EVOLUTION & DEVELOPMENT 11:1, 50–68 (2009)
DOI: 10.1111/j.1525-142X.2008.00302.x

The deep evolution of metazoan microRNAs

Benjamin M. Wheeler,[a] Alysha M. Heimberg,[b] Vanessa N. Moy,[b] Erik A. Sperling,[c] Thomas W. Holstein,[d] Steffen Heber,[e] and Kevin J. Peterson[b,*]

[a]Department of Computer Science, North Carolina State University, Raleigh, NC 27695, USA
[b]Department of Biological Sciences, Dartmouth College, Hanover, NH 03755, USA
[c]Department of Geology and Geophysics, Yale University, P.O. Box 208109, New Haven, CT 06520, USA
[d]Molekulare Evolution und Genomik, Heidelberger Institut für Zoologie, Neuenheimer Feld 230, 69120 Heidelberg, Germany
[e]Bioinformatics Research Center, North Carolina State University, Raleigh, NC 27695, USA
*Author for correspondence (email: kevin.j.peterson@dartmouth.edu)

SUMMARY microRNAs (miRNAs) are approximately 22-nucleotide noncoding RNA regulatory genes that are key players in cellular differentiation and homeostasis. They might also play important roles in shaping metazoan macroevolution. Previous studies have shown that miRNAs are continuously being added to metazoan genomes through time, and, once integrated into gene regulatory networks, show only rare mutations within the primary sequence of the mature gene product and are only rarely secondarily lost. However, because the conclusions from these studies were largely based on phylogenetic conservation of miRNAs between model systems like *Drosophila* and the taxon of interest, it was unclear if these trends would describe most miRNAs in most metazoan taxa. Here, we describe the shared complement of miRNAs among 18 animal species using a combination of 454 sequencing of small RNA libraries with genomic searches. We show that the evolutionary trends elucidated from the model systems are generally true for all miRNA families and metazoan taxa explored: the continuous addition of miRNA families with only rare substitutions to the mature sequence, and only rare instances of secondary loss. Despite this conservation, we document evolutionary stable shifts to the determination of position 1 of the mature sequence, a phenomenon we call seed shifting, as well as the ability to post-transcriptionally edit the 5′ end of the mature read, changing the identity of the seed sequence and possibly the repertoire of downstream targets. Finally, we describe a novel type of miRNA in demosponges that, although shows a different pre-miRNA structure, still shows remarkable conservation of the mature sequence in the two sponge species analyzed. We propose that miRNAs might be excellent phylogenetic markers, and suggest that the advent of morphological complexity might have its roots in miRNA innovation.

INTRODUCTION

microRNAs (miRNAs) are a recently discovered group of small RNA regulatory genes that have captured the attention of investigators interested in the control of cellular differentiation (Ambros 2004; Zhao and Srivastava 2007; Hobert 2008; Makeyev and Maniatis 2008; van Rooij et al. 2008; Yi et al. 2008), its misregulation (Lu et al. 2005; Meltzer 2005; Calin and Croce 2006; Esquela-Kerscher and Slack 2006; He et al. 2007; Sevignani et al. 2007; Barbarotto et al. 2008; Medina and Slack 2008; Yang et al. 2008), and its evolution (Sempere et al. 2006; Heimberg et al. 2008; Pierce et al. 2008). With the demonstration that morphologically complex animals like vertebrates and fruit flies possess a protein-coding genetic tool kit that is largely conserved across Metazoa (Peterson and Sperling 2007; Putnam et al. 2007; Ryan et al. 2007; Simionato et al. 2007; Yamada et al. 2007; Larroux et al. 2008), investigators began to look at other components of gene regulatory networks for molecules that might show a different evolutionary pattern. Initial surveys of miRNAs across the animal kingdom demonstrated a very compelling feature of miRNA evolution when compared with the evolution of the protein-coding repertoire: miRNA families are continuously being added to bilaterian lineages through evolutionary time such that vertebrates, for example, were characterized by the possession of miRNA families not found in arthropods, and vice versa (Hertel et al. 2006; Sempere et al. 2006; Prochnik et al. 2007; Heimberg et al. 2008). Further, these authors showed that miRNAs, once integrated into the genomic regulatory circuitry, are only rarely secondarily lost, and the mature miRNA sequences are under intense negative selection.

Because of these features, evolutionary biologists have attempted to reconstruct the phylogenetic history of the miRNAs found in vertebrates, insects, and nematodes, and noted the following. First, known miRNAs neither could be found nor could be detected in sponges (Sempere et al. 2006; Prochnik et al. 2007). Second, bilaterians had a greatly

expanded repertoire of miRNAs as compared with cnidarians (Sempere et al. 2006; Prochnik et al. 2007), and vertebrates had a greatly expanded set of miRNAs as compared with bilaterian invertebrates (Hertel et al. 2006; Heimberg et al. 2008), and both of these advents correlated with increases to morphological complexity (Sempere et al. 2006; Lee et al. 2007; Niwa and Slack 2007; Heimberg et al. 2008). Third, only a small subset of the bilaterian miRNAs were detected in acoel flatworms, consistent with both their relatively simple morphology and their inferred phylogenetic position (Sempere et al. 2006, 2007; Baguñà et al. 2008). Hence, miRNAs could be very useful tools to explore hypotheses centered around the molecular basis of morphological complexity (Sempere et al. 2006; Lee et al. 2007; Niwa and Slack 2007; Heimberg et al. 2008), the phenomenon of developmental canalization (Hornstein and Shomron 2006; Cui et al. 2007; Peterson 2008), the control of the tempo of morphological evolution (Peterson et al. 2007), and the interrelationships among metazoan taxa (Sempere et al. 2007; Sperling and Peterson in press).

However, all of these evolutionary surveys necessarily relied on comparing the conserved miRNAs between the taxon of interest, whether annelid, agnathan, or acoel flatworm, with the known miRNA complement of mammals, arthropods, and/or nematodes because the complements of only these three model taxa had been explored independently of phylogenetic conservation (Lagos-Quintana et al. 2001; Lau et al. 2001; Lee and Ambros 2001). Thus, if a taxon increased its rate of mutation or secondarily lost many miRNAs, then the phylogenetic history of that particular miRNA would be inaccurately reconstructed, leading to potentially spurious claims about the import of miRNAs with respect to metazoan macroevolution. Although a few more recent studies have explored the miRNA repertoire of additional taxa independent of phylogenetic conservation (Palakodeti et al. 2006; Fu et al. 2008), an independent test of these claims through an examination of the evolutionary dynamics of miRNAs has yet to be explored in a taxon where both phylogenetic position and time of origin were carefully controlled.

Here, we explore the evolutionary dynamics of miRNAs across metazoan phylogeny and through deep evolutionary time by combining 454 sequencing of miRNA libraries with genomic searches. By sequencing a library from a taxon that is a sister lineage to a second taxon with a sequenced genome, the conserved set of miRNAs shared between these two taxa can be determined. Further, if the divergence time between these two taxa is known from the fossil record and/or estimated via a molecular clock (Peterson et al. 2008), then the rate of miRNA acquisition in this clade can be addressed, and compared with other similarly dated clades. Using this approach, we have identified virtually all known and phylogenetically-ancient miRNA families, as well as discovered numerous new miRNA families, in many taxa hitherto unsampled for their miRNA complements, including demosponges, cnidarians, annelids, gastropod molluscs, hemichordates, and echinoderms (see Table 1 for all taxonomic nomenclature used throughout this article). All of these clades have unique miRNA families not found anywhere else in the animal kingdom, including demosponges, which appear to possess a novel type of miRNA, a type that more closely resembles plant miRNAs than known animal miRNAs. We show that protostomes and deuterostomes have a greatly expanded number of miRNAs as compared with either acoel flatworms or to cnidarians, and that the rate of acquisition of these novel miRNAs is not seen anywhere else in the animal tree except at the base of the vertebrates (Heimberg et al. 2008). Importantly, all of the miRNAs, including the sponge miRNAs, show remarkable conservation of the mature gene sequence. When substitutions do occur, they occur at predictable positions along the mature sequence, with the highest frequency of changes localized to the 3′ end of the mature gene product. In a few cases, the identification of nucleotide 1 of the mature sequence of a specific miRNA in two or more taxa has been moved toward either 5′ or 3′, a phenomenon we call "seed shifting" because the identity of positions 2–8, and hence the seed sequence, was also changed with the change in the identification of position 1. We also document phylogenetically conserved edits to the 5′ end involving nucleotide substitutions, nucleotide insertions, and/or nucleotide deletions within the seed region itself, both of which again changes the identity of nucleotides 2–8. Finally, almost all nodes within Metazoa are characterized by the addition of at least one novel miRNA family, and these novel miRNA families are only rarely secondarily lost in descendent taxa. Our data are consistent with the idea that miRNAs are

Table 1. Taxonomic nomenclature

Metazoa—Demospongia+Eumetazoa
Demospongia (Haplosclerida)—*Haliclona*+*Amphimedon*
Eumetazoa—Cnidaria+Triploblastica
Cnidaria—*Nematostella* (Anthozoa)+*Hydra* (Hydrozoa)
Triploblastica—*Symsagittifera* (Acoela)+Nephrozoa
Nephrozoa—Protostomia+Deuterostomia
Protostomia—Ecdysozoa+Eutrochozoa
Ecdysozoa—*Priapulus* (Priapulida)+Arthropoda
Arthropoda—*Ixodes* (Chelicerata)+Pancrustacea
Pancrustacea—*Daphnia* (Crustacea)+*Drosophila* (Insecta)
Eutrochozoa—*Cerebratulus* (Nemertea)+Neotrochozoa
Neotrochozoa—Annelida+Mollusca
Annelida—*Nereis*+*Capitella*
Mollusca (Gastropoda)—*Haliotis*+*Lottia*
Deuterostomia—Ambulacraria+Chordata
Ambulacraria—Echinodermata+*Saccoglossus* (Hemichordata)
Echinodermata (Eleutherozoa)—*Strongylocentrotus* (Echinoidea)+*Henricia* (Asteroidea)
Chordata—*Branchiostoma* (Cephalochordata)+*Homo* (Vertebrata)

continuously being added to metazoan genomes through time, that once acquired they show few nucleotide substitutions, and they are only rarely secondarily lost. These three features suggest that miRNAs might be excellent phylogenetic markers, and that metazoan morphological complexity and constraint might have their roots in miRNA innovation.

MATERIALS AND METHODS

Total RNA extraction and northern analysis

Total RNA was extracted from whole adult animals using standard Trizol method (Invitrogen, Carlsbad, CA, USA) from *Cerebratulus lacteus*, *Saccoglossus kowalevskii*, *Nereis diversicolor*, and *Haliclona* sp. (all of which were purchased from MBL, Woods Hole, MA, USA); *Strongylocentrotus purpuratus* and *Haliotis rufescens* (both purchased from The Cultured Abalone, Santa Barbara, CA, USA), *Henricia sanguinolenta* (purchased from Gulf of Maine Inc., Pembroke, ME, USA); *Branchiostoma floridae* (purchased from the Gulf Specimen Marine Laboratory, Panacea, FL, USA); *Capitella* sp. (a gift from J. Grassle and C. Noji, Rutgers University); *Priapulus caudatus* (a gift from A. Wallberg, Uppsala University); and *Symsagittifera roscoffensis* (a gift from J. Baguñà and P. Martinez, University of Barcelona). *Hydra magnipapillata* strain 105 and *Nematostella vectensis* were cultured as described (Guder et al. 2006; Kusserow et al. 2005). Poriferan samples used in Northern analyses are described (E. A. Sperling, unpublished data). Northern analyses using Starfire probes (IDT) (sequences available upon request) were performed as previously described (Sempere et al. 2004).

Small RNA library construction

Small RNA libraries were constructed as described (Lau et al. 2001) with only minor modifications. To isolate small RNAs, fluorescein-labeled DNA oligonucleotides equivalent to 21 and 27 nucleotides (nt) in molecular weight were combined with 200–500 μg of total RNA and electrophoresed on a 15% urea-polyacrylamide gel. Following the 3′ linker ligation, 31 and 43 nt fluorescein markers were combined with the ligated RNA just before electrophoresis and used to guide the excision of the 3′ ligated RNAs (between 35 and 41 nt in size). Following the 5′ linker ligation a 51 nt fluorescein marker was used in the same manner and the gel was excised above the marker to include the 5′ and 3′ ligated RNAs (between 52 and 58 nt in size). The 3′ and 5′ linker-ligated small RNAs were reverse transcribed to make a small RNA cDNA. The small RNA cDNA was amplified via PCR using the following conditions: an initial denaturation at 96°C for 1 min; 33 cycles at 96°C (10 sec), 50°C (1 min), and 72°C (15 sec); a final extension time of 5 min; and then held indefinitely at 10°C. The PCR primers included the 454 primers plus a unique 4 nt barcode (sequences available upon request) so that the source of the sequence could be identified after sequencing. The resultant PCR amplicons were electrophoresed through a 3% agarose gel and bands at approximately 100 nt were excised, gel extracted (Qiagen QIAquick Gel Extraction Kit; Qiagen, CA, USA), and concentrations were measured using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). Libraries with different barcodes were pooled for a total of 100 ng and submitted to 454 Life Sciences (Branford, CT, USA) and the Yale Center for Genomics and Proteomics Sequencing Facility for sequencing.

miRMiner

A program was developed, named miRMiner, to discover known miRNAs in new taxa, and to discover novel miRNAs from all taxa. First, the 5′ and 3′ PCR primers were removed by applying a 21 nt cutoff on either end of each sequence read. The resulting sequence reads were organized by taxon, based on the removed barcode, and a 17–25 nt length cutoff enforced. Shorter or longer reads, and those reads without a matching 5′ and 3′ barcode, were removed from the data set. The overall quality of each read was calculated by averaging the quality value of each nucleotide base over the entire sequence. Reads whose average quality was below 20 were discarded. Within each species, duplicate reads were eliminated, and the number of duplicates annotated as the read's frequency count. In each nonredundant set, reads that were identical to reads with a higher frequency count when ignoring differences on the 5′ and 3′ end, and allowing one gap or mismatch, were grouped. From each such group, the read with the highest frequency count was chosen as representative, and the others excluded from further analysis.

Known miRNAs were annotated by identifying homologous mature and star miRNA sequences in miRBase (Griffith-Jones et al. 2007) release 10.1. Standalone BLAST (blastn, version 2.2.27) was used to generate a list of candidate identities. This list was filtered using three criteria, evaluated on an ungapped global alignment of the read and the hit sequence beginning at the 5′ end: (1) sequence length must match within 2 nt; (2) positions 2–7 of the seed sequence must be identical; and (3) the remainder of the alignment may contain no more than three mismatches. Sequence reads that matched a known miRNA or miRNA star sequence within the above criteria were annotated and removed from the data set.

Known non-miRNA reads were identified by comparison with NCBI's "nt" nucleotide database using Standalone MEGABLAST (version 2.2.17). Reads matching a known RNA molecule with percent identity $>95\%$ were removed from the data set, and the remaining sequence reads were then investigated for phylogenetic conservation. Reads from all species were combined, and those that "matched" a read with a higher frequency count were grouped. Matches were determined using the three criteria used to identify known miRNAs given above (similar length, seed sequence identity, and nonseed sequence similarity). Reads conserved across multiple taxa were grouped, and groups were ranked by the frequency count of the most frequently occurring sequence. Reads not conserved across multiple taxa were divided by taxon and ranked by frequency count. This completed the automated analysis by miRMiner, resulting in a list of conserved reads across all taxa and lists of unique reads for each taxon. The number of reads per taxon, including the total number of annotated known miRNAs and novel miRNAs (new miRNA families discovered herein), are listed for the 13 species analyzed in Table 2.

Sequence reads with a frequency count two or greater were investigated as potentially novel miRNAs. First, blastn (performed at NCBI with default settings using the BLAST network service) was used to identify potential orthologs (generally anything with an

Table 2. Numerical analysis of library reads from the 13 taxa analyzed with 454 sequencing

Taxon	Total reads[1]	Parsed reads[2]	Known miRNAs[3]	Novel miRNAs[4]	Total miRNAs
Branchiostoma floridae	77,672	58,583	25,287	0	25,287
Capitella sp.	75,105	47,459	37,056	530	37,586
Cerebratulus lacteus	75,220	13,415	7041	162	7203
Haliclona sp.	27,059	11,195	0	250	250
Haliotis rufescens	183,827	36,923	16,206	1166	17,372
Henricia sanguinolenta	75,880	22,848	10,786	402	11,188
Hydra magnipapillata	9964	2397	0	4	4
Nematostella vectensis	67,864	11,462	31	10	41
Nereis diversicolor	73,146	18,159	8430	351	8781
Priapulus caudatus	82,639	53,771	14,080	25	14,105
Saccoglossus kowalevskii	48,525	19,973	6,838	299	7137
Strongylocentrotus purpuratus	58,142	19,822	9477	1171	10,648
Symsagittifera roscoffensis	47,521	6890	304	0	304
Summary	902,564	323,167	135,536	4370	139,906

[1]Total reads is the number of raw 454 reads corresponding to the taxon as indicated by the four nucleotide barcode at the beginning of each sequence. Note that the total amount of total RNA used to build the small RNA library varied by $<2\times$ across all taxa, and that the same mass of PCR product was sent for sequencing. We made two separate libraries for *H. rufescens* (foot, and visceral mass+larvae)—these numbers reflect the combined reads for both libraries.

[2]Parsed reads are the number of 454 reads investigated in this study. Reads were chosen as the section of the total reads whose length was between 17 and 25 nt and that met minimum quality criteria.

[3]Known miRNA refer to reads corresponding to known miRNA families listed in mirBase 10.1.

[4]Novel miRNA refer to reads corresponding to new miRNA families discovered herein. Note that these numbers only reflect only the shared novel miRNAs (supporting information Table S1) and not the numerous novel miRNAs restricted to each individual taxon—these novel miRNAs will be reported elsewhere.

E value of <0.1) in the genomic traces deposited at GenBank of the following taxa: *Amphimedon queenslandica* (haplosclerid demosponge), *B. floridae* (cephalochordate), *Capitella* sp. (polychaete annelid), *Daphnia pulex* (crustacean arthropod), *H. magnipapillata* (hydrozoan cnidarian), *Ixodes scapularis* (chelicerate arthropod), *Lottia gigantea* (gastropod mollusc), *N. vectensis* (anthozoan cnidarian), *S. kowalevskii* (hemichordate), and *S. purpuratus* (echinoid echinoderm). Potential orthologs were then folded using mfold (Zuker et al. 1999) and putative sequences were annotated as miRNAs if they met the structural criteria outlined by Ambros et al. (2003), and did not blast to any known miRNAs deposited at miRBase. Sequences where both the genomic fold and library reads were obtained were submitted to miRBase. In some cases, miRNA genes were found in genomic traces but library reads were not (e.g., miR-219 in the deuterostomes), or alternatively numerous high-quality reads were found in the small RNA library, but a genomic read was not (e.g., miR-190, miR-193, and miR-281 in *S. purpuratus*); in the latter case, these miRNAs were not submitted to miRBase.

Sequence evolution

To ascertain the extent of nucleotide substitutions in miRNAs, the mature sequences of all known and novel miRNAs analyzed herein (supporting information Table S1) from the 14 nephrozoan taxa shown in Fig. 2 were aligned, and the number of substitutions was determined using the most parsimonious interpretation of the data given in the taxonomic topology. For example, both of the annelids analyzed (*Capitella* sp. and *N. diversicolor*) differ at positions 9 and 11 in the mature sequence of miR-100 as compared with the 12 other nephrozoan taxa (see supporting information Table S1), and because both annelid have the same nucleotide substitutions and they are sister taxa (Fig. 2), these were counted as a single change at position 9, and a single change at position 11. In contrast, in miR-76 both the annelid *Capitella* and the gastropod *Haliotis* have an uracil at position 11, whereas all other taxa, including their sister species, *Nereis* and *Lottia*, respectively, have an adenine in this same position, and thus this was counted as two separate substitutions. We only analyzed the first 23 nucleotides of the mature sequence as there are relatively few miRNAs that are primitively 24 nt or longer (6 of 90 analyzed genes).

To compare the amount of nucleotide change between miRNAs and 18S rDNA, complete or nearly complete 18S rDNA gene sequences were downloaded for the 14 nephrozoan species (or closely related species within the genus depending on availability) shown in Fig. 2. The sequences were aligned using Muscle (Edgar 2004) on the EMBL-EBI web server, resulting in an alignment of 2788 nucleotides. The alignment was trimmed for conserved blocks using Gblocks (Castresana 2000), using the default settings except that the "Allowed Gap Positions" was set to "With Half," resulting in a final alignment of 1410 conserved positions. The rate of nucleotide change was then calculated by dividing the total number of characters by the tree length as determined using the Tree Window feature in the program MacClade (Maddison and Maddison 2005) for the topology shown in Fig. 2.

Genome walking

Genomic DNA was isolated using the DNAeasy Blood and Tissue kit (Qiagen), and the final concentration of genomic DNA was adjusted to 0.1 μg/μl. Genome walking was performed using

the GenomeWalker Universal kit (Clontech, Mountain View, CA, USA). The primers used for the following novel miRNAs are as follows: miR-1984 (5′-CCCTGCCCTATCCGTCAGGAACTGTG -3′); miR-1985 (5′-CCCTGCCATTTTTATCAGTCACTGTG-3′); miR-1986 (5′-CCCTGGATTTCCCAAGATCCGTGAT-3′); miR-1988 (5′-AGTGGTTTTCCGTTGCACATGC-3′); miR-1990-5p (5′-AGTAAGTTGATGGGGTCCCAGG-3′); miR-1990-3p (5′-GCAAGTAGTTGACGTAGTCCCG-3′); miR-1991 (5′-CTTACC CTGTTAATCGGAGAAGT-3′); miR-1992 (5′-CCCCACATCA GTGGTACAACTGCTGA); miR-1994 (5′-CCCCGAGGGAGG ACACACTGTCTCA-3′); and miR-2002 (5′- CCCATGGAACC AGCAGATGTATTC-3′). Using genomic DNA from the gastropod mollusc *H. rufescens* the novel miRNAs miR-1984, miR-85, and miR-86 were amplified as follows: 94°C for 25 sec, 50°C for 25 sec, and 72°C for 3 min (8 ×), followed by 94°C for 25 sec, 60°C for 25 sec, and 72°C for 3 min (33 ×), followed by a final extension (72°C) for 7 min. The secondary PCR used 1 μl of the primary PCR as the template, and cycled 27 times (94°C for 25 sec, 60°C for 25 sec, 72°C for 3 min), followed by a final extension (72°C) for 7 min. miR-1990 and miR-1991 were amplified from *H. rufescens* using a modified touchdown protocol. The starting annealing temperature was set at T_m +3°C for 25 sec and after three cycles dropped 1°C for another three cycles, all the way to T_m − 1°C, followed by a final 33 cycles at the T_m. The denaturation and extension temperature and times were 94°C for 25 sec and 72°C for 3 min, respectively, followed by a final extension (72°C) for 7 min. The same touchdown protocol was used to amplify miR-1992 and miR-1994 from the nemertean *C. lacteus*, and miR-2002 from the sea star *H. sanguinolenta*. All PCR reactions used Taq DNA polymerase (Roche Applied Science, Indianapolis, IN, USA) instead of the Advantage 2 Polymerase mix suggested in the GenomeWalker kit. All amplified bands were gel purified using the QIAquick Gel Purification kit (Qiagen), cloned into the pGEM T-easy vector (Promega, San Luis Obispo, CA, USA), plasmid purified using QIAprep Spin Miniprep kit (Qiagen) and sequenced on an ABI Model 3100 genetic analyzer. Sequences were aligned and analyzed for folding using mfold as above.

RESULTS

miRNA discovery

Just over 900,000 sequences of small RNAs were obtained using 454 sequencing technology (Margulies et al. 2005) from 13 taxa spanning metazoan evolution (Table 2): the demosponge *Haliclona* sp., the anthozoan cnidarian *N. vectensis*, the hydrozoan cnidarian *H. magnipapillata*, the acoel flatworm *S. roscoffensis*, the enteropneust hemichordate *S. kowalevskii*, the echinoid echinoderm *S. purpuratus*, the asteroid echinoderm *H. sanguinolenta*, the cephalochordate *B. floridae*, the priapulid *P. caudatus*, the nemertean *C. lacteus*, the polychaete annelid *N. diversicolor*, the polychaete annelid *Capitella* sp., and the gastropod mollusc *H. rufescens*. After identifying known miRNAs, miRMiner's list of potential miRNAs were then blasted to the genomic traces of *A. queenslandica*, *N. vectensis*, *I. scapularis*, *D. pulex*, *Capitella* sp., *L. gigantea*, *S. purpuratus*, *S. kowalevskii*, and *B. floridae* (see "Materials and methods"). Because many of these genomes are not assembled, it is not possible to reliably identify and map all non-miRNA reads to identify the genic source of the transcript. However, we did not find any conserved transcripts that were not either a miRNA or a degraded piece of a known noncoding RNA (e.g., rRNA, tRNA, etc.). Beyond miRNAs, there does not appear to be any other type of small noncoding RNAs conserved across these taxa, given the length and structural restrictions imposed by our protocols (see "Materials and methods").

Using this approach, we mapped the phylogenetic distribution and evolution of all previously described miRNA families that evolved deep in animal history, as well as discovered 41 new miRNA families that are shared between two or more species considered herein (the numerous new miRNA families that were restricted to just a single taxon will be reported elsewhere). Most of these novel miRNAs were found to be restricted to taxa that until now have not been examined for their miRNA complements (e.g., Eutrochozoa, Ambulacraria, etc.), and these are described below. Nonetheless, two new miRNAs were discovered for nodes previously characterized: a novel nephrozoan miRNA (miR-2001) was found in virtually all systems *except* the insects and the vertebrates, and a novel protostome miRNA (miR-1993) was found in all protostome taxa *except* arthropods (Fig. 1A). The alignments of the mature sequences and an example of a structure for all discovered miRNAs reported herein, in addition to the source of evidence of the miRNA for each taxon considered (whether found in the genome, recovered from our small RNA libraries, or both), are given in supporting information Table S1.

Continuous addition of miRNAs with minimal substitutions and rare secondary loss

Figure 2 shows the acquisitional history of the previously identified and newly discovered miRNA families, including the conserved expressed star sequences, throughout 18 metazoan taxa. These same data are also given in Table 3 (and the details of all data are given in supporting information Table S1). Aside from Deuterostomia, all nodes examined are characterized by the acquisition of at least one new miRNA family (shown in blue in Fig. 2). Eutrochozoans added two novel miRNA families (miR-1992 and miR-1994; see Fig. 3A) and the expression of the star sequence of miR-1175 (= miR-958; Table 3) at equal frequency to the mature sequence (Fig. 1B). miR-1175* reads were not found in the sister lineage, the priapulid *P. caudatus* (but it was expressed at roughly 6% of the mature sequence in *D. melanogaster*; Okamura et al. 2008). Neotrochozoans added one novel family (miR-1989; Fig. 1C) and the expression of the star sequence of miR-133 (Fig. 1B), and again star reads were not found in the sister

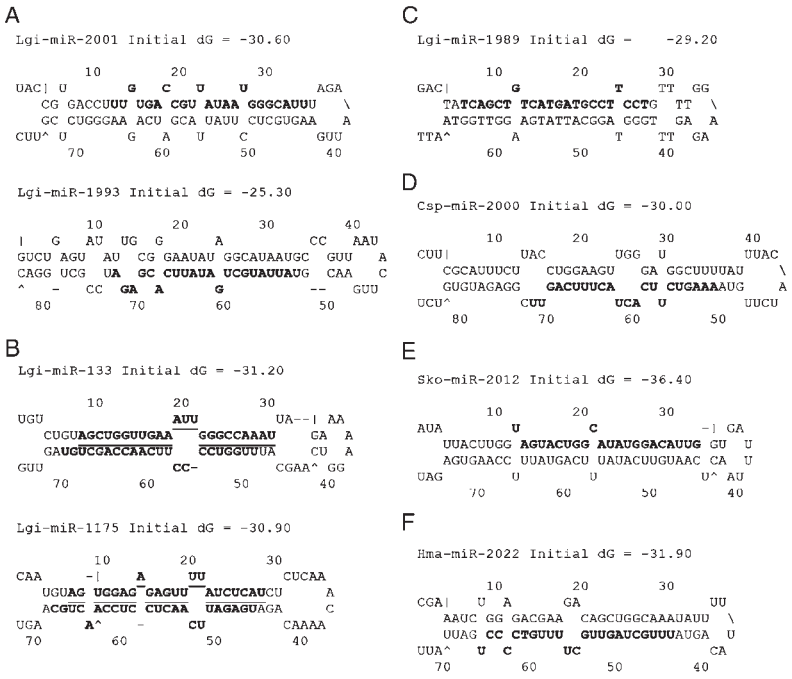


Fig. 1. Examples of novel microRNAs (miRNAs). The mature sequences are shown in bold. (A) Two novel miRNAs for nodes previously characterized, Nephrozoa (miR-2001, top) and Protostomia (miR-1993, bottom). Both of these genes have been secondarily lost in all of three miRNA model systems (dipteran flies, vertebrates, and *Caenorhabditis elegans*). (B) Examples of conserved expression of star sequences (bold underline) for the neotrochozoan miR-133 and the eutrochozoan miR-1175. (C–F) Examples of clade-specific miRNAs: Neotrochozoa (miR-1989, C), Annelida (miR-2000, D), Ambulacraria (miR-2012, E), and Cnidaria (miR-2022, F). See also supporting information Table S1 for examples of structures for all miRNAs reported herein. Taxonomic abbreviations: Csp, *Capitella* sp.; Hma, *Hydra magnipapillata*; Lgi, *Lottia gigantea*; Sko, *Saccoglossus kowalevskii*.

taxon, the nemertean *C. lacteus*. Both of the clades Annelida (Fig. 1D) and Gastropoda (Fig. 3B) added seven novel families each, two of which in the gastropod are derived from a single locus (miR-1990). Ecdysozoa added one previously characterized family (miR-993), plus a paralog of an existing family (miR-13, a member of the miR-2 family). Arthropods added four known families, and pancrustaceans one known family. Although no known families characterize Deuterostomia, within deuterostomes ambulacrarians acquired five novel families (Fig. 1E), two of which were derived from the same locus (miR-2008), and eleutherozoan echinoderms acquired 10 novel families, including two from a single locus (miR-2006) and one star sequence (miR-2011*). Finally, chordates and cnidarians are characterized by the addition of four known families and one novel family (miR-2022; Fig. 1F), respectively.

Importantly, there are only 11 suggested losses of miRNA families (not expressed in a small RNA library and not detected in genomic traces) shared by at least two of the taxa considered herein, seven of which were lost in the arthropods and the chordates (Fig. 2, red; Table 3). In addition to miR-2001 (Fig. 1A), both arthropods and chordates each lost miR-242, a miRNA originally described from nematodes but found here in the ambulacrarians and neotrochozoan taxa. There are only two suggested losses in eutrochozoans. First, in the two annelid taxa analyzed, we find no evidence for the presence of the protostome-specific miRNA miR-76, a miRNA found in the genomes of *D. pulex* and *L. gigantea*, and in the library reads of the nemertean *C. lacteus*. And second, in the two gastropods we found no evidence for the presence of miR-365, a miRNA family thought to be vertebrate specific (Heimberg et al. 2008), but discovered here in the ambulacrarian and annelid systems (and secondarily lost in the ecdysozoans as well). The same miRNA being lost in multiple taxa was also seen with miR-315, a gene secondarily lost in the echinoderms and in the chordates. In fact, four miRNAs (miR-242, miR-315, miR-365, and miR-2001) were each lost twice independently, accounting for eight of the 11 secondary losses. Thus, secondary loss is skewed both taxonomically and toward particular miRNAs.

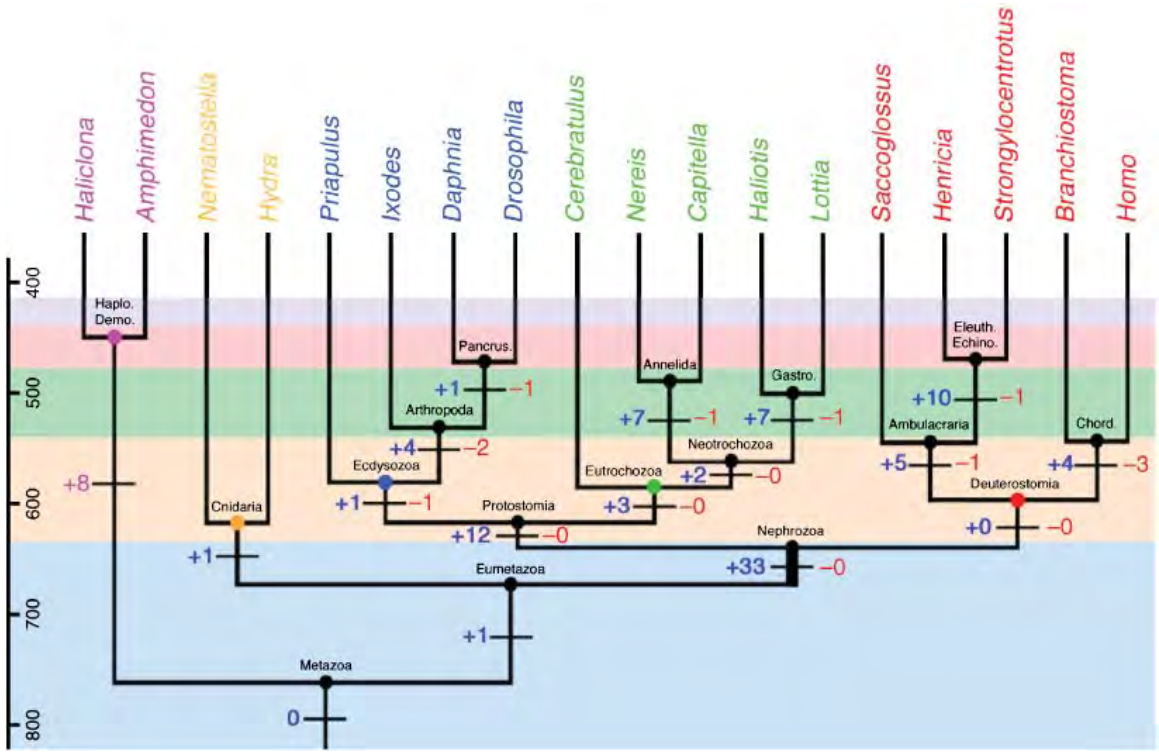


Fig. 2. Continuous acquisition of microRNA (miRNA) families with minimal secondary loss. The acquisition (blue) and secondary loss (red) of miRNA families are shown at each node. Sponges are shown in magenta, cnidarians in orange, ecdysozoans in blue, eutrochozoans in green, and deuterostomes in red. All nodes are also labeled (Haplo. Demo., Haplosclerid Demosponges; Pancrus., Pancrustacea; Gastro., Gastropoda; Eleuth. Echino., Eleutherzoan Echinoderms; Chord., Chordata; see Table 1). Divergence times are taken from a molecular clock in conjunction with the fossil record (Peterson et al. 2008). Because the divergence time of acoel flatworm is not known, they are not shown in this figure. However, their miRNA complement is entirely consistent with them being the sister taxon of the nephrozoans (indicated with a thick line; see text). Note that the phylogenetic topology (Peterson and Eernisse 2001), although derived independently from miRNAs, is entirely consistent with the acquisitional history of novel miRNA families, with only Deuterostomia not characterized by the presence of at least one novel miRNA family. Geological ages are as follows: Cryogenian (ice blue—850–635 Ma); Ediacaran (brown—635–542 Ma); Cambrian (dark green—542–488 Ma); Ordovician (salmon pink—488–444 Ma); and Silurian (lavender—444–416 Ma).

Star sequences

The importance of star sequences, those sequences that are the rarer partner of the mature sequence, has recently been recognized, both in terms of increasing the complexity of regulatory networks and in governing miRNA and messenger RNA evolution (Okamura et al. 2008). In addition to the star sequences of miR-133 in neotrochozoans and miR-1175 in eutrochozoans (Fig. 1B), we found several other instances of relatively high expression (>0.1% of the mature level) of star sequences (see supporting information Table S1 for all cloned instances of star sequences). For example, *Haliotis* expressed the star sequences of both miR-33 and miR-281, the former at 156% of the mature, and the latter at 6% of the mature. The starfish *Henricia* expressed the star sequence of miR-200 at 12% mature level whereas the sea urchin *S. purpuratus* expressed this same star sequence at 0.4% mature level. Further, four of the novel miRNAs discovered herein are derived from two independent loci: miR-1990 in the gastropod *H. rufescens* and miR-2006 in the echinoderms. The other two instances of expressed star sequences in the echinoderms, miR-2008 and miR-2011, are discussed below. As expected (Okamura et al. 2008), these conserved star sequences were almost perfectly conserved in sequence, similar to mature sequences.

Isolation of miRNA loci in a nongenome taxon

Because the miRNA complements of six of the taxa considered herein were derived solely from small RNA library reads, we asked if at least a few of these miRNAs do indeed derive from a miRNA locus as they do in the sister species with a sequenced genome. We chose to examine the novel

Table 3. Evolutionary acquisition of miRNA families

Taxon	miRNA family gains[1]	Inferred miRNA family losses[2]	Total number of miRNA families
Haplosclerida	8:2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021	?	8
Eumetazoa	1: (10, 100)	?	1
Cnidaria	1: 2022	0	2
Triploblastica	8: 1, 31, 34, 79, 92, 124, 219, (252a, 252b)	0	9
Nephrozoa	26: let7, 7, (8, 141, 200), 9, (22, 745, 980), (29, 83, 285), 33, 71, (96, 182, 183, 263), (125, lin4), 133, 137, 153, 184, 190, 193, 210, (216, 283, 747), 242, 278, 281, 315, 365, 375, 2001	0	34
Protostomia	12: Bantam, (2, 13), 12, (67, 307), (76, 981), 87, 277, 279, 317, 750, (958, 1175), 1993	0	46
Eutrochozoa	3: 958*, 1992, 1994	0	49
Neotrochozoa	2: 133*, 1989	0	51
Annelida	7: 1987, 1995, 1996, 1997, 1998, 1999, 2000	1: 76	57
Gastropoda	7: 1984, 1985, 1986, 1988, 1990-5p, 1990-3p, 1991	1: 365	57
Ecdysozoa	1: 993	1: 365	46
Arthropoda	4: 275, 276, iab4-3p, iab4-5p	2: 242,1993	48
Pancrustacea	1: 965	1: 2001	48
Deuterostomia	0	0	34
Ambulacraria	5: 2008-3p, 2008-5p, 2011, 2012, 2013	1: (216, 283)	38
Eleutherozoa	10: (8, 141, 200)*, 2002, 2004, 2005, 2006-3p, 2006-5p, 2007, 2009, 2010, 2011*	1: 315	46
Chordata	4: (103, 107), 129, 135, 217	3: 242, 315, 2001	35

[1]Families are designated parenthetically. In some cases the same gene was given at least two different names (e.g., miR-22 = miR-745 = miR-980), whereas in other cases there were gene duplications generating at least two copies of the gene in an individual taxon's genome (e.g., miR-10 family; miR-252 family; miR-96 family). See supporting information Table S1 for gene duplications versus gene nomenclature problems—each individual family member is given a unique row (e.g., miR-96, miR-182, miR-183), whereas nomenclature mistakes are combined into a single row with all the names given for this gene indicated at the top (e.g., miR-182 = miR-263b).

[2]Question marks (?) indicate that it is not possible at the moment to reconstruct losses for this node.

eutrochozoan miRNA genes (miR-1992 and miR-1994) in the nemertean *C. lacteus*, and the novel gastropod-specific miRNA genes (miR-1984, miR-1985, miR-1986, miR-1988, miR-1990, and miR-1991) in the gastropod mollusc *H. rufescens*. Using the mature sequence as the gene-specific primer (see "Materials and methods"), we were able to amplify the precursor hairpin sequence for both of the eutrochozoan-specific miRNA genes from the genome of *C. lacteus* (Fig. 3A), and for all six of the gastropod-specific miRNA loci from the genome of *H. rufescens*, two of which (miR-1985 and miR-1986) are shown in Fig. 3B. Thus, at least in these eight cases, the library reads do indeed derive from a *bona fide* miRNA locus.

Substitutional profile of miRNAs through time

We next explored the mutational history of mature miRNA sequences. Ninety-three miRNA genes, which were shared among at least two of the 14 nephrozoan taxa considered herein, were systematically analyzed for substitutions and insertion/deletion events (see "Materials and methods"). Amazingly, the substitution rate of all known and novel miRNAs across these 14 taxa, whose independent evolutionary history spans just over 7800 million years, is only 3.5% (584 total substitutions out of 16,729 nucleotides analyzed). Moreover, there were only 21 indel events, and all were localized in the middle or in the 3′ half of the miRNA. To compare this rate with 18S rDNA, one of the most conserved genes in the metazoan genome, we aligned this gene from the same 14 taxa, and with the unalignable regions removed, calculated a substitution rate of 7.3%. Hence, miRNAs evolve more than twice as slowly as the most conserved positions in a gene that is often used for reconstructing the deepest nodes in the tree of life.

In terms of the spatial location of substitutions within the mature gene product, most changes occurred at the 3′ end of the mature sequence as expected, but other regions of the gene, especially nucleotide 1 and nucleotide 10, showed a relatively high percentage of substitutions (Fig. 4, top). The two most infrequent places for substitutions to occur are the seed region (positions 2–8) and the 3′ complementarity region spanning nucleotides 13–16, especially position 15. Therefore, the mutational profile of the miRNA mature sequence matches the importance of these two regions to base pair with the 3′UTR of targets (Grimson et al. 2007; Filipowicz et al. 2008) (Fig. 4, bottom).

"Seed shifts"

Despite the extreme conservation of miRNAs documented above, diversity of miRNA sequence is still achieved by several different types of processes. Because modifications to the 3′ end have already been extensively documented (Landgraf

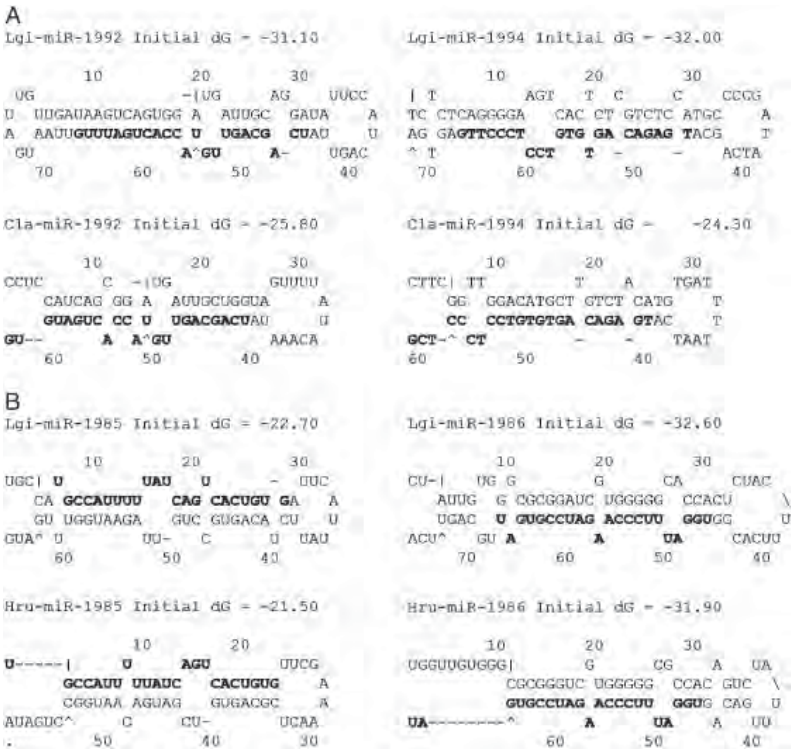


Fig. 3. Using "Genome-walking" to clone microRNA (miRNA) loci from taxa without a sequenced genome. Using the mature sequence as a primer (bold, see "Materials and methods") we isolated two genomic hairpin structures from the nemertean *Cerebratulus lacteus* (Cla), miR-1992 and miR-1994 (A), and six genomic hairpin structures from the gastropod mollusc *Haliotis rufescens* (Hru), two of which are shown (miR-1985 and miR-1986) (B). Note that these sequences fold into canonical hairpins with significant similarity to their respective orthologs in *Lottia gigantea* (Lgi).

et al. 2007; Ruby et al. 2007) (and confirmed here for numerous miRNAs for almost all taxa investigated, but the details will be reported elsewhere), we chose to focus instead on conserved modifications to the 5′ end, a more interesting category as changes to this end of the mature miRNA change the identification of the seed sequence (positions 2–8) and presumably the target profile of the miRNA (Habig et al. 2007; Kawahara et al. 2007, 2008). We found several instances where the mature sequence of a given miRNA is moved either 1–2 nt 3′ or 1–2 nt 5′, is phylogenetically conserved between two or more taxa, and is evolutionary derived. We call these evolutionary stable movements "seed shifts" because the determination of position 1 of the mature sequence, and hence positions 2–8 of the seed, is moved either 5′ or 3′ (Fig. 5A).

Two different types of seed shifts were found. First, a dominant transcript from a single locus that differs in the position of nucleotide 1 with respect to other taxa was seen with, for example, miR-22. In the two echinoderm taxa, a single nucleotide has been added 5′ so that the original position 1 was now position 2 (Fig. 5A). miR-22 also serves as an example of the second type of shift, a shift in the seed sequence of a paralogous gene that differs in the determination of position 1 with respect to other copies of the gene, both in the same taxon as well as in other taxa. Both the annelid *Capitella* and the mollusc *Lottia* have two miR-22 (= miR-745; Table 3) loci. In *Capitella*, *Nereis*, and *Haliotis*, miR-22a is seed shifted 3′ so that ancestral position 2, which characterized miR-22b, was now position 1 (Fig. 5A). In both cases, concomitant changes were seen at the 3′ end so that the length of the mature read remained approximately 22 nt (Fig. 5A). Interestingly, changes to the establishment of position 1 can be accounted for by simply noting that in all cases the mature sequence starts within about 2 nt of the loop (Fig. 5A, bold nucleotides), suggesting that structural considerations might partially underlie these shifts, given what is known concerning nuclear miRNA processing (Kim 2005). Other instances of shared seed shifts included miR-210 in the two annelid species, miR-133 in the two echinoderm species, a miR-2 paralog in the three eutrochozoan phyla, and a miR-10 paralog in the two neotrochozoan phyla (see supporting information Table S1).

5′ end editing

5′ end posttranscriptional editing was also observed, such that a single locus (as assessed by querying genomic trace archives or genome walking) gives rise to at least two different tran-

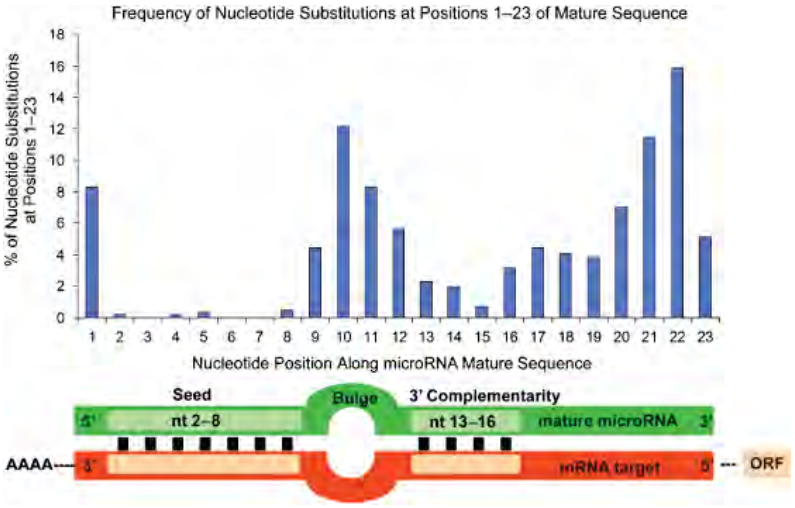


Fig. 4. Mutational profile of microRNA (miRNA) mature sequences. Shown at the top is the percentage of nucleotide substitutions from positions 1 to 23 (for the few miRNAs with longer mature sequences these additional nucleotides are not shown). Note that two areas in particular have relatively high conservation, the seed (positions 2–8) and positions 13–16. This could be explained by the interaction between the mature sequence of the miRNA and its targets (bottom) (Grimson et al. 2007; Filipowicz et al. 2008).

scripts with different seed sequences. For example, in the two annelid taxa, two different mature miR-67 transcripts were cloned, and they differ only with respect to the start of the mature read, one with the 5′ uracil, the primitive condition, and a second read that is missing the 5′ uracil and instead starts with the cytosine, the ancestral position 2 (supporting information Table S1). In both taxa, both sequences are the same outside of this shift, but the former was more highly expressed than the latter (*Nereis*: 78 reads vs. 6; *Capitella*: 60 vs. 1). Two very interesting examples of this sort of 5′ editing were seen with two of the miRNAs specific to ambulacrarians. Both miR-2008 and miR-2011 are 5′ edited such that, like miR-67 in the two annelid taxa, two different mature transcripts were derived from a single genomic locus. But in these two cases the star sequence (or other arm of the hairpin) was also expressed, resulting in the expression of at least three different gene products (supporting information Table S1).

A second type of 5′ edit was seen whereby the gene is characterized by an adenine in the seed region whereas the library read has instead a guanine in this same position. This is the hallmark of inosine editing, whereby adenine is posttranscriptionally edited to an inosine, a reaction catalyzed by adenosine deaminases and known to occur in miRNA sequences, both pre- and mature (Luciano et al. 2004; Blow et al. 2006; Kawahara et al. 2007). We found a single instance of this phenomenon that involved the seed region of a miRNA and was phylogenetically conserved, miR-200 (= miR-8; Table 3) in the deuterostomes (Fig. 5B). In all four deuterostome taxa, one set of mature sequences had a guanine in position 5 whereas in another set of mature sequences position 5 was characterized by an adenine, as were all genomic miR-200 loci (Fig. 5B, arrowhead). As expected from a statistical analysis of A-I edits (Kawahara et al. 2008), the edited adenine is 3′ of a uracil and mismatched with a cytosine (supporting information Table S1). This edit, because it occurs in the seed sequence, would be expected to alter the downstream repertoire of miR-200 in these taxa (Kawahara et al. 2007, 2008).

Interestingly, unlike the hemichordate though, in *B. floridae*, *S. purpuratus*, and *H. sanguinolenta*, this inosine edit was associated with an *insertion* of an adenine between positions 1 and 2, as assessed from the more highly expressed version of the miRNA gene and the genomic sequence from each respective taxon (Fig. 5B, arrow), and with respect to other modified miRNAs. In fact, both echinoderm taxa expressed a mature sequence of miR-200 that had the insert, but was not inosine edited (Fig. 5B mature, read [M] 2). Further, in both echinoderms the star sequence was also expressed (supporting information Table S1), and thus, the sea urchin, for example, appears to express four different transcripts, each with a unique seed sequence, from a single miR-200 gene.

Similar to miR-200, and unlike the described situation with the mouse let-7 gene, which showed constrained editing in the seed region (Reid et al. 2008), a fair number of the miRNA transcripts sequenced herein had an adenine insertion in the seed sequence, usually inserted between positions 1 and 2. Examples included miR-71, miR-184, miR-2002, and let-7 (supporting information Table S1). In all cases, these transcripts were sequenced more than once in at least one taxon, suggesting that these adenine insertions were not sequencing errors. Further, the fact that in no case was a corresponding copy found in any genome, whether *S. purpuratus*, *B. floridae*,

or *Capitella* sp., suggests that these inserted sequences are not coming from some unsequenced genomic locus. Consistent with this idea is the fact that in the star fish both mature sequences of miR-2002 had identical sequences, but with two nucleotides that differ with respect to *S. purpuratus*, coupled with an indel (Fig. 5C, bracket). If there were two copies of miR-2002 in the genomes of these two taxa, then both copies would have to have changed in exactly the same way to accommodate both the same two nucleotide substitutions (either T to C in the star fish, or C to T in the sea urchin), and the indel event, which is unlikely. Further, using genome walking to amplify miR-2002 from the genome of the sea star *H. sanguinolenta* (see "Materials and methods"), we again only find a single locus orthologous to the single gene found in the *S. purpuratus* genome (Fig. 5C).

Although an insertion between positions 1 and 2 was the most common insertion edit, two genes showed an adenine insertion between positions 3 and 4—miR-29 and miR-31 (supporting information Table S1), and one gene, miR-10, showed an adenine insertion between positions 6 and 7 (Fig. 5D). miR-10 was particularly interesting as it showed the plethora of ways a taxon can increase its miRNA diversity. For example, amphioxus has duplicated this gene several times to give rise to three paralogs; one copy (miR-10a) is seed shifted such that one set of mature transcripts had a guanine 5′ of ancestral position 1 (Fig. 5D, double arrowhead). This same locus is also 5′ edited such that a second set of transcripts (Blfa-M2) starts at the ancestral position 1. A second gene (miR-10b) is also 5′ edited with one set of mature transcripts having an adenine inserted between positions 6 and 7, similar to what was found in the sea urchin and star fish (Fig. 5D, arrow) (as well as the priapulid; supporting information Table S1). Thus, amphioxus is able to generate at least five different miR-10 transcripts from three different miR-10 genes. The sea urchin showed three different mature sequences (similar to the sea star) from a single miR-10 locus, with the third set of reads characterized by a cytosine to uracil edit (shown as a thymidine in Fig. 5D, arrowhead) at position 4 in the seed sequence. Cytosine to uracil editing is well known and associated with know flanking sequences around the edit (Blanc and Davidson 2003), which appear to be present in the miR-10 locus of the sea urchin as well (supporting information Table S1), but we have no experimental data supporting this hypothesis. We also observed other types of nucleotide edits including a guanine to uracil edit, again at position 4, in let-7 of amphioxus, starfish, sea urchin, *H. rufescens*, and *Capitella* (supporting information Table S1), but to our knowledge this type of edit has not been experimentally verified.

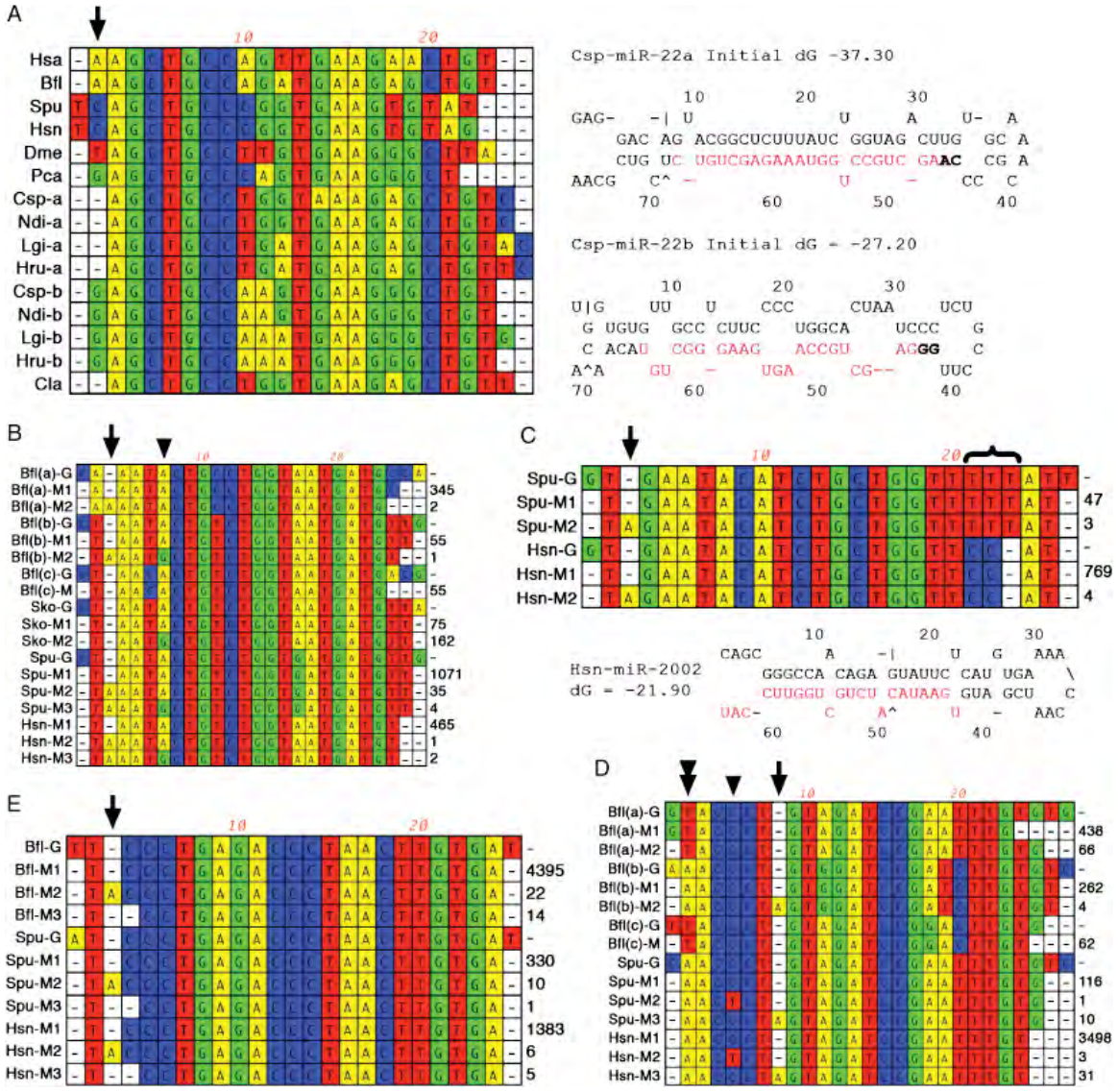
Finally, a few miRNAs showed evidence of nucleotide deletions in the seed region. Again, using the deuterostomes as an example, amphioxus and the two echinoderms

Fig. 5. Seed shifting and 5′ editing. (A) Examples of seed shifts with the miR-22 gene. Assuming the adenine is the ancestral position 1 (arrow), which is the most parsimonious interpretation of the data, the two echinoderm taxa have added an additional nucleotide 5′ so that the ancestral position 1 is now position 2. The annelids and the molluscs each have two copies of miR-22 (= miR-745; Table 3). One copy (paralog "a") is seed shifted 3′ so that ancestral position 2 is now position 1 (as is the nermertean [Cla]), whereas the second copy ("b") shows the presumed ancestral condition. Note that there is an accompanying shift to the 3′ end keeping the length of the mature read approximately 22 nt long. Despite these seed shifts, in all cases the mature read (red) starts within about 2 nt (bold in structures) from the loop, suggesting that structural considerations might be at least partially underlying these shifts. (B–E) 5′ editing. Genomic reads are indicated with a "G" and library reads are indicated with an "M". The number of reads for each mature sequence is given on the right side of each alignment; genomic reads are indicated with a dash. (B) Inosine editing and nucleotide insertions associated with miR-200 in deuterostomes. All four deuterostome taxa have mature sequences (with the number of occurrences in our small RNA libraries indicated on the left of the alignment) characterized by a guanine at position 5 rather than the adenine, which was found in the genomic data for all sequenced taxa (arrowhead). Further, amphioxus *(Branchiostoma)*, the sea urchin *(Strongylocentrotus)* and the sea star *(Henricia)* also have mature sequences with both the inosine edit (arrowhead) and an adenine inserted between positions 1 and 2 (arrow, as determined with comparison with other miRNAs, see the text and panels C and D), as well as another set of mature sequences that show the adenine insertion (arrow), but not the inosine edit (arrowhead). (C) Despite finding only a single miR-2002 locus in the genome of the sea urchin (Spu-G), and amplifying only a single locus from the sea star (Hsn-G, see structure below the alignment) two different mature sequences (M1 and M2) were found in the small libraries from each taxon, and in both cases the second mature sequence has an adenine inserted between positions 1 and 2. Although there could be two loci in these genomes, note that this would require both copies in one of the two taxa to make the same substitutions near the 3′ end—either the two T's were changed to C's or vice versa. Moreover, this would also require an indel event in the same area in both copies (Fig. 5C, bracket). We instead suggest that this in an example of 5′ editing with an adenine inserted into the transcript posttranscriptionally. (D) Seed shifts, nucleotide editing, and adenine insertions associated with miR-10 of deuterostomes. Amphioxus has three copies of miR-10, one of which (a) is seed shifted (ancestral position 1 is indicated with the double arrowhead) and the second (b) shows an adenine insertion between nucleotides 6 and 7 (arrow). Further, both the sea urchin and sea star show a cytosine to uracil edit at position 4 (arrowhead, shown as a thymidine on the figure). (E) Nucleotide insertions and deletions associated with miR-125 in deuterostomes. Amphioxus, sea urchin, and the sea star each have a set of mature sequences with an adenine insertion (M2) and a cytosine deletion (M3) (arrow). Although position 2 is reconstructed as the deletion, it is formally possible that either cytosine 2, 3, or 4 is deleted. Taxonomic abbreviations: Bfl, *Branchiostoma floridae*; Csp, *Capitella* sp.; Cla, *Cerebratulus lacteus*; Dme, *Drosophila melanogaster*; Hru, *Haliotis rufescens*; Has, *Homo sapiens*; Hsn, *H. sanguinolenta*; Lgi, *Lottia gigantea*; Ndi, *Nereis diversicolor*; Pca, *Priapulus caudatus*; Sko, *Saccoglossus kowelevskii*; Spu, *Strongylocentrotus purpuratus*.

showed evidence for both nucleotide insertion (adenine between positions 1 and 2 as above; Fig. 5E, arrow) as well as nucleotide deletion (Fig. 5E, arrow) in the miR-125 mature sequences. Thus, a variety of transcripts, each with a different seed sequence, are generated from a single genic locus using a combination of nucleotide indels and nucleotide editing, and presumably each can regulate a unique and potentially nonoverlapping set of target mRNAs (Blow et al. 2006; Kawahara et al. 2007, 2008; Ruby et al. 2007).

Expansion of the nephrozoan repertoire of miRNAs and miRNA families

It had previously been recognized that there was a major expansion of the miRNA family-level complement at the base of the nephrozoans (Hertel et al. 2006; Sempere et al. 2006; Prochnik et al. 2007), which is confirmed here. We find evidence for the presence of 34 miRNA families in the last common ancestor of protostomes and deuterostomes, and because we controlled for time, rates can be determined for each of the nodes considered herein. Importantly, aside from



the base of the vertebrates (Heimberg et al. 2008), the rate of acquisition at the base of the nephrozoans is at least five times, and often an order of magnitude or more, higher than anywhere else on the tree (Fig. 2). The contrast is especially striking when one compares Nephrozoa versus Cnidaria, two sister lineages that diversified roughly at the same point in geologic time, near the beginning of the Ediacaran 635 Ma (Peterson et al. 2008) (Fig. 2)—the former is characterized by the addition of 33 novel families since its divergence from cnidarians for a rate of one novel miRNA family per million years, whereas we find evidence for only a single novel family shared between the two cnidarian taxa *Nematostella* and *Hydra* (miR-2022, Fig. 1F) for a rate of 0.03 miRNA families per million years.

Two of these novel nephrozoan miRNA families duplicated to give rise to multiple paralogs by the time protostomes diverged from deuterostomes. First, the miR-96 family, which consists of miR-96, miR-182, miR-183, and miR-263 (Pierce et al. 2008), had three copies present in the last common ancestor of nephrozoans: miR-96, miR-182 (= miR-263b), and miR-183 (= miR-263a, miR-228), and all three genes were found in several taxa including the ambulacrarians, the neotrochozoans, *Branchiostoma*, and *Daphnia* (supporting information Table S1). Further, our data suggest that miR-96 was lost in the fly lineage after it split from *Daphnia* as orthologs were cloned or found in all of the ecdysozoan taxa analyzed herein (supporting information Table S1). The second example is miR-252. There was only one copy of miR-252 in the genome of both the fly and *C. elegans* (and no copies in any known vertebrate), but interestingly these represent each of the two copies of the original miR-252 gene such that the copy found in fly (here called miR-252a) was found throughout Nephrozoa including *Branchiostoma* (supporting information Table S1), but was not found in the nematode, and the copy found in the *Caenorhabditis elegans* (here called miR-252b) was again found throughout Nephrozoa, including *Branchiostoma* and all of the ecdysozoans systems explored herein (supporting information Table S1), but was not present in the fly. With these gene duplications, coupled with the gene duplication leading to the origin of miR-10 from miR-100 (which was the only known miRNA found or cloned in either cnidarian), gives a total of 38 evolutionary long-lived miRNA genes present in the last common ancestor of nephrozoans, compared with only two (miR-100 and miR-2022) in the last common ancestor of cnidarians.

Conserved miRNAs in acoel flatworms

Sempere et al. (2006, 2007) explored the miRNA complement of acoel flatworms via northern analysis and saw evidence for five miRNAs families: miR-31, miR-34, miR-92, miR-124, and miR-219. However, because just a few mutations to the primary sequence would dramatically affect the hybridization kinetics, and hence potentially result in nondetection of the mature miRNA (Sempere et al. 2006; Pierce et al. 2008), we built a small RNA library from the acoel flatworm *S. roscoffensis* to more fully characterize its conserved miRNA complement. Except for miR-34, transcripts of all of these miRNAs were detected in the small RNA library of *S. roscoffensis*. We also found evidence for only three additional known miRNA families in our small RNA library: miR-1, miR-79, and miR-252b. Both miR-1 and miR-79 could not be detected with standard probes as there are several changes to the mature sequence in both the middle and at the 3′ ends (supporting information Table S1); miR-252 was not examined by Sempere et al. (2006, 2007) as it is not present in vertebrates and was only recently discovered in fly (Ruby et al. 2007; Stark et al. 2007). Hence, of the 33 novel miRNA families found in nephrozoans as compared with cnidarians, eight are found in the relatively simple acoel flatworms. Further, if the absence of miR-252a is confirmed, and is indeed a primitive absence (vs. a secondary loss), then this suggests that the gene duplication event giving rise to two copies in the last common ancestor of nephrozoans occurred after acoels split from the nephrozoan branch. Together these data are consistent with the hypothesis that acoels branched off from the remaining triploblasts before the last common ancestor of protostomes and deuterostomes (Baguñà et al. 2008) (thick branch on Fig. 2).

Novel miRNAs in demosponges

No miRNAs were held in common between the sponges and eumetazoans, as expected (Sempere et al. 2006; Prochnik et al. 2007). Nonetheless, eight miRNAs were discovered that were shared between the two haplosclerid demosponges under consideration, *Haliclona* sp. and *A. queenslandica*. The structures of two of these novel miRNAs are shown in Fig. 6A with the mature sequence cloned out of the *Haliclona* miRNA library in bold (all eight are given in supporting information Table S1). Like eumetazoan miRNAs, the sequences are highly conserved—in seven of the eight miRNAs there are no substitutions in the mature sequence between *Haliclona* and *Amphimedon* despite sharing a last common ancestor some 450 Ma; only miR-2021 had a single substitution near the 3′ end (supporting information Table S1). Nonetheless, these miRNAs are structurally very different from all known drosha-processed eumetazoan miRNAs—in eumetazoans the mature sequence is usually within a few nucleotides of the loop (Kim 2005) (see Figs. 1, 3, and 5), whereas in all eight haplosclerid miRNAs the mature sequence was at least 10 nt, and often more than 30 nt, away from the loop (Fig. 6A). Further, the position of the mature sequence within the hairpin structure varied with respect to

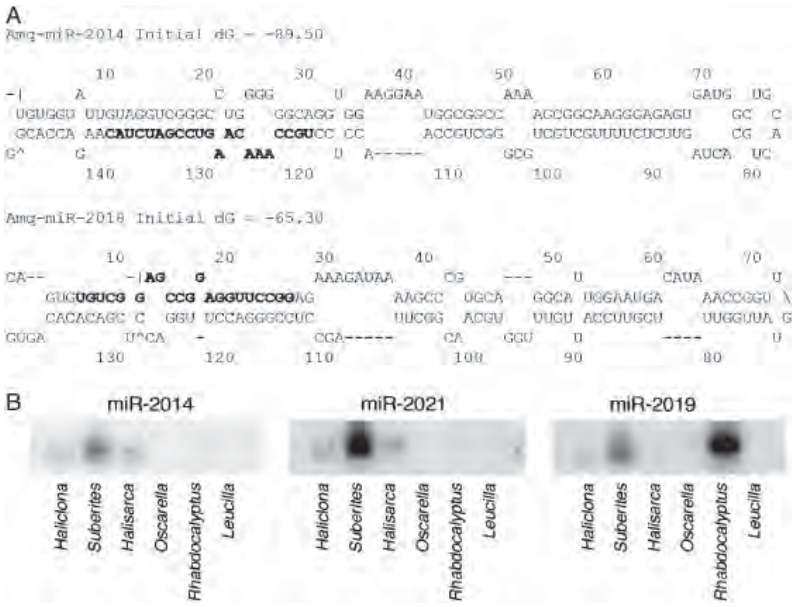


Fig. 6. Novel microRNA (miRNA) families in demosponges. (A) Two examples of representative miRNAs cloned out of a small RNA library from the demosponge *Haliclona* sp. and then found in the genomic traces of *Amphimedon queenslandica* (Amq). These structures are unusual with respect to eumetazoan miRNAs (compare with Figs. 1, 3 and 5, and supporting information Table S1) in that the mature sequence (shown in bold) is far removed from the loop (approximately 33 nucleotides (nt) in the two structures figured) as opposed to approximately 2 nt from the loop in eumetazoan drosha-processed miRNAs (Kim 2005). The mature sequence, however, is always approximately 22 nt long. (B) Representative northern analyses of novel demosponge miRNAs throughout "Porifera." miR-2014 and miR-2021 were detected in all three demosponges (*Haliclona*, *Suberites*, *Halisarca*), but was not detected in other sponge taxa including the homoscleromorph (*Oscarella*), the hexactinellid (*Rhabdocalyptus*), and the calcisponge (*Leucilla*) (left). miR-2019, however, was strongly detected, albeit at a slightly higher size, in the hexactinellid, and not detected in the G2 demosponge *Halisarca*.

the loop among the different miRNAs (supporting information Table S1).

To address the phylogenetic distribution of these eight novel miRNAs, we assayed total RNA from five other sponge species spanning the range of poriferan evolution using northern analysis. Demosponges are divided into four major taxa, G1–G4 (Borchiellini et al. 2004). Both *Haliclona* and *Amphimedon* are members of the G3 clade, and the last common ancestor of all living demosponges would be the last common ancestor of these two G3 taxa, the G4 taxa (represented here by *Suberites*), and the G1+G2 taxa (represented here by the G2 taxon *Halisarca*). The three other major groups of sponges include the homoscleromorphs (represented here by *Oscarella*), hexactinellids (*Rhabdocalyptus*), and calcisponges (*Leucilla*). We detected transcripts from six of the eight novel miRNAs, and in all cases the miRNA was detected in the G4 taxon *Suberites*, and four of these six miRNAs (miR-2014 [Fig. 6B, left], miR-2015, miR-2016, and miR-2021 [Fig. 6B, middle]) were also detected in the G2 taxon *Halisarca*. One novel miRNA was detected in the G3 and G4 taxa, but transcripts were also detected, albeit at a slightly higher size, in the hexactinellid (miR-2019; Fig. 6B, right). Given that the divergences of these sponge lineage all occurred long before the Cambrian (Peterson et al. 2008), these are very ancient and clade-specific miRNAs.

DISCUSSION

As expected (Sempere et al. 2006), the miRNA system, when explored across Metazoa and independent of phylogenetic conservation with model systems, shows both few examples of secondary loss and very low levels of nucleotide substitutions to the primary sequence. Indeed, miRNAs evolve more than twice as slowly as the most conserved positions in one of the most conserved genes in the metazoan genome, 18S rDNA. Despite this conservation, a remarkable diversity of genic products can be produced using a variety of processes including gene duplication followed by seed shifts, and 5′ editing. Importantly, all but one node explored herein can be characterized by the possession of at least one novel miRNA family, and thus miRNAs are continuously being added to

the metazoan genome through geologic time. Further, aside from the base of vertebrates (Heimberg et al. 2008), the extent of the miRNA family-level expansion at the base of the Nephrozoa is unique, at least with respect to the taxa analyzed herein (Hertel et al. 2006; Sempere et al. 2006, 2007; Prochnik et al. 2007). Morphologically simple taxa like cnidarians and haplosclerid demosponges evolved relatively few novel miRNA families as compared with the early evolution of the more complex and organ-bearing nephrozoans, which interestingly, at least for Cnidaria, spans about the same interval and the same amount of geologic time (Fig. 2).

miRNAs and metazoan phylogeny

Aside from Deuterostomia, all nodes considered herein are characterized by the possession of at least one unique miRNA family (Fig. 2). We find unequivocal and independent support for the clades haplosclerid demosponges, eumetazoans, cnidarians, triploblasts, nephrozoans, protostomes, ecdysozoans, arthropods, pancrustaceans, eutrochozoans, neotrochozoans, annelids, gastropod molluscs, ambulacrarians, eleutherozoan echinoderms, and chordates, nodes that are generally recovered from both morphological and molecular phylogenetic studies (Peterson and Eernisse 2001; Eernisse and Peterson 2004). Nonetheless, Dunn et al. (2008) have proposed a series of hypotheses based on an extensive EST data set that are in stark contrast to these results. First, these authors suggested that nemerteans and annelids are sister taxa with respect to molluscs. Second, they suggested that acoel flatworms are lophotrochozoans closely related to rotifers and other platyzoans including the true flatworms, the Platyhelminthes. And third, they suggested that cnidarians and sponges are sister taxa. All of these hypotheses are in conflict with the miRNA data obtained herein. First, if nemerteans and annelids are indeed sister taxa, then they must have had (at least primitively) the novel miRNA miR-1989, second copies of miR-10 and miR-22 (both of which have seed shifts; Fig. 5, supporting information Table S1), as well as express the star sequence of miR-133 (Table 3). Further, one would predict that they would also possess a subset of the annelid-specific miRNAs. Instead, we find that nemerteans not only lack these neotrochozoan-specific miRNAs, they also do not share any annelid- (or mollusc)-specific miRNAs, consistent with them being eutrochozoans, but not neotrochozoans (Peterson and Eernisse 2001).

Their second result, the inclusion of acoel flatworms into the platyzoan lophotrochozoans, is dubious given the numerous and independent studies that have found them to be the sister taxon to the nephrozoans (Baguñà and Riutort 2004; Wallberg et al. 2007; Baguñà et al. 2008). If acoel flatworms were indeed nested within the lophotrochozoans, then one would expect them to have at least one protostome-specific miRNA family. Instead, they have only eight of the 33 nephrozoan-specific miRNAs families, and none of the 12 protostome-specific miRNA families. Further, no sequences in our *S. roscoffensis* small RNA library, aside from these known miRNAs and degraded rRNAs, tRNAs, and snoRNAs, appeared in either of the two flatworm genomes, *Schistosoma masoni* or *Schmidtea mediterranea*, a most unexpected result if indeed acoel flatworms are closely related to the true plathelminthes.

As outlined by Sempere et al. (2006, 2007), initial gain followed by some secondary loss and/or high substitution frequency would result in a mosaic pattern of the miRNA complement. This is nicely demonstrated with ascidian urochordates where their phylogenetic position with respect to other invertebrate phyla is unquestioned. Ascidians have a few triploblast, nephrozoan, and chordate miRNAs, but appear to have lost numerous miRNAs as well; importantly though, they also share three miRNAs with vertebrates not seen outside of this clade, miR-126, miR-135, and miR-155 (Norden-Krichmaer et al. 2007; Heimberg et al. 2008). Thus, despite clear secondary loss, ascidians are still correctly reconstructed as the sister taxon of vertebrates based solely on their miRNA complement (Sperling and Peterson in press). In contrast, acoels share with Plathyhelminthes only a small set of the basal complement of miRNA families, the same miRNA genes found throughout ecdysozoans, eutrochozoans, and deuterostomes (Palakodeti et al. 2006; Sempere et al. 2007). Given the strength of the previous molecular phylogenetic results (Baguñà and Riutort 2004; Baguñà et al. 2008), the unequivocal nature of the miRNA data obtained herein, and the fact that another EST analysis found acoels to be basal deuterostomes (Philippe et al. 2007), we would suggest that these EST results are inaccurate reconstructions of metazoan phylogeny.

This inaccuracy is nowhere more apparent than at the base of the metazoan tree. Dunn et al. (2008) suggested that ctenophores are basal metazoans followed by a clade of Porifera+Cnidaria, albeit with low support, as the sister taxon of the triploblasts. Although at times the clade Porifera+Cnidaria is recovered by analyses based on mitochondrial DNA (Lavrov et al. 2008; Erpenbeck et al. 2006), morphological and molecular phylogenetic analyses generally find Cnidaria+Triploblastica (Eernisse and Peterson 2004; Halanych 2004). Indeed, genome-based phylogenetic analyses unequivocally recovered Cnidaria+Triploblastica, a clade called Eumetazoa (Table 1), with respect to the sponge *A. queenslandica* (Putnam et al. 2007; Srivastava et al. 2008). Again, this same clade, Eumetazoa, is recovered from a consideration of miRNAs alone—cnidarians share miR-100 with triploblasts, but cnidarians do not appear to possess any of the demosponge-specific miRNAs, and demosponges do not possess the cnidarian-specific miRNA miR-2022 nor any other eumetazoan miRNA.

More importantly, the miRNAs we recovered from the two haploscerlids, and that showed broad conservation throughout demosponges, are unique both in sequence and in secondary structure with respect to the miRNAs recovered from eumetazoans. In eumetazoans, Drosha measures approximately 22 nt from the loop, and through its action precisely determines one end of the mature miRNA (Kim 2005). Hence, the mature sequence resides within approximately 2 nt of the loop (see Figs. 1, 3, and 5, and supporting information Table S1). However, the miRNAs recovered from the two haplosclerids resemble plant miRNAs (Reinhart et al. 2002; Jones-Rhoades et al. 2006) in that the mature gene sequence is often 30 or more nucleotides away from the loop, and no miRNAs were found in sponges whose mature sequence was within 10 nt of the loop. We did not recover a single potential example of a eumetazoan miRNA whose mature sequence lay beyond just a few nucleotides of the inferred loop, in contrast to what we found with *all* of the demosponge-specific miRNAs. Because of the pronounced difference in the pre-miRNA structures between demosponges and eumetazoans, suggestive of significant differences between the nuclear processing machinery of the two, and indeed perhaps indicative of independent origins of their respective miRNAs, suggests that a sister group relationship between cnidarians and sponges is highly unlikely.

miRNAs and metazoan complexity

The continuous acquisition of miRNAs through geologic time in all metazoan lineages is unique with respect to transcription factors. Metazoan transcription factor families, whether it be *Fox*, *Hox*, *Sox*, or *T-box*, are ancient innovations (Wray et al. 2003), which were all present in the metazoan genome by the time the last common ancestor of demosponges and eumetazoans evolved (Peterson and Sperling 2007; King et al. 2008; Larroux et al. 2008). And aside from a few family-level expansions (e.g., *Hox*, Chourrout et al. 2006), few, if any, transcription factor families have evolved within Metazoa. miRNAs, on the other hand, show a fundamentally different pattern: no families are known to exist in the last common ancestor of metazoans, and continuous innovation of novel miRNA families occurred within each metazoan lineage, to say nothing of duplications, seed shifts, and/or 5′ edits of preexisting miRNAs. Thus, transcription factor disparity was generated early in metazoan history, with only increases to diversity occurring over the last 650 Myr, whereas both miRNA diversity *and* disparity continuously increased in every metazoan lineage through geologic time. This would be like only arthropods having *Hox* genes, annelids *Fox* genes, echinoderms *Sox* genes, and vertebrates *T-box* genes—all arbitrary chosen groups of developmentally important transcription factors. Needless to say, our image of how development works, and how it evolves, would be remarkably different if this was indeed the case. Instead, conservation of transcription factors is realized, but the networks continuously acquire lineage-specific miRNAs through time with only rare instances of secondary loss or changes to the sequence of the mature gene product.

Despite differences in mode and location of action (Hobert 2008), the analogy between miRNA and transcription factor families is not superficial or trivial as each is characterized as collections of independently derived monophyletic groups of *trans*-acting factors that recognize distinct sequence-specific *cis* motifs. But because the metazoan genome is continuously acquiring novel miRNAs, the regulatory networks involving both sets of *trans*-acting factors become more complex through time as more and more of the metazoan messenger RNA tool kit comes under the regulatory control of miRNAs, and as new families of miRNAs come under the control of ancient transcription factor families (Chen and Rajewsky 2007; Shalgi et al. 2007).

Not only do nephrozoan genomes, at least, continue to acquire novel miRNA families, they use several other means to increase the miRNA diversity of their developmental tool kit, allowing for even more translational regulation of cellular differentiation and homeostasis. First, star sequences become incorporated into gene regulatory networks, and hence come under the intense negative selection associated with the mature read of the miRNA gene (Okamura et al. 2008). Second, diversity of miRNAs is increased by duplicating previously evolved miRNA genes to generate new paralogs. Although in general these paralogs have the same seed sequence, suggesting that they might not have acquired new targets, two in particular, miR-2 and miR-22, showed evidence of gene duplication followed by a seed shift, whereby the start of the mature sequence was moved either 5′ or 3′, resulting in changing the identity of positions 2–8, which presumably changes the target profile of the miRNA gene. This could effectively create a totally different and nonoverlapping set of targets from the primitive miRNA locus. And third, 5′ editing creates transcripts that differ with respect to their seed sequences suggesting that a single locus can influence the translation of several potentially nonoverlapping sets of target genes depending on which transcript is present at any given time.

Given that morphological complexity also increases through geologic time with the evolution of novel cell types (Valentine et al. 1994), and that miRNAs seem to play important roles in the differentiation and regulation of these cell types (Ambros 2004; Zhao and Srivastava 2007; Hobert 2008; Makeyev and Maniatis 2008; van Rooij et al. 2008; Yi et al. 2008), it seems likely that the evolution of novel miRNA families is intimately tied to the evolution of novel cell types and hence morphological complexity (Sempere et al. 2006; Lee et al. 2007; Niwa and Slack 2007; Peterson et al. 2007;

Heimberg et al. 2008). Whether miRNAs cause the evolution of novel cell types, or whether the appearance of novel cell types exerts a strong selective pressure for the integration of novel miRNAs, remains an open question. Nonetheless, the data reported herein and elsewhere (Hertel et al. 2006; Sempere et al. 2006, 2007; Prochnik et al. 2007; Heimberg et al. 2008) strongly suggest that the relationship between morphological complexity and miRNAs is not simply correlative, but instead is causal, although in a manner that remains to be more fully elucidated with further experimental and theoretical work.

Acknowledgments

We thank B. Wiegmann for introducing K. J. P. to the North Carolina State Bioinformatics group, and E. Seaver for introducing K. J. P. to J. Grassle. We extend our gratitude to L. Sempere for the conception and the sequences of the fluorescein-labeled markers, to V. Ambros for his assistance and advice in making the small RNA libraries, and to all of the people who supplied animals. We thank Dr. Kenneth Nelson and the Yale Center for Genomics and Proteomics Sequencing Facility for sequencing assistance. We also thank P. Donoghue, the two referees, and the members of the Peterson laboratory for their comments on the manuscript. Finally, we thank G. Soukup for his help with the miR-96 family, and for identifying miR-183 in the hemichordate. This work was supported by the National Science Foundation.

REFERENCES

Ambros, V., et al. 2003. A uniform system for microRNA annotation. *RNA* 9: 277–279.

Ambros, V. 2004. The functions of animal microRNAs. *Nature* 431: 350–355.

Baguñà, J., Martinez, P., Paps, J., and Riutort, M. 2008. Back in time: a new systematic proposal for the bilateria. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363: 1481–1491.

Baguñà, J., and Riutort, M. 2004. The dawn of bilaterian animals: the case of acoelomorph flatworms. *Bioessays* 26: 1046–1057.

Barbarotto, E., Schmittgen, T. D., and Calin, G. A. 2008. MicroRNAs and cancer: profile, profile, profile. *Int. J. Cancer* 122: 969–977.

Blanc, V., and Davidson, N. O. 2003. C-to-U RNA editing: mechanisms leading to genetic diversity. *J. Biol. Chem.* 278: 1395–1398.

Blow, M. J., et al. 2006. RNA editing of human microRNAs. *Genome Biol.* 7: R27.

Borchiellini, C., Chombard, C., Manuel, M., Alivon, E., Vacelet, J., and Boury-Esnault, N. 2004. Molecular phylogeny of demspongiae: implications for classification and scenarios for character evolution. *Mol. Phylogenet. Evol.* 32: 823–837.

Calin, G. A., and Croce, C. M. 2006. MicroRNA signatures in human cancers. *Nat. Rev. Cancer* 6: 857–866.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17: 540–552.

Chen, K., and Rajewsky, N. 2007. The evolution of gene regulation by transcription factors and microRNAs. *Nat. Rev. Genet.* 8: 93–103.

Chourrout, D., et al. 2006. Minimal ProtoHox cluster inferred from bilaterian and cnidarian Hox complements. *Nature* 442: 684–687.

Cui, Q., Yu, Z., Purisima, E. O., and Wang, E. 2007. MicroRNA regulation and interspecific variation of gene expression. *Trends Genet.* 23: 372–375.

Dunn, C. W., et al. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 397: 707–710.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32: 1792–1797.

Eernisse, D. J., and Peterson, K. J. 2004. The history of animals. In J. Cracraft and M. J. Donoghue (eds.). *Assembling the Tree of Life*. Oxford University Press, Oxford, pp. 197–208.

Erpenbeck, D., et al. 2006. Mitochondrial diversity of early-branching Metazoa is revealed by the complete mt genome of a haplosclerid demosponge. *Mol. Biol. Evol.* 24: 19–22.

Esquela-Kerscher, A., and Slack, F. J. 2006. Oncomirs—microRNAs with a role in cancer. *Nat. Rev. Cancer* 6: 259–269.

Filipowicz, W., Bhattacharyya, S. N., and Sonenberg, N. 2008. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat. Rev. Genet.* 9: 102–114.

Fu, X., Adamski, M., and Thompson, E. M. 2008. Altered miRNA repertoire in the simplified chordate, *Oikopleura dioica*. *Mol. Biol. Evol.* 25: 1067–1080.

Griffiths-Jones, S., Saini, H. K., Dongen, S. v., and Enright, A. J. 2007. miRBase: tools for microRNA genomics. *Nucleic Acids Res.* 36: D154–D158.

Grimson, A., Farh, K. K.-H., Johnston, W. K., Garrett-Engele, P., Lim, L. P., and Bartel, D. P. 2007. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell* 27: 91–105.

Guder, C., Pinho, S., Nacak, T. G., Schmidt, H. A., Hobmayer, B., Niehrs, C., and Holstein, T. 2006. An ancient Wnt-Dickkopf antagonism in *Hydra*. *Development* 133: 901–911.

Habig, J. W., Dale, T., and Bass, B. L. 2007. miRNA editing—we should have inosine this coming. *Mol. Cell* 25: 792–793.

Halanych, K. M. 2004. The new view of animal phylogeny. *Ann. Rev. Ecol. Syst.* 35: 229–256.

He, L., He, X., Lowe, S. W., and Hannon, G. J. 2007. microRNAs join the p53 network—another piece in the tumour-suppression puzzle. *Nat. Rev. Cancer* 7: 819–822.

Heimberg, A. M., Sempere, L. F., Moy, V. N., Donoghue, P. C. J., and Peterson, K. J. 2008. MicroRNAs and the advent of vertebrate morphological complexity. *Proc. Natl. Acad. Sci. USA* 105: 2946–2950.

Hertel, J., et al. 2006. The expansion of the metazoan microRNA repertoire. *BMC Genom.* 7: 25.

Hobert, O. 2008. Gene regulation by transcription factors and microRNAs. *Science* 319: 1785–1786.

Hornstein, E., and Shomron, N. 2006. Canalization of development by microRNAs. *Nat. Genet.* 38 (suppl): S20–S24.

Jones-Rhoades, M. W., Bartel, D. P., and Bartel, B. 2006. MicroRNAs and their regulatory roles in plants. *Ann. Rev. Plant Biol.* 57: 19–53.

Kawahara, Y., Zinshteyn, B., Sethupathy, P., Iizasa, H., Hatzigeorgiou, A. G., and Nishikura, K. 2007. Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science* 315: 1137–1140.

Kawahara, Y., et al. 2008. Frequency and fate of microRNA editing in human brain. *Nucleic Acids Res.* 36: 5270–5280.

Kim, V. N. 2005. MicroRNA biogenesis: coordinated cropping and dicing. *Nat. Rev. Mol. Cell Biol.* 6: 376–385.

King, N., et al. 2008. The genome of the choanoflagellate *Monisiga brevicollis* and the origin of metazoans. *Nature* 451: 783–788.

Kusserow, A., Pang, K., Sturm, C., Hrouda, M., Lentifer, J., Schmidt, H. A., Technau, U., von Haeseler, A., Hobmayer, B., Martindale, M. Q., and Holstein, T. W. 2005. Unexpected complexity of the Wnt gene family in a sea anemone. *Nature* 433: 156–160.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. 2001. Identification of novel genes coding for small expressed RNAs. *Science* 294: 853–858.

Landgraf, P., et al. 2007. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* 129: 1401–1414.

Larroux, C., Luke, G. N., Koopman, P., Rokhsar, D., Shimeld, S. M., and Degnan, B. M. 2008. Genesis and expansion of metazoan transcription factor gene classes. *Mol. Biol. Evol.* 25: 980–996.

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

Lavrov, D. V., Wang, X., and Kelly, M. 2008. Reconstructing ordinal relationships in the Demospongiae using mitochondrial genomic data. *Mol. Phylogenet. Evol.*: 111–124.

Lee, C.-T., Risom, T., and Strauss, W. M. 2007. Evolutionary conservation of microRNA regulatory circuits: an examination of microRNA gene complexity and conserved microRNA–target interactions. *DNA Cell Biol.* 26: 209–218.

Lee, R. C., and Ambros, V. 2001. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* 294: 862–864.

Lu, J., et al. 2005. MicroRNA expression profiles classify human cancers. *Nature* 435: 834–838.

Luciano, D. J., Mirsky, H., Vendetti, N. J., and Maas, S. 2004. RNA editing of a miRNA precursor. *RNA* 10: 1174–1177.

Maddison, D. R., and Maddison, W. P. 2005. *MacClade 4*. Sinauer Associates, Sunderland.

Makeyev, E. V., and Maniatis, T. 2008. Multilevel regulation of gene expression by microRNAs. *Science* 319: 1789–1790.

Margulies, M., et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 376–380.

Medina, P. P., and Slack, F. J. 2008. microRNAs and cancer. *Cell Cycle* 7: 2485–2492.

Meltzer, P. S. 2005. Small RNAs with big impacts. *Nature* 435: 745–746.

Niwa, R., and Slack, F. J. 2007. The evolution of animal microRNA function. *Curr. Opin. Genet. Dev.* 17: 145–150.

Norden-Krichmaer, T. M., Holtz, J., Pasquinelli, A. E., and Gaasterland, T. 2007. Computational prediction and experimental validation of *Ciona intestinalis* microRNA genes. *BMC Genom.* 8: 445.

Okamura, K., Phillips, M. D., Tyler, D. M., Duan, H., Chou, Y.-t., and 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.

Palakodeti, D., Smielewska, M., and Graveley, B. R. 2006. MicroRNAs from the planarian *Schmidtea mediterranea*: a model system for stem cell biology. *RNA* 12: 1–10.

Peterson, K. J. 2008. Molecular paleobiology and the Cambrian explosion: 21st century answers to 19th century problems. In P. H. Kelley and R. K. Bambach (eds.). *From Evolution to Geobiology: Research Questions Driving Paleontology at the Start of a New Century*. Paleontological Society, New Haven, pp. 105–116.

Peterson, K. J., Cotton, J. A., Gehling, J. G., and Pisani, D. 2008. The Ediacaran emergence of bilaterians: congruence between the genetic and geologic fossil records. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363: 1435–1443.

Peterson, K. J., and Eernisse, D. J. 2001. Animal phylogeny and the ancestry of bilaterians: inferences from morphology and 18S rDNA gene sequences. *Evol. Dev.* 3: 170–205.

Peterson, K. J., and Sperling, E. A. 2007. Poriferan ANTP genes: primitively simple or secondarily reduced? *Evol. Dev.* 9: 405–408.

Peterson, K. J., Summons, R. E., and Donoghue, P. C. J. 2007. Molecular Paleobiology. *Palaeontology* 50: 775–809.

Philippe, H., Brinkmann, H., Martinez, P., Riutort, M., and Baguñà, J. 2007. Acoel flatworms are not Platyhelminthes: evidence from phylogenomics. *PLoS ONE* 2: e717.

Pierce, M. L., Weston, M. D., Fritzsch, B., Gabel, H. W., Ruvkun, G., and Soukup, G. A. 2008. MicroRNA-183 family conservation and ciliated neurosensory organ expression. *Evol. Dev.* 10: 106–113.

Prochnik, S. E., Rokhsar, D., and Aboobaker, A. A. 2007. Evidence for a microRNA expansion in the bilaterian ancestor. *Dev. Genes Evol.* 217: 73–77.

Putnam, N. H., et al. 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* 317: 86–94.

Reid, J. G., et al. 2008. Mouse let-7 miRNA populations exhibit RNA editing that is constrained in the 5′-seed/cleavage/anchor regions and stabilize predicted mmu-let7a:mRNA duplexes. *Genome Res* 18: 1571–1581.

Reinhart, B. J., Weinstein, E. G., Rhoades, M. W., Bartel, B., and Bartel, D. P. 2002. MicroRNAs in plants. *Genes Dev.* 16: 1616–1626.

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

Ryan, J. F., et al. 2007. Pre-bilaterian origins of the Hox cluster and Hox code: evidence from the sea anemone, *Nematostella vectensis*. *PLoS ONE* 1: e153.

Sempere, L. F., Cole, C. N., McPeek, M. A., and Peterson, K. J. 2006. The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. *J. Exp. Zool. (Mol. Dev. Evol.)* 306B: 575–588.

Sempere, L. F., Freemantle, S., Pitha-Rowe, I., Moss, E., Dmitrovsky, E., and Ambros, V. 2004. Expression profiling of mammalian mi croRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. *Genome Biol.* 5: R13.

Sempere, L. F., Martinez, P., Cole, C., Baguñà, J., and Peterson, K. J. 2007. Phylogenetic distribution of microRNAs supports the basal position of acoel flatworms and the polyphyly of Platyhelminthes. *Evol. Dev.* 9: 409–415.

Sevignani, C., et al. 2007. MicroRNA genes are frequently located near mouse cancer susceptibility loci. *Proc. Natl. Acad. Sci. USA* 104: 8017–8022.

Shalgi, R., Lieber, D., Oren, M., and Pilpel, Y. 2007. Global and local architecture of the mammalian microRNA-transcription factor regulatory network. *PLoS Comp. Biol.* 3: 1291–1304.

Simionato, E., et al. 2007. Origin and diversification of the basic helix–loop–helix gene family in metazoans: insights from comparative genomics. *BMC Evol. Biol.* 7: 33.

Sperling, E. A., and Peterson, K. J. 2009. microRNAs and metazoan phylogeny: big trees from little genes. In M. J. Telford and D. T. J. Littlewood (eds.). *Animal Evolution - Genomes, Trees and Fossils*. Oxford University Press, Oxford. In press.

Srivastava, M., et al. 2008. The *Trichoplax* genome and the nature of placozoans. *Nature* 454: 955–960.

Stark, A., et al. 2007. Systematic discovery and characterization of fly microRNAs using 12 *Drosophila* genomes. *Genome Res.* 17: 1865–1879.

Valentine, J. W., Collins, A. G., and Meyer, C. P. 1994. Morphological complexity increase in metazoans. *Paleobiology* 20: 131–142.

van Rooij, E., Liu, N., and Olson, E. N. 2008. MicroRNAs flex their muscles. *Trends Genet.* 24: 159–166.

Wallberg, A., Curini-Galletti, M., Ahmadzadeh, A., and Jondelius, U. 2007. Dismissal of acoelomorpha: acoela and nemertodermatida are separate early bilaterian clades. *Zool. Scr.* 36: 509–523.

Wray, G. A., et al. 2003. The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* 20: 1377–1419.

Yamada, A., Pang, K., Martindale, M. Q., and Tochinai, S. 2007. Surprisingly complex T-box gene complement in diploblastic metazoans. *Evol. Dev.* 9: 220–230.

Yang, N., Coukos, G., and Zhang, L. 2008. MicroRNA epigenetic alterations in human cancer: one step forward in diagnosis and treatment. *Int. J. Cancer* 122: 963–968.

Yi, R., Poy, M. N., Stoffel, M., and Fuchs, E. 2008. A skin microRNA promotes differentiation by repressing 'stemness'. *Nature* 452: 225–229.

Zhao, Y., and Srivastava, D. 2007. A developmental view of microRNA function. *Trends Biochem. Sci.* 32: 189–197.

Zuker, M., Mathews, D. H., and Turner, D. H. 1999. Algorithms and thermodynaics for RNA secondary structure prediction: a practical guide. In J. Barciszewski and B. F. C. Clark (eds.). *RNA Biochemistry and Biotechnology*. Kluwer Academic Publishers, Dordrecht.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table S1. Alignments of the mature sequences and an example of a structure for all discovered miRNAs reported herein, in addition to the source of evidence of the miRNA for each taxon considered (whether found in the genome, recovered from our small RNA libraries, or both).

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.