An Abundant Class of Tiny RNAs with Probable Regulatory Roles in *Caenorhabditis elegans*

Nelson C. Lau, Lee P. Lim, Earl G. Weinstein, David P. Bartel†

Whitehead Institute for Biomedical Research

and Department of Biology, MIT

9 Cambridge Center

Cambridge, MA 02142

†To whom correspondence should be addressed.

phone: 617-258-5287; fax: 617-258-6768; e-mail: dbartel@wi.mit.edu

Two small temporal RNAs (stRNAs), *lin-4* and *let-7*, control developmental timing in *Caenorhabditis elegans*. We find that these two regulatory RNAs are members of a large class of 21–24-nucleotide non-coding RNAs, called microRNAs (miRNAs). We report on 55 novel miRNAs in *C. elegans*. The miRNAs have diverse expression patterns during development: a *let-7* paralog is temporally co-expressed with *let-7*; miRNAs encoded in a single genomic cluster are co-expressed during embryogenesis; still other miRNAs are expressed constitutively throughout development. Potential orthologs of several novel miRNA genes were identified in *Drosophila* and human genomes. The abundance of these tiny RNAs, their expression patterns, and their evolutionary conservation imply that, as a class, miRNAs have broad regulatory functions in animals.

Two types of short RNAs, both about 21–25 nt in length, serve as guide RNAs to direct posttranscriptional regulatory machinery to specific mRNA targets. Small temporal RNAs (stRNAs) control developmental timing in *Caenorhabditis elegans* (*1-3*). They pair to sites within the 3´-untranslated region (3´ UTR) of target mRNAs, causing translational repression of these mRNAs, thereby triggering the transition to the next developmental stage (*1-5*). Small interfering RNAs (siRNAs), which direct mRNA cleavage during RNA interference (RNAi) and related processes, are the other type of short regulatory RNAs (*6-12*). Both stRNAs and siRNAs are generated by processes requiring Dicer, a multidomain protein with tandem RNAse III domains (*13-15*). Dicer cleaves within the double-stranded portion of precursor molecules to yield the 21–25 nt guide RNAs.

lin-4 and *let-7* have been the only two stRNAs identified, and so the extent to which this type of small non-coding RNA normally regulates eukaryotic gene expression is only beginning to be understood (*1-5*). RNAi-related processes protect against viruses or mobile genetic elements, yet these processes are known to normally regulate only one other mRNA, that of *Drosophila Stellate* (*16-20*). To investigate whether RNAs resembling stRNAs or siRNAs might be playing a more general role in gene regulation, we isolated and cloned endogenous *C. elegans* RNAs that have the expected features of Dicer products. Tuschl and colleagues showed that such a strategy is feasible when they fortuitously cloned endogenous *Drosophila* RNAs while cloning siRNAs processed from exogenous dsRNA in an embryo lysate (*12*). Furthermore, other efforts focusing on longer RNAs have recently uncovered many novel non-coding RNAs (*21, 22*).

Dicer products, such as stRNAs and siRNAs, can be distinguished from most other oligonucleotides that might be present in *C. elegans* by three criteria: a length of about 22 nt, a 5´-terminal monophosphate, and a 3´-terminal hydroxyl group (*12, 13, 15*). Accordingly, a procedure was developed for isolating and cloning *C. elegans* RNAs with these features (*23*). Of the clones sequenced, 330 matched *C. elegans* genomic sequence, including 10 representing *lin-4* RNA and one representing *let-7* RNA. Another 182

corresponded to *E. coli* genomic sequence. *E. coli* RNA clones were expected because the worms were cultured with *E. coli* as the primary food source.

Three hundred of the 330 *C. elegans* clones have the potential to pair with nearby genomic sequence to form fold-back structures resembling those thought to be needed for Dicer processing of *lin-4* and *let-7* stRNAs (Fig. 1) (*24*). These 300 clones with predicted fold-backs represent 54 unique sequences: *lin-4*, *let-7*, and 52 other RNAs (Table 1). Thus, *lin-4* and *let-7* RNAs appear to be members of a larger class of non-coding RNAs that are about 20–24 nt in length and processed from fold-back structures. We and the two other groups with concurrent reports refer to this class of tiny RNAs as microRNAs, abbreviated miRNAs, with individual miRNAs and their genes designated miR-# and *mir-#*, respectively (*25, 26*).

We propose that most of the miRNAs are expressed from independent transcription units, previously unidentified because they do not contain an open reading frame (ORF) or other features required by current gene-recognition algorithms. No miRNAs matched a transcript validated by an annotated *C. elegans* expressed-sequence tag (EST), and most were at least 1 kb from the nearest annotated sequences (Table 1). Even the miRNA genes near predicted coding regions or within predicted introns are probably expressed separately from the annotated genes: If most miRNAs were expressed from the same primary transcript as the predicted protein, their orientation would be predominantly the same as the predicted mRNA, but no such bias in orientation was observed (Table 1). Likewise, other types of RNA genes located within *C. elegans* intronic regions are usually expressed from independent transcription units (*27*).

Whereas both *lin-4* and *let-7* RNAs reside on the 5′ arm of their fold-back structures (*1, 3*), only about a quarter of the other miRNAs lie on the 5′ arm of their proposed fold-back structures, as exemplified by miR-84 (Table 1; Fig. 1A). All the others reside on the 3′ arm, as exemplified by miR-1 (Table 1; Fig. 1B). This implies that the stable product of Dicer processing can reside on either arm of the precursor and that features of the miRNA or its precursor, other than the loop connecting the two arms, must determine which side of the fold-back contains the stable product.

When compared to the RNA fragments cloned from *E. coli*, the miRNAs had unique length and sequence features (Fig. 2). The *E. coli* fragments had a broad length distribution, ranging from 15 to 29, which reflects the size-selection limits imposed during the cloning procedure (*23*). In contrast, the miRNAs had a much tighter length distribution, centering on 21–24 nt, coincident with the known specificity of Dicer processing (Fig 2A). The miRNA sequence composition preferences were most striking at the 5′ end, where there was a strong preference for U and against G at the first position and then a deficiency of Us at positions 2–4 (Fig. 2B). miRNAs were also generally deficient in C, except at position 4. These composition preferences were not present in the clones representing *E. coli* RNA fragments.

The expression of 20 cloned miRNAs was examined, and all but two (miR-41 and miR-68) were readily detected on Northern blots (Fig. 3). For these 18 miRNAs with detectable expression, the dominant form was the mature 20–24 nt fragment(s), though for most, a longer species was also detected at the mobility expected for the fold-back precursor RNA. Fold-back precursors for *lin-4* and *let-7* have also been observed, particularly at the stage in development when the stRNA is first expressed (*1, 14, 15*).

Because the miRNAs resemble stRNAs, their temporal expression was examined. RNA from wild-type embryos, the four larval stages (L1–L4), and young adults was probed. RNA from *glp-4 (bn2)* young adults, which are severely depleted in germ cells (*28*), was also probed because miRNAs might have critical functions in the germ line, as suggested by the finding that worms deficient in Dicer have germ line defects and are sterile (*14, 29*). Many miRNAs have intriguing expression patterns during development (Fig. 3). For example, the expression of miR-84, an miRNA with 77% sequence identity to *let-7* RNA, was found to be indistinguishable from that of *let-7* (Fig. 3). Thus, it is tempting to speculate that miR-84 is an stRNA that works in concert with *let-7* RNA to control the larval-to-adult transition, an idea supported by the identification of plausible binding sites for miR-84 in the 3´ UTRs of appropriate heterochronic genes (*30*).

Nearly all of the miRNAs appear to have orthologs in other species, as would be expected if they had evolutionarily conserved regulatory roles. About 85% percent of the novel miRNAs had recognizable homologs in the available *C. briggsae* genomic sequence, which at the time of our analysis included about 90% of the *C. briggsae* genome (Table 1). Over 40% of the miRNAs appeared to be identical in C. briggsae, as is seen *lin-4* and *let-7* RNAs, (*1, 3*). Those miRNAs not absolutely conserved between *C. briggsae* and *C. elegans* might still have important functions, but might have more readily co-varied with their target sites because, for instance, they might have fewer target sites. It is noteworthy that when the sequence of the miRNA differs from that of its homologs, there is usually a compensatory change in the other arm of the fold-back to maintain pairing, providing phylogenetic evidence for the existence and importance of the fold-back secondary structures. *let-7*, but not *lin-4*, has discernable homologs in more distantly related organisms, including *Drosophila* and human (*31*). At least seven other miRNA genes (*mir-1*, *mir-2*, *mir-34*, *mir-60*, *mir-72*, *mir-79*, and *mir-84*) appear to be conserved in *Drosophila*, and most of these (*mir-1*, *mir-34*, *mir-60*, *mir-72*, and *mir-84*) appear to be also conserved in human (*24*). The most highly conserved novel miRNA, miR-1, is expressed throughout *C. elegans* development (Fig. 3) and so is unlikely to control developmental timing but might instead control tissue-specific events.

The distribution of miRNA genes within the *C. elegans* genome is not random (Table 1). For example, clones for six miRNA paralogs clustered within an 800-bp fragment of chromosome II (Table 1). Computer folding readily identified the fold-back structures for the six cloned miRNAs of this cluster, and predicted the existence of a seventh paralog, miR-39 (Fig. 1D). Northern analysis confirmed the presence and expression of miR-39 (Fig. 3). The homologous cluster in *C. briggsae* appears to have eight related miRNAs. Some of the miRNAs in the *C. elegans* cluster are more similar to each other than to those of the *C. briggsae* cluster, and vice versa, indicating that the size of the cluster has been quite dynamic over a short evolutionary interval, with expansion and perhaps also contraction since the divergence of these two species.

Northern analysis of the miRNAs of the *mir-35–41* cluster showed that these miRNAs are highly expressed in the embryo and in young adults (with eggs), but not at other developmental stages (Fig. 3). For the six detectable miRNAs of this cluster, longer species with mobilities expected for the respective fold-back RNAs also appear to be expressed in the germ line, as indicated by the observation that L4 animals, which have developing gonads but not embryos, express these longer RNAs, whereas germ line-deficient adults do not (Fig. 3) (*30*).

The close proximity of the miRNA genes within the *mir-35–41* cluster (Fig. 1D) suggests that they are all transcribed and processed from a single precursor RNA, an idea supported by the coordinate expression of these genes (Fig. 3). This operon-like organization and expression brings to mind several potential models for miRNA action. For example, each miRNA of the operon might target a different member of a gene family for translational repression. At the other extreme, they all might converge on the same target, just as *lin-4* and *let-7* RNAs potentially converge on the 3´ UTR of *lin-14* (*3*).

Another four clusters were identified among the sequenced miRNA clones (Table 1). Whereas the clones from one cluster were not homologous to clones from other clusters, the clones within each cluster were usually related to each other, as seen with the *mir-35–41* cluster. The last miRNA of the *mir-42–44* cluster is also represented by a second gene, *mir-45*, which is not part of the cluster. This second gene appears to enable more constitutive expression of the miRNA (miR-44/45) as compared to the first two genes of the *mir-42–44* cluster, which are expressed predominantly in the embryo (Fig. 3).

Dicer processing of stRNAs differs from that of siRNAs in its asymmetry: RNA from only one arm of the fold-back precursor accumulates, while the remainder of the precursor quickly degrades (*15*). This asymmetry extends to nearly all the miRNAs. For the 35 miRNAs yielding more than one clone, in only one case, miR-56, were RNAs cloned from both arms of a hairpin (Fig. 1C, Table 1). The functional miRNA appears to be miR-56 and not miR-56*, as indicated by analysis of sequence conservation between *C. elegans* and *C. briggsae* orthologs, analogy to the other constituents of the *mir-54–56* cluster, and Northern blots detecting RNA from only the 3´ arm of the fold-back (*30*).

We were surprised to find that few, if any, of the cloned RNAs had the features of siRNAs. No *C. elegans* clones matched the antisense of annotated coding regions. Of the 30 *C. elegans* clones not classified as miRNAs, 15 matched fragments of known RNA genes, such as tRNA and ribosomal RNA. Of the remaining 15 clones, the best candidate for a natural siRNA is GGAAAACGGGUUGAAAGGGA. It was the only *C. elegans* clone perfectly complementary to an annotated EST, hybridizing to the 3´ UTR of gene ZK418.9, a possible RNA-binding protein. Even if this and a few other clones do represent authentic siRNAs, they would still be greatly outnumbered by the 300 clones representing 54 different miRNAs. Our cloning protocol is not expected to preferentially exclude siRNAs; it was similar to the protocol that efficiently cloned exogenous siRNAs from *Drosophila* extracts (*12*). Instead, we propose that the preponderance of miRNAs among our clones indicates that in healthy, growing cultures of *C. elegans*, regulation by miRNAs normally plays a more dominant role than does regulation by siRNAs.

Irrespective of the relative importance of miRNAs and siRNAs in the normal regulation of endogenous genes, our results show that small RNA genes, of the type exemplified by *lin-4* and *let-7*, are more abundant in *C. elegans* than previously appreciated. Results from a parallel effort that directly cloned small RNAs from *Drosophila* and HeLa cells demonstrates that the same is true in other animals (*25*), a conclusion further supported by the orthologs to the *C. elegans* miRNAs that we identified through database searching. Many of the miRNAs that we identified are represented by only a single clone (Table 1), suggesting that our sequencing has not reached saturation and that there are over a hundred miRNA genes in *C. elegans*.

We presume that there is a reason for the expression and evolutionary conservation of these small non-coding RNAs. Our favored hypothesis is that these novel miRNAs, together with *lin-4* and *let-7* RNAs, constitute an important and abundant class of riboregulators, pairing to specific sites within mRNAs to direct the posttranscriptional regulation of these genes (*32*). The abundance and diverse expression patterns of miRNA genes implies that they function in a variety of regulatory pathways, in addition to their known role in the temporal control of developmental events.

References and Notes

1. R. C. Lee, R. L. Feinbaum, V. Ambros, *Cell* **75**, 843 (1993).
2. B. Wightman, I. Ha, G. Ruvkun, *Cell* **75**, 855 (1993).
3. B. J. Reinhart, et al., *Nature* **403**, 901 (2000).
4. E. G. Moss, R. C. Lee, V. Ambros, *Cell* **88**, 637 (1997).
5. F. J. Slack, et al., *Mol Cell* **5**, 659 (2000).
6. A. J. Hamilton, D. C. Baulcombe, *Science* **286**, 950 (1999).
7. S. M. Hammond, E. Bernstein, D. Beach, G. J. Hannon, *Nature* **404**, 293 (2000).
8. P. D. Zamore, T. Tuschl, P. A. Sharp, D. P. Bartel, *Cell* **101**, 25 (2000).
9. S. Parrish, J. Fleenor, S. Xu, C. Mello, A. Fire, *Mol. Cell* **6**, 1077 (2000).
10. D. Yang, H. Lu, J. W. Erickson, *Curr Biol* **10**, 1191 (2000).
11. C. Cogoni, G. Macino, *Curr Opin Genet Dev* **10**, 638 (2000).
12. S. M. Elbashir, W. Lendeckel, T. Tuschl, *Genes Dev* **15**, 188 (2001).
13. E. Bernstein, A. A. Caudy, S. M. Hammond, G. J. Hannon, *Nature* **409**, 363 (2001).
14. A. Grishok, et al., *Cell* **106**, 23 (2001).
15. G. Hutvágner, et al., *Science* **293**, 834 (2001).
16. F. G. Ratcliff, S. A. MacFarlane, D. C. Baulcombe, *Plant Cell* **11**, 1207 (1999).
17. R. F. Ketting, T. H. Haverkamp, H. G. van Luenen, R. H. Plasterk, *Cell* **99**, 133 (1999).
18. H. Tabara, et al., *Cell* **99**, 123 (1999).
19. S. Malinsky, A. Bucheton, I. Busseau, *Genetics* **156**, 1147 (2000).
20. A. A. Aravin, et al., *Cur. Biol.* **11**, 1017 (2001).
21. A. Huttenhofer, et al., *Embo J* **20**, 2943 (2001).
22. K. M. Wassarman, F. Repoila, C. Rosenow, G. Storz, S. Gottesman, *Genes Dev* **15**, 1637 (2001).
23. Short endogenous *C. elegans* RNAs were cloned using a protocol inspired by Elbashir et al. (*12*), but modified to make it specific for RNAs with 5´-terminal

phosphate and 3´-terminal hydroxyl groups. In our protocol (*24*), gel-purified 18–26 nt RNA from mixed-stage worms was ligated to a pre-adenylylated 3´-adaptor oligonucleotide in a reaction using T4 RNA ligase but without ATP. Ligated RNA was gel-purified, then ligated to a 5´-adaptor oligonucleotide in a standard T4 RNA ligase reaction. Products from the second ligation were gel-purified, then reverse transcribed and amplified by using the primers corresponding to the adaptor sequences. To achieve ligation specificity for RNA with a 5´-terminal phosphate and 3´-terminal hydroxyl, phosphatase and phosphorylase treatments, useful for preventing circularization of Dicer products (*12*), were not included our protocol. Instead, circularization was avoided by using the pre-adenylylated 3´-adaptor oligonucleotide and omitting ATP during the first ligation reaction.

24. Supplemental material describing methods and predicted fold-back secondary structures for the miRNAs of Table 1 and some of their homologs in other species is available at *Science* Online at www.sciencemag.org/???????

25. M. Lagos-Quintana, R. Rauhut, W. Lendeckel, T. Tuschl, *Science* **in press** (2001).

26. R. C. Lee, V. Ambros, *Science* **in press** (2001).

27. *C. elegans* Sequencing Consortium, *Science* **282**, 2012 (1998).

28. M. J. Beanan, S. Strome, *Development* **116**, 755 (1992).

29. S. W. Knight, B. L. Bass, *Science* **in press** (2001).

30. L. C. Lau, L. P. Lee, E. Weinstein, D. P. Bartel, data not shown

31. A. E. Pasquinelli, et al., *Nature* **408**, 86 (2000).

32. This begs the question as to why more riboregulators have not been found previously. Perhaps they had not been identified biochemically because of a predisposition towards searching for protein rather than RNA factors. They could be identified genetically—this was how *lin-4* and *let-7* were discovered (*1-3*)—however, when compared to mutations in protein-coding genes, point substitutions in these short RNA genes would be less likely and perhaps less disruptive of function. Furthermore, mutations that map to presumed intergenic regions with no associated RNA transcript detectable on a standard RNA blot might be put aside in favor of other mutants.

33. Wormbase is available on the Web at www.wormbase.org.

34. Sequencing traces representing 2.5–3-fold average coverage of the *C. briggsae* genome were obtained at www.ncbi.nlm.nih.gov/Traces

35. I. L. Hofacker, et al., *Monatshefte f. Chemie* **125**, 167 (1994).

36. We thank C. Ceol for guidance in culturing and staging *C. elegans*, R. Horvitz for the use of equipment and facilities, T. Tuschl, G. Ruvkun, B. Reinhart, A. Pasquinelli, A. Ensminger, and C. Mello for helpful discussions, and P. Zamore, T. Orr-Weaver, and M. Lawrence for comments on this manuscript, and T. Tuschl, V. Ambros, G. Ruvkin, R. Horvitz, P. Sharp, and J. Hodgkin for discussions on nomenclature.

Table 1. miRNAs cloned from *C. elegans*. 300 RNA clones represented 54 different miRNAs. Also included are miR-39, miR-65, and miR-69, three miRNAs predicted based on homology and/or proximity to cloned miRNAs. miR-39 and miR-69 have been validated by Northern analysis (Fig. 3), whereas miR-65 is not sufficiently divergent to be readily distinguished by Northern analysis. All *C. elegans* sequence analysis relied on WormBase, release WS45 (*33*). Some miRNAs were represented by clones of different lengths, due to heterogeneity at the miRNA 3´-terminus. The observed lengths are indicated, as is the sequence of the most abundant length. Comparison to *C. briggsae* shotgun sequencing traces revealed miRNA orthologs with 100% sequence identity (+++) and potential orthologs with >90% (++) and >75% (+) sequence identity (*24, 34*). Five miRNA genomic clusters are indicated with square brackets. Naming of miRNAs was coordinated with the Tuschl and Ambros groups (*25, 26*).

miRNA gene	Number of clones	miRNA Sequence	Length	*C. briggsae* homology	Fold-back arm	Chromosome and distance to nearest gene
lin-4	10	UCCCUGAGAC CUCAAGUGUG A	21	+++	5'	II
let-7	1	UGAGGUAGUA GGUUGUAUAG UU	22	+++	5'	X
mir-1	9	UGGAAUGUAA AGAAGUAUGU A	21	+++	3'	I 3.7 kb from start of T09B4.3, antisense
mir-2	24	UAUCACAGCC AGCUUUGAUG UGC	22–23	+++	3'	I 0.6 kb from start of M04C9.6b
mir-34	3	AGGCAGUGUG GUUAGCUGGU UG	22	+++	5'	X 2.1 kb from end of Y41G9A.4, antisense
mir-35	9	UCACCGGGUG GAAACUAGCA GU	22	+	3'	II 1.3 kb from end of F54D5.12, antisense
mir-36	1	UCACCGGGUG AAAAUUCGCA UG	22	+	3'	II 1.2 kb from end of F54D5.12, antisense
mir-37	2	UCACCGGGUG AACACUUGCA GU	22	++	3'	II 1.1 kb from end of F54D5.12, antisense
mir-38	1	UCACCGGGAG AAAAACUGGA GU	22	+	3'	II 1.0 kb from end of F54D5.12, antisense
mir-39	0	UCACCGGGUG UAAAUCAGCU UG	predicted	++	3'	II 0.8 kb from end of F54D5.12, antisense
mir-40	2	UCACCGGGUG UACAUCAGCU AA	22	+	3'	II 0.7 kb from end of F54D5.12, antisense
mir-41	2	UCACCGGGUG AAAAAUCACC UA	22	+	3'	II 0.6 kb from end of F54D5.12, antisense
mir-42	1	CACCGGGUUA ACAUCUACAG	20	+++	3'	II 1.2 kb from end of ZK930.2, antisense
mir-43	1	UAUCACAGUU UACUUGCUGU CGC	23	+++	3'	II 1.1 kb from end of ZK930.2, antisense
mir-44	3	UGACUAGAGA CACAUUCAGC U	21	+++	3'	II 1.0 kb from end of ZK930.2, antisense
mir-45				+++	3'	II 0.7 kb from end of K12D12.1, antisense
mir-46	2	UGUCAUGGAG UCGCUCUCUU CA	22	+++	3'	III 3.0 kb from end of ZK525.1, antisense
mir-47	6	UGUCAUGGAG GCGCUCUCUU CA	22	+++	3'	X 1.8 kb from end of K02B9.2, antisense
mir-48	11	UGAGGUAGGC UCAGUAGAUG CGA	22–24	+++	5'	V 6.1 kb from start of Y49A3A.4
mir-49	1	AAGCACCACG AGAAGCUGCA GA	22	+++	3'	X 2.7kb from end of F19C6.1, antisense
mir-50	2	UGAUAUGUCU GGUAUUCUUG GGUU	24	++	5'	I in intron of Y71G12B.11a
mir-51	6	UACCCGUAGC UCCUAUCCAU GUU	23	++	5'	IV 0.4 kb from end of F36H1.6, antisense
mir-52	47	CACCCGUACA UAUGUUUCCG UGCU	22–25	+++	5'	IV 4.6 kb from end of Y37A1B.6, antisense
mir-53	2	CACCCGUACA UUUGUUUCCG UGCU	24	–	5'	IV 1.9 kb from end of F36H1.6, antisense
mir-54	2	UACCCGUAAU CUUCAUAAUC CGAG	24	+	3'	X 5.5 kb from end of F09A5.2, antisense
mir-55	5	UACCCGUAUA AGUUUCUGCU GAG	23	+	3'	X 5.3 kb from end of F09A5.2, antisense
mir-56	5	UACCCGUAAU GUUUCCGCUG AG	22	+	3'	X 5.2 kb from end of F09A5.2, antisense
	2	UGGCGGAUCC AUUUUGGGUU GUA	23	+	5'	X 5.2 kb from end of F09A5.2, antisense
mir-57	9	UACCCUGUAG AUCGAGCUGU GUGU	24	+++	5'	II 0.9 kb from start of AF187012-1.T09A5
mir-58	31	UGAGAUCGUU CAGUACGGCA AU	21–22	+++	3'	IV in intron of Y67D8A.1
mir-59	1	UCGAAUCGUU UAUCAGGAUG AUG	23	+	3'	IV 1.8 kb from start of B0035.1a, antisense
mir-60	1	UAUUAUGCAC AUUUUCUAGU UCA	23	++	3'	II 1.5 kb from end of C32D5.5
mir-61	1	UGACUAGAAC CGUUACUCAU C	21	+	3'	V 0.4 kb from end of F55A11.3, antisense
mir-62	1	UGAUAUGUAA UCUAGCUUAC AG	22	+++	3'	X in intron of T07C5.1
mir-63	1	UAUGACACUG AAGCGAGUUG GAAA	24	–	3'	X 1.7 kb from start of C16H3.2, antisense
mir-64	2	UAUGACACUG AAGCGUUACC GAA	23	–	5'	III 0.25 kb from start of Y48G9A.1
mir-65	0	UAUGACACUG AAGCGUAACC GAA	predicted	+	5'	III 0.10 kb from start of Y48G9A.1
mir-66	10	CAUGACACUG AUUAGGGAUG UGA	23–24	–	5'	III in coding sequence of Y48G9A.1
mir-67	2	UCACAACCUC CUAGAAAGAG UAGA	24	+++	3'	III 4.4 kb from end of EGAP1.1
mir-68	1	UCGAAGACUC AAAAGUGUAG A	21	–	3'	IV 3.3 kb from start of Y51H4A.22
mir-69	0	UCGAAAAUUA AAAAGUGUAG A	predicted	–	3'	IV 0.6 kb from start of Y41D4B.21, antisense
mir-70	1	UAAUACGUCG UUGGUGUUUC CAU	23	+	3'	V in intron of T10H9.5
mir-71	5	UGAAAGACAU GGGUAGUGA	19, 20, 22	+++	5'	I 7.8 kb from start of M04C9.6b
mir-72	9	AGGCAAGAUG UUGGCAUAGC	20, 21, 23	–	3'	II 0.21 kb from end of F53G2.4, antisense
mir-73	2	UGGCAAGAUG UAGGCAGUUC AGU	23	++	3'	X 2.9 kb from start of T24D8.6, antisense
mir-74	7	UGGCAAGAAA UGGCAGUCUA CA	22	++	3'	X 3.2 kb from start of T24D8.6, antisense
mir-75	2	UUAAAGCUAC CAACCGGCUU CA	22	++	3'	X 3.5 kb from start of F47G3.3
mir-76	1	UUCGUUGUUG AUGAAGCCUU GA	22	++	3'	III 3.0 kb from start of C44B11.3, antisense
mir-77	1	UUCAUCAgGC CAUAGCUGUC CA	22	+++	3'	II 1.5 kb from start of T21B4.9, antisense
mir-78	2	UGGAGGCCUG GUUGUUUGUG C	21	–	3'	IV 2.0 kb from start of Y40H7A.3, antisense
mir-79	1	AUAAAGCUAG GUUACCAAAG CU	22	+++	3'	I 2.3 kb from end of C12C8.2
mir-80	25	UGAGAUCAUU AGUUGAAAGC CGA	23	+++	3'	III 4.7 kb from end of F44E2.2, antisense
mir-81	7	UGAGAUCAUC GUGAAAGCUA GU	22	+++	3'	X in intron of T07D1.2, antisense
mir-82	6	UGAGAUCAUC GUGAAAGCCA GU	22	+++	3'	X 0.11 kb from start of T07D1.2
mir-83	1	UAGCACCAUA UAAAUUCAGu AA	22	++	3'	IV 5.0 kb from start of C06A6.2
mir-84	3	UGAGGUAGUA UGUAAUAUUG UA	22, 24	+	5'	X 0.8 kb from end of B0395.1, antisense
mir-85	1	UACAAAGUAU UUGAAAAGUC GUGC	24	++	3'	II in intron of F49E12.8, antisense
mir-86	6	UAAGUGAAUG CUUUGCCACA GUC	23	+++	5'	III in intron of Y56A3A.7

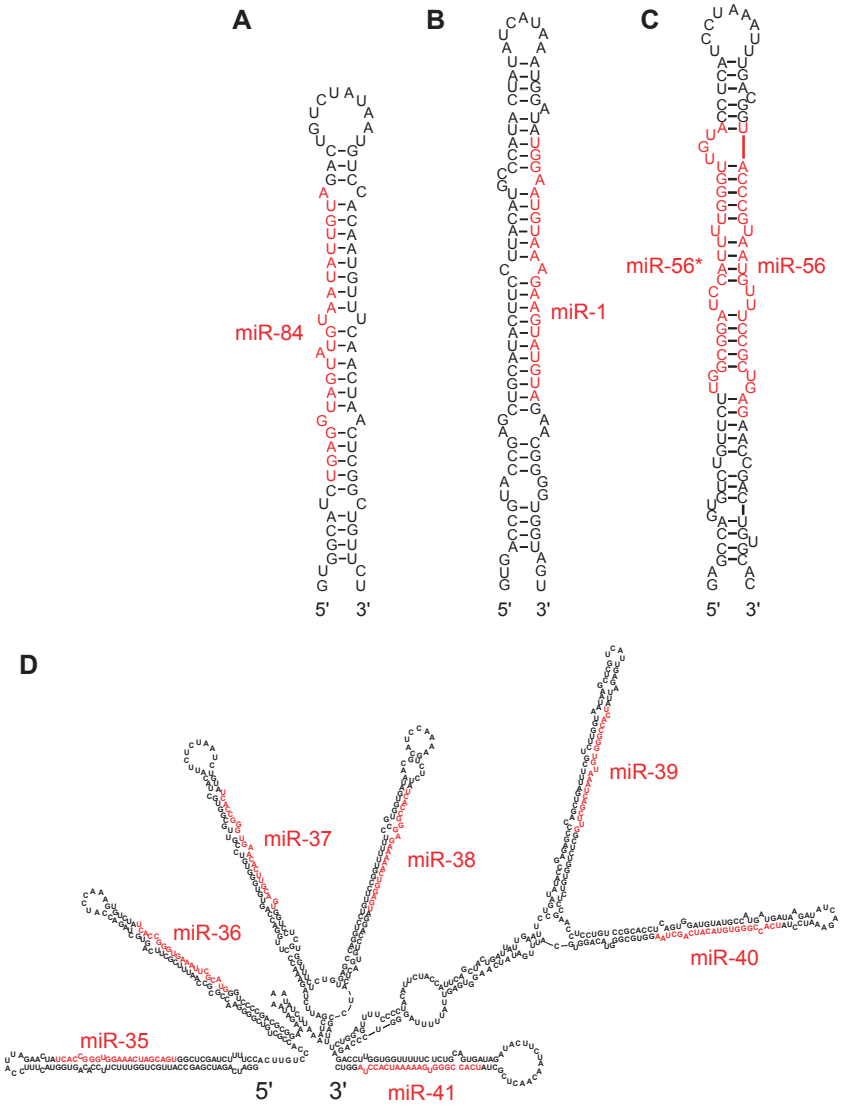
Figure Legends

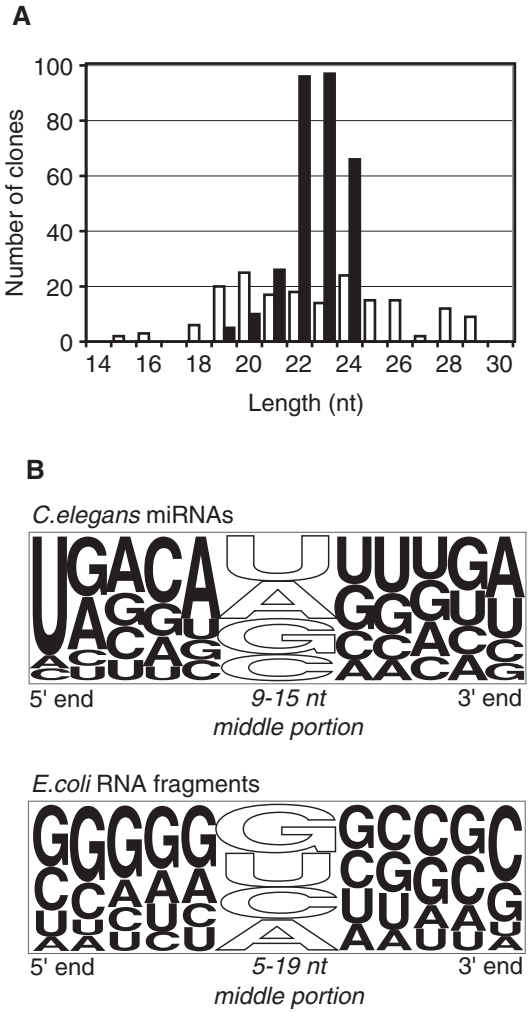
Fig. 1. Fold-back secondary structures involving miRNAs (red) and their flanking sequences (black), as predicted computationally using RNAfold (*35*). (**A**) miR-84, an miRNA with similarity to *let-7* RNA. (**B**) miR-1, an miRNA highly conserved in evolution. (**C**) miR-56 and miR-56*, the only two miRNAs cloned from both sides of the same fold-back. (**D**) The *mir-35–41* cluster.

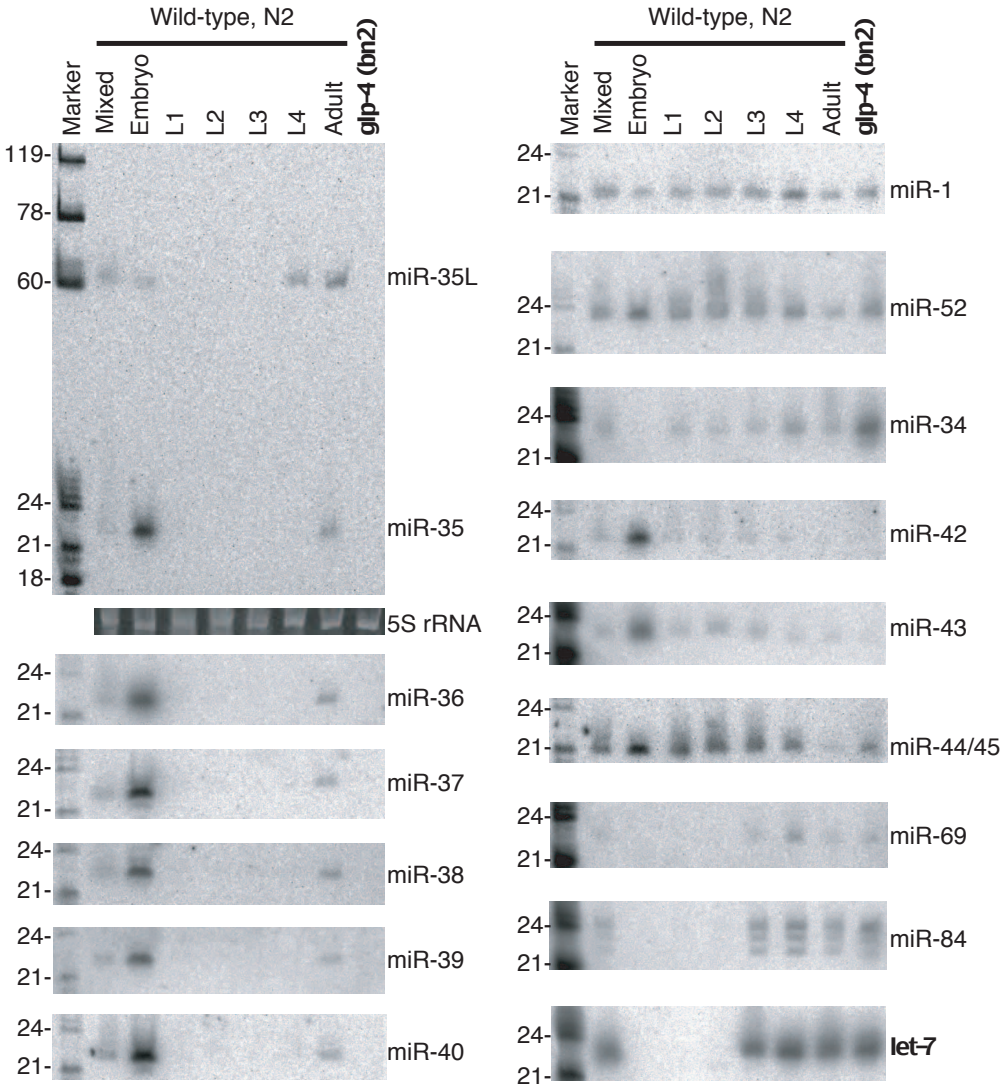
Fig. 2. Unique sequence features of the miRNAs. (**A**) Length distribution of the clones representing *E. coli* RNA fragments (white bars) and *C. elegans* miRNAs (black bars). (**B**) Sequence composition of the unique clones representing *C. elegans* miRNAs and *E. coli* RNA fragments. To avoid over-representation from groups of related miRNAs in this analysis, each set of paralogs was represented by its consensus sequence.

Fig. 3. Expression of novel miRNAs and *let-7* RNA during *C. elegans* development. Northern blots probe total RNA from mixed-stage worms (Mixed), worms staged as indicated, and *glp-4 (bn2)* adult worms (*24*). Specificity controls ruled out cross-hybridization among probes for miRNAs from the *mir-35–41* cluster (*24*). Other blots indicate that, miR-46/47, miR-56, miR-64/65, miR-66, and miR-80 are expressed constitutively throughout development (*30*).

Lau et al.
Figure 1.




A
100
80
60
40
20
0
Number of clones
14
16
18
20
22
24
26
28
30
Length (nt)
B
C.elegans miRNAs
5' end
9-15 nt
3' end
middle portion
E.coli RNA fragments
5' end
5-19 nt
3' end
middle portion


Wild-type, N2
Marker
Mixed
Embryo
L1
L2
L3
L4
Adult
glp-4 (bn2)
119-
78-
60-
24-
21-
18-
miR-35L
miR-35
5S rRNA
miR-36
miR-37
miR-38
miR-39
miR-40
miR-1
miR-52
miR-34
miR-42
miR-43
miR-44/45
miR-69
miR-84
let-7