RESEARCH COMMUNICATION

‘Inc-miRs’: functional intron-interrupted microRNA genes

Huibin Zhang,1 Jay M. Maniar,1 and Andrew Z. Fire1,2,3

1Department of Genetics, Stanford University, Stanford, California 94305, USA; 2Department of Pathology, Stanford University, Stanford, California 94305, USA

The discovery of microRNAs (miRNAs) lin-4 and let-7 as temporal regulators in Caenorhabditis elegans led to broader searches for novel miRNAs and their biological roles. Unlike protein-coding genes and some long noncoding RNAs, canonical metazoan miRNAs are not known to contain introns within their genomic precursor sequences. Because the short length of miRNAs complicates a statistically definitive assignment of split genes in RNA sequencing data sets, we took an experimental approach toward testing the compatibility of splicing and functional miRNA biogenesis. To definitively evaluate the possibility that miRNAs could derive from interrupted genes, we constructed intron-interrupted variants of C. elegans lin-4 and assayed for their miRNA-encoding capability and biological activity in the developing organism. Our studies indicate that (1) intron-containing miRNAs (inc-miRs) can be efficiently spliced and processed to produce miRNAs with normal termini, and (2) these miRNAs can be functional in full rescue of developmental phenotypes in null mutants lacking endogenous lin-4. This study provides the first evidence to support the ability of intron-interrupted miRNA precursors to produce functional regulators and identifies an additional modality available for metazoan miRNA production.

Supplemental material is available for this article.

Received April 12, 2011; revised version accepted June 16, 2011.

MicroRNAs (miRNAs) were first discovered with the identification of lin-4 as a regulator of temporal development in Caenorhabditis elegans [Lee et al. 1993; Wightman et al. 1993]. Although a handful of other miRNAs have been identified through genetics means [Reinhart et al. 2000; Johnston and Hobert 2003], the number of miRNAs discovered using forward genetics has remained small. The majority of C. elegans miRNA knockouts do not display overt phenotypes, indicating a combination of overlapping phenotypes (partial or complete redundancy) and subtle biological effects in this system [Miska et al. 2007; Alvarez-Saavedra and Horvitz 2010].

With the advent of better small RNA capture methods and high-throughput sequencing, thousands of miRNAs have now been identified in a variety of organisms [Ambros and Lee 2004; Pfeffer et al. 2005; Kloosterman et al. 2006; Rajagopalan et al. 2006]. Some of the criteria used in novel miRNA annotation include the prediction of a fold-back precursor structure, detection of the partially complementary “star” strand, and phylogenetic conservation of the candidate miRNA [Ambros et al. 2003]. Although a large proportion of miRNAs still have unknown functions [Bushati and Cohen 2007], much headway has been made in understanding miRNA biogenesis, as well as identifying the important players in this pathway [Kim et al. 2009].

miRNA genes may be transcribed either from their own promoters, as part of a cluster of miRNA genes, or as part of a host gene [Fig. 1A; Kim et al. 2009]. These long transcripts are capped and spliced to form primary miRNAs [pri-miRNAs] in metazoans [Cai et al. 2004; Lee et al. 2004]. Pri-miRNAs are processed by Drosha to produce ~70-nucleotide (nt) stem–loop intermediates, known as precursor miRNAs [pre-miRNAs] [Kim et al. 2009]. Dicer then acts on pre-miRNAs to generate ~22-nt mature miRNAs [Kim et al. 2009]. In the case of mirtrons, pre-miRNAs are spliced directly out of small introns, therefore bypassing the Drosha processing step [Okamura et al. 2007; Ruby et al. 2007]. Furthermore, in cases where pre-miRNAs of mirtrons are derived from longer introns, the exosome has been found to mediate 3’ end trimming to yield the pre-miRNA [Fig. 1A; Flynt et al. 2010].

Protein-coding genes often require splicing to remove introns that interrupt both coding and noncoding sequences [Soller 2006]. Intron removal has also been observed with long noncoding RNAs [Mattick and Makunin 2006]. In contrast, metazoan miRNA genes are not known to contain introns within the pre-miRNA or mature miRNA sequences [Fig. 1B]. None of the few miRNAs and pre-miRNAs identified through “forward genetics” contains an intron. The bulk of characterized miRNAs, identified first through high-throughput sequencing, also contains no intron interruptions. It should be pointed out that the standard methodology of novel miRNA discovery was not designed to find such intron-interrupted entities, since algorithms that have been used for miRNA annotation apply an initial filter of requiring a perfect match to the relevant genome [Lagos-Quintana et al. 2001; Lau et al. 2001; Lee and Ambros 2001]. It is certainly conceivable that such endogenous intron-containing miRNAs (inc-miRs) may exist; both protein-coding and miRNA genes are transcribed by RNA polymerase II, unlike other noncoding small RNA genes such as tRNAs [Lee et al. 2004]. In addition, the studies on the biogenesis of mirtrons and other miRNAs that reside in introns of host genes show that splicing can be integrated into the biogenesis pathway to generate mature miRNAs and, in the case of mirtrons, plays an essential role [Kim and Kim 2007; Okamura et al. 2007; Ruby et al. 2007]. Intriguingly, there is one instance in plants [miR444] in which splicing signals close to the loop region of a miRNA precursor are conserved and where splicing could contribute to the production of a canonical Dicer substrate [Sunkar et al. 2005; Lu et al. 2008; Thieme et al. 2011]. A computational
search for intron-interrupted miRNA genes in C. elegans small RNA data sets (Maniar and Fire 2011) was carried out, but the short length of these RNAs complicates the establishment of genomic origins. Indeed, by sheer numbers of possibilities for start sites and intron locations, a substantial fraction of random 21-mers is assignable to interrupted genomic segments. Additional inherent ambiguities in post-transcriptional modifications—such as RNA editing (Luciano et al. 2004), 3' nucleotide additions (Li et al. 2005), and mRNA degradation—further complicate any definitive assignment of intron-interrupted miRNA genes based on sequence data alone. These considerations led us to address the possibility of spliced miRNAs more directly, using a defined assay system.

In this study, we created intron-interrupted lin-4 constructs and assayed for their lin-4 rescue activities using the well-characterized lin-4 phenotypic rescue assay in C. elegans. The introns in these constructs interrupt the pre-miRNA sequence. The rescue of the lin-4-null phenotype and subsequent high-throughput sequencing and analyses of captured mature lin-4 miRNAs in these transgenic animals indicate that inc-miRs can be accurately spliced to produce functional miRNAs efficiently.

Results and Discussion

In vivo function of intron-containing lin-4 constructs

Rescue of null mutations in the well-characterized C. elegans miRNA lin-4 provides a definitive assay for miRNA functionality. This heterochronic miRNA plays an important role in temporal control of development, especially during the L1-to-L2 transition (Chalfie et al. 1981). In animals that lack lin-4, development does not occur in a timely fashion and specific tissues fail to terminally differentiate, lin-4(e912) animals fail to form adult alae, lack a functional vulva, and essentially never lay eggs (Chalfie et al. 1981). We examined the functionality of lin-4 when interrupted by a class II intron and introduced into populations of lin-4-null mutant animals using DNA-mediated transformation.

Twelve intron-containing lin-4 constructs [int_1 to int_12] were made using a synthetic 51-nt intron [Fig. 2A, panel i] previously used in GFP constructs for C. elegans expression (Fire et al. 1998). The 695-nt lin-4 genomic sequence used in these constructs [Fig. 2A, panel ii] has been shown to rescue the lin-4-null phenotype (Lee et al. 1993). The 51-nt intron was inserted at different locations within the lin-4 precursor sequence (indicated by arrows in Fig. 2B). These intron-containing lin-4 constructs were introduced into lin-4(e912) animals and the transgenic strains were initially assayed for lin-4 rescue activity by assessing vulva formation at the L4 larval and adult stages.

As shown in Figure 2C, several intron-interrupted constructs showed rescue of vulval development at levels comparable with those seen with the parental [wild-type] plasmid. Different degrees of rescue were observed for the various insertion points tested. To provide more accurate assessments of lin-4 rescue activities, the transgenic animals were categorized into four different phenotypes: (1) egg-laying, (2) protruding vulva [pvl], (3) bursting through a partially formed vulva [bursting], and (4) vulvaless. Both pvl and vulvaless animals retain fertilized eggs and eventually become “bags of worms” when the progeny hatch inside them. Animals with a “bursting vulva” phenotype usually do not survive to produce progeny. As there is some variation in exogenous DNA expression between transgenic lines (due to mosaicism, copy number variation, and transgene silencing), we observed a range of rescue phenotypes associated with vulva formation. By considering all three phenotypic categories that indicated signs of vulva formation [egg-laying, pvl, and bursting] and comparison with transgenic animals expressing a wild-type lin-4 construct [no intron, 48%], five constructs [int_3, int_5, int_6, int_9, and int_10] showed robust lin-4 rescue activities (ranging from 30%–73%) [Fig. 2C]. The remaining intron-containing lin-4 constructs had varying limited levels of rescue activity, between 2% and 17% of assayed transgenic animals displayed restoration of vulva formation [Fig. 2C]. A detailed summary of the percentages of transgenic animals in each phenotypic category for each transgenic line is found in Supplemental Figures 1 and 2.

In addition to vulval formation and egg-laying, a lack of lin-4 also causes a number of other developmental defects, including defects in cuticle patterning [a loss of adult cuticular structures called alae] and whole-animal growth rate [about a half day longer in the time needed to develop from embryo to L4 at 25°C from our observations] (Chalfie et al. 1981). We assayed transgenic animals from lines with robust rescue [int_3, int_5, int_55m, int_8, and int_10] and found that adult alae were present in these populations [27%–96%, compared with 31% for parental [wild-type] lin-4 construct]. Populations of rescued animals also exhibited rescued growth rates that were comparable with wild-type N2 populations [alleviating the approximately half-day lag observed in lin-4-null mutants], together with the rescue of vulval development and egg-laying, these observations strongly suggest that intron-interrupted lin-4 constructs can provide functionality in developmental progression.
Splicing is required for rescue activity of intron-containing lin-4 constructs

To test whether splicing was required for mature lin-4 production from intron-interrupted lin-4 constructs, we mutated the splice sites of two intron-interrupted lin-4 constructs [int_3 and int_5] previously shown to have robust lin-4 rescue activities. The synthetic intron, built with the consensus class II intron structure <gt...ag> (Blumenthal and Steward 1997; Soller 2006), was changed by mutation at the junctions to yield <ag...ct> (Fig. 3A). Transgenic animals expressing these mutated intron-interrupted lin-4 constructs [int_3sm and int_5sm] were assayed for their lin-4 rescue activities.

As shown in Figure 3B, when compared with rescue percentages from the previous experiments [int_3 and int_5], splice site mutations in construct int_3 prevented effective rescue of the lin-4 mutation, while equivalent splice site mutations in construct int_5 retained rescue activity. A detailed summary of the percentages of transgenic animals in each phenotypic category for each transgenic line is found in Supplemental Figures 1 and 3. The contrasting results between int_3sm and int_5sm may be due to the differences in intron location within the miRNA precursor, the intron in int_3 is located within the double-stranded stem, whereas the intron in int_5 is located in the loop region (Fig. 2B). These results suggest a flexibility in the loop region that may be absent in the regions destined to comprise the mature miRNA.

Generation of ‘wild-type’ lin-4 miRNA from intron-containing lin-4 constructs

We prepared small RNA libraries from transgenic lin-4(e912) lines for high-throughput sequencing using the Illumina platform. These libraries were prepared using the 5’ phosphate-dependent protocol to enrich for miRNAs (Pfeffer et al. 2005). The small RNA reads were aligned to both the wild-type lin-4 precursor and the corresponding intron-containing lin-4 precursor in order to capture all lin-4 variants.

As indicated in Figure 4A, both mature lin-4 and lin-4* were observed in four transgenic lines with robust rescue activities [int_5, int_5sm, int_8, and int_10], as well as in two transgenic lines with limited rescue activities [int_1 and int_11]. We note that the more active inc-miR constructs [int_5, int_8, and int_10] produce mature lin-4 at levels in a range that is within several-fold of that observed in wild-type animals (Fig. 4A). Although measurement of absolute miRNA levels through RNA-seq is subject to sequence-specific variations in capture and sequencing efficiencies, a number of studies validates the use of differences in sequence count ratios in measuring relative miRNA levels as they change.

Figure 2. In vivo function of inc-miR constructs. (A) Key features used in creating intron-interrupted lin-4 constructs. (Panel i) Sequence of the 51-nt synthetic intron used is shown here, with the consensus splice donor (GT) and acceptor (AG) in red. (Panel ii) Relative position of the 695-nt lin-4(e912) genomic fragment (black line) to host gene F59G1.4 (above black line) is illustrated. The lin-4(d4)-containing genomic fragment overlaps with both the intron [red line] and exon [blue box] of F59G1.4. (B) Sites of intron insertion in the lin-4 precursor sequence. The lin-4 precursor sequence is shown, with the mature lin-4 sequence highlighted in orange. Colored arrows indicate sites of intron insertion within the precursor sequence; different colors indicate different levels of lin-4 rescue activity [limited or robust] associated with each intron-interrupted lin-4 construct. (C) Percentages of transgenic animals with lin-4 rescue activity for intron-interrupted lin-4 constructs. Percentages of transgenic animals with vulva (including egg-laying, pvI, and bursting phenotypes) for each intron-interrupted lin-4 construct are represented by a series of box plots. Multiple transgenic lines were assayed for each intron-interrupted lin-4 construct. The average percentage of transgenic animals displaying specific lin-4 rescue activity across all transgenic lines for each construct is represented by the band in the middle of the box plot [intersection of blue and red boxes]. The top and bottom of each box plot represent the standard deviation between the different transgenic lines assayed. The black dashes within each box plot represent the standard error of mean for that construct. Maximum and minimum percentages of transgenic animals with specific lin-4 rescue activity observed for each construct are shown by the top and bottom whiskers of the box plot. The number above each box plot is the total number of transgenic animals assayed for each construct, and the number in parentheses is the number of transgenic lines assayed for each construct. The percentages for uninjected control animals [N2 and lin-4(e912)] are represented by single black lines. (D) Intron-containing let-7 constructs functionally rescue the let-7(n2853) mutation. The let-7 precursor sequence is shown, with mature let-7 sequence highlighted in orange. Red arrows indicate sites of intron insertion within the precursor sequence. The table summarizes the percentages of transgenic animals showing rescue of vulval bursting phenotype for each intron-interrupted let-7 construct at 25°C.
in different conditions [Lui et al. 2007; Marioni et al. 2008; Subramanian et al. 2008; Xu et al. 2010]. Thus, sequence counts for lin-4 relative to total miRNA counts provide a metric for relative changes in lin-4 abundance between different transgenic strains. The fact that some mature lin-4 and lin-4* were captured from transgenic lines with limited rescue activities from int_3 and int_5 [Fig. 4A] leaves either insufficient expression, inefficient splicing, or inefficient engagement of the resulting small RNA in the miRNA machinery as possible explanations of limited lin-4 rescue activities in these intron-interrupted lin-4 constructs. An analysis of start and end positions of captured lin-4 reads shows that the majority of lin-4 reads start and end at similar positions as wild-type mature lin-4 captured from N2 animals [Fig. 4B; Supplemental Fig. 4]. Together, these results demonstrate that intron-containing lin-4 constructs were accurately spliced to generate wild-type lin-4.

Consistent with the retention of rescue activity by int_5sm [Fig. 3B], sequencing of small RNA populations in transgenic int_5sm lin-4(e912) animals yielded wild-type lin-4 and lin-4* [Fig. 4A]. The majority of lin-4 reads start and end at similar positions as wild-type mature lin-4 captured from N2 animals [Fig. 4C]. We did not capture other small RNA sequences that originate from the rest of unspliced int_5sm precursor, suggesting that the observed robust rescue is not likely a result of random “dicing” of unspliced int_5sm precursor to give sufficient levels of functional lin-4. Another possible explanation would be that mature wild-type lin-4 was stabilized, while other Dicer products of the int_5sm precursor were rapidly degraded.
In vivo function of intron-containing let-7 constructs

To further assess the functionality of inc-miRs, intron-interrupted versions of a different *C. elegans* miRNA, let-7, were constructed and assayed for phenotypic rescue of the let-7(n2853) mutation. The heterochronic miRNA let-7 is highly conserved across species (Pasquinelli et al. 2000) and is important for temporal development, especially during the larval-to-adult transition in *C. elegans* (Reinhart et al. 2000). The temperature-sensitive let-7(n2853) mutation causes adult lethality through vulval bursting at higher temperatures (Reinhart et al. 2000). Functional rescue of the bursting phenotype in transformed let-7(n2853) animals at nonpermissive temperatures (25°C) allows us to test the functionality of intron-interrupted let-7 constructs. As shown in Figure 2D, the four intron-interrupted let-7 constructs tested were able to rescue adult lethality associated with vulval bursting at 25°C in transgenic let-7(n2853) animals. These results demonstrate that the capacity of intron-interrupted miRNAs to function in biological contexts is not limited to *lin-4*, suggesting instead that this capacity may be a general property of at least a substantial subset of miRNA effectors.

Implications of functional inc-miRs

In this study, we demonstrate that functional metazoan miRNAs can be processed from intron-interrupted precursors. This analysis highlights both the potential existence of spliced miRNAs in physiological systems and the potential utility of intron-interrupted miRNA transgene constructs as biotechnology tools. Since previous studies advancing sequence-based discovery of miRNAs have all required a contiguous alignment to the genome, intron-interrupted entities would have been missed (Ambros et al. 2003). The small size of mature miRNAs, combined with inherent variability in intron length, sequence, and structure, confounds any statistically and biologically definitive attempt to assign any given miRNA to a specific intron-interrupted genome segment. Nonetheless, our work argues that such entities may (and are likely to) exist. With the wealth of sequencing data coming from various high-throughput sequencing platforms, a significant portion of sequence reads are discarded as “unaligned.” Although many different mechanisms may account for these sequence reads, the functionality of the inc-miR constructs indicates that unaligned sequences cannot simply be ignored. Together with the recent discovery of mirtrons (Okamura et al. 2007, Ruby et al. 2007), our observations illustrate the diversity in miRNA biogenic mechanisms and encourage consideration of all possibilities in the discovery of novel miRNAs.

Materials and methods

Transgene-based expression of inc-miR constructs

Intron-containing *lin-4* constructs *lin-4* DNA and marker plasmid [pC1φpha-1(+)](Granato et al. 1994) were mixed and introduced into populations carrying the *lin-4* null mutation e912, and transgenic populations were then assayed for their temporal progression phenotypes. We note that transgenic strains with wild-type *lin-4* are subject to differing degrees of partial rescue and mosaicism, with perfect rescue [full pheno-typic rescue and egg-laying] observed in only a fraction of transgenic animals (Supplemental Fig. 1). Based on a titration of wild-type *lin-4* plasmid pHZ018, we found optimal rescuing activity in a range encompassing a ratio of *lin-4* construct to marker plasmid of 1:4 or 1:8. Each mutant construct was then assayed for rescue for both ratios. A complete description of the constructs, concentrations, and strains used in the transgenic experiments is provided in the Supplemental Material.

Intronic-containing let-7 constructs *let-7* DNA and marker plasmid [pPD117.01/gene-7-gfp](Chálie et al. 1994; Fire et al. 1998) were introduced into MT7626[let-1026] (Reinhart et al. 2000), and resulting transgenic populations [animals with green fluorescent touch receptor neurons] were assayed for rescue of vulval bursting phenotype. A complete description of the constructs and concentrations used in the transgenic experiments is provided in the Supplemental Material.

Splice site mutations of intron-containing *lin-4* constructs

Due to their robust *lin-4* rescue activities and locations of introns outside the *lin-4* mature sequence, the possibility that either int_3 or int_5 could have produced functional *lin-4* without splicing was of particular interest. We mutated the splice sites of intron-interrupted *lin-4* constructs int_3 and int_5 by substituting AG for GT at the splice donor site and CT for AG at the splice acceptor site (Fig. 3A), yielding constructs int_3sm [pHZ133] and int_5sm [pHZ136], respectively. Transgenic lines were derived in a similar manner by coinjection with *pha-1(+)* into PD7143 as described above and assayed for rescue of *lin-4(e912)* vulvalless phenotype.

Preparation of small RNA libraries for Illumina sequencing

Small RNA libraries were made from the following strains: PD1473 [line 183.3.1.1], PD1416 [line 140.2.2.1], PD1472 [line 150.4.1.1], PD1465 [line 174.1.3.1], PD1445 [line 179.1.4.1], PD1412 [line 25.1.3], PD1412 [line 25.1.3], N2, and *lin-4(e912)*. A detailed description on library preparation is presented in the Supplemental Material.

Sequence analysis of small RNA libraries

Thirty-six-nucleotide reads were generated from the small RNA libraries using the Illumina Genome Analyzer system. The captured small RNA sequences, after removal of linker and adaptor sequences, were aligned using BLAT against a reference list of known *C. elegans* mature miRNA sequences downloaded from mirBase [http://www.mirbase.org]. The number of captured small RNA sequences that perfectly matched mature *lin-4* was obtained for each library. We also searched the small RNA sequence data for examples of aberrantly spliced products (i.e., for sequences that could have been derived from cryptic donor or acceptor sites within the transgene-produced precursor RNA); no such aberrantly spliced small RNAs were observed in these analyses.

Acknowledgments

We thank Karen Artilas, Sam Gu, Ayelet Lamm, and Julia Pak for useful discussions. This work was supported by NIH grant R01GM37706. H.Z. was supported by A’STAR, Singapore. J.M.M. was supported by NIH (T32HG00044) and the National Science Foundation Graduate Research Fellowship Program. The MT7626 strain was provided by the Caenorhabditis Genetics Center, which is funded by the NIH National Center for Research Resources [NCRR].

References

Alvarez-Saavedra E, Horvitz HR. 2010. Many families of *C. elegans* microRNAs are not essential for development or viability. *Curr Biol* 20:367–373.

Ambros V, Lee RC. 2004. Identification of microRNAs and other tiny noncoding RNAs by cDNA cloning. *Methods Mol Biol* 265:131–158.

Ambros V, Bartel B, DP Bartel, Burge CB, Carrington JC, Chen X, Dreyfuss G, Eddy SR, Griffiths-Jones S, Marshall M, et al. 2003. A uniform system for microRNA annotation. *RNA* 9:277–279.

Blumenthal T, Steward K. 1997. *cis*-splicing in worms. In *C. elegans II* [ed. DK Riddle et al.], pp. 117–145. Cold Spring Harbor Laboratory Press, Plainview, NY.
'Inc-miRs': functional intron-interrupted miRNA genes

Huibin Zhang, Jay M. Maniar and Andrew Z. Fire

*Genes Dev.* 2011, 25:
Access the most recent version at doi:10.1101/gad.2058711

---

**Supplemental Material**

http://genesdev.cshlp.org/content/suppl/2011/08/09/25.15.1589.DC1

**References**

This article cites 37 articles, 13 of which can be accessed free at:
http://genesdev.cshlp.org/content/25/15/1589.full.html#ref-list-1

**License**

**Email Alerting Service**

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.