

Identification of Novel Genes Coding for Small Expressed RNAs

Mariana Lagos-Quintana, Reinhard Rauhut, Winfried Lendeckel, Thomas Tuschl*

In *Caenorhabditis elegans*, *lin-4* and *let-7* encode 22- and 21-nucleotide (nt) RNAs, respectively, which function as key regulators of developmental timing. Because the appearance of these short RNAs is regulated during development, they are also referred to as small temporal RNAs (stRNAs). We show that many 21- and 22-nt expressed RNAs, termed microRNAs, exist in invertebrates and vertebrates and that some of these novel RNAs, similar to *let-7* stRNA, are highly conserved. This suggests that sequence-specific, posttranscriptional regulatory mechanisms mediated by small RNAs are more general than previously appreciated.

Two distinct pathways exist in animals and plants in which 21- to 23-nt RNAs function as posttranscriptional regulators of gene expression. Small interfering RNAs (siRNAs) act as mediators of sequence-specific mRNA degradation in RNA interference (RNAi) (1–

5), whereas stRNAs regulate developmental timing by mediating sequence-specific repression of mRNA translation (6–11). siRNAs and stRNAs are excised from double-stranded RNA (dsRNA) precursors by Dicer (12–14), a multidomain ribonuclease

III protein, thus producing RNA species of similar sizes. However, siRNAs are believed to be double-stranded (2, 5, 12), whereas stRNAs are single-stranded (8).

We previously developed a directional cloning procedure to isolate siRNAs after processing of long dsRNAs in *Drosophila melanogaster* embryo lysate (2). Briefly, 5' and 3' adapter molecules were ligated to the ends of a size-fractionated RNA population, followed by reverse transcription polymerase chain reaction (PCR) amplification, concatamerization, cloning, and sequencing. This method, originally intended to isolate siRNAs, led to the simultaneous identification of 16 novel 20- to 23-nt short RNAs, which are encoded in the *D. melanogaster* genome and are expressed in 0- to 2-hour embryos (Table 1). The method was adapted to clone RNAs in a similar size range from HeLa cell total RNA (15), which led to the

Department of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, D-37077 Göttingen, Germany.

*To whom correspondence should be addressed. E-mail: ttuschl@mpibpc.gwdg.de

Fig. 1. Expression of miRNAs. Representative examples of Northern blot analysis are depicted (27). The position of 76-nt val-tRNA is indicated on the blots; 5S rRNA serves as a loading control. **(A)** Northern blots of total RNA isolated from staged populations of *D. melanogaster*, probed for the indicated miRNA. E, embryo; L, larval stage; P, pupa; A, adult; S2, Schneider-2 cells. **(B)** Northern blots of total RNA isolated from HeLa cells, mouse kidneys, adult zebrafish, frog ovaries, and S2 cells, probed for the indicated miRNA.

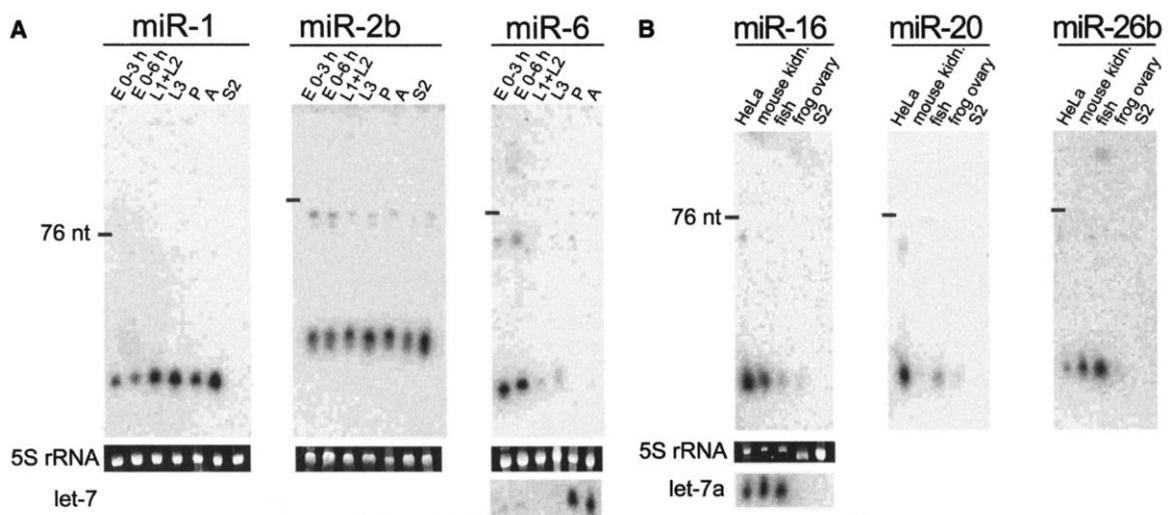
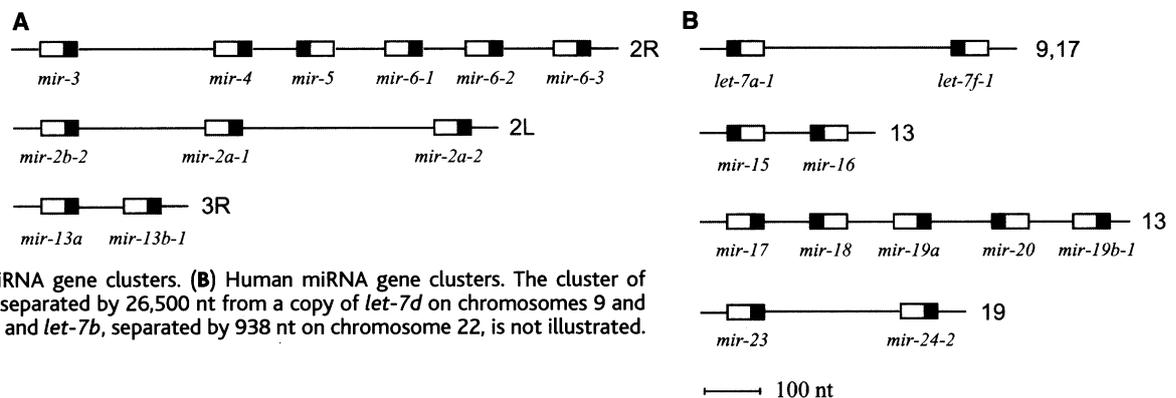


Fig. 2. Genomic organization of miRNA gene clusters. The precursor structure is indicated as a box, and the location of the miRNA within the precursor is shown in black; the chromosomal location is also indicated to the right. **(A)** *D. melanogaster* miRNA gene clusters. **(B)** Human miRNA gene clusters. The cluster of *let-7a-1* and *let-7f-1* is separated by 26,500 nt from a copy of *let-7d* on chromosomes 9 and 17. A cluster of *let-7a-3* and *let-7b*, separated by 938 nt on chromosome 22, is not illustrated.



REPORTS

identification of 21 novel human micro-RNAs (Table 2), thus providing further evidence for the existence of a large class of small RNAs with potential regulatory roles. Because of their small size, and in agreement with the authors of two related papers in this issue (16, 17), we refer to these novel RNAs as micro-RNAs (miRNAs). The miRNAs we studied are abbreviated as miR-1 to miR-33, and the genes encoding miRNAs are named *mir-1* to *mir-33*. Highly homologous miRNAs are referred to by the same gene number, but followed by a lowercase letter; multiple genomic copies of a *mir* gene are annotated by adding a dash and a number.

The expression and size of the cloned, endogenous short RNAs were also examined by Northern blotting (Fig. 1 and Tables 1 and 2). For analysis of *D. melanogaster* RNAs, total RNA was prepared from different developmental stages, as well as from cultured Schneider-2 (S2) cells, which were originally derived from 20- to 24-hour *D. melanogaster* embryos (18) (Fig. 1 and Table 1). miR-3 to miR-7 are expressed only during embryogenesis and not at later developmental stages. The temporal expression of miR-1, miR-2, and miR-8 to miR-13 was less restricted. These miRNAs were observed at all developmental stages, and significant variations in the expression levels were

sometimes observed. Interestingly, miR-1, miR-3 to miR-6, and miR-8 to miR-11 were completely absent from cultured S2 cells, whereas miR-2, miR-7, miR-12, and miR-13 were present in S2 cells, therefore indicating cell type-specific miRNA expression. miR-1, miR-8, and miR-12 expression patterns are similar to those of *lin-4* stRNA in *C. elegans*, as their expression is strongly up-regulated in larvae and sustained to adulthood (19). miR-9 and miR-11 are present at all stages but are strongly reduced in the adult, which may reflect a maternal contribution from germ cells or expression in one sex only.

The *mir-3* to *mir-6* genes are clustered (Fig.

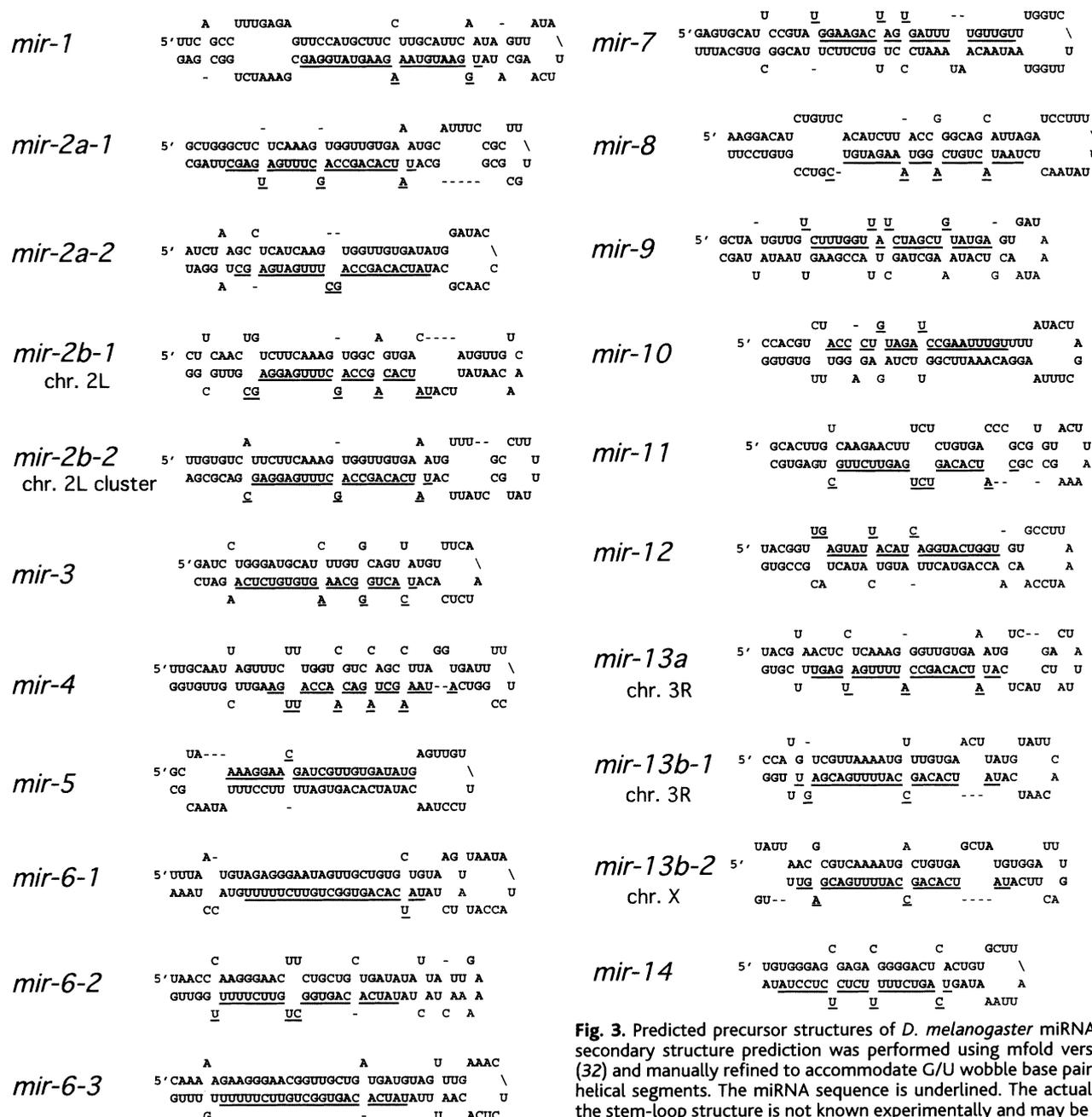


Fig. 3. Predicted precursor structures of *D. melanogaster* miRNAs. RNA secondary structure prediction was performed using mfold version 3.1 (32) and manually refined to accommodate G/U wobble base pairs in the helical segments. The miRNA sequence is underlined. The actual size of the stem-loop structure is not known experimentally and may be slightly shorter or longer than represented. Multicopy miRNAs and their corresponding precursor structures are also shown.

REPORTS

let-7a-1
chr. 9,17

```

      U      UAGG ACA C
5' UGGGA GAGGUAGUAGGUUUAUAGU GUC CCCA C
      AUCCU UUCUGUCAUCAUCAUCAAA UAG GGGU A
      U-----A--C
    
```

let-7a-2
chr. 11

```

      UU G U      UGAAUUAAC AA
5' AGG GAG UAG AGGUUUAUAGU AUC G
      UCC UUC AUC UCCGACAUUCUAA UAG G
      U- G C -----AG
    
```

let-7a-3
chr. 22

```

      U      ----- U
5' GGG GAGGUAGUAGGUUUAUAGU UGGGGC \
      UCC UUCUGUCAUCAUCAUCAAA GUCCCC C
      U      UAGGGUAUC U
    
```

let-7b

```

      U      - A-----UG
5' CGGGG GAGGUAGUAGGUUUGUGUUU UC GGGCAG \
      GUCCC UUCUGUCAUCAUCAUCAAA AG CCGGUU A
      - U AAGGCUC GU
    
```

let-7c

```

      A      UU G U      UA G A AC
5' GC UCCGGG GAG UAG AGGUUUAUAGU GA UU C \
      CG AGGUUC UUC AUC UCCAACAUGUCAAA UU AG G C
      - CU G U -- G U UC
    
```

let-7d

```

      A      C      UUA-----GG
5' CCUAGGA GAGGUAGUAGGUU AUAGUU GGGCAG \
      GGAUUCU UUCUGUCAUCAUCAUCAAA CCGGUU A
      - A UGGAGGAACA UU
    
```

let-7e

```

      C CU G      U GGA----A
5' CC GGG GAG UAGGAGGUUUAUAGU GA GG C
      GG CCC UUC AUCCUCCGCAUUAUCA CU CC A
      A CU G - AGAGGAA C
    
```

let-7f-1
chr. 9,17

```

      AGU      -----UG
5' UCAG GAGGUAGUAGAUUUAUAGU GGGUAG \
      AGUC UUCUGUCAUCAUCAUCAAA UCCCAU A
      CC- GAGGACTUG UU
    
```

let-7f-2
chr. X

```

      U      ----UCAU
5' CUGUGGA GAGGUAGUAGAUUUAUAGU UUAAGG A
      GGCAACCU UUCUGUCAUCAUCAUCAAA GGUUCU C
      - UAGA ACCC
    
```

mir-15

```

      GAGUAAAGUA      UA      GA U
5' CCTUG GCAGCACA AUGGUUUGU UUU \
      GGAAC CGUCUGUU UAACGGACGU AAA G
      AUAAAAACUC UA GG A
    
```

mir-16

```

      AG C      - A      CGUUA UCUA
5' GUCAGC UGC UUAAGCAGCAC GU AAUAUUGG AGAU \
      CAGUUG AUG AGUCGUCGUG CA UUAUGACC UCUA A
      GA A U A -----UUA
    
```

mir-17

```

      GA      CA-      A G      G - AUA
5' GUCA AUAUUGU AAGUGCUU CA UGCAG UAG UG \
      CAGU UAUUACG UUCACGGA GU ACGUC AUC AC U
      GG AUG A G - U GUG
    
```

mir-18

```

      CU      U      C      U      A      UGAA AG
5' UGUU AAGG GCAU UAG GCAG UAG GU A
      ACGG UUCC CGUG AUC CGUC AUC CG U
      UC U A C - UA--AU
    
```

mir-19a

```

      U      U      -- ---AGA
5' GCAG CC CUGUAGUUUUGCAUAG UUGCAC UACA \
      CGUC GG GGUAGUCAAAACGU AACGUG AUGU A
      C U U A UUG AAG
    
```

mir-19b-1
chr. 13

```

      UU      - - UC UGUGUG
5' CACUG CUUAGGUUAGUUUUGCA GG UUUGCA CAGC \
      GUGAU GUGUACAGUCAAAACGU CC AAACGU GUCG A
      -- A U -----UCUUAU
    
```

mir-19b-2
chr. X

```

      CUAC      - - UUCA U
5' ACAUUG UTAACAUAUAGUUUUGCA GG UUGCAU GCGUAUA A
      UGUAAU AGUUAUAGUCAAAACGU CC AAACGU UGUAAU U
      ---- A U UCGG G
    
```

mir-20

```

      C ^ A ^      ^ ^ - UU ^ ^ ^
5' GUAG ACU AAGUCUUUAUAGUGCAG UAG UG U \
      CGUC UGA UUCACGAGUAUAUCGUC AUC AU A
      A AA - U UG
    
```

mir-21

```

      A A A U AA
5' UGUCGGGUAGCUTUAUC GACTUG UGUUG CUGU G \
      ACAGUCUGUCGGGUAG CUGAC ACAAC GGUA C U
      - C - - UC
    
```

mir-22

```

      U CC      - A U CCUG
5' GGC GAG GCAUAGUUCUUCAG UGGCA GCUUUA GU \
      CCG CUC CGUUGUCAAGAAUGU ACCGU CGAAAU CG A
      U C- G - - ACCC
    
```

mir-23

```

      C C - G G CUUC
5' GG CGG UGGGG UUCUCUGG GAUG GAUUGU C
      CC GCC ACCU AGGGACC UUA CUAAAC U
      A A U G A ACUG
    
```

mir-24-1
chr. 9

```

      G G A      UA UCUCAU
5' CUCC GU CCU CUGAGCUGA UCAGU \
      GAGG CA GGA GACUUGACU GGUCA U
      A A C C- CACAUU
    
```

mir-24-2
chr. 19

```

      CC CG CU- AA-- UU
5' CUCUG UCC UGC ACUGAGCUG ACACAG \
      GGGAC AGG ACG UGACUCUGU UGUUUU G
      A- -- ACU CACA UG
    
```

mir-25

```

      A AG G UU G UG ACG
5' GGCC GUGUUG AGGC GAGAC G GCAAU CUGG C
      CCG CGUGAC UCUG CUCUG C CGUUA GGUU U
      C AG G - UU A CG CCG
    
```

mir-26a

```

      G      U      U      GCAG
5' GUG CCUCGU CAAGUAA CCAGAUAGGCUU G
      CGC GGGCA GUUCAU GGUUCUAUCCGGUA U
      A C - ACCC
    
```

mir-26b

```

      GA - U UC UUGU
5' CCGG CCC AGU CAAGUAAU AGGAUAGGUU \
      GGCC GGG UCG GUUCAUA UCUUUGCCGAC C
      AG C - CC CUGU
    
```

mir-27

```

      A A A      U      G UCCAC
5' CUG GG GC GGGCUUAGCUGUC GUGAGCA GG \
      GAC CC CG CUUGAAUCGGUA CACUUG CU A
      C C C - G GAAAC
    
```

mir-28

```

      C      A      U      ----CC
5' GGU CUUGCCUC AGGACUCACAGUCUA UG AGUUA U
      UCA GGACGGGAG UCCUCGAGUUAUAGAU AC UCAGU U
      C G C C UCUU CU
    
```

mir-29

```

      UUU      C UCAAU
5' AUGACUGAUUUC UGGUGUU AGAG \
      UAUUGGCUAAG ACCACGA UCTU A
      UCU - UUAUU
    
```

mir-30

```

      A      UC      ----A
5' GCG CUGUAAACAUC GACUGGAAGCU GUG A
      CGU GACGUUUUAGG UUGACTUUUCG CAC G
      C -- GUAGA C
    
```

mir-31

```

      A      G      C      U-      GAA
5' GGAGAGG GGCAA AUG UGGCAUAGC GUU C
      CCUUUCU CCGUU UAC ACCGUUUCG CAA U
      A A A UC GGG
    
```

mir-32

```

      U      - UU C
5' GGAGAUUAUUGCAU ACTUAAGUUGCAU G GU A
      CUUUUAUAGUGUGUG UGAUUUAACGUA C CG C
      - A UC G
    
```

mir-33

```

      A UU      UUCU UG
5' CUGUGUGCAUUGU G GCAUUGCAU GG \
      GACACUACGUGACA C UGUAAAGUAC CC G
      C UU ----AU
    
```

Fig. 4. Predicted precursor structures of human miRNAs. For legend, see Fig. 3.

REPORTS

Table 1. *D. melanogaster* miRNAs. The sequences given represent the most abundant, and typically longest, miRNA sequence identified by cloning; miRNAs frequently vary in length by one or two nucleotides at their 3' termini. From 222 short RNAs sequenced, 69 (31%) corresponded to miRNAs, 103 (46%) to already characterized functional RNAs (rRNA, 75L RNA, and tRNA), 30 (14%) to transposon RNA fragments, and 20 (10%) sequences had no database entry. The frequency for cloning a particular

miRNA as a percentage relative to all identified miRNAs is indicated. Results of Northern blotting of total RNA isolated from staged populations of *D. melanogaster* are summarized. E, embryo; L, larval stage; P, pupa; A, adult; S2, Schneider-2 cells. The strength of the signal within each blot is represented from strongest (+++) to undetected (-). *let-7* stRNA was probed as the control. GenBank accession numbers and homologs of miRNAs identified by database searching in other species are provided in (21).

miRNA	Sequence (5' to 3')	Freq. (%)	E 0 to 3 hours	E 0 to 6 hours	L1 + L2	L3	P	A	S2
miR-1	UGGAAUGUAAAGAAGUAUGGAG	32	+	+	+++	+++	++	+++	-
miR-2a*	UAUCACAGCCAGCUUUGAUGAGC	3							
miR-2b*	UAUCACAGCCAGCUUUGAGGAGC	3	++	++	++	+++	++	+	+++
miR-3	UCACUGGGCAAAGUGUGUCUCA	9	+++	+++	-	-	-	-	-
miR-4	AUAAAGCUAGACAACCAUUGA	6	+++	+++	-	-	-	-	-
miR-5	AAAGGAACGAUCGUUGUGAUUAUG	1	+++	+++	+/-	+/-	-	-	-
miR-6	UAUCACAGUGGCUGUUCUUUUU	13	+++	+++	+/-	+/-	-	-	-
miR-7	UGGAAGACUAGUGAUUUUGUUGU	4	+++	++	+/-	+/-	+/-	+/-	+/-
miR-8	UAAUACUGUCAGGUAAAGAUGUC	3	+/-	+/-	+++	+++	+	+++	-
miR-9	UCUUUGGUUAUCUAGCUGUAUGA	7	+++	++	+++	+++	+++	+/-	-
miR-10	ACCCUGUAGAUCGGAUUUUGU	1	+	+	++	+++	+/-	+	-
miR-11	CAUCACAGUCUGAGUUCUUGC	7	+++	+++	+++	+++	+++	+	-
miR-12	UGAGUAAUACAUCAGGUACUGGU	7	+	+	++	++	+	+++	+/-
miR-13a*	UAUCACAGCCAUUUUGACGAGU	1	+++	+++	+++	+++	+	+++	+++
miR-13b*	UAUCACAGCCAUUUUGAUGAGU	0							
miR-14	UCAGUCUUUUUCUCUCUCCUA	1	-	-	-	-	-	-	-
let-7	UGAGGUAGUAGGUUGUAUAGUU	0	-	-	-	-	+++	+++	-

*Similar miRNA sequences are difficult to distinguished by Northern blotting because of potential cross-hybridization of probes.

2A), and *mir-6* is present as triple repeat with slight variations in the *mir-6* precursor sequence but not in the miRNA sequence itself. The expression profiles of miR-3 to miR-6 are highly similar (Table 1), which suggests that a single embryo-specific precursor transcript may give rise to the different miRNAs or that the same enhancer regulates miRNA-specific promoters. Several other fly miRNAs are also found in gene clusters (Fig. 2A).

The expression of HeLa cell miR-15 to miR-33 was examined by Northern blotting using HeLa cell total RNA, in addition to total RNA prepared from mouse kidney, adult zebrafish, *Xenopus laevis* ovary, and *D. melanogaster* S2 cells (Fig. 1B and Table 2). miR-15 and miR-16 are encoded in a gene cluster (Fig. 2B) and are detected in mouse kidney, adult zebrafish, and very weakly in frog ovary, which may result from miRNA expression in somatic ovary tissue rather than in oocytes. *mir-17* to *mir-20* are also clustered (Fig. 2B) and are expressed in HeLa cells and adult zebrafish, but undetectable in mouse kidney and frog ovary (Fig. 1 and Table 2), and therefore represent a likely case of tissue-specific miRNA expression.

The majority of vertebrate and invertebrate miRNAs identified in this study are not related by sequence, but a few exceptions do exist and are similar to results previously reported for *let-7* RNA (8). Sequence analysis of the *D. melanogaster* miRNAs revealed four such instances of sequence conservation between invertebrates and vertebrates. miR-1 homologs are encoded in the genomes of *C. elegans*, *C. briggsae*, and humans and are found in cDNAs from zebrafish, mice, cows, and humans. The expres-

sion of *mir-1* was detected by Northern blotting in total RNA from adult zebrafish and *C. elegans*, but not in total RNA from HeLa cells or mouse kidney (Table 2) (20). Interestingly, although *mir-1* and *let-7* are both expressed in adult flies (Fig. 1A) (8) and are both undetected in S2 cells, only *let-7* is detectable in HeLa cells. This represents another case of tissue-specific expression of an miRNA and indicates that miRNAs may play a regulatory role not only in developmental timing but also in tissue specification. miR-7 homologs were found by database searches of the mouse and human genomes and of expressed sequence tags (ESTs). Two mammalian miR-7 variants are predicted by sequence analysis in mice and humans and were detected by Northern blotting in HeLa cells and adult zebrafish, but not in mouse kidney (Table 2). Similarly, we identified mouse and human miR-9 and miR-10 homologs by database searches but only detected *mir-10* expression in mouse kidney.

The identification of evolutionarily related miRNAs, which have already acquired multiple sequence mutations, was not possible by standard bioinformatic searches. Direct comparison of the *D. melanogaster* miRNAs with the human miRNAs identified an 11-nt segment shared between *D. melanogaster* miR-6 and HeLa miR-27, but no further relationships were detected. It is possible that most miRNAs only act on a single target and therefore allow for rapid evolution by covariation. Highly conserved miRNAs may act on more than one target sequence and therefore have a reduced probability for evolutionary drift by covariation (8). An alternative interpretation is that the sets

of miRNAs from *D. melanogaster* and humans are fairly incomplete and that many more miRNAs remain to be discovered, which will provide the missing evolutionary links.

lin-4 and *let-7* stRNAs were predicted to be excised from longer transcripts that contain stem-loop structures about 30 base pairs in length (6, 8). Database searches for newly identified miRNAs revealed that all miRNAs are flanked by sequences that have the potential to form stable stem-loop structures (Figs. 3 and 4). In many cases, we were able to detect the predicted precursors (about 70 nt) by Northern blotting (Fig. 1). Some miRNA precursor sequences were also identified in mammalian cDNA (EST) databases (21), indicating that primary transcripts longer than 70-nt stem-loop precursors also exist. We never cloned a 22-nt RNA complementary to any of the newly identified miRNAs, and it is as yet unknown how the cellular processing machinery distinguishes between an miRNA and its complementary strand. Comparative analysis of the precursor stem-loop structures indicates that the loops adjacent to the base-paired miRNA segment can be located on either side of the miRNA sequence (Figs. 3 and 4), suggesting that neither the 5' nor the 3' location of the stem-closing loop is the determinant of miRNA excision. It is also unlikely that the structure, length, or stability of the precursor stem is the critical determinant because the base-paired structures are frequently imperfect and interspersed by G/U wobbles and less stable, non-Watson-Crick base pairs such as G/A, U/U, C/U, and A/A. Therefore, a sequence-specific recognition process is a likely determinant for miRNA excision, perhaps me-

REPORTS

Table 2. Human miRNAs. From 220 short RNAs sequenced, 100 (45%) corresponded to miRNAs, 53 (24%) to already characterized functional RNAs (rRNA, snRNA, and tRNA), and 67 (30%) of the sequences had no database entry. Results of Northern blotting of total RNA isolated from different vertebrate species and S2 cells are indicated. For legend, see Table 1.

miRNA	Sequence (5' to 3')	Freq. (%)	HeLa cells	Mouse kidney	Adult fish	Frog ovary	S2
let-7a*	UGAGGUAGUAGGUUGUAUAGUU	10	+++	+++	+++	-	-
let-7b*	UGAGGUAGUAGGUUGUGUGUU	13					
let-7c*	UGAGGUAGUAGGUUGUAUGUU	3					
let-7d*	AGAGGUAGUAGGUUGCAUAGU	2	+++	+++	+++	-	-
let-7e*	UGAGGUAGGAGGUUGUAUAGU	2	+++	+++	+++	-	-
let-7f*	UGAGGUAGUAGAUUGUAUAGUU	1					
miR-15	UAGCAGCACAUAAUGGUUUUGUG	3	+++	++	+	+/-	-
miR-16	UAGCAGCACGUAUUUUGGCG	10	+++	+	+/-	+/-	-
miR-17	ACUGCAGUGAAGGCACUUGU	1	+++	-	-	-	-
miR-18	UAAGGUGCAUCUAGUGCAGUA	2	+++	-	-	-	-
miR-19a*	UGUGCAAUCUAUGCAAAACUGA	1	+++	-	+/-	-	-
miR-19b*	UGUGCAAUCUAUGCAAAACUGA	3					
miR-20	UAAAGUGCUUUAUGUCAGGUA	4	+++	-	+	-	-
miR-21	UAGCUUAUCAGACUGAUGUUGA	10	+++	+	++	-	-
miR-22	AAGCUGCCAGUUGAAGAACUGU	10	+++	+++	+	+/-	-
miR-23	AUCACAUUGCCAGGUAUUCC	2	+++	+++	+++	+	-
miR-24	UGGCUCAGUUCAGCAGGAACAG	4	++	+++	++	-	-
miR-25	CAUUGCACUUGUCUCGGUCUGA	3	+++	+	++	-	-
miR-26a*	UUCAAGUAAUCCAGGAUAGGCU	2	+	++	+++	-	-
miR-26b*	UUCAGUAAUCCAGGAUAGGCU	1					
miR-27	UUCACAGUGGCUAAGUUCGCU	2	+++	+++	++	-	-
miR-28	AAGGAGCUCACAGUCUAUUGAG	2	+++	+++	-	-	-
miR-29	CUAGCACCUCUGAAAUCGGUU	2	+	+++	+/-	-	-
miR-30	CUUUCAGUCGGAUGUUUGCAGC	2	+++	+++	+++	-	-
miR-31	GGCAAGAUUCGCGCAUAGCUG	2	+++	-	-	-	-
miR-32	UAUUGCACAUACUAAGUUGC	1	-	-	-	-	-
miR-33	GUGCAUUGUAGUUGCAUUG	1	-	-	-	-	-
miR-1	UGGAAUGUAAAGAAGUAUGGAG	0	-	-	+	-	-
miR-7	UGGAAGACUAGUAUUUUGUUGU	0	+	-	+/-	-	+/-
miR-9	UCUUUGGUUAUCUAGCUGUAUGA	0	-	-	-	-	-
miR-10	ACCCUGUAGAUCCGAAUUUGU	0	-	+	-	-	-

*Similar miRNA sequences are difficult to distinguish by Northern blotting because of potential cross-hybridization of probes.

diated by members of the Argonaute (RDE-1/AGO1/PIWI) protein family. Two members of this family, ALG-1 and ALG-2, have recently been shown to be critical for stRNA processing in *C. elegans* (13). Members of the Argonaute protein family are also involved in RNAi and posttranscriptional gene silencing. In *D. melanogaster*, these include Argonaute2, a component of the siRNA-endonuclease complex (RISC) (22), and its relative Aubergine, which is important for silencing of repeat genes (23). In other species, these include RDE-1 in *C. elegans* (24); Argonaute1 in *Arabidopsis thaliana* (25); and QDE-2 in *Neurospora crassa* (26). In addition to the RNase III Dicer (12, 13), the Argonaute family represents another evolutionary line between RNAi and miRNA maturation.

Despite advanced genome projects, computer-assisted detection of genes encoding functional RNAs remains problematic (27). Cloning of expressed, short functional RNAs, similar to EST approaches (RNomics), is a powerful alternative and probably the most efficient method for identification of such novel gene products (28–31). The number of functional RNAs has been widely underestimated and is expected to grow rapidly because of the development of new functional RNA cloning methodologies.

The challenge for the future is to define the function and the potential targets of these novel miRNAs by using bioinformatics as well as genetics and to establish a complete catalog of time- and tissue-specific distribution of the already identified and yet to be uncovered miRNAs. *lin-4* and *let-7* stRNAs negatively regulate the expression of proteins encoded by mRNAs in which 3' untranslated regions contain sites of complementarity to the stRNA (9–11). Because these interaction domains are only 6 to 10 base pairs long and often contain small bulges and G/U wobbles (9–11), the prediction of miRNA target mRNAs represents a challenging bioinformatic and/or genetic task. A profound understanding of the expression, processing, and action of miRNAs may enable the development of more general methods to direct the regulation of specific gene targets and may also lead to new ways of reprogramming tissues.

References and Notes

1. S. M. Elbashir *et al.*, *Nature* **411**, 494 (2001).
2. S. M. Elbashir, W. Lendeckel, T. Tuschl, *Genes Dev.* **15**, 188 (2001).
3. A. J. Hamilton, D. C. Baulcombe, *Science* **286**, 950 (1999).
4. S. M. Hammond, E. Bernstein, D. Beach, G. J. Hannon, *Nature* **404**, 293 (2000).
5. P. D. Zamore, T. Tuschl, P. A. Sharp, D. P. Bartel, *Cell* **101**, 25 (2000).

6. R. C. Lee, R. L. Feinbaum, V. Ambros, *Cell* **75**, 843 (1993).
7. B. J. Reinhart *et al.*, *Nature* **403**, 901 (2000).
8. A. E. Pasquinelli *et al.*, *Nature* **408**, 86 (2000).
9. V. Ambros, *Curr. Opin. Genet. Dev.* **10**, 428 (2000).
10. E. G. Moss, *Curr. Biol.* **10**, R436 (2000).
11. F. Slack, G. Ruvkun, *Annu. Rev. Genet.* **31**, 611 (1997).
12. G. Hutvagner, J. McLachlan, É. Bálint, T. Tuschl, P. D. Zamore, *Science* **293**, 834 (2001).
13. A. Grishok *et al.*, *Cell* **106**, 23 (2001).
14. E. Bernstein, A. A. Caudy, S. M. Hammond, G. J. Hannon, *Nature* **409**, 363 (2001).
15. Cloning of 19- to 24-nt RNAs from *D. melanogaster* 0- to 2-hour embryo lysate was performed as described (2). For cloning of HeLa miRNAs, 1 mg of HeLa total RNA was separated on a 15% denaturing polyacrylamide gel, and RNA of 19- to 25-nt size was recovered. A 5' phosphorylated 3' adapter oligonucleotide (5' pUUUaacccggaattccagc: uppercase, RNA; lowercase, DNA; p, phosphate; x, 4-hydroxymethylbenzyl) and a 5' adapter oligonucleotide (5' acggaattcctactAAA: uppercase, RNA; lowercase, DNA) were ligated to the short HeLa cell RNAs. Reverse transcription PCR was performed with 3' primer (5' GACTAGCTGGAATTCGGCGTTAAA) and 5' primer (5' CAGCCAACCGAATTCCTCACTAAA), followed by concatamerization after Eco RI digestion and T4 DNA ligation (2). After ligation of concatamers into pCR2.1 TOPO vectors, about 100 clones were selected and subjected to sequencing.
16. N. C. Lau, L. P. Lim, E. G. Weinstein, D. P. Bartel, *Science* **294**, 858 (2001).
17. R. C. Lee, V. Ambros, *Science* **294**, 862 (2001).
18. I. Schneider, *J. Embryol. Exp. Morphol.* **27**, 353 (1972).
19. R. Feinbaum, V. Ambros, *Dev. Biol.* **210**, 87 (1999).
20. M. Lagos-Quintana, R. Rauhut, W. Lendeckel, T. Tuschl, data not shown.

21. Supplementary Web material is available on Science Online at www.sciencemag.org/cgi/content/full/294/5543/853/DC1.
22. S. M. Hammond, S. Boettcher, A. A. Caudy, R. Kobayashi, G. J. Hannon, *Science* **293**, 1146 (2001).
23. A. A. Aravin *et al.*, *Curr. Biol.* **11**, 1017 (2001).
24. H. Tabara *et al.*, *Cell* **99**, 123 (1999).
25. M. Fagard, S. Boutet, J. B. Morel, C. Bellini, H. Vaucheret, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11650 (2000).
26. C. Catalanotto, G. Azzalin, G. Macino, C. Cogoni, *Nature* **404**, 245 (2000).
27. S. R. Eddy, *Curr. Opin. Genet. Dev.* **9**, 695 (1999).
28. K. M. Wassarman, F. Repoila, C. Rosenow, G. Storz, S. Gottesman, *Genes Dev.* **15**, 1637 (2001).
29. J. Cavaille *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14311 (2000).
30. A. Hüttenhofer *et al.*, *EMBO J.* **20**, 2943 (2001).
31. L. Argaman *et al.*, *Curr. Biol.* **11**, 941 (2001).
32. D. H. Mathews, J. Sabina, M. Zuker, D. H. Turner, *J. Mol. Biol.* **288**, 911 (1999).
33. We acknowledge J. Martinez, S. M. Elbashir, C. Will, R. Rivera-Pomar, S. Baskerville, B. Reinhart, and

P. D. Zamore for comments on the manuscript; G. Hernandez for the gift of total RNA isolated from staged fly populations; H. Brahms, C. Schneider, P. Kempkes, E. Raz, and A. Mansouri for providing tissues or cells; B. Reinhart for advice on Northern blot analysis; G. Dowe for sequencing; P. Bucher for bioinformatic consultations; and R. Lührmann for support. Funded by a BMBF Biofuture grant.

31 July 2001; accepted 14 September 2001

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*

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- to 24-nucleotide noncoding RNAs, called microRNAs (miRNAs). We report on 55 previously unknown miRNAs in *C. elegans*. The miRNAs have diverse expression patterns during development: a *let-7* paralog is temporally coexpressed with *let-7*; miRNAs encoded in a single genomic cluster are coexpressed during embryogenesis; and still other miRNAs are expressed constitutively throughout development. Potential orthologs of several of these 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 to 25 nucleotides (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 and 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 ribonuclease III (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 noncoding 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 play 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 previously unknown noncoding 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 1 representing *let-7* RNA. Another 182 corresponded to the *Escherichia 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 sequences 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 noncoding RNAs that are about 20–24 nt in length and are processed from fold-back structures. We and the two other groups reporting in this issue of the journal 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 and Fig. 1A). All the others are on the 3' arm, as exemplified by miR-1 (Table 1 and 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

Whitehead Institute for Biomedical Research, and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, MA 02142, USA.

*To whom correspondence should be addressed. E-mail: dbartel@wi.mit.edu