Introduction

More than a decade ago Victor Ambros and colleagues, working on a gene called lin-4 that controlled the timing of C. elegans larval development, observed that the gene did not encode a protein but, instead, produced two small RNAs. Moreover, the larger of the two RNAs produced by this gene, a 61nt sequence, folded into a stem-loop structure which seemed to be the precursor of the shorter, 21nt, RNA. It was also found that the lin-4 RNA displayed antisense complementarity with several sites in the 3’-UTR of another gene called lin-14 [1]. The 3’UTR of lin-14 was previously suspected of being the precise target of regulation by lin-4 and a model was proposed in which the lin-4 RNA regulates lin-14 by binding to complementary sequences in the 3’-UTR of the lin-14 mRNA. This mechanism serves to suppress lin-14 message.

For several years the lin-4 suppression story languished because no additional lin-4-like RNAs were found, even in C. elegans. Then, in 2000, a second C. elegans gene known to be involved in larval development was discovered to also produce a 22nt riboregulatory RNA [2]. Called let-7, this gene encodes a 73nt precursor RNA that forms a stem-loop structure. More importantly, unlike lin-4 before it, let-7 homologs were soon identified in both human and Drosophila genomes and the let-7 small RNA itself was found in more than a dozen bilateral metazoans including human, mouse, chicken, frog, zebrafish, fruit fly, sea urchin, and abalone. The let-7 RNA was not found in Cnidarians (corals and jellyfish), Poriferans (sponges), yeast, E. coli, or Aribidopsis.

Following this discovery, several labs, cloning RNAs from human, worm, and fly cells, discovered more than 100 similar RNAs; 20 in Drosophila, 30 in human, and 60 in C. elegans [3]. In each case, the RNA products were 22nt endogenously expressed RNAs derived from a longer precursor RNA displaying a characteristic stem-loop structure. Expanded searches revealed that the genes encoding these RNAs tended to be conserved in closely related species and some were broadly conserved in the manner of
let-7. This growing class of small RNAs was named microRNA, or miRNA. Since 2001 miRNAs have become a major class of riboregulatory nucleotides and an miRNA Registry has been established through the Sanger Institute [4]. (http://www.sanger.ac.uk/Software/Rfam/mirna/index.shtml). New miRNAs are cataloged and named using recently adopted standards. Release 5.0 lists 1,345 miRNAs from nine species and a set of guidelines for miRNA annotation is now in place [5].

**Structure and Function of miRNAs**

The growth of information on miRNAs permits some generalizations. Most miRNAs, about 75% to date, map to genomic regions distinct from previously annotated genes which implies that they derive from independent transcription units. Roughly one-fourth of known miRNAs have been found in the introns of genes. Often, but not always, the miRNA regulates expression of that gene product. Conserved relationships between an miRNA and a “host” mRNA have been routinely observed. For example, one of these conserved relationships, the miRNA miR-7 in the last intron of the gene encoding heterogeneous nuclear ribonucleoprotein-K, hnRNP-K (Figure 1A), extends to *Drosophila*, mouse, human, *Canis familiaris*, and *Anopheles gambiae*. Over expression of hnRNP-K has been observed in cancer cells but a role of miR-7 in this process is, as yet, unknown. Sequence analyses using both GenBank and the miRNA Registry show that the mature miR-7 sequence is completely conserved over a wide range of species.

A.

```
Gtcattttaagtttagctttgtttagctttatacataactaaaccccttttcctctcttaattga
ttttttctcttttagaggacatgtgctcaaaccttttgagaagctccatttacccacatcttgca
cttaattttctctgtcctctttttgttaaatatgtaagagctttttcttttagaagccatggtgc
cttaattttttctttctttttctttgtaattttttttttttttttttttttttttttttttttttttttttttttttt
```

B.

```
Homo sapiens    UGGAAGACUAGUGAUUUUGU−
Canis familiaris UGGAAGACUAGUGAUUUUGU−
Mus musculus    UGGAAGACUAGUGAUUUUGU−
Rattus norvegicus UGGAAGACUAGUGAUUUUGU−
Gallus gallus    UGGAAGACUAGUGAUUUUGU−
Danio rerio      UGGAAGACUAGUGAUUUUGU−
Drosophila melanogaster UGGAAGACUAGUGAUUUUGU−
Drosophila pseudoobscura UGGAAGACUAGUGAUUUUGU−
```

Figure 1. A. The sequence of intron 13 of human hnRNP-K from chromosome 9q21.13-q21.33 containing miR-7 is shown. The 110nt pre-miRNA sequence is in blue with the 22nt mature miRNA sequence in red. Transcription orientation of this miRNA is the same as the transcription orientation of the hnRNP-K gene. Thus, they are transcribed together and miR-7 is subsequently processed. B. A Clustal alignment of the mature miR-7 of several species.
A number of miRNAs are clustered with an arrangement and expression pattern suggesting a single multi-cistronic primary transcript that is subsequently processed into individual miRNAs. Though this arrangement was primarily observed in the *Drosophila* genome, highly conserved miRNA clusters are known from all genomes so far surveyed. Among miRNA clusters that have recently been examined in detail are the miR17 cluster [6], the miRNAs imbedded in the *Hox* gene cluster [7], and the large imprinted miRNA cluster in the mouse *Dlk1-Gtl2* domain [8]. Interestingly, some miRNAs organized in this way are related but, again, there are exceptions. In general, miRNA clusters appear to evolve in much the same way as other multi-gene clusters, through duplication, differentiation, and some loss. The miR17 cluster is particularly informative in this regard. Tanzer and Stadler show that the history of this cluster, composed of miR-17, miR-18, miR-19a, miR19b, miR-20, miR-25, miR-92, miR-93, miR-106a, and miR-106b, can be traced through tandem duplications and loss of individual members in vertebrates from teleost fish to human and some homologs in the cluster are found in both *Drosophila* and *C. elegans* [6]. At the present time, however, functional relationships among the members of the miR17 cluster are not known.

Microarray technology has led to a number of interesting findings with regard to miRNA expression. Many of the *C. elegans* miRNAs display stage-specific expression patterns. This is also true in *Drosophila* and in humans, particularly with regard to tissue-specific expression patterns. The growth of annotated miRNAs in all species so far examined suggests that it is possible that every metazoan cell type may have a distinct miRNA expression pattern at every level of development. In addition, some miRNAs stand out via their sheer abundance in some cells. The *C. elegans* miRNAs miR-2, miR-52, and miR-58 are found to have more than 50,000 copies per adult cell. This abundance is greater than is seen for the crucial spliceosome component U6 snRNA. Other miRNAs, like miR-124, are found in very low copy numbers which could be due to limited expression in only certain cells. *miR-124*, for example, is only expressed in neurons.

Regardless of expression levels or patterns, one thing about miRNAs is becoming apparent. These small, non-coding RNAs are intimately involved in regulating gene expression. From the very first examples in *C. elegans* to the most recent explosion of identified sequences in virtually all metazoans, miRNAs preferentially associate with genes that control some aspect of cellular differentiation. It is also seen that some miRNAs can regulate more than one target. In *C. elegans*, for example, the venerable *lin-4* and *let-7* miRNAs regulate at least two genes each that are involved in larval developmental transitions with *lin-4* expressed in early transitions affecting both *lin-14* and *lin-28* and *let-7* expressed in late transitions affecting both *lin-41* and *hbl-1* [1, 9, 10, 11, 12, 13, 14]. Recent bioinformatics analyses of mature microRNA sequences suggest that the proportion of genes in all genomes under some form of microRNA regulation may be very large. Lewis et al. have predicted that as much as 30% of all human genes may be under miRNA regulation [15]. This also suggests that there may be far more...
miRNA loci than previously predicted, a conclusion reached by several recent studies including Berezikov et al. [16].

**Plant miRNAs**

The discovery of miRNAs in plants, including *Arabidopsis* and *Oryza*, has raised several interesting issues. Reinhart et al. believe that miRNAs pre-date the divergence of plants and metazoans [17]. Some support for this position can be seen in the parallels between plant and metazoan miRNAs;

1. plant miRNAs are endogenously expressed small RNAs processed from a larger precursor RNA.
2. plant miRNAs are generally conserved.
3. plant miRNAs are found mostly in genomic regions distinct from annotated genes.

Data so far obtained suggest that the proportion of plant gene regulation via miRNA intervention in plants may be even greater than that in metazoans. However, plants are not animals and data accumulating from plant genomes also show several substantial differences in plant miRNAs compared to animals. First, plant pre-miRNAs have larger and more variable stem-loop structures than do metazoan pre-miRNAs (Figure 2). Second, mature plant miRNAs generally present with fewer mismatches with their target sites than do average animal miRNAs (see below). Third, mature miRNAs in plants tend more toward 21nt as a rule as opposed to the 22nt-23nt mature miRNAs in animals. Finally, so far there is a stronger preference for 5’-terminal uridines in plant miRNAs than in animal miRNAs.

![Figure 2](image)

**Figure 2. Examples of the variation commonly found among plant pre-microRNAs.** Source: Reinhart et al. [17].
To date, it is not possible to definitively say whether miRNA gene regulation predates the divergence of plants and animals as there are no miRNAs yet found to be conserved across the plant-metazoan boundary (see discussion below).

Returning to miRNA targeting in plants, as noted there is a tendency in plant miRNAs to pair with mRNA targets with near-perfect complementarity. In addition, the targets of plant miRNAs tend to be in coding sequences rather than in 3'-UTRs which is the norm for animal miRNAs. The specificity of plant miRNA targeting in coding sequences suggests that plant miRNAs may act more like siRNAs than do animal miRNAs. Specific miRNA::mRNA pairing is conserved in plants over the few species so far investigated, particularly Arabidopsis and Oryza. Plant miRNAs characterized to date appear to have a pronounced preference for transcription factor gene families. The combination of precise targeting and gene family preferences makes target prediction for newly discovered miRNAs somewhat easier in plants than in animals.

In general, then, plant miRNAs are shorter, 21nt versus 22nt-23nt, but have larger and more variable precursors. They also tend to target single sites in the coding regions of mRNAs rather than multiple sites in 3'-UTRs.

miRNA Biosynthesis

The mechanism through which miRNAs are generated is dictated by their secondary structures. The primary miRNA transcript, called the pri-miRNA, appears to be processed as a capped and polyadenylated structure that can serve as an mRNA [18]. This mRNA-like structure folds into a hairpin that becomes the substrate for an enzyme complex composed of an RNase III enzyme called DROSHA [19] and a partnering molecule called PASHA, for partner of Drosha, or DCRG [20, 21]. This releases the hairpin in the form of the pre-miRNA. Once released from the pri-miRNA, the pre-miRNA is transported out of the nucleus via a TRANSPORTIN mechanism. In the cytoplasm, the pre-miRNA is a substrate for the processing enzyme DICER. Fragmentation of the precursor by DICER releases the 21bp – 23bp mature miRNA fragment. This mature miRNA fragment then becomes a single stranded species that will bind to the mRNA target site. Like siRNAs such binding creates a double stranded RNA that induces the formation of the RISC which cleaves the RNA and suppresses the message. A synopsis of this generalized mechanism is shown in Figure 3.

Since some tendencies and commonalities have been observed in both plant and metazoan miRNAs several bioinformatics efforts to predict the existence of miRNA loci from genomic sequences and to identify their targets have been made [22, 23, 24]. So far, the success of these predictive methods has been limited but, where they have worked, they have worked pretty well.
Evolution of miRNAs

Clearly the study of miRNAs is in its infancy and there is far more that is unknown about them than is known. Relatively few miRNA sequences have been validated and, in general, the mechanism of their action is just now being worked out in first-order detail. However, the opportunity to speculate about the evolution of miRNAs is irresistible, particularly in view of some of the early comparative data that have been reported.

Conservation of the let-7 miRNA over a broad range of metazoan species presented by Pasquinelli et al. is a case in point [2]. The let-7 precursor sequences in Homo sapiens, Drosophila melanogaster, and Caenorhabditis elegans are shown by Pasquinelli et al. to contain the same mature miRNA sequence in the same place. Let-7 targets in the lin-41 3’-UTR were confirmed in D. melanogaster, C. elegans, and, importantly, in the zebrafish, Danio rerio. Phylogenetically organized let-7 expression assays indicate that this miRNA is expressed in virtually every major branch of the metazoan (see [2]; Figure 4). Only Cnidarians (corals and jelly fish) and Poriferans (sponges) failed to exhibit let-7 expression. These data suggest that the let-7 miRNA evolved with and was instrumental in the evolution of bilaterian metazoans, a development that is on the order of 600 to 630 millions years old [25, 26].

Not to be outdone, Floyd and Bowman reported on two plant miRNAs, miR165 and miR166, that specifically target genes in the class-III homeodomain-leucine zipper (HD-Zip) gene family [27]. This gene family is crucial in establishing meristem formation, adaxial domains during leaf development, and the proper patterning and differentiation of vascular bundles in flowering plants. Both miR165 and miR166 are well characterized in Arabidopsis and Oryza but Floyd and Bowman extended the analysis of these RNAs to include bryophytes, lycopods, and ferns as well as seed plants. Their results show that the mature miR166 miRNA sequence is conserved in 18 of 21 positions in species ranging from angiosperms to liverworts and hornworts. Phylogenetic studies of these plants indicate that the last common ancestor of the land plants is at least 400 million years old. Thus, Floyd and Bowman conclude that the mi166 miRNA dates back to this common ancestor.

It is probably not chance that the two oldest conserved miRNAs so far found are both involved in crucial symmetry and differentiation events during plant and animal development. Indeed, such evolutionary conservation is expected for loci involved in essential cellular functions. The importance of this for assessing the evolution of miRNAs, however, is that the more fundamental the process is that is regulated by an miRNA, the more likely it is that a common ancestor can be found. The key questions here are, 1. what is the origin of miRNA regulation; i.e., did it derive from an siRNA-like process or is there some other track, and 2. did miRNA regulation in plants and metazoans develop before or after the divergence of these two fundamentally different types of multicellular life; i.e., are we looking at parallel or convergent evolution or what? There are clearly likely places to look for these answers. In the metazoan world,
Figure 3. Schematic representation of miRNA formation in both plants and metazoans. This scheme is only a very general outline of the major events.
the Cniderians and Poriferans are one place to look. In the plant world, the discovery of organisms like *Mesostigma viride* which may predate the separation of the streptophyta (land plants) and chlorophyta (green algae) provide an obvious hunting ground [28]. Finally, examination of the so-called “early” eukaryotes might lead to some interesting discoveries as well [29].

A recent paper by Allen et al. does present a tantalizing model for miRNA origination based upon an analysis of some sequences in *A. thaliana* [30]. As discussed above, plant miRNAs display extensive sequence complementarity to their mRNA targets but the fold-back precursor sequences are highly variable. Allen et al. observed that pre-miRNAs might arise through inverted duplications in their future target genes. If this is true, then very young miRNAs might retain vestiges of their target sequences outside the mature miRNA sequence. A synopsis of this hypothesis is shown below (Figure 4).

Some evidence of this model in *A. thaliana* is seen in miRNAs like miR161 and miR163, reported by Allen et al., that retain target sequence complementarity outside the mature miRNA sequence itself [30]. Another locus, ASRP1729, that produces at least two small RNAs, has several features similar to an miRNA precursor but, as yet, has not become an miRNA locus. Both miR161 and miR163 interact with HEN-1 and DCL-1 but ASRP1729 interacts only with HEN-1. Both miR161 and miR163 display greater than expected sequence similarities with potential target genes in two gene families; pentatricopeptide repeat proteins (PPRs) and S-adenosylmethionine-dependent methyltransferases (SAMT’s) respectively. Members of both gene families evolved via gene duplication and map as a cluster near the miRNA loci that appear to regulate their expression. ASRP 1729 also lies very close to a duplication cluster of genes representing a class known as divergent C1 domain proteins (DC1) but there is no evidence of miRNA-like regulation. Thus, these three loci, according to Allen et al., represent step 2 and 3 in the evolutionary scheme shown in Figure 4.

**Figure 4. Synopsis of the Allen et al. (2004) model for the formation of miRNA loci.** An inverted repeat in the target gene results in the formation of a fold-back structure which can be processed by the DICER complex into siRNAs. As the sequence ages, the fold-back structure diverges until only the specific miRNA is left to complement the target sequence. HEN-1 is a
novel helix-loop-helix protein involved in miRNA production in *A. thaliana* and DCL-X and DCL-1 refer to DICER-like enzymes (Source: [31])

**Conclusions**

The microRNA story is still in its very early stages but there are a number of general conclusions that can be stated. First, miRNAs are a new and potentially powerful class of riboregulatory elements. Processes that are regulated tend to be those involved in development and cellular differentiation. The mechanism of miRNA processing is tied to that involved in siRNA processing in that the pre-miRNA is cleaved by the same enzyme complex that is involved in RNAi. A number of differences between plant and metazoan miRNAs suggest that they may have shared the same proto-miRNA mechanisms but evolved independently in parallel.

**References and Resources**


**microRNA databases**

A number of useful web sites have already been established for microRNA research. Some of these are listed below. New sites will be added to this list as they come on-line.

- miRNA Registry- Noted above, this registry catalogs miRNAs for a wide range of species and is a clearing house for assigning names to new miRNAs prior to publication. [www.sanger.ac.uk/Software/Rfam/mirna/](http://www.sanger.ac.uk/Software/Rfam/mirna/)

- MicroRNAdb- A year-old database of microRNAs. Includes some features not found in the miRNA Registry. [www.166.111.30.65/micrornadb/](http://www.166.111.30.65/micrornadb/)

- miRanda- An algorithm for finding genomic targets for microRNAs [www.microrna.org/miranda.html](http://www.microrna.org/miranda.html)

- TargetScan- A computational method for finding microRNA genes. [www.genes.mit.edu/targetscan/](http://www.genes.mit.edu/targetscan/)

- MiRscan- Software that assigns a score to potential hairpin structures based upon a set of 50 experimentally verified microRNA hairpins from *C. elegans* and *C. briggsae*. [www.genes.mit.edu/mirs Cann](http://www.genes.mit.edu/mirs Can).