Developmental Cell, Vol. 5, 337–350, August, 2003, Copyright ©2003 by Cell Press

# The Small RNA Profile during *Drosophila melanogaster* Development

**Alexei A. Aravin,[1] Mariana Lagos-Quintana,[2] Abdullah Yalcin,[3] Mihaela Zavolan,[4] Debora Marks,[5] Ben Snyder,[4] Terry Gaasterland,[4] Jutta Meyer,[2] and Thomas Tuschl[2,*]**

[1]Department of Animal Molecular Genetics
Institute of Molecular Genetics
Kurchatov sq. 2
Moscow 123182
Russia

[2]Laboratory for RNA Molecular Biology
The Rockefeller University
1230 York Avenue, Box 186
New York, New York 10021

[3]Department for Cellular Biochemistry
Max Planck Institute for Biophysical Chemistry
Am Fassberg 11
37077 Göttingen
Germany

[4]Laboratory for Computational Genomics
The Rockefeller University
1230 York Avenue, Box 250
New York, New York 10021

[5]Center for Computational Biology and Bioinformatics
Columbia University Genome Center
Russ Berrie Pav., Room 520b
1150 St. Nicholas Avenue
New York, New York 10032

## Summary

**Small RNAs ranging in size between 20 and 30 nucleotides are involved in different types of regulation of gene expression including mRNA degradation, translational repression, and chromatin modification. Here we describe the small RNA profile of *Drosophila melanogaster* as a function of development. We have cloned and sequenced over 4000 small RNAs, 560 of which have the characteristics of RNase III cleavage products. A nonredundant set of 62 miRNAs was identified. We also isolated 178 repeat-associated small interfering RNAs (rasiRNAs), which are cognate to transposable elements, satellite and microsatellite DNA, and *Suppressor of Stellate* repeats, suggesting that small RNAs participate in defining chromatin structure. rasiRNAs are most abundant in testes and early embryos, where regulation of transposon activity is critical and dramatic changes in heterochromatin structure occur.**

## Introduction

One of the recent surprises in molecular and cellular biology was the identification of 20 to 30 nucleotide (nt) RNA molecules that guide transcriptional and posttranscriptional gene silencing (for reviews see Denli and Hannon, 2003; Grishok and Mello, 2002; Moss, 2002; Voinnet, 2002; Wassenegger, 2002; Zamore, 2002). The small RNAs are derived from double-stranded RNA (dsRNA) precursors that are processed by the ribonuclease type III enzyme Dicer (Bernstein et al., 2001; Elbashir et al., 2001b; Grishok et al., 2001; Hutvágner et al., 2001; Ketting et al., 2001; Park et al., 2002; Provost et al., 2002; Reinhart et al., 2002; Zhang et al., 2002). At least three types of small RNAs have been described: (1) Small interfering RNAs (siRNAs) (Caplen et al., 2001; Elbashir et al., 2001a, 2001b), (2) microRNAs (miRNAs) (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001), and (3) repeat-associated small interfering RNAs (rasiRNAs) (Djikeng et al., 2001; Elbashir et al., 2001b; Hamilton et al., 2002; Llave et al., 2002a; Mette et al., 2002; Reinhart and Bartel, 2002).

siRNAs and miRNAs are derived from different sources of dsRNA and act in common pathways interchangeably depending on the degree of complementarity with their target RNA sequence (Doench et al., 2003; Hutvágner and Zamore, 2002; Llave et al., 2002b; Rhoades et al., 2002; Tang et al., 2003). Sources for siRNAs are dsRNAs that are typically hundreds of base pairs (bp) long, e.g., replication intermediates of RNA viruses (for reviews see Plasterk, 2002; Waterhouse et al., 2001). miRNAs originate from genes that encode short 20 to 30 bp dsRNA hairpins and represent a growing class of noncoding RNAs that are found in plants and animals (Ambros et al., 2003b; Dostie et al., 2003; Grad et al., 2003; Lagos-Quintana et al., 2001, 2002, 2003; Lau et al., 2001; Lee and Ambros, 2001; Lee et al., 1993; Lim et al., 2003a, 2003b; Llave et al., 2002a; Mette et al., 2002; Mourelatos et al., 2002; Park et al., 2002; Reinhart et al., 2000, 2002). Many miRNAs are conserved in sequence between distantly related organisms, suggesting that the interactions between these miRNAs and their targets constitute essential processes (Pasquinelli et al., 2000).

miRNA expression is regulated in a spatial and temporal manner. The miRNA transcripts are trimmed in the nucleus to a dsRNA precursor and subsequently exported to the cytoplasm for Dicer processing (Lee et al., 2002). Most of our current knowledge of miRNA function derives from the two *Caenorhabditis elegans* miRNAs lin-4 and let-7, originally named small temporal RNAs (stRNAs) for their role in developmental timing (Lee et al., 1993; Reinhart et al., 2000). Investigation of *lin-4* and *let-7* genetic interactions identified genes containing sequences in their 3′ UTRs partially complementary to these miRNAs (Abrahante et al., 2003; Lin et al., 2003; Pasquinelli and Ruvkun, 2002). lin-4 and let-7 miRNAs form bulged RNA duplexes with their target mRNAs, which leads to repression of protein synthesis after translation initiation while the mRNA remains intact (Olsen and Ambros, 1999; Seggerson et al., 2002; Wightman et al., 1993). Recently, two *D. melanogaster* miRNAs, bantam (Brennecke et al., 2003) and miR-14 (Xu et al., 2003), were genetically characterized and some of the respective target mRNAs were identified. However, the precise mechanism of target repression by

*Correspondence: ttuschl@rockefeller.edu

these *D. melanogaster* miRNAs remains to be characterized.

rasiRNAs are presumably derived from long dsRNAs and match to repetitive sequence element in sense and antisense orientation (Djikeng et al., 2001; Llave et al., 2002a; Reinhart and Bartel, 2002). Repetitive sequences are often associated with regions of heterochromatin. Transposable elements (TEs), for example, may produce dsRNA upon random integration nearby transcriptionally active sequences. The high density of transposons in particular genomic regions makes it more likely that dsRNA is generated by transcription from adjacent opposing promoters of transposons rather than cellular genes. rasiRNAs are likely to function as guide RNAs during the establishment and/or maintenance of heterochromatin in plants (Hamilton et al., 2002; Llave et al., 2002a; Mette et al., 2002), *Trypanosoma brucei* (Djikeng et al., 2001), *Drosophila melanogaster* (Aravin et al., 2001; Pal-Bhadra et al., 2002), and fission yeast (Hall et al., 2002; Reinhart and Bartel, 2002; Volpe et al., 2002). Small RNAs and proteins related to RNA silencing have also been identified in *Tetrahymena thermophila* and are thought to participate in guiding programmed DNA elimination of dispersed sequence elements in order to form the transcriptionally active macronucleus after sexual conjugation (Mochizuki et al., 2002; Tautz et al., 1988; Taverna et al., 2002).

Two distinct classes of small RNAs were found in plants. These classes are 21 nt and 24 nt in size and mediate posttranscriptional and transcriptional gene silencing (Hamilton et al., 2002; Mallory et al., 2002; Tang et al., 2003). siRNA product inhibition experiments suggest that different homologs of Dicer are responsible for production of these two types of small RNAs (Tang et al., 2003). siRNAs and miRNAs are near 21 nt in size (Llave et al., 2002a; Reinhart et al., 2002), whereas rasiRNAs are 24 nt long (Hamilton et al., 2002; Llave et al., 2002a). Certain viral suppressors of RNA silencing in plants specifically affect the accumulation of 24 nt rasiRNAs (Hamilton et al., 2002). It was recently discovered that plants with mutations in the *Argonaute 4* gene were impaired in transcriptional silencing at certain loci (Zilberman et al., 2003). These mutants lacked accumulation of 24 nt siRNAs of the AtSN1 retroelement and derepressed a normally silent gene because of loss of DNA methylation and histone H3-lysine 9 methylation (Zilberman et al., 2003). Plant miRNAs are mostly 21 nt consistent with their role in posttranscriptional regulation (Llave et al., 2002a; Park et al., 2002; Reinhart et al., 2002).

A natural case in which repetitive DNA sequences function to silence a cellular gene was first described in *D. melanogaster* (Livak, 1984, 1990; Palumbo et al., 1994). Deletion of *Suppressor of Stellate* [*Su(Ste)*] repeats located on the Y chromosome results in derepression of *Stellate* located on the X chromosome leading to meiotic abnormalities and male infertility due to crystallization of overexpressed Stellate protein in sperm cells. Derepression of *Stellate* is also observed in mutants of the *spn-E* DE-H helicase (Aravin et al., 2001; Stapleton et al., 2001) and the Argonaute *aub* (Schmidt et al., 1999). Interestingly, *spn-E* mutants accumulate transcripts from retrotransposons in the germline (Aravin et al., 2001; Kogan et al., 2003) and both *spn-E* and *aub* mutants are defective for initiation of RNAi in the female germline (Kennerdell et al., 2002). *Stellate* silencing is correlated with the presence of 25–27 nt siRNAs deriving from both strands of *Su(Ste)* (Aravin et al., 2001). A study investigating multicopy transgene silencing in *D. melanogaster* implicated another Argonaute gene, *piwi*, in posttranscriptional and transcriptional gene silencing (Pal-Bhadra et al., 2002). Links between protein factors involved in RNAi and the silencing of endogenous transposable elements have also been made in other species like *C. elegans* (Ketting et al., 1999; Tabara et al., 1999) and *Chlamydomonas reinhardtii* (Wu-Scharf et al., 2000).

The Argonautes have been shown to be the largest conserved class of proteins that link the various pathways of RNA silencing. Various Argonaute family members function in both posttranscriptional and transcriptional gene silencing in protists, fungi, plants, and animals (for reviews see Carmell et al., 2002). Biochemical characterization of the RNAi machinery identified a ribonucleoprotein complex termed RISC (RNA induced silencing complex) that contains the siRNAs as guide RNAs (Elbashir et al., 2001b; Hammond et al., 2000; Zamore et al., 2000). In *D. melanogaster*, RISC was found to contain Ago2 (Hammond et al., 2001), and, in human cells, RISC contains the Argonaute proteins eIF2C1 and/or eIF2C2 (Martinez et al., 2002). Interestingly, eIF2C2 was also identified in a protein complex associated with miRNAs in humans (Mourelatos et al., 2002), and it was subsequently shown that this complex was able to cleave miRNA-complementary target mRNAs (Hutvágner and Zamore, 2002). The reciprocal case, that siRNAs can also function as miRNAs, was also shown (Doench et al., 2003). Therefore, small RNAs associated with Argonaute proteins can either act as siRNA or miRNAs depending on the degree of complementarity to the target mRNA.

Here we cloned and sequenced the short RNAs present in *D. melanogaster* at different stages of development as well as in adult testes. Several different classes of dsRNA-derived short RNAs were identified. We validated 62 unique miRNAs and annotated 178 rasiRNAs. The distinct features of the two classes of small RNAs suggest different regulatory mechanisms and function for the two classes. This study also indicated the fundamental importance of recording the small RNA profile for understanding gene regulation in eukaryotic cells.

## Results

### Cloning of *D. melanogaster* Small RNAs

Small RNAs in the size range of 16–29 nt were cloned from defined developmental stages of *D. melanogaster*. In addition, small RNAs were also cloned from microdissected adult testis. To obtain the small RNA fraction, total RNA was isolated using acid guanidinium thiocyanate-phenol-chloroform (Chomczynski and Sacchi, 1987) adapted to maximize recovery of small RNAs (Lagos-Quintana et al., 2001, 2002). The small RNAs were then size fractionated on denaturing polyacrylamide gels and excised from the gel directly below the 2S rRNA band that is detectable by UV shadowing. In contrast to many other species, *D. melanogaster* ribosomal RNA (rRNA) is composed of four individual RNAs, 28S, 18S, 5.8S, and 2S, the latter of which is 30 nt in

Table 1. Composition of Small RNA cDNA Libraries Prepared from Different Developmental Stages and Testes of *D. melanogaster*

| | Embryo Stage (hr) | | | | | Larva Stages | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Type | 0–2 | 2–4 | 4–6 | 6–12 | 12–24 | L1 | L1 + L2 | L3 | Pupa | Adult | Testes | Total (clones) | Total (%) |
| rRNA | 161 | 119 | 160 | 85 | 23 | 193 | 111 | 293 | 91 | 438 | 451 | 2160 | 53.0 |
| tRNA[a] | 25 | 14 | 25 | 1 | 2 | 35 | 64 | 53 | 9 | 99 | 25 | 354 | 8.7 |
| miRNA | 112 | 42 | 25 | 7 | 3 | 12 | 8 | 8 | — | 121 | 41 | 382 | 9.4 |
| rasiRNA | 79 | 39 | 13 | | 1 | 2 | 3 | — | — | 12 | 28 | 178 | 4.4 |
| mRNA | 3 | 7 | 11 | 2 | 1 | 26 | 30 | 32 | 1 | 16 | 13 | 143 | 3.5 |
| snRNA/snoRNA[b] | 4 | — | 4 | 2 | 3 | 1 | 2 | 2 | — | 16 | 1 | 35 | 0.9 |
| Other ncRNA[c] | 8 | 1 | 7 | 1 | — | 1 | 3 | 1 | 1 | 13 | 3 | 40 | 1.0 |
| *S. cerevisiae* | 45 | 101 | 49 | 22 | 19 | 53 | 61 | 26 | 1 | 141 | 7 | 525 | 12.9 |
| Bacteria, plants | 1 | 1 | 2 | — | — | — | 6 | 4 | 1 | 2 | — | 17 | 0.4 |
| *DCV* | — | — | — | — | — | 1 | 1 | 11 | 1 | 3 | — | 17 | 0.4 |
| Unknown[d] | 31 | 14 | 7 | 2 | 7 | 15 | 32 | 25 | 8 | 56 | 25 | 223 | 5.5 |
| Total | 469 | 338 | 303 | 122 | 59 | 339 | 321 | 455 | 113 | 920 | 594 | 4074 | 100 |

The number of sequenced clones is indicated according to matches to the annotation provided within the various public databases including the annotation of the *D. melanogaster* genome (version 3.1 from http://www.bdgp.org), a dataset of *D. melanogaster* sequences from GenBank (http://www.fruitfly.org/sequence/sequence_db/na_gb.dros), a database of transposable elements (http://www.bdgp.org/p_disrupt/datasets/VERSION3/ALL_SEQUENCES_dmel_RELEASE3.FASTA.ALL.v3) and canonical sequences (http://www.bdgp.org/p_disrupt/datasets/NATURAL_TRANSPOSABLE_ELEMENTS.fa), a database of *D. melanogaster* tRNA sequences (http://rna.wustl.edu/GtRDB/Dm/Dm-seqs.html), a database of small RNA sequences provided by A. Hüttenhofer, and a database of miRNAs (http://www.sanger.ac.uk/Software/Rfam/ftp.shtml) and predicted miRNA sequences (Lim et al., 2003). The assignment of annotation was performed in a hierarchical manner. Perfect matches of small RNA sequences to the genomes of *S. cerevisiae* were classified as *S. cerevisiae* sequences followed by matches to *D. melanogaster* rRNA, tRNA, and then snRNAs/snoRNAs and other ncRNAs. The remaining sequences were then checked against euchromatic and heterochromatic *D. melanogaster* genomic sequences as well as GenBank sequences of other organisms and classified as mRNAs, *Drosophila C Virus* (*DCV*), bacterial, and plant genomes.

[a] The annotation for small RNAs as tRNA breakdown products was performed using the tRNA sequences provided at the Genomic tRNA Database at http://rna.wustl.edu/tRNAdb/ as well as annotated GenBank sequences.

[b] snRNA and snoRNA hits were assigned by using the sequence sets provided in (Yuan et al., 2003).

[c] The noncoding RNAs (ncRNAs) were provided from A. Hüttenhofer and contained longer RNA sequences (>40 nt) without annotation or assigned function.

[d] Unknown describes a category for which most sequences do not match to *S. cerevisiae* or *D. melanogaster*. This category also contains some sequences with near perfect match to regions of *D. melanogaster* with no available annotation and no evidence for a repetitive character or a miRNA fold-back structure.

size (Tautz et al., 1988). The small RNAs were then cloned and sequenced (Pfeffer et al., 2003).

A total of 4074 clone sequences were obtained and current public databases were used to annotate 95.6% of these sequences; the residual sequences could not be annotated because they did not match to any of the sequenced genomes in the database (183 clones) or because they matched to a region of the *D. melanogaster* genome for which no functional or sequence homology assignment could be made (40 clones). The largest class of cloned RNAs represents breakdown products of abundant noncoding (or nonmessenger) RNAs (rRNA, tRNA, snRNA, snoRNA, and others) of *D. melanogaster* (63.5%), followed by breakdown product sequences from *Saccharomyces cerevisiae* (12.9%), which constitutes the preferred diet of *D. melanogaster*. A few bacterial and plant rRNA fragments from salmonella, cereal, and hops were also found, presumably because the baking yeast that was used for fly food was brewing yeast. A small fraction of *D. melanogaster* mRNA breakdown products (3.5%) was also identified (Table 1). The residual 577 sequences (14.1%) fell into the following three classes: miRNAs (9.3%), repeat-associated small RNAs derived from sense and antisense strands of repetitive elements (4.4%), and small RNAs from *Drosophila* virus C (0.4%).

The proportion of small RNAs with regulatory function relative to rRNA breakdown products varied drastically for the different developmental stages. Early embryos and adults showed the highest content of small regulatory RNAs (between 15% and 40%) while in late embryo and larvae the number was significantly lower. From pupae, almost all cloned small RNAs represented rRNA and tRNA breakdown products. Presumably, extensive apoptosis during metamorphic tissue reorganization caused extensive rRNA breakdown and made it impossible to clone small RNAs at high frequency. Also, we found that pupae- and adult-specific miRNAs, such as let-7, were still present in pupae RNA preparations and were readily detectable by Northern blotting (Figure 1B).

miRNAs and siRNAs are generated by Dicer RNase III cleavage from dsRNA precursors and therefore contain 5′ phosphates and 2′,3′-hydroxyl termini (Bernstein et al., 2001; Elbashir et al., 2001b; Grishok et al., 2001; Hutvágner et al., 2001; Nykänen et al., 2001). In contrast, breakdown products of longer RNAs that are generated by single-strand specific RNases or by simple hydrolysis should contain 5′-hydroxyl and 2′,3′ cyclic phosphate or 2′ or 3′ monophosphate termini. In theory, it should be possible to discriminate between small regulatory RNAs and RNA breakdown products by the presence or absence of a 5′ phosphate. The cloning method described by Lau et al. (2001), which takes advantage of the presence of a 5′ phosphate, should therefore discriminate against RNA breakdown products and enrich for RNase III processing products. However, when we compared the two cloning protocols using identical sources of starting material, we found that the content

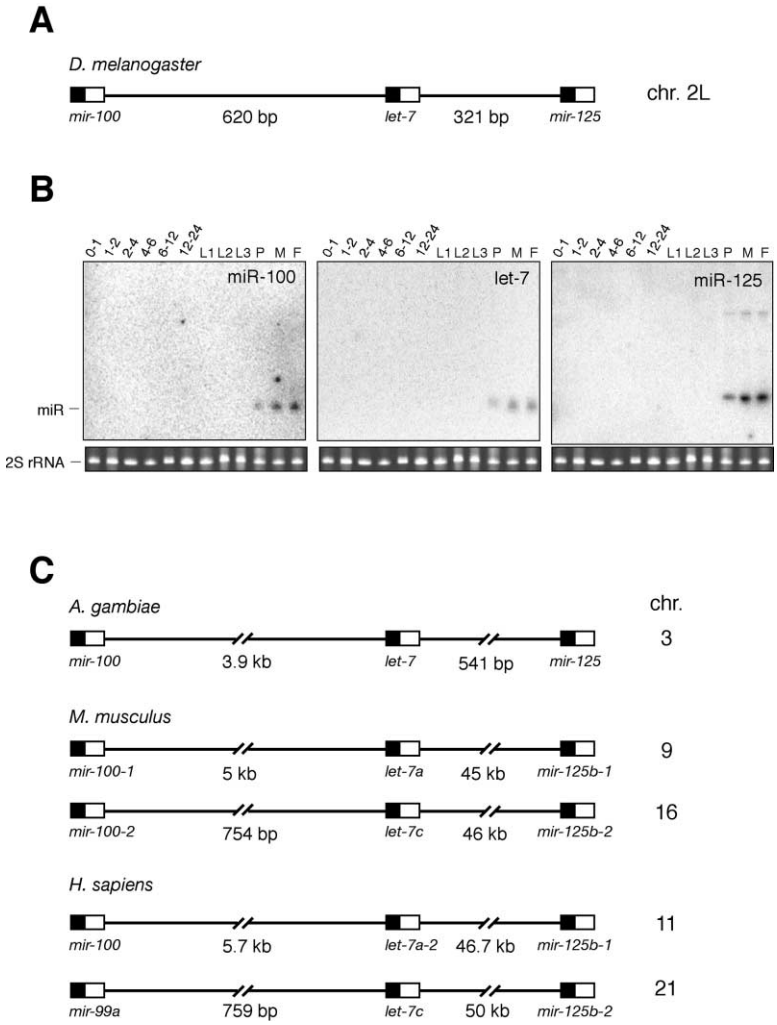


Figure 1. miRNA Genes *mir-100*, *let-7* and the *lin-4* Homolog *mir-125* Are Clustered

(A) Arrangement of the miRNA genes in *D. melanogaster*. The 70 nt fold-back precursor is indicated as box, the position of the miRNA in the precursor is shown in black. The chromosome location is indicated to the right.

(B) Northern blots confirming the coexpression of miR-100, let-7, and miR-125 in pupa and adult. As control for loading the 30 nt 2S rRNA band was visualized by ethidium bromide staining of the polyacrylamide gel before transfer of the RNA to the blotting membrane. Development stages of embryos are indicated in hours after egg laying. Larval stages are indicated as L1, L2, and L3. P indicates pupal stage; M, adult male; F, adult female.

(C) The miRNA clustering is conserved between invertebrates and vertebrates but gene duplication occurred in mammals and the spacing between miRNA precursors increased with increasing genome size.

of rRNA and tRNA breakdown RNAs was practically the same. Both methods yielded 15% of miRNA sequences using 0 to 2 hr embryo small RNAs and between 2% and 3% miRNAs from combined L1 and L2 stage small RNAs after sequence analysis of more than 100 clones for each RNA source and each cloning method. This suggests that small RNAs without a 5′ phosphate are rapidly phosphorylated in the cell and become indistinguishable from Dicer RNase III processing products. Rapid addition of 5′ phosphates to nonphosphorylated synthetic siRNA duplexes as well as rapid dephosphorylation of 3′ phosphates of synthetic siRNA duplexes was observed previously in *D. melanogaster* and human cell extracts (Martinez et al., 2002; Nykänen et al., 2001). However, it was not previously examined if other single- or double-stranded RNAs were also substrates for these kinases and phosphatases. We therefore continued using our original cloning protocol (Lagos-Quintana et al., 2001, 2002; Pfeffer et al., 2003).

### Identification of Distinct Classes of Small RNAs

Cloning of small RNAs provides an overview of the RNAs expressed in a cell or organism. If a sufficiently large number of clones were examined, complete contigs of all nonmessenger RNAs could have been assembled from partly overlapping small RNA segments similar to shotgun genome sequencing. In contrast to dsRNA processing products, the breakdown products of noncoding RNAs and mRNAs show no significant length bias and are broadly distributed throughout the size window used for cloning. miRNAs and rasiRNAs show a much tighter size distribution (Supplemental Table S1 [http://www.developmentalcell.com/cgi/content/full/5/2/337/DC1]), which centers on distinct sizes for both classes. About 70% of the identified sequences of miRNAs are between 21 and 23 nt in size with an average of 22.0 nt. This size distribution is similar to the length distribution of siRNAs generated by in vitro processing of dsRNA in *D. melanogaster* embryo lysate (Elbashir et al., 2001a). About 65% of the rasiRNAs are between 23 and 26 nt in size with an average of about 23.6 nt (Supplemental Table S1).

miRNAs, siRNAs, and rasiRNAs have a strong preference for pyrimidine residues, uridine in particular, at their 5′-most position. The average nucleotide composition for the first, second, and last nucleotide of the various RNA classes is shown in Supplemental Table S2. Only the 5′-most position of RNase III processing products displays this sequence bias. Additional support

for the presence of dsRNA precursors for rasiRNAs is provided by the presence of sense as well as antisense oriented sequences cognate to repeat elements of LTR and non-LTR retrotransposons.

### Absence of Evidence for Antisense mRNA Regulation

Bioinformatic analysis indicated that a significant number of genes are partially overlapping in opposite orientation in various organisms (Merino et al., 1994; Shendure and Church, 2002; Yelin et al., 2003). Partly complementary transcripts may participate in gene regulation by antisense or RNAi-related mechanisms. We cloned 143 small RNAs that matched to cDNAs or predicted cDNA sequences in the *D. melanogaster* genome. One hundred forty-one of these sequences were in sense (+) orientation to the mRNA without overlap to other genes, suggesting that these small RNAs are breakdown products of mRNAs. This is supported by the broad length distribution similar to rRNA and tRNA fragments and the absence of a 5′ pyrimidine sequence bias characteristic for dsRNA processing products (Supplemental Table S2). Only two sequences mapped to the coding region of predicted genes in antisense (−) orientation but without any evidence for the presence of overlapping transcripts. These findings are in strong contrast to observations in *C. elegans*, where 50%–90% of cloned small RNAs with a corresponding match to a protein-coding sequence were in antisense orientation to the open reading frame (Ambros et al., 2003b; Lim et al., 2003b).

### Small RNAs Derived from *Drosophila C virus* RNA

*Drosophila C virus* (*DCV*) belongs to the class of *Dicistroviridae* and the genus *Cripavirus* (Cricket paralysis-like virus) and contains a positive-sense genomic RNA of 9264 nucleotides in length (Johnson and Christian, 1998). *DCV* is phylogenetically related to mammalian picornaviruses. *DCV* is commonly associated with *Drosophila* in nature and in laboratory culture. Infections are not usually associated with a noticeable disease state although they commonly reduce life expectancy of infected individuals. We isolated 17 small RNAs from *DCV*, 16 of which were in + orientation and one in − orientation. Considering the broad size distribution of the + orientation *DCV* small RNAs, which is similar to the sense mRNA fragments described above, it is likely that most of the + strand *DCV* small RNAs are not RNase III cleavage products. Although the possibility remains that the 21 nt *DCV* antisense RNA is a breakdown derived from the − strand of DCV, it is also conceivable that the 21 nt antisense *DCV* small RNA is a true siRNA that derived from the dsRNA replication intermediate of *DCV*.

### microRNA Identification

We identified 382 clones that were derived from 60 miRNA genes (Table 2). Some of the cloned miRNAs, such as miR-1 through miR-14, let-7, and bantam were described previously (Brennecke et al., 2003; Lagos-Quintana et al., 2001; Pasquinelli et al., 2000). According to the convention for miRNA annotation (Ambros et al., 2003a), new miRNA gene names were assigned based on the evidence of cloning the small RNA from our cDNA libraries and their phylogenetic conservation as a fold-back precursor structure in other species (Supplemental Tables S3 and S4). miRNAs that closely resemble in sequence previously described miRNAs were miR-9b, 9c, 31a, 31b, 34, 79, 92a, 92b, 124, 184, and 210. These were named based on their evolutionary relationships to preidentified miRNAs. Based on the cloned sequences of bantam miRNAs, its previously deduced sequence (Brennecke et al., 2003) needs to be revised. All cloned bantam miRNAs begin with the same 5′ uridine residue encoded one nucleotide downstream of the proposed sequence.

For some miRNAs, we cloned the strand opposite to the accumulating and conserved miRNA. We refer to these sequences as the miR* sequence (Lau et al., 2001) (Table 2). In most cases, miR* is clearly less abundant, but for miR-10*, miR-13a*, and miR-281-2*, too few sequences were cloned to make any conclusion on the relative abundance. A similar situation was found for two small RNAs that are excised from a hairpin residing in the noncoding RNA transcript *iab-4* of the Bithorax complex. *iab-4* contributes to proper formation of abdominal segments (Mattick and Gagen, 2001).

miRNA genes are often found in close proximity to each other forming larger miRNA gene clusters (Lagos-Quintana et al., 2001; Lau et al., 2001). We found 11 gene clusters in the *D. melanogaster* genome containing on average three miRNAs, with the longest cluster containing eight miRNAs (Table 2). Some clusters are found within intergenic regions, while others are located within the intronic regions of protein-coding genes (Supplemental Table S3). Some clusters contain only highly homologous miRNAs, such as *mir-92a* and *mir-92b* or *mir-281-1* and *mir-281-2*, suggestive of rather recent gene duplication. For example, the *mir-2*/*mir-13* family is composed of eight genes that are encoded in four different genomic regions in *D. melanogaster*. There are two single copy genes and two clusters containing three gene copies each. In *Anopheles gambiae*, only five members of this gene family are found, and they are all within one cluster.

Because positional clustering of miRNAs is a common genomic feature of miRNAs, we examined the regions adjacent to miRNA genes for the presence of additional miRNAs that may have escaped the nonsaturating cloning and sequencing protocol. By identifying fold-back structures as well as sequence homologs of known miRNAs and only considering those candidates that were also conserved in other insect genomes, we identified eight more miRNAs. These include *mir-283* clustered with *mir-12* and *mir-304*, *mir-100* clustered with *let-7* and *mir-125*, *mir-313* and *mir-310* clustered with *mir-311* and *mir-312*, and *mir-2c* clustered with *mir-13a* and *mir-13b*. Nonclustered conserved fold-back structures were identified for *mir-87* and *mir-133*. The expression of miR-87, -100, -125, -133, -283, -310, and -312 was confirmed by Northern analysis (Figure 2). The validation of the predicted miR-2c was not attempted because of predictable problems of cross-hybridization of the Northern blotting probe to miR-2a and miR-2b. To date, including the predicted clustered and/or conserved miRNAs, a total of 62 unique miRNA sequences encoded by 68 genes have been identified in *D. melanogaster*.

Table 2. MicroRNA Profile During *D. melanogaster* Development

| Gene | Mature miRNA and miRNA* Sequence | Size Range (nt) | Chr | Position, Orientation | Embryo (hr) | | | | | Larva | | | A | T | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | 0–2 | 2–4 | 4–6 | 6–12 | 12–24 | 1 | 1 + 2 | 3 | | | |
| *mir*-303 | UUUAGGUUUCACAGGAAACUGGU | 23 | X | 4106836…58,− | | | | | | | | | | 1 | 1 |
| *mir*-31b | UGGCAAGAUGUCGGAAUAGCUG | 22 | X | 8732382…403,− | | | | | | | | | 1 | | 1 |
| *mir*-13b-2§ | UAUCACAGCCAUUUUGACGAGU | 17–24 | X | 8830259…80,+ | 3 | | | | | 1 | | | | | 4 |
| *mir*-283[p,n] | UAAAUAUCAGCUGGUAAUUCU | 21[p] | X | 15238989…9009,+ | | | | | | | | | | | |
| *mir*-304 | UAAUCUCAAUUUGUAAAUGUGAG | 21–23 | X | 15239984…40006,+ | 2 | | | | | | | | 1 | 4 | 7 |
| *mir*-12 | UGAGUAUUACAUCAGGUACUGGU | 21–24 | X | 15240493…515,+ | 5 | | 3 | | | | | | 2 | 3 | 13 |
| *mir*-210 | UUGUGCGUGUGACAGCGGCU | 20–23 | X | 17859230…49,+ | | | | | | | | | 3 | | 3 |
| *mir*-275 | UCAGGUACCUGAAGUAGCGCGCG | 19–25 | 2L | 7418084…106,+ | 4 | 1 | | | | 1 | | | | | 6 |
| *mir*-305 | AUUGUACUUCAUCAGGUGCUCUG | 19–23 | 2L | 7418210…32,+ | 1 | | | | | | | | 1 | | 2 |
| *mir*-2b-1§ | UAUCACAGCCAGCUUUGAGGAGCG | 19–24 | 2L | 8250853…76,− | 4 | 2 | | 1 | | | | | 5 | | 12 |
| *mir*-87[p,n] | UUGAGCAAAAUUUCAGGUGUG | 21 | 2L | 9942846…65,− | | | | | | | | | | | |
| *mir*-9c | UCUUUGGUAUUCUAGCUGUAGA | 22 | 2L | 16676020…41,+ | 1 | 1 | | | 1 | | | | | | 3 |
| *mir*-306 | UCAGGUACUUAGUGACUCUCAA | 22 | 2L | 16676502…22,+ | | | | | | | | | 1 | 1 | 2 |
| | *GGGGGUCACUCUGUGCCUGUGC | 22 | 2L | 16676539…60,+ | | 1 | | | | | | | | | 1 |
| *mir*-79 | UAAAGCUAGAUUACCAAAGCAU | 22 | 2L | 16676697…718,+ | 1 | | | | | | | | | | 1 |
| *mir*-9b | UCUUUGGUGAUUUUAGCUGUAUG | 21–23 | 2L | 16676842…64,+ | 1 | 1 | | | | | | | | | 2 |
| *mir*-124[n] | UAAGGCACGCGGUGAAUGCCAAG | 19–23 | 2L | 17544509…31,+ | | | 2 | | | 2 | | 1 | 1 | | 6 |
| *mir*-100[p,n] | AACCCGUAAAUCCGAACUUGUG | 22 | 2L | 18449528…49,+ | | | | | | | | | | | |
| *let*-7[n] | UGAGGUAGUAGGUUGUAUAGU | 20–21 | 2L | 18450127…147,+ | | | | | | | | | 3 | 2 | 5 |
| *mir*-125[p,n] | UCCCUGAGACCCUAACUUGUGA | 22 | 2L | 18450426…47,+ | | | | | | | | | | | |
| *mir*-2a-2[n] | UAUCACAGCCAGCUUUGAUGAGCU | 20–24 | 2L | 19547578…601,− | 3 | 3 | 2 | | | 1 | | | 4 | 2 | 15 |
| | *CCUCAUCAAGUGGUUGUGAUA | 21 | 2L | 19547617…37,− | 2 | | | | | | | | 2 | | 4 |
| *mir*-2a-1[n] | UAUCACAGCCAGCUUUGAUGAGCU | 20–24 | 2L | 19547984…8007,− | 3 | 3 | 2 | | | 1 | | | 4 | 2 | 15 |
| *mir*-2b-2§ | UAUCACAGCCAGCUUUGAGGAGCG | 19–24 | 2L | 19548271…94,− | 4 | 2 | | 1 | | | | | 5 | | 12 |
| *mir*-1[n] | UGGAAUGUAAAGAAGUAUGGAG | 20–27 | 2L | 20457238…59,+ | 23 | 1 | 4 | | | 4 | 2 | 3 | 34 | 3 | 74 |
| *mir*-133[p,n] | UUGGUCCCCUUCAACCAGCUGU | 22 | 2L | 20586415…36,+ | | | | | | | | | | | |
| *mir*-14 | UCAGUCUUUUUCUCUCUCCUA | 19–25 | 2R | 4614434…54,+ | 2 | | 2 | | | | | | | | 4 |
| *mir*-307 | UCACAACCUCCUUGAGUGAG | 20 | 2R | 4681896…915,− | | | | | | | | | 1 | | 1 |
| *mir*-281-2§ | UGUCAUGGAAUUGCUCUCUUUGU | 23 | 2R | 7234876…98,− | | | | | | | | | 1 | | 1 |
| | *AAGAGAGCUAUCCGUCGACAGU | 22 | 2R | 7234922…43,− | | | | | | | | | 1 | | 1 |
| *mir*-281-1§ | UGUCAUGGAAUUGCUCUCUUUGU | 23 | 2R | 7235096…118,− | | | | | | | | | 1 | | 1 |
| *mir*-184 | UGGACGGAGAACUGAUAAGGGC | 19–23 | 2R | 8394134…55,− | 5 | 2 | 1 | | 1 | 1 | | | 12 | 4 | 26 |
| | *CCUUAUCAUUCUCUCGCCCCG | 21 | 2R | 8394174…94,− | | | | | | | | | 1 | | 1 |
| *mir*-308 | AAUCACAGGAUUAUACUGUGAG | 18–22 | 2R | 9279528…49,− | 4 | 1 | | | | | | | 1 | | 6 |
| *mir*-278[n] | UCGGUGGGACUUUCGUCCGUUU | 20–22 | 2R | 10720847…68,+ | | | | | | | | 1 | 1 | 6 | 8 |
| *mir*-8[n] | *AUCUUACCGGGCAGCAUUAGA | 21 | 2R | 11895173…93,+ | | | | | | | | | 1 | | 1 |
| | UAAUACUGUCAGGUAAAGAUGUC | 17–25 | 2R | 11895209…31,+ | 2 | 1 | 1 | | | 1 | 1 | 2 | 5 | 1 | 14 |
| *mir*-31a | UGGCAAGAUGUCGGCAUAGCUGA | 22–23 | 2R | 12847576…98,− | 1 | | | | | | 1 | | | | 2 |
| *mir*-6-3§[n] | UAUCACAGUGGCUGUUCUUUUU | 21–22 | 2R | 14724442…63,− | 9 | 1 | | | | | | | | | 10 |
| *mir*-6-2§[n] | UAUCACAGUGGCUGUUCUUUUU | 21–22 | 2R | 14724594…615,− | 9 | 1 | | | | | | | | | 10 |
| *mir*-6-1§[n] | UAUCACAGUGGCUGUUCUUUUU | 21–22 | 2R | 14724732…53,− | 9 | 1 | | | | | | | | | 10 |

(*continued*)

Table 2. Continued

| Gene | Mature miRNA and miRNA* Sequence | Size Range (nt) | Chr | Position, Orientation | Embryo (hr) 0–2 | 2–4 | 4–6 | 6–12 | 12–24 | Larva 1 | 1 + 2 | 3 | A | T | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *mir*-5 | AAAGGAACGAUCGUUGUGAUAUG | 22–23 | 2R | 14724920…42,− | 3 | 5 | 1 | | | | | | | | 9 |
| *mir*-4[n] | AUAAAGCUAGACAACCAUUGAA | 21–22 | 2R | 14725020…41,− | 5 | 2 | | | | | | | | | 7 |
| | *CUUUGGUCGUCCAGCCUUAGGUGA | 24–25 | 2R | 14725053…76,− | | 2 | | | | | | | | | 2 |
| *mir*-286 | UGACUAGACCGAACACUCGUGCU | 22–24 | 2R | 14725155…77,− | 1 | 7 | 3 | 2 | | | | | 1 | | 14 |
| *mir*-3[n] | UCACUGGGCAAAGUGUGUCUCA | 18–22 | 2R | 14725320…41,− | 7 | 2 | | | | | | | | | 9 |
| *mir*-309 | GCACUGGGUAAAGUUUGUCCUA | 22 | 2R | 14725430…51,− | | 2 | | | | | | | | | 2 |
| *mir*-310[p,n] | UAUUGCACACUUCCCGGCCUUU | 22 | 2R | 15647477…99,− | | | | | | | | | | | |
| *mir*-311[n] | UAUUGCACAUUCACCGGCCUGA | 22–23 | 2R | 15647600…21,− | 1 | 1 | | | | | | | | | 2 |
| *mir*-312[n] | UAUUGCACUUGAGACGGCCUGA | 22–23 | 2R | 15647770…91,− | 2 | 1 | | | | | | | | | 3 |
| *mir*-313[p] | UAUUGCACUUUUCACAGCCCGA | 22 | 2R | 15647905…27,− | | | | | | | | | | | |
| *mir*-7[n] | UGGAAGACUAGUGAUUUUGUUGUU | 21–24 | 2R | 15669793…816,+ | 3 | 1 | 1 | | | | | | | | 5 |
| *bantam* | UGAGAUCAUUUUGAAAGCUGAUU | 20–23 | 3L | 622888…910,+ | | | 2 | | | 1 | 2 | | 3 | | 8 |
| *mir*-276b | UAGGAACUUAAUACCGUGCUCU | 22 | 3L | 10277375…96,+ | | | | | | | | | 1 | | 1 |
| *mir*-276a | UAGGAACUUCAUACCGUGCUCU | 22 | 3L | 10322810…31,+ | | | | 2 | | | | | 6 | | 8 |
| *mir*-314 | UAUUCGAGCCAAUAAGUUCGG | 21 | 3L | 11730629…49,+ | 1 | | | | | | | | 2 | 1 | 4 |
| *mir*-285 | UAGCACCAUUCGAAAUCAGUGC | 22 | 3L | 11903655…76,− | | | | | | | | | 1 | | 1 |
| *mir*-263b | CUUGGCACUGGGAGAAUUCAC | 21 | 3L | 15792339…59,− | | | | | | | | | 1 | | 1 |
| *mir*-315 | UUUUGAUUGUUGCUCAGAAAGC | 22 | 3L | 18809861…82,+ | | | | | 1 | | | | | | 1 |
| *mir*-9a[n] | UCUUUGGUUAUCUAGCUGUAUGA | 21–23 | 3L | 19515097…119,+ | 5 | 1 | 2 | 2 | | | 1 | | | 1 | 12 |
| *mir*-316 | UGUCUUUUUCCGCUUACUGGCG | 20–22 | 3L | 21586044…65,− | | | | | | | | 1 | | 1 | 2 |
| *mir*-10[n] | *CAAAUUCGGUUCUAGAGAGGUUU | 23 | 3R | 2635235…57,− | | | | | | | 1 | | | | 1 |
| | ACCCUGUAGAUCCGAAUUUGU | 21 | 3R | 2635277…97,− | 1 | | | | | | | | | | 1 |
| *mir*-317 | UGAACACAGCUGGUGGUAUCCAGU | 20–24 | 3R | 5916921…44,+ | | | | | | | | | 5 | 3 | 8 |
| *mir*-277 | UAAAUGCACUAUCUGGUACGACA | 21–23 | 3R | 5925820…42,+ | | | | | | | | | 13 | | 13 |
| *mir*-34 | UGGCAGUGUGGUUAGCUGGUUGUG | 17–25 | 3R | 5926692…715,+ | | | | | | | | | 6 | 7 | 13 |
| *mir*-318 | UCACUGGGCUUUGUUUAUCUCA | 22 | 3R | 6234084…105,+ | | | | | | | | | 3 | | 3 |
| *mir*-13b-1§ | UAUCACAGCCAUUUUGACGAGU | 17–24 | 3R | 11243152…73,− | 3 | | | | | 1 | | | | | 4 |
| *mir*-13a | UAUCACAGCCAUUUUGAUGAGU | 22 | 3R | 11243287…308,− | | | | | | | | | | | |
| | *CUCCUCAAAGGGUUGUGAAAUG | 22 | 3R | 11243328…49,− | | 1 | | | | | | | | | 1 |
| *mir*-2c[p] | UAUCACAGCCAGCUUUGAUGGGC | 22 | 3R | 11243500…23,− | | | | | | | | | | | |
| *mir*-iab-4-5p | ACGUAUACUGAAUGUAUCCUGA | 22 | 3R | 12682018…39,+ | | 1 | | | | | | | | | 1 |
| *mir*-iab-4-3p | CGGUAUACCUUCAGUAUACGUAAC | 24 | 3R | 12682052…75,+ | | | 1 | | | | | | | | 1 |
| *mir*-11 | AUCACAGUCUGAGUUCUUGC | 20–23 | 3R | 17439190…209,− | 6 | 1 | | | | | | | | | 7 |
| *mir*-92a | CAUUGCACUUGUCCCGGCCUAU | 22–25 | 3R | 21461647…68,+ | 2 | 1 | | | | | | | | | 3 |
| *mir*-92b | AAUUGCACUAGUCCCGGCCUGC | 22 | 3R | 21466544…65,+ | 1 | | | | | | | | | | 1 |
| *mir*-279 | UGACUAGAUCCACACUCAUUAA | 22 | 3R | 25030739…60,+ | 1 | | | | | | | | | | 1 |

Clusters of miRNAs are represented by vertical bars between the columns that indicate the chromosome location. An asterisk is used to denote small RNAs that are derived from the strand opposite to the miRNA strand within the fold-back precursor. For the iab-4 derived miRNAs, 5p and 3p indicate 5′ and 3′ location within the conserved fold-back sequence of the iab-4 transcript.

§More copies of this miRNA are found in the genome and the clone numbers indicated cannot be assigned to a unique locus.

[p]Predicted miRNA based on phylogenetic conservation or vicinity to other clustered miRNAs.

[n]Expression of miRNA was also confirmed by Northern blotting (Figures 1 and 2).

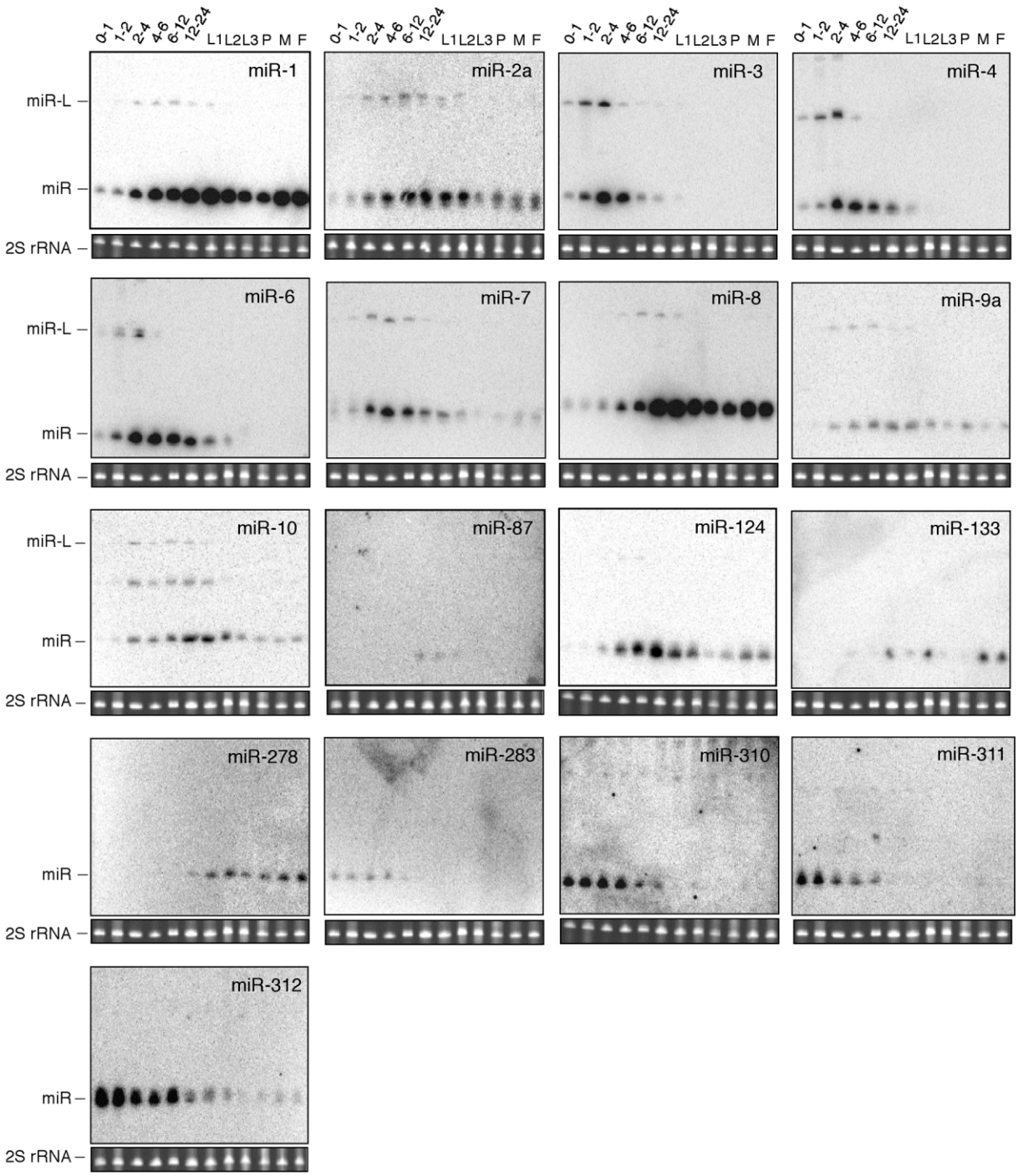


Figure 2. Expression Profiles of a Subset of Identified miRNA Genes

Northern blotting was performed as described in the legend to Figure 1B.

Although we did not clone miR-125, the lin-4 homolog of fly, we previously showed that *mir-125* was expressed in pupa and adult (Lagos-Quintana et al., 2002). As expected from their close proximity, the expression pattern of *mir-125* was identical to that of *D. melanogaster let-7* (Pasquinelli et al., 2000). A third miRNA gene that is homologous to mammalian *mir-100* was found in close proximity to *mir-125* and *let-7* (Figure 1A). We confirmed that *D. melanogaster* miR-100 is coexpressed with miR-125 and let-7 by Northern blotting (Figure 1B). Coexpression of miR-125 and let-7 and induction by the steroid hormone ecdysone at the onset of metamorphosis has been reported recently (Bashirullah et al., 2003; Sempere et al., 2003), as well as the coregulated expression of miR-100 (Sempere et al., 2003). The polycistronic expression of this miRNA cluster was also confirmed by

RT-PCR analysis, which detected a long primary transcript comprising all the hairpin precursor sequences (A.A., unpublished data). The clustering for *mir-100*, *mir-125*, and *let-7* is conserved in *A. gambiae*, although the distance between the genes has increased (Figure 1C). In mouse and human, the gene cluster underwent duplication and the distance between miRNA genes was increased further (Figure 1C). In *C. elegans*, neither *lin-4* nor its paralog *mir-237* are spatially linked with *let-7* family members (Lim et al., 2003b). The *lin-4* gene is expressed at an earlier stage than the *let-7* gene (Reinhart et al., 2000), but *mir-237* shows a very similar expression pattern to that of *let-7* (Lim et al., 2003b). Evidence further supporting the conservation of *lin-4* and *let-7* coregulation in nematodes comes from the identification of both lin-4 and let-7 RNA binding sites in the 3′ UTR of *lin-14*, *lin-28*, *lin-41*, and *lin-57* transcripts (Abrahante et al., 2003; Lin et al., 2003; Reinhart et al., 2000). Homologs to miR-100 were not identified in *C. elegans*.

The 5′ portion of the miRNAs shows the highest degree of conservation and miRNAs can be grouped in families based on this criterion (Ambros et al., 2003b; Lim et al., 2003b). It was also noticed that the 5′ ends of some miRNAs are complementary to sequence motifs in the 3′ UTR of some mRNA that are subject to posttranscriptional regulation (Lai, 2002). It was suggested that these motifs represent miRNA binding sites, which has yet to be confirmed. The 62 unique miRNA sequences can be grouped into 43 families, 14 of which are universally conserved (Supplemental Table S3). For three families, the conservation extends to mammals only, and for six families the conservation extends to nematodes only. This indicates that *D. melanogaster* is a good model system to bridge the gap for studying miRNA function in mammals and nematodes.

### Repeat-Associated Small Interfering RNAs

A surprisingly large number of small RNAs (178 clones) map to repetitive sequence elements of the *D. melanogaster* genome. The relative ratio of miRNAs to rasiRNAs varies as a function of development. Whereas about equal proportions of miRNAs to rasiRNAs are found in the early embryo, the fraction of rasiRNAs drops significantly during the transition to adults, in which only about 10% of the small regulatory RNAs are represented as rasiRNAs. The content of rasiRNAs in microdissected testes is comparable to that of early embryos. Together, these findings suggest that the taming of transposable elements and the establishment of chromatin structure may be initiated in germline tissue and during early embryonic development in an RNA-dependent manner.

The rasiRNAs contain sequences from all known forms of repetitive sequence elements, such as retrotransposons, DNA transposons, satellite, and microsatellite DNA sequences, complex repeats such as the *Su(Ste)* locus, as well as vaguely characterized repetitive sequence motifs (Table 3). The transposable elements are categorized by their mode of transposition and fall into two major subgroups. The class I elements transpose via an RNA intermediate, while the class II elements transpose through a DNA intermediate. Class I elements are again subdivided into long-terminal repeat (LTR) retrotransposons and non-LTR transposons, also referred to as long-interspersed nuclear elements (LINE) or poly(A)-type retrotransposons. Class II elements are characterized by a terminal-inverted repeat (TIR). We have cloned small RNAs from 38 different transposable elements corresponding to 40% of all known transposable elements in *D. melanogaster* (Kaminker et al., 2002). The most frequently cloned rasiRNAs were derived from the LTR transposon *roo*, which is also the most abundant transposable element in the euchromatic portion of the *D. melanogaster* genome. This indicates a role for RNAi in controlling the mobility of transposable elements in *D. melanogaster*.

The production of small sense and antisense RNAs from the repetitive *Su(Ste)* locus sequences was shown previously (Aravin et al., 2001), indicating a role for dsRNA in the homology-dependent regulation of *Stellate*. In agreement with this, we cloned several siRNAs in sense and antisense orientation from the *Su(Ste)* locus in testes tissue. The identification of rasiRNAs from other not yet annotated repetitive sequences suggests that systems similar to that of *Ste/Su(Ste)* might exist in *D. melanogaster*. Therefore, small RNA cloning is a viable strategy for identification of such cases. A complete list of all cloned repeat-associated sequences is provided in Supplemental Table S5.

Some repeat sequences such as HeT-A, TART, and subterminal minisatellites are restricted to telomeres and required for telomere maintenance. These elements are absent from euchromatic regions (Kaminker et al., 2002) and therefore link small RNAs to regulation of heterochromatin. Small RNAs of the GAGAA microsatellite family were cloned from testes and originated probably from transcripts of the fully heterochromatic Y chromosome (Lohe et al., 1993).

Another group of 10 rasiRNAs could be assigned to a 200 kb region from chromosome 2R (band 42AB), which is composed of numerous different types of transposable elements, many of which are damaged and have diverged from canonical sequences. There are no predicted genes in this region, but several lethal mutations have been described that map to this region (Bender, 1996). The reasons for this lethality are unclear, but it is conceivable that this region is an important source for small RNAs involved in some aspect of heterochromatin regulation.

## Discussion

### Two Size Classes of Small RNAs in *D. melanogaster*

Here we describe the small RNA profile of the various stages of *D. melanogaster* development. Similar to plants (Hamilton et al., 2002; Llave et al., 2002a), *D. melanogaster* produces two distinct classes of small RNAs from dsRNA precursors. The two predominant classes of small RNAs are 21 and 22 nt miRNAs and 24–26 nt rasiRNAs. The strong bias for uridine at the 5′ termini and the match to repetitive elements in sense and antisense orientation is indicative of dsRNA-specific RNase III processing (Bernstein et al., 2001) and suggests a dsRNA origin for rasiRNAs. We also identified one antisense siRNA derived from the common *Drosophila C virus*. The propagation of the positive RNA strand *Flock House*

Table 3. Repeat-Associated Small Interfering RNA Profile During *D. melanogaster* Development

| Repeat | Name | Size Range (nt) | Embryo (hr) 0–2 | 2–4 | 4–6 | 6–24 | L1–L3 | P | A | T | Total | Orientation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LINE | aurora-element | 22–25 | | | 1 | | 1 | | 1 | 1 | 4 | 3(+)/1(−) |
| | baggins | 24 | 1 | 1 | | | | | | | 2 | 2(+)/0(−) |
| | BS | 22–25 | 1 | | | | | | | 1 | 2 | 1(+)/1(−) |
| | Cr1a | 22 | 1 | | | | | | | | 1 | 1(+)/0(−) |
| | Doc3-element | 23 | 1 | | | | | | | | 1 | 0(+)/1(−) |
| | F-element | 24–27 | | | 1 | | | | 1 | 1 | 3 | 0(+)/3(−) |
| | G-element | 24 | | | | | | | | 1 | 1 | 0(+)/1(−) |
| | HeT-A element | 24–25 | 2 | 1 | | | | | | | 3 | 2(+)/1(−) |
| | Ivk | 25 | | 1 | | | | | | | 1 | 0(+)/1(−) |
| | R1-element | 24–29 | 1 | 1 | | | | | | 1 | 3 | 1(+)/2(−) |
| | Rt1a | 23–27 | 4 | 2 | | | | | | | 6 | 0(+)/6(−) |
| | Rt1b | 22–27 | 2 | 3 | 1 | | 1 | | | | 7 | 2(+)/5(−) |
| | Rt1c, diver | 16 | | 1 | | | | | | | 1 | 0(+)/1(−) |
| | TART-element | 21–27 | 2 | 3 | | | | | 1 | | 6 | 3(+)/3(−) |
| | X-element | 21–24 | 2 | | | | | | | 1 | 3 | 0(+)/3(−) |
| LTR | 17.6 | 22–24 | 2 | | | 1 | | | | | 3 | 0(+)/3(−) |
| | 1731 | 19–26 | | 1 | | | | | | 1 | 2 | 1(+)/1(−) |
| | 3S18 | 25 | | 1 | | | | | | | 1 | 1(+)/0(−) |
| | 412 | 27 | | | 1 | | | | | | 1 | 0(+)/1(−) |
| | accord | 28 | | 1 | | | | | | | 1 | 0(+)/1(−) |
| | blood | 26 | | | 1 | | | | | | 1 | 0(+)/1(−) |
| | copia | 24–25 | 1 | | | | | | | 2 | 3 | 2(+)/1(−) |
| | GATE | 24–26 | 1 | | 1 | | | | | | 2 | 2(+)/0(−) |
| | gypsy3, springer | 23 | | 1 | | | | | | | 1 | 0(+)/1(−) |
| | HMS-Beagle | 23–26 | | 1 | | | | | 1 | | 2 | 0(+)/2(−) |
| | invader1 | 20–23 | 2 | | | | | | 1 | | 3 | 2(+)/1(−) |
| | invader4 | 21 | 1 | | | | | | | | 1 | 0(+)/1(−) |
| | Max-element | 19–27 | 6 | 2 | | | | | | | 8 | 4(+)/4(−) |
| | McClintock | 26 | | 1 | | | | | | | 1 | 0(+)/1(−) |
| | mdg3 | 26 | 1 | | | | | | | | 1 | 1(+)/0(−) |
| | opus | 19–25 | 2 | | | | | | 1 | | 3 | 2(+)/1(−) |
| | Quasimodo | 21 | 1 | | | | | | | | 1 | 0(+)/1(−) |
| | roo | 18–27 | 16 | 5 | 1 | | | | 1 | | 23 | 4(+)/19(−) |
| | rooA | 24 | 1 | | | | | | | | 1 | 0(+)/1(−) |
| | Stalker | 16–24 | | | 1 | | | | 1 | | 2 | 0(+)/2(−) |
| TIR | 1360 | 17–25 | 1 | | | | | | | 3 | 4 | 4(+)/0(−) |
| | hopper2 | 24–26 | 3 | | | | | | | | 3 | 3(+)/0(−) |
| | transib4 | 26 | | | 1 | | | | | | 1 | 0(+)/1(−) |
| Su(Ste) | Su(Ste) | 23–26 | | | | | | | | 5 | 5 | 1(+)/4(−) |
| Other Heterochromatic Repeats | 1.688 satellite DNA | 19–28 | 2 | | | | 2 | | | | 4 | undef. |
| | microsatellite | 17–24 | | | | | | | | 7 | 7 | undef. |
| | subterminal minisatellite at telomere of 2L | 24 | | 1 | | | | | | | 1 | undef. |
| | repeat on chr. X | 21 | 1 | | | | | | | | 1 | undef. |
| | 42 AB region on chr. 2R | 23–28 | 6 | 3 | 1 | | | | | | 10 | undef. |
| | unspecified | 16–28 | 15 | 9 | 3 | | 1 | 1 | 4 | 4 | 37 | undef. |

The developmental stages are larva 1, L1; larva 2, L2; larva 3, L3; pupa, P; adult, A. and testes, T. Transposable elements are abbreviated as LINE, long interspersed nuclear element, LTR, long terminal repeat retrotransposons, and TIR, terminal inverted repeat DNA transposons. The orientation in + indicates the same orientation as the open reading frame within the transposable element. The orientation in *Su(Ste)* is given with respect to the coding sequence of *stellate*.

*virus* (*FHV*) in *D. melanogaster S2* cells is associated with the formation of 21 and 22 nt siRNA (Li et al., 2002), further supporting that RNA viruses can elicit an RNAi response. The 21 and 22 nt class of small RNAs (siRNAs and miRNAs) is associated with mRNA degradation and translational regulation, and both small RNAs interact with common components of the gene-silencing machinery (Doench et al., 2003; Hutvágner and Zamore, 2002). In plants, the longer class of small RNAs has been suggested to mediate repression of retroelements via histone H3 methylation as well as asymmetric DNA methylation (Zilberman et al., 2003) and systemic silencing (Hamilton et al., 2002).

The identification of two size classes of small RNAs in plants and *D. melanogaster* suggests that at least two distinct dsRNA-processing enzymes or enzyme complexes are involved in their production. *Arabidopsis thaliana* encodes at least 4 Dicer-like (dcl) proteins (Schauer et al., 2002). Complete loss of *dcl-1* function is lethal (Golden et al., 2002). Plants homozygous for

the weak loss-of-function allele *dcl-1-9* are strongly impaired in miRNA precursor processing (Park et al., 2002; Reinhart et al., 2002), yet retain activity for mediating RNAi (Finnegan et al., 2003). This suggested that another member of the Dicer family generates siRNAs, although it cannot be excluded that dcl-1 may also be able to generate 21 nt siRNAs (Finnegan et al., 2003). Addition of long siRNAs into wheat germ extracts competitively inhibited the production of long siRNAs from dsRNA (Tang et al., 2003). However, the addition of short siRNAs did not inhibit the production of either long or short siRNAs from dsRNA, indicating the presence of distinct Dicer-like enzymes.

*D. melanogaster* encodes two members of Dicer, *dcr1* and *dcr2* (Bernstein et al., 2001). In vitro processing of dsRNA in cytoplasmic extracts from 0–2 hr embryos or late embryo-derived *Schneider 2* (*S2*) cells resulted only in the formation of the 21 and 22 nt siRNAs (Bernstein et al., 2001; Zamore et al., 2000), although EST data suggest that both Dicer proteins are expressed in *S2* cells. Similarly, nuclear extracts prepared from *S2* cells only generated 21 and 22 nt siRNAs (M.L.-Q., unpublished data). The presence of long siRNAs was however documented by Northern blotting for *Su(Ste)* derived siRNAs (Aravin et al., 2001). It is possible that cofactors needed for production of longer siRNAs were missing in the embryonic and *S2* cell extracts.

## miRNAs

The number of miRNA genes identified in various organisms is still increasing. The miRNA profile of *C. elegans* is probably the most complete, with the number of validated miRNA genes being approximately 95 (Ambros et al., 2003b; Lim et al., 2003b). Independent estimates for the number of miRNA genes in *C. elegans* are 90–120 genes (Lim et al., 2003b) and 140–300 genes (Grad et al., 2003), although the lower estimate is based on more sensitive computational analyses and likely to be more accurate. For *D. melanogaster*, we have identified 62 unique miRNAs that are encoded by 68 genes. Cloning identified 55 of the unique miRNAs. The predicted miRNAs were validated by Northern blotting. Cloning of miRNAs from different developmental stages provided information on the developmental regulation of miRNAs. However, the cloning frequencies were generally not high enough to accurately quantify the level of expression. Therefore, to further our understanding of the developmental expression patterns, we performed Northern blots for a sample set consisting of 20 miRNAs (Figures 1 and 2). The spatial expression pattern of miRNAs was not examined because methods for in situ hybridization of miRNAs remain to be developed.

Intergenic miRNAs are under the control of their own promoter. miRNAs that are localized in introns might be expected, in some cases, to show the same expression pattern as the protein-coding gene, since all 18 of the identified intronic miRNAs are in the same orientation as the protein coding sequence. miR-11 is located in one of the introns of the transcription factor *E2f*, which is predominantly expressed during early development but not in adult (Dynlacht et al., 1994). Consistent with the expression pattern of *E2f*, miR-11 was cloned from early embryos, and Northern blot analysis confirmed that expression was absent from adults (Lagos-Quintana et al., 2001). The cluster including miRNAs 9c, P310, 79, and 9b is distributed over two introns of the protein serine/threonine kinase *grp* that is involved in the developmentally controlled DNA replication/damage checkpoint (Fogarty et al., 1997). It is possible that these miRNAs are also involved in regulating irradiation damage response because the *grp*/miRNA gene structure is conserved in *Drosophila* and *A. gambiae*. The most striking case of intronic conservation is presented by *mir-7*, which is found within the last intron of the *heterogeneous nuclear ribonucleoprotein K* (*hnRNP K*) gene of *Drosophila*, *A. gambiae*, mouse, and human. Overexpression of *hnRNP K* (also referred to as *transformation upregulated nuclear protein* [*TUNP*]) has been observed in cancer cells (Bomsztyk et al., 1997; Dejgaard et al., 1994), but so far the attention focused only on the encoded protein and the role of the encoded miR-7 was ignored. These examples highlight the importance of understanding miRNA gene structure and their genomic distribution and also indicate the need for comprehensive annotation of noncoding RNAs in sequenced genomes.

## rasiRNAs

Our set of 178 rasiRNAs represents the largest collection of repeat-associated small RNAs (rasiRNAs) to date. It provides insights into the complexity of RNA-guided regulation of heterochromatin. A comprehensive annotation of rasiRNAs is difficult because methods for reliable assembly of contigs containing highly repetitive sequences are only currently being developed (Hoskins et al., 2002). Many rasiRNAs are derived from classical regions of heterochromatin, such as the telomeres, or they match segments of unassembled genomic sequences classified as heterochromatin. rasiRNAs are very rich in sequences cognate to all classes of transposable elements. By analogy to the plant world, it is presumed that this longer class of siRNAs is also mediating transcriptional regulation. A role for rasiRNAs in systemic silencing can be ruled out because systemic silencing does not exist in *D. melanogaster* (Roignant et al., 2003).

*D. melanogaster* contains five members of the Argonaute protein family, three of which (*Ago1*, *Ago2, piwi*) have been implicated in mechanisms involving siRNAs and miRNAs (Carmell et al., 2002; Hammond et al., 2001; Pal-Bhadra et al., 2002; Williams and Rubin, 2002). *Ago3*, *aub*, and *piwi* are predominantly expressed in germ cells and the early embryo (Williams and Rubin, 2002). The increased abundance of rasiRNAs in embryo and testes strongly suggests that these Argonaute members are specific binding partners for rasiRNAs. Mutation of *aub* as well as of the RNA helicase *spn-E* leads to a loss of the *Su(Ste)* function (Aravin et al., 2001; Schmidt et al., 1999; Stapleton et al., 2001). In addition, *spn-E* but not *aub* mutants show derepression of some transposable elements in the germline (Aravin et al., 2001). The transcriptional repression of certain multiple-copy transgenes in *D. melanogaster* requires *piwi* (Pal-Bhadra et al., 2002). In fission yeast only a single Argonaute protein has been identified, which is

required for establishment and maintenance of heterochromatin (Hall et al., 2002; Volpe et al., 2002). In addition, some small RNAs have been found in fission yeast that are homologous to these regions (Reinhart and Bartel, 2002). Together with the biochemical evidence for Argonaute proteins as binding partners of siRNAs (Hammond et al., 2001; Martinez et al., 2002), a picture for RNA-guided sequence-specific regulation of chromatin structure is gradually emerging.

In summary, small RNAs are not only involved in guiding RNA degradation and translational control but presumably also contribute to chromatin modification and transcriptional silencing. Recording the small RNA profile will become as important as recording mRNA expression profiles to understand how cells regulate and modulate gene expression. *D. melanogaster* is an excellent model organism for investigating the function of the small RNAs in these processes.

### Experimental Procedures

RNA preparation was performed as previously described (Lagos-Quintana et al., 2002). Total RNA was isolated from embryos at different times after egg laying as well as from larval stages 1, 2, and 3 stages, from pupae, adults, and microdissected adult testes. Cloning of miRNAs was performed as described (Pfeffer et al., 2003). Northern blots were stripped and reprobed several times. Before reprobing, it was confirmed that the stripping was complete by phosphorimaging of the stripped membrane.

### Acknowledgments

We thank G. Dowe and M. Killian for sequencing, W. Lendeckel for help on cloning, A. Hüttenhofer for providing *D. melanogaster* noncoding RNA sequences to facilitate sequence annotation, Y. Dorsett, M. Landthaler, S. Pfeffer, J. Silverman, and J. Martinez for comments on the manuscript, H. Jäckle, R. Rivera-Pomar, and R. Lührmann for support. This work was funded by a Bundesministerium und Forschung (BMBF) Biofuture grant number 0311856, Program Physico-Chemical Biology RAS, and an EMBO short term fellowship ASTF 9963 to A.A.

Received: June 16, 2003
Revised: June 23, 2003
Accepted: June 27, 2003
Published online: July 2, 2003

### References

Abrahante, J.E., Daul, A.L., Li, M., Volk, M.L., Tennessen, J.M., Miller, E.A., and Rougvie, A.E. (2003). The Caenorhabditis elegans hunchback-like gene lin-57/hbl-1 controls developmental time and is regulated by microRNAs. Dev. Cell *4*, 625–637.

Ambros, V., Bartel, B., Bartel, D.P., Burge, C.B., Carrington, J.C., Chen, X., Dreyfuss, G., Eddy, S.R., Griffiths-Jones, S., Marshall, M., et al. (2003a). A uniform system for microRNA annotation. RNA *9*, 277–279.

Ambros, V., Lee, R.C., Lavanway, A., Williams, P.T., and Jewell, D. (2003b). MicroRNAs and Other Tiny Endogenous RNAs in C. elegans. Curr. Biol. *13*, 807–818.

Aravin, A.A., Naumova, N.M., Tulin, A.V., Vagin, V.V., Rozovsky, Y.M., and Gvozdev, V.A. (2001). Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. Curr. Biol. *11*, 1017–1027.

Bashirullah, A., Pasquinelli, A.E., Kiger, A.A., Perrimon, N., Ruvkun, G., and Thummel, C.S. (2003). Coordinate regulation of small temporal RNAs at the onset of Drosophila metamorphosis. Dev. Biol. *259*, 1–8.

Bender, M. (1996). New lethal loci in the 42AB region of the Drosophila melanogaster 2nd chromosome. Drosoph. Inf. Serv. *77*, 143.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363–366.

Bomsztyk, K., Van Seuningen, I., Suzuki, H., Denisenko, O., and Ostrowski, J. (1997). Diverse molecular interactions of the hnRNP K protein. FEBS Lett. *403*, 113–115.

Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B., and Cohen, S.M. (2003). bantam encodes a developmentally regulated microrna that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. Cell *113*, 25–36.

Caplen, N.J., Parrish, S., Imani, F., Fire, A., and Morgan, R.A. (2001). Specific inhibition of gene expression by small double-stranded RNAs in invertebrate and vertebrate systems. Proc. Natl. Acad. Sci. USA *98*, 9742–9747.

Carmell, M.A., Xuan, Z., Zhang, M.Q., and Hannon, G.J. (2002). The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. Genes Dev. *16*, 2733–2742.

Chomczynski, P., and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. *162*, 156–159.

Dejgaard, K., Leffers, H., Rasmussen, H.H., Madsen, P., Kruse, T.A., Gesser, B., Nielsen, H., and Celis, J.E. (1994). Identification, molecular cloning, expression and chromosome mapping of a family of transformation upregulated hnRNP-K proteins derived by alternative splicing. J. Mol. Biol. *236*, 33–48.

Denli, A.M., and Hannon, G.J. (2003). RNAi: an ever-growing puzzle. Trends Biochem. Sci. *28*, 196–201.

Djikeng, A., Shi, H., Tschudi, C., and Ullu, E. (2001). RNA interference in Trypanosoma brucei: cloning of small interfering RNAs provides evidence for retroposon-derived 24–26-nucleotide RNAs. RNA *7*, 1522–1530.

Doench, J.G., Petersen, C.P., and Sharp, P.A. (2003). siRNAs can function as miRNAs. Genes Dev. *17*, 438–442.

Dostie, J., Mourelatos, Z., Yang, M., Sharma, A., and Dreyfuss, G. (2003). Numerous microRNPs in neuronal cells containing novel microRNAs. RNA *9*, 180–186.

Dynlacht, B.D., Brook, A., Dembski, M., Yenush, L., and Dyson, N. (1994). DNA-binding and trans-activation properties of Drosophila E2F and DP proteins. Proc. Natl. Acad. Sci. USA *91*, 6359–6363.

Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001a). Duplexes of 21-nucleotide RNAs mediate RNA interference in mammalian cell culture. Nature *411*, 494–498.

Elbashir, S.M., Lendeckel, W., and Tuschl, T. (2001b). RNA interference is mediated by 21 and 22 nt RNAs. Genes Dev. *15*, 188–200.

Finnegan, E.J., Margis, R., and Waterhouse, P.M. (2003). Posttranscriptional gene silencing is not compromised in the Arabidopsis CARPEL FACTORY (DICER-LIKE1) mutant, a homolog of Dicer-1 from Drosophila. Curr. Biol. *13*, 236–240.

Fogarty, P., Campbell, S.D., Abu-Shumays, R., Phalle, B.S., Yu, K.R., Uy, G.L., Goldberg, M.L., and Sullivan, W. (1997). The Drosophila grapes gene is related to checkpoint gene chk1/rad27 and is required for late syncytial division fidelity. Curr. Biol. *7*, 418–426.

Golden, T.A., Schauer, S.E., Lang, J.D., Pien, S., Mushegian, A.R., Grossniklaus, U., Meinke, D.W., and Ray, A. (2002). SHORT INTEGUMENTS1/SUSPENSOR1/CARPEL FACTORY, a Dicer homolog, is a maternal effect gene required for embryo development in Arabidopsis. Plant Physiol. *130*, 808–822.

Grad, Y., Aach, J., Hayes, G.D., Reinhart, B.J., Church, G.M., Ruvkun, G., and Kim, J. (2003). Computational and experimental identification of C. elegans microRNAs. Mol. Cell *11*, 1253–1263.

Grishok, A., and Mello, C.C. (2002). RNAi (Nematodes: Caenorhabditis elegans). Adv. Genet. *46*, 339–360.

Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G., and Mello, C.C. (2001). Genes and mechanisms related to RNA interference regulate expression of the

small temporal RNAs that control *C. elegans* developmental timing. Cell *106*, 23–34.

Hall, I.M., Shankaranarayana, G.D., Noma, K., Ayoub, N., Cohen, A., and Grewal, S.I. (2002). Establishment and maintenance of a heterochromatin domain. Science *297*, 2232–2237.

Hamilton, A., Voinnet, O., Chappell, L., and Baulcombe, D. (2002). Two classes of short interfering RNA in RNA silencing. EMBO J. *21*, 4671–4679.

Hammond, S.M., Bernstein, E., Beach, D., and Hannon, G.J. (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. Nature *404*, 293–296.

Hammond, S.M., Boettcher, S., Caudy, A.A., Kobayashi, R., and Hannon, G.J. (2001). Argonaute2, a link between genetic and biochemical analyses of RNAi. Science *293*, 1146–1150.

Hoskins, R.A., Smith, C.D., Carlson, J.W., Carvalho, A.B., Halpern, A., Kaminker, J.S., Kennedy, C., Mungall, C.J., Sullivan, B.A., Sutton, G.G., et al. (2002). Heterochromatic sequences in a Drosophila whole-genome shotgun assembly. Genome Biol. *3*, RESEARCH0085.1–0085.16.

Hutvágner, G., McLachlan, J., Bálint, É., Tuschl, T., and Zamore, P.D. (2001). A cellular function for the RNA interference enzyme Dicer in small temporal RNA maturation. Science *93*, 834–838.

Hutvágner, G., and Zamore, P.D. (2002). A microRNA in a multiple-turnover RNAi enzyme complex. Science *297*, 2056–2060.

Johnson, K.N., and Christian, P.D. (1998). The novel genome organization of the insect picorna-like virus Drosophila C virus suggests this virus belongs to a previously undescribed virus family. J. Gen. Virol. *79*, 191–203.

Kaminker, J.S., Bergman, C.M., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., Frise, E., Wheeler, D.A., Lewis, S.E., Rubin, G.M., et al. (2002). The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective. Genome Biol. *3*, RESEARCH0084.1–0084.20.

Kennerdell, J.R., Yamaguchi, S., and Carthew, R.W. (2002). RNAi is activated during Drosophila oocyte maturation in a manner dependent on aubergine and spindle-E. Genes Dev. *16*, 1884–1889.

Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J., and Plasterk, R.H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes Dev. *15*, 2654–2659.

Ketting, R.F., Haverkamp, T.H., van Luenen, H.G., and Plasterk, R.H. (1999). Mut-7 of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. Cell *99*, 133–141.

Kogan, G.L., Tulin, A.V., Aravin, A.A., Abramov, Y.A., Kalmykova, A.I., Maisonhaute, C., and Gvozdev, V.A. (2003). The GATEretrotransposon in Drosophila melanogaster: mobility in heterochromatin and aspects of its expression in germline tissues. Mol. Genet. Genomics *269*, 234–242.

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

Lagos-Quintana, M., Rauhut, R., Meyer, J., Borkhardt, A., and Tuschl, T. (2003). New microRNAs from mouse and human. RNA *9*, 175–179.

Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., and Tuschl, T. (2002). Identification of tissue-specific microRNAs from mouse. Curr. Biol. *12*, 735–739.

Lai, E.C. (2002). Micro RNAs are complementary to 3′ UTR sequence motifs that mediate negative post-transcriptional regulation. Nat. Genet. *30*, 363–364.

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.

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

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843–854.

Lee, Y., Jeon, K., Lee, J.T., Kim, S., and Kim, V.N. (2002). MicroRNA maturation: stepwise processing and subcellular localization. EMBO J. *21*, 4663–4670.

Li, H., Li, W.X., and Ding, S.W. (2002). Induction and suppression of RNA silencing by an animal virus. Science *296*, 1319–1321.

Lim, L.P., Glasner, M.E., Yekta, S., Burge, C.B., and Bartel, D.P. (2003a). Vertebrate microRNA genes. Science *299*, 1540.

Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B., and Bartel, D.P. (2003b). The microRNAs of *Caenorhabditis elegans*. Genes Dev. *17*, 991–1008.

Lin, S.Y., Johnson, S.M., Abraham, M., Vella, M.C., Pasquinelli, A., Gamberi, C., Gottlieb, E., and Slack, F.J. (2003). The C. elegans hunchback homolog, hbl-1, controls temporal patterning and is a probable microRNA target. Dev. Cell *4*, 639–650.

Livak, K.J. (1984). Organization and mapping of a sequence on the Drosophila melanogaster X and Y chromosomes that is transcribed during spermatogenesis. Genetics *107*, 611–634.

Livak, K.J. (1990). Detailed structure of the Drosophila melanogaster stellate genes and their transcripts. Genetics *124*, 303–316.

Llave, C., Kasschau, K.D., Rector, M.A., and Carrington, J.C. (2002a). Endogenous and silencing-associated small RNAs in plants. Plant Cell *14*, 1605–1619.

Llave, C., Xie, Z., Kasschau, K.D., and Carrington, J.C. (2002b). Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. Science *297*, 2053–2056.

Lohe, A.R., Hilliker, A.J., and Roberts, P.A. (1993). Mapping simple repeated DNA sequences in heterochromatin of Drosophila melanogaster. Genetics *134*, 1149–1174.

Mallory, A.C., Reinhart, B.J., Bartel, D., Vance, V.B., and Bowman, L.H. (2002). A viral suppressor of RNA silencing differentially regulates the accumulation of short interfering RNAs and micro-RNAs in tobacco. Proc. Natl. Acad. Sci. USA *99*, 15228–15233.

Martinez, J., Patkaniowska, A., Urlaub, H., Lührmann, R., and Tuschl, T. (2002). Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. Cell *110*, 563–574.

Mattick, J.S., and Gagen, M.J. (2001). The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. Mol. Biol. Evol. *18*, 1611–1630.

Merino, E., Balbas, P., Puente, J.L., and Bolivar, F. (1994). Antisense overlapping open reading frames in genes from bacteria to humans. Nucleic Acids Res. *22*, 1903–1908.

Mette, M.F., van der Winden, J., Matzke, M., and Matzke, A.J. (2002). Short RNAs can identify new candidate transposable element families in Arabidopsis. Plant Physiol. *130*, 6–9.

Mochizuki, K., Fine, N.A., Fujisawa, T., and Gorovsky, M.A. (2002). Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. Cell *110*, 689–699.

Moss, E.G. (2002). MicroRNAs: hidden in the genome. Curr. Biol. *12*, R138–140.

Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M., and Dreyfuss, G. (2002). miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes Dev. *16*, 720–728.

Nykänen, A., Haley, B., and Zamore, P.D. (2001). ATP requirements and small interfering RNA structure in the RNA interference pathway. Cell *107*, 309–321.

Olsen, P.H., and Ambros, V. (1999). The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. Dev. Biol. *216*, 671–680.

Pal-Bhadra, M., Bhadra, U., and Birchler, J.A. (2002). RNAi related mechanism affect both transcriptional and posttranscriptional transgene silencing in *Drosophila*. Mol. Cell *9*, 315–327.

Palumbo, G., Bonaccorsi, S., Robbins, L.G., and Pimpinelli, S. (1994). Genetic analysis of Stellate elements of Drosophila melanogaster. Genetics *138*, 1181–1197.

Park, W., Li, J., Song, R., Messing, J., and Chen, X. (2002). CARPEL

FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. Curr. Biol. *12*.

Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.D., Degnan, B., Müller, P., et al. (2000). Conservation of the sequence of *let-7* heterochronic regulatory RNA. Nature *408*, 86–89.

Pasquinelli, A.E., and Ruvkun, G. (2002). Control and developmental timing by microRNAs and their targets. Annu. Rev. Cell Dev. Biol. *18*, 495–513.

Pfeffer, S., Lagos-Quintana, M., and Tuschl, T. (2003). Cloning of small RNA molecules. In Current Protocols in Molecular Biology, F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidmann, J.A. Smith, and K. Struhl, eds. (New York: Wiley and Sons), 26.24.21–26.24.16.

Plasterk, R.H. (2002). RNA silencing: the genome's immune system. Science *296*, 1263–1265.

Provost, P., Dishart, D., Doucet, J., Frendewey, D., Samuelsson, B., and Radmark, O. (2002). Ribonuclease activity and RNA binding of recombinant human Dicer. EMBO J. *21*, 5864–5874.

Reinhart, B.J., and Bartel, D.P. (2002). Small RNAs correspond to centromere heterochromatic repeats. Science *297*, 1831.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature *403*, 901–906.

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

Rhoades, M., Reinhart, B., Lim, L., Burge, C., Bartel, B., and Bartel, D. (2002). Prediction of plant microRNA targets. Cell *110*, 513.

Roignant, J.Y., Carre, C., Mugat, B., Szymczak, D., Lepesant, J.A., and Antoniewski, C. (2003). Absence of transitive and systemic pathways allows cell-specific and isoform-specific RNAi in Drosophila. RNA *9*, 299–308.

Schauer, S.E., Jacobsen, S.E., Meinke, D.W., and Ray, A. (2002). DICER-LIKE1: blind men and elephants in Arabidopsis development. Trends Plant Sci. *7*, 487–491.

Schmidt, A., Palumbo, G., Bozzetti, M.P., Tritto, P., Pimpinelli, S., and Schafer, U. (1999). Genetic and molecular characterization of sting, a gene involved in crystal formation and meiotic drive in the male germ line of Drosophila melanogaster. Genetics *151*, 749–760.

Seggerson, K., Tang, L., and Moss, E.G. (2002). Two genetic circuits repress the Caenorhabditis elegans heterochronic gene lin-28 after translation initiation. Dev. Biol. *243*, 215–225.

Sempere, L.F., Sokol, N.S., Dubrovsky, E.B., Berger, E.M., and Ambros, V. (2003). Temporal regulation of microRNA expression in Drosophila melanogaster mediated by hormonal signals and broad-Complex gene activity. Dev. Biol. *259*, 9–18.

Shendure, J., and Church, G.M. (2002). Computational discovery of sense-antisense transcription in the human and mouse genomes. Genome Biol. *3*, RESEARCH0044.1–0044.14.

Stapleton, W., Das, S., and McKee, B.D. (2001). A role of the Drosophila homeless gene in repression of Stellate in male meiosis. Chromosoma *110*, 228–240.

Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., and Mello, C.C. (1999). The rde-1 gene, RNA interference, and transposon silencing in *C. elegans*. Cell *99*, 123–132.

Tang, G., Reinhart, B.J., Bartel, D.P., and Zamore, P.D. (2003). A biochemical framework for RNA silencing in plants. Genes Dev. *17*, 49–63.

Tautz, D., Hancock, J.M., Webb, D.A., Tautz, C., and Dover, G.A. (1988). Complete sequences of the rRNA genes of Drosophila melanogaster. Mol. Biol. Evol. *5*, 366–376.

Taverna, S.D., Coyne, R.S., and Allis, C.D. (2002). Methylation of histone h3 at lysine 9 targets programmed DNA elimination in tetrahymena. Cell *110*, 701–711.

Voinnet, O. (2002). RNA silencing: small RNAs as ubiquitous regulators of gene expression. Curr. Opin. Plant Biol. *5*, 444–451.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., and Martienssen, R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science *297*, 1833–1837.

Wassenegger, M. (2002). Gene silencing-based disease resistance. Transgenic Res. *11*, 639–653.

Waterhouse, P.M., Wang, M.B., and Lough, T. (2001). Gene silencing as an adaptive defence against viruses. Nature *411*, 834–842.

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell *75*, 855–862.

Williams, R.W., and Rubin, G.M. (2002). ARGONAUTE1 is required for efficient RNA interference in *Drosophila* embryos. Proc. Natl. Acad. Sci. USA *99*, 6889–6894.

Wu-Scharf, D., Jeong, B., Zhang, C., and Cerutti, H. (2000). Transgene and transposon silencing in *Chlamydomonas reinhardtii* by a DEAH-Box RNA helicase. Science *290*, 1159–1163.

Xu, P., Vernooy, S.Y., Guo, M., and Hay, B.A. (2003). The Drosophila microRNA Mir-14 suppresses cell death and is required for normal fat metabolism. Curr. Biol. *13*, 790–795.

Yelin, R., Dahary, D., Sorek, R., Levanon, E.Y., Goldstein, O., Shoshan, A., Diber, A., Biton, S., Tamir, Y., Khosravi, R., et al. (2003). Widespread occurrence of antisense transcription in the human genome. Nat. Biotechnol. *21*, 379–386.

Yuan, G., Klambt, C., Bachellerie, J.P., Brosius, J., and Hüttenhofer, A. (2003). RNomics in Drosophila melanogaster: identification of 66 candidates for novel non-messenger RNAs. Nucleic Acids Res. *31*, 2495–2507.

Zamore, P.D. (2002). Ancient pathways programmed by small RNAs. Science *296*, 1265–1269.

Zamore, P.D., Tuschl, T., Sharp, P.A., and Bartel, D.P. (2000). RNAi: Double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell *101*, 25–33.

Zhang, H., Kolb, F.A., Brondani, V., Billy, E., and Filipowicz, W. (2002). Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. EMBO J. *21*, 5875–5885.

Zilberman, D., Cao, X., and Jacobsen, S.E. (2003). ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science *299*, 716–719.

**Note Added in Proof**

A recent study computationally identified numerous *D. melanogaster* miRNAs and used Northern blotting for validation (Lai, E.C., Tomancak, P., Williams, R.W., and Rubin, G.M. [2003]. Computational identification of *Drosophila* microRNA genes. Genome Biol. *4*, R42. Published online: June 30, 2003). Thirty-two previously unknown miRNA genes were identified in common, eight novel miRNA genes were not found in our study, and twelve novel miRNAs were absent from the prediction list. The 5′ ends of 13 commonly identified miRNAs were misidentified in the prediction between +5 and −2 nucleotides according to cloned sequences. The miRNAs derived from the *mir-281-1* and *mir-281-2* gene cluster were predicted antisense to cloned sequences, and miR-276a and miR-276b were predicted from the strand opposite to the cloned sequences. The curated set of 76 miRNA genes will become available at http://www.sanger.ac.uk/Software/Rfam/mirna/index.shtml.