2+3+4	8179	-21212.42	-21212.42	0	1	-
1	3644	-6668.46	-6668.46	0	1	-
2	933	-3604.17	-3604.17	0	1	-
3	508	-1027.43	-1024.18	6.50	1.65E-01	-
4	3094	-9380.63	-9379.95	1.36	8.51E-01	-

[1]lnLi$_{NMCCT}$: NonMolecular Clock Canonical Topology.
[2]lnLi$_{NMCML}$: NonMolecular Clock Maximum Likelihood Topology.
H$_O$: topologies are equivalent between maximum likelihood topology and the canonical topology.
Likelihood Ratio Test (2(lnLi$_{NMCCT}$ − lnLi$_{NMCML}$)) is approximately Chi-Square distributed with *d.f.* = number of taxa − 2, in this case, *d.f.* = 4.

evolutionary change in some of these lineages and, furthermore, within different areas of the microRNA hairpin might be acting under distinct evolutionary processes. The subset of 94 shared pre-miRNA hairpin sequences from six species (i.e., chicken, platypus, wallaby, opossum, mouse, and human) were used to reconstruct a phylogenetic tree by maximum likelihood (Fig. **2**). The resulting phylogenetic topology conforms to the canonical mammalian phylogeny, with both eutherians and marsupials forming sister clades and the monotreme as a more basal group, closer to the out group chicken. In the different partitions of the pre-miRNA sequence that correspond to the mature and star sequences (Fig. **2B**), the loop (2C), the fold back (2D), and other hairpin sequence (2E), the arrangement of the monotreme with respect to the eutherian or marsupial clades deviates from the canonical topology with respect to partitions 3 or 4 but neither deviation is statistically significant by likelihood ratio test (Table **2**). Furthermore, the nodes where partitions 3 and 4 deviate from the canonical mammalian topology show low bootstrap support, Fig. (**2D**) and (**2E**).

The question of whether there is evidence for deviations from a molecular clock hypothesis was addressed with the entire data set as well as with each partition by performing a likelihood ratio test on a molecular clock versus a non-molecular clock canonical mammalian topology (Table **3**). All partitions together as well as partitions 1, 2 and 4 show highly significant deviations from the molecular clock hypothesis. By contrast, the fold back segment, or partition 3, of the microRNAs, shows no significant deviation from the molecular clock hypothesis, but this might be due to a low

phylogenetic signal within this small sized partition. To test whether the dataset for partition 3 contained low phylogenetic information we used a likelihood ratio test to distinguish between the canonical mammalian phylogeny and an unstructured topology or a "star" tree. Following the naming conventions in Tables **2** and **3**, the star tree had a lnLi$_{ST}$ = -1052.55 when compared to the lnLi$_{MCCT}$, which results in a LRT = 42.88 and P = 1.10E-08. Therefore the fold back partition contains sufficient phylogenetic information to distinguish between an unstructured versus a structured topology, suggesting that the failure to reject the molecular clock hypothesis is not due to low information content, but perhaps due to fewer deviations from clock-like assumptions.

Relative Rate Tests

Due to the fairly consistent phylogenetic signal with the microRNAs and significant rejection of a molecular clock hypothesis, relative rate tests were performed between pairwise members of sister taxa with a single outgroup forming the root. First, chicken was used as the outgroup to compare among the three mammalian lineages. Using the unpartitioned data set, 4 of the 8 comparisons resulted in significant rate differences between lineages (Supplemental File **3**). The first of these, where platypus is fast and human is slow, is also seen in partition 4, but not site partitions 1, 2 or 3. The second, where platypus is fast and opossum is slow, is also seen in site partitions 2 and 4, but not 1 and 3. The third, where mouse is fast and opossum is slow, is also seen across all site partitions 1, 2, 3 and 4. Finally, where mouse is fast and wallaby is slow, the result is seen in site partition 2, but

Table 3. **Tests of Molecular Clock Hypotheses for the Pre-miRNA Subset (1+2+3+4) and for Each Partition: Mature miRNA and "Star" Sequences (1), Loop Sequences (2), Foldback Sequences (3), and Remaining Hairpin Sequences (4)**

Data Partition	Base Positions	lnLi$_{MCCT}$[1]	lnLi$_{NMCCT}$[2]	2(lnLi$_{NMCCT}$ − lnLi$_{MCCT}$)	P	Sig[3]
1+2+3+4	8179	-21265.87	-21212.42	106.90	7.62E-11	***
1	3644	-6681.52	-6668.46	26.12	2.99E-05	**
2	933	-3617.68	-3604.17	27.02	1.97E-05	**
3	508	-1031.11	-1027.43	7.36	1.18E-01	-
4	3094	-9415.20	-9380.63	69.14	6.84E-11	***

[1]lnLi$_{NMCCT}$: NonMolecular Clock Canonical Topology.
[2]lnLi$_{NMCCT}$: NonMolecular Clock Maximum Likelihood Topology.
[3]**p < 0.01; ***p < 0.001.
H$_O$: Branch lengths are ultrametric across the entire 6 taxon topology.
Likelihood ratio test (2(lnLi$_{NMCCT}$ − lnLi$_{MCCT}$)) is approximately Chi-Square distributed with *d.f.* = number of taxa − 2, in this case, *d.f.* = 4.

not in partitions 1, 3, or 4. Additional rate differences not seen overall can be seen in some of the partitions alone, such as site partition 1 shows evidence of mouse fast and platypus slow while partition 2 shows evidence of human fast and wallaby slow.

Similar patterns for relative rate tests are seen when examining only the portion of the mammalian phylogeny among eutherians and marsupials using platypus as the outgroup root (Supplemental File **4**). Again, using the entire unpartitioned dataset, 3 of the 4 comparisons resulted in significant rate differences. Patterns of rate differences are again heterogeneous across site partitions where, in the first, where human is fast and opossum is slow, the same result is seen in site partition 2, but not 1, 3 or 4. The second, where mouse is fast and opossum is slow, is seen in site partitions 1, 2 and 4, but not site partition 3. Finally, in the third, where mouse is fast and wallabny is slow, the result is seen in site partitions 2 and 4, but not 1 and 3. Again, additional rate differences not seen overall can be found, such as in partition 2 which shows evidence of human fast and wallaby slow.

Finally, comparing rates within eutherians and within marsupials is possible by using a marsupial as the outgroup root for the eutherians and a eutherian as the outgroup root for the marsupials. With the entire unpartitioned dataset both of the comparisons resulted in significant rate differences (Supplemental File **5**). Where wallaby is fast and opossum is slow overall (using human as the outgroup root), the result is seen in site partitions 1 and 4, but not 2 and 3. Where mouse is fast and human is slow overall (using opossum as the outgroup root), the result is seen in site partitions 1, 2 and 4, but not 3. None of the rate comparisons among eutherians, among marsupials or in comparisons between eutherians and marsupials show significant differences in rate in site partition 3. Even at the level of all mammals, the fold back region, site partition 3, only shows a single deviation from the null hypothesis of equality of rates, and this is where mouse is fast and opossum is slow, a result found for all site partitions for those lineages.

Rates of microRNA Change Per Site Per Million Years

It is possible to approximate the substitution rate per site per unit time, by first taking the branch lengths of a phylogenetic topology (Supplemental File **6**), and dividing the substitution rate per site by the fossil record-derived time of the most recent common ancestor (MRCA) of that bifurcation event (Supplemental File **7**). While the assumptions of molecular clock-like and the equality of relative rates between lineages are clearly violated for the overall microRNA dataset as well as site partitions 1, 2 and 4, partition 3 is most closely clock-like and contains only a single rate comparison that deviates from the null expectation. So, the values presented in Supplemental Files **6** and **7** are complete for the dataset entirely as well as all site partitions separately, but it is only site partition 3 that these numbers are estimated from data that do not violate basic assumptions. If the fold back segment of the microRNA is truly a neutral and clock-like segment of the molecule, then an average substitution rate of 1.33E-04 substitutions per site per million years for the fold back sites would be the best estimate. However, it is clear that there is substantial and statistically significant evidence that the majority of the sites within these microRNA se-

quences are under distinct lineage specific or site specific evolutionary regimens such that they violate several assumptions for those approximations to be meaningful. However, we can consider the base substitution rates shown in Supplemental File **7**, with estimates as fast as 2.23E-03 substitutions per site per million years to as slow as 3.47E-05 substitutions per site per million years, as first order approximations of what the rate variation might be among lineages and among sites overall as well as within the site partitions. Further, it is possible to put confidence intervals on these estimates by accounting for the confidence in the branch substitution rates as well as the fossil time points, but we suggest that a confidence interval in this case would lead to a false sense of precision in the face of the various violated assumptions.

DISCUSSION

A Triassic Mammal miRNA Profile

The early evolutionary history of the Mammalia is primarily defined on the basis of cranial and dental features such as the relative size and position of the quadrate and squamosal bones of the skull, the articulation of the jaw, the emergence of distinct inner ear ossicles, and clearly differentiated teeth with multiple roots [18]. While much of the evidence supporting Upper Triassic (225 – 205 MYA) and Lower Jurassic (205 – 180 MYA) phylogenies was fragmentary, several crucial discoveries have firmly established that the basic mammalian body plan was present at least 220 MYA and that there was substantial morphologic variation among the Mammalia at the beginning of the Jurassic [19-22]. Molecular analyses of extant mammals, including monotremes and marsupials, suggest that the monotremetherian divergence occurred around 210 MYA and the metatherian (marsupial mammal)-eutherian (placental mammal) divergence occurred around 180 MYA [23, 24]. Thus, by comparing the conserved microRNA content of monotreme, marsupial, and placental mammal genomes, we suggest that a minimal miRNA profile of a Late Triassic common mammalian ancestor would have had to include at least those miRNAs conserved between non-eutherian and eutherian mammals.

The Sempere *et al.* [3] evolutionary analysis of more than 300 miRNAs over more than 25 sequenced genomes included a taxonomic history of 292 non-paralogous miRNAs. Not surprisingly, none of the 174 miRNAs presented here for *M. domestica* appear on the listing for Anthropoda, Insecta, Diptera, or the *Drosophilae*. However, excluding these taxa, nearly all of the remaining miRNAs are present in *M. domestica* until the taxon Eutheria, where only four miRNAs (*miR-147*, miR-*152*, *miR-212*, and *miR-425*) were found in the opossum genome. Among those non-paralogous miRNAs listed under the taxa Hominidae and *H. sapiens*, none are present in the opossum. Sempere *et al.* [3] listed a total of 115 eutherian, hominid, and human miRNAs that we did not find in opossum. We expanded the list *via in silico* homology search to include all miRBase Release 10.0 mouse and human miRNAs. Within the expanded list only *miR-632, miR-768, miR-801, miR-873,* and *miR-875* were found in the opossum genome and all five were also found in the wallaby genome, the platypus genome or in both genomes. This brought the final total to the 171 conserved miRNAs (three

of the four Mdo-miR-147s are not counted in the total) presented in Table **1**. The fact that this number leaves more than two-thirds of known mammalian miRNAs exclusive to placental mammals is consistent with the previously noted observations that major developmental events, such as the advent of bilateral symmetry, the emergence of the vertebrates, and the placental reproduction strategy among mammals, were each accompanied by major expansions in the microRNA repertoire of the relevant taxa [4-6]. While there will doubtless be lineage-specific miRNAs identified in monotreme and marsupial genomes (we have so far found fifteen such miRNAs in the opossum), the 171 miRNAs identified here as shared among all mammalian genomes can be taken to constitute a core conserved miRNA complement of the Mammalia and, by inference, would have been present in the Late Triassic mammal common ancestor that lived prior to the divergence of the three therian clades.

Mammals, microRNAs, Phylogenetic Topologies, Rates and Clocks

Within the group of 94 miRNAs shared among the five mammalian and one avian species examined here, there is good molecular phylogenetic evidence that, overall, miRNAs track their expected evolutionary lineages within the mammals. However, we have shown that there is substantial evidence of different rates of substitution among these lineages and that these rates are not constant when considered across putatively functionally different regions of pre-miRNA hairpins. It is speculative, but tantalizing, to envision selective forces acting on various regions of miRNAs during specific time points that can account for these differences. From the degree of conservation of pre-miRNA and, certainly, mature miRNA sequences, it is likely that purifying selection has overall had the greatest influence on removing variation from miRNAs. However, over the time period examined here, the possibility of positive directional selection of miRNAs to adopt novel functionality during the evolution of mammals is of great interest.

Branches Defining Sites Predicting Targets

Substitutions that can be accounted for by branch B-C in Fig. (**2A**), would have the closest association with the divergence and subsequent evolution of eutherian mammals. While any base substitution could account for functional shifting of a miRNA to its regulatory targets, either through miRNA expression, miRNA processing or the interaction of the miRNA with its final targets, it is easiest to conceptualize the substitutions that occur in the mature miRNA as having the most obvious influence on regulatory interactions of the miRNA with the target mRNAs. Thus, it is easiest to conceive that a specific pattern of nucleotide substitution at a site that might be causal as displaying a high degree of conservation in non-eutherians followed by change in the eutherian lineage with subsequent conservation of the substitution. Therefore, a nucleotide substitution pattern of identity in chicken, platypus, wallaby, and opossum and then a difference between these genomes and those of placental mammals but with identity within the placental mammals is the scenario most closely approximating the expectation of a eutherian-specific substitution event. We see this substitution pattern at 22 sites in the mature sequences of 17 of the 94 miRNAs forming our shared miRNA subset (Supplemental File **8**). Among these 17 mature

sequences are two that show two changes and 15 that show one base change each. Of the 22 total base changes, only two (miR-222 and let-7d) occur within the position 2 to 9 seed region of the nature miRNA, the region known to be crucial to miRNA target specificity [21-23]. Of course, whether any of these 22 sites are relevant at all in eutherian evolution or whether they are simply neutral variation that associate with one expected pattern of mammalian evolution is clearly unknown but, again, there are 3 distinct patterns within our subset of pre-miRNAs that appear to be emergent in early evolution of mammalian miRNAs: 1) substantial evolutionary rate differences among conserved miRNAs, suggesting fluctuations in selective regimen in miRNAs that are preserved over time, 2) the emergence of novel miRNAs in eutherian mammals, presumably by multiple mechanisms, which can then adopt additional functions, and 3) small changes in mature miRNA sequences in some lineages that may alter the regulatory relationship between the miRNA and its targets. A fourth possible factor that might be acting is the potential modulation of which strand or strands become mature miRNAs, but at the present time evidence for mature miRNAs on both strands of mammalian pre-miRNA hairpins, the so-called "star" sequences, is primarily restricted to the eutherian species mouse and human, and while we applied these site partitions across all lineages, they may not be true. Deep sequencing being carried out in the recently completed platypus genome may shed some very important light on this issue [25, 26].

miRNA Expression: New Targets and New Functions?

Several papers have reported on expression of various miRNAs in different placental mammalian tissues [27-30]. Among these are nearly all of the 171 miRNAs identified in the non-eutherian mammalian genomes (Table **4**). Given that these 171 miRNAs represent a core mammalian miRNA complement, finding them among commonly expressed miRNAs is not surprising. For example, the *miR-302* cluster, shown to be highly expressed in the heart, would be expected to be similarly expressed in marsupials and monotremes. Other well known heart-expressed miRNA are present as well [31]. Those miRNAs known to play a major role in skeletal muscle development and differentiation [32, 33] (i.e., *miR-1*, *miR-133*, and *miR-206*) are also represented. The same is true for those miRNAs expressed brain, liver, and kidney.

An interesting group of miRNAs in Table **4** are those present in our presumed common mammalian ancestor that are known to be expressed in human placental tissue [27, 29]. Among these miRNAs, more than half (26 of 57) were observed to display their highest level of expression in placenta among the tissues examined. While there are dozens of miRNAs expressed in eutherian placenta that have evolved since the metatherian-eutherian divergence, notably, for example, the extensive cluster discovered on human chromosome 19 that is exclusively expressed in primate placental tissue [16], expression of so many ancestral miRNAs in a tissue that was, at best, only rudimentarily developed prior to that divergence confirms that tissue- and, perhaps, target mRNA-, specificity is fluid among miRNAs over evolutionary time. This is certainly not a new idea given the multiplicity of mRNA targets identified for various miRNAs and the multiplicity of miRNAs that target the same mRNA. However, the observation that some miRNAs conserved in non-

Table 4.   Conserved MicroRNA Expression in Placental mammals

Tissue	microRNA
Brain	let-7, 7, 9, <u>10a</u>, 15, 16, 18, 20, 21, 22, 23a, 23b, 24, 25[b], 26, 27, 29a, 29b, 29c, 30c, 31, 34, 92, 93, 96, 100, 103, 106, <u>124</u>, 125, 128, <u>129</u>,132[a], 133, 137, <u>143</u>, 145, 181, 183, 191, 196, <u>199</u>, 204, 212, <u>214</u>, 218, <u>219</u>, 221, 222, 224
Heart	let-7, 1, 15, 16, 23a, 24, 30c, 100, 126, 133, 145, 206[b], <u>302a</u>, <u>302b</u>, <u>302c</u>, <u>302d</u>, <u>367</u>, <u>499</u>
Lung	let-7, 15, 16, 23a, 23b, 24, 26a, 26b, 27a, 27b, 29a, 29b, 29c, 30a, 30b, 30c, 100, 126, 145
Liver	let-7, 9, 10, 15, 16, 18, 21, 22, 23a, 23b, 24, 25[b], 27, 29, 30c, 31, 34, 92, 93, 103, 106, 107, <u>122a</u>, 125, 128, 133, 135, 140, 141, 142, 145, 148, 152, 181, 183, 186, 191, 192, 194, 196, 199, 200c[c], 204, 205, 212, 214[b], 218, 221, 222, 223
Kidney	let-7, 16, 23a, 23b, 24, 30c, 100, 126, 145
Skeletal Muscle	let-7, <u>1</u>, 7, 16, 23b, 30c, 100, <u>133</u>, 144, 145, 181, 206[b]
Placenta	**let-7**, 9, 10, **15**, **16**, **18**, **19**, **20**, **21**, **22**, **23a**, 23b, 24, **25[b]**, 26, **27**, 29, **30**, 31, **32**, 34, 92, **93**, 96, **100**, **103**, 106, **107[b]**, 125, 126, 128, 129, 130, 133, 136, 140, **141**, 142, **143**, 145, 152, **181**, <u>184</u>, 186[a], **191**, 192[b], 193, 194, 196, 199, **200c[c]**, 204, 214[b], 218, **221**, **222**, 223[b], **301**, **302**

Reported expression in tissues of placental mammals of microRNAs conserved in opossum (*Monodelphis domestica)*, tammar wallaby (*Macropus eugenii*), and platypus (*Ornithorhynchus anatinus*) genomes. Expression of microRNAs identified in marsupial and monotreme genomes that are known to be expressed in the mammalian placenta are also shown. MicroRNAs shown in BOLD type were reported by Barad *et al*. [27] to have their highest expression levels in placenta among the tissues they examined. Underlined microRNAs are those shown by Liang *et al*. [29] to have the highest levels of differential expression in that tissue among the tissues they examined. In many cases the specific subtype of the microRNA was not given in the original source and microRNAs with a miRBase identification number above 302 were only reported in [29].
Sources: [27-30].
[a]miRNA as yet identified in opossum and wallaby only.
[b]miRNA as yet identified in opossum and platypus only.
[c]miRNA as yet identified in opossum only.

placental mammals are highly expressed in eutherian placenta may create an opportunity to understand how miRNAs accomplish their multitasking by looking at the specific miRNAs and their targets in placental and non-placental tissues in all three of the mammalian reproductive strategies. It should be noted that several of the miRNAs identified above as containing eutherian-specific origin patterns of nucleotide substitution and conservation (Supplemental File **8**), are among those expressed in placenta. The opportunity to gain evolutionary insights into microRNA function by further examining such patterns of miRNA expression will surely not be overlooked.

CONCLUSIONS

Molecular paleontology is a means to gain insight on molecular phenomena that no longer exist and have not existed for a very long time. Here, we have applied the principles of comparative genomics to assemble a plausible minimal miRNA profile of the last common ancestor of the Mammalia, an animal that lived 230 million years ago [18]. That profile is composed of 171 miRNAs shared among the genomes of the three mammalian reproductive clades; monotremes, marsupials, and placentals. We further examined the tempo and mode of evolutionary change in the precursor sequences of 94 of these 171 miRNAs and find consistency between the overall pattern of nucleotide substitution and the canonical phylogenetic history of these lineages. We also find abundant evidence for heterogeneity among different lineages and, by partitioning the pre-miRNA hairpin into four structure/function regions, among the different segments of the pre-miRNA. Unexpectedly, the one region of the pre-miRNA that appears to be least influenced by variation in the rates of evolutionary change is the fold back sequence of the hairpin directly opposite the mature miRNA. Using our pre-miRNA partition scheme we are able to calibrate a molecular clock and find that the fold back region is most closely conforming to molecular clock-like behavior

with an overall substitution rate of approximately 1.33E-04 substitutions per site per million years.

Finally, we have shown that most of these 171 miRNAs are well known and are expressed in major tissues such as brain, heart, lung, liver, kidney, and skeletal muscle. We have also shown that many of these miRNAs are highly expressed in eutherian placenta and that a small number contain lineage-specific patterns of substitution in their mature sequences suggesting the (very speculative) possibility that they may have acquired somewhat different regulatory targeting in the lineage leading to the placental mammals. Taken together, our observations suggest that studying the specific functions of these placentally-expressed miRNAs in all three mammalian clades has the potential of addressing the question of how miRNAs evolve and acquire new targets and new functions.

ADDITIONAL INFORMATION[§]

Supplemental File 1. List of 189 *M. domestica* miRNAs, 171 conserved and 18 new, including chromosome coordinates.

Supplemental File 2. Map of the *M. domestica* genome indicating the relative position of 189 miRNAs and the five major miRNA processing proteins Dicer, DROSHA, Exportin 5, DGCR8, and AGO2.

Supplemental File 3. Relative rates tests of pre-miRNA sequence substitutions: intra-mammalian, inter: monotreme-marsupial-eutherian using chicken as the outgroup root.

Supplemental File 4. Relative rates tests of pre-miRNA sequence substitutions: intra-non-monotreme mammal, inter: marsupial-eutherian using platypus as the outgroup root.

Supplemental File 5. Relative rates tests of pre-miRNA sequence substitutions: inter-marsupial with human as the

_____
[§] Available online.

outgroup root and inter-mammalian with opossum as the outgroup root.

Supplemental File 6. Branch length estimates under a non-molecular clock model.

Supplemental File 7. Base substitution rates.

Supplemental File 8. miRNAs showing a non-eutherian/eutherian mature miRNA substitution pattern.

REFERENCES

[1] Canestro C, Yokoi H, Postlethwait JH. Evolutionary developmental biology and genomics. Nat Rev Genet 2007; 8: 932-42.
[2] Raff RA. Written in stone: fossils, genes and evo-devo. Nat Rev Genet 2007; 8: 911-20.
[3] Sempere LF, Cole CN, McPeek MA, *et al*. The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. J Exp Zool B Mol Dev Evol 2006; 306: 575-88.
[4] Heimberg AM, Sempere LF, Moy VN, *et al*. MicroRNAs and the advent of vertebrate morphological complexity. Proc Natl Acad Sci USA 2008; 105: 2946-50.
[5] Hertel J, Lindemeyer M, Missal K, *et al*. The expansion of the metazoan microRNA repertoire. BMC Genomics 2006; 7: 25.
[6] Prochnik SE, Rokhsar DS, Aboobaker AA. Evidence for a microRNA expansion in the bilaterian ancestor. Dev Genes Evol 2007; 217: 73-7.
[7] Devor EJ, Samollow PB. *In vitro* and in silico annotation of conserved and nonconserved microRNAs in the genome of the marsupial Monodelphis domestica. J Hered 2008; 99: 66-72.
[8] Griffiths-Jones S, Saini HK, van Dongen S. miRBase: tools for microRNA genomics. Nucleic Acids Res 2008; 36: D154-8.
[9] Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994; 22: 4673-80.
[10] Okamura K, Phillips MD, Tyler DM, *et al*. The regulatory activity of microRNA* species has substantial influence on microRNA and 3' UTR evolution. Nat Struct Mol Biol 2008; 15: 354-63.
[11] Felsenstein J. PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics 1989; 5: 164-6.
[12] Robinson-Rechavi M, Huchon D. RRTree: relative-rate tests between groups of sequences on a phylogenetic tree. Bioinformatics 2000; 16: 296-7.
[13] Li P, Bousquet J. Realtive-rate test for the nucleotide substitutions between two lineages. Mol Biol Evol 1992; 9: 1185-9.
[14] Sanderson MJ. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 2003; 19: 301-2.
[15] Berezikov E, Guryev V, van de Belt J, *et al*. Phylogenetic shadowing and computational identification of human microRNA genes. Cell 2005; 120: 21-4.
[16] Bentwich I, Avniel A, Karov Y, *et al*. Identification of hundreds of conserved and nonconserved human microRNAs. Nat Genet 2005; 37: 766-70.
[17] Tomari Y, Zamore PD. Perspective: machines for RNAi. Genes Dev 2005; 19: 517-29.
[18] Kermack DM, Kermack KA. The evolution of mammalian characters. Sydney: Croom Helm, Kapitian Szabo Publisher 1984.
[19] Lucas S, Heckert A, Harris J, *et al*. Mammal-like tooth from the Upper Triassic of Germany. J Vert Paleontol 2001; 21: 397-9.
[20] Lucas S, Luo Z. *Adelobasileus* from the Upper Triassic of West Texas: the oldest mammal. J Vert Paleontol 1993; 13: 309-34.
[21] Luo ZX. Transformation and diversification in early mammal evolution. Nature 2007; 450: 1011-9.
[22] Martin T. Paleontology. Early mammalian evolutionary experiments. Science 2006; 311: 1109-10.
[23] Nilsson MA, Arnason U, Spencer PB, *et al*. Marsupial relationships and a timeline for marsupial radiation in South Gondwana. Gene 2004; 340: 189-96.
[24] Wakefield MJ, Graves JA. Marsupials and monotremes sort genome treasures from junk. Genome Biol 2005; 6: 218.
[25] Murchison EP, Kheradpour P, Sachidanandam R, *et al*. Conservation of small RNA pathways in platypus. Genome Res 2008; 18: 995-1004.
[26] Warren WC, Hillier LW, Marshall Graves JA, *et al*. Genome analysis of the platypus reveals unique signatures of evolution. Nature 2008; 453: 175-83.
[27] Barad O, Meiri E, Avniel A, *et al*. MicroRNA expression detected by oligonucleotide microarrays: system establishment and expression profiling in human tissues. Genome Res 2004; 14: 2486-94.
[28] Landgraf P, Rusu M, Sheridan R, *et al*. A mammalian microRNA expression atlas based on small RNA library sequencing. Cell 2007; 129: 1401-14.
[29] Liang Y, Ridzon D, Wong L, *et al*. Characterization of microRNA expression profiles in normal human tissues. BMC Genomics 2007; 8: 166.
[30] Sempere LF, Freemantle S, Pitha-Rowe I, *et al*. Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. Genome Biol 2004; 5: R13.
[31] Sood P, Krek A, Zavolan M, *et al*. Cell-type-specific signatures of microRNAs on target mRNA expression. Proc Natl Acad Sci USA 2006; 103: 2746-51.
[32] Chen JF, Mandel EM, Thomson JM, *et al*. The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. Nat Genet 2006; 38: 228-33.
[33] Latronico MV, Catalucci D, Condorelli G. Emerging role of microRNAs in cardiovascular biology. Circ Res 2007; 101: 1225-36.