Genetic and genomic analysis of small RNA pathways in nematodes

Genetic and genomic analysis of small RNA pathways in nematodes

A dissertation presented

By

Zhen Shi

to

The department of Biological Sciences in Dental Medicine

in partial fulfillment of the requirement

for the degree of

Doctor of Philosophy

in the subject of

Biological Sciences in Dental Medicine

Harvard University

Cambridge, Massachusetts

August 2013

Dissertation Advisor: Prof. Gary Ruvkun Author: Zhen Shi

Genetic and genomic analysis of small RNA pathways in nematodes

ABSTRACT

Small noncoding RNAs, including microRNAs (miRNAs), piwi-interacting RNAs (piRNAs), and endogenous small-interfering RNAs (endo-siRNAs), regulate developmental and defense pathways in animals. While many small RNA silencing protein cofactors have been identified, much more is to be learned from a dynamic and quantitative perspective to reveal the underlying mechanisms and designing principles of each pathway. In this dissertation, I present studies that examine the temporal dynamics of small RNA pathways – one from an evolutionary time scale among the nematode species, and one from finely staged *Caenorabditis elegans* during the first larval stage. I also describe works identifying new cofactors functions in the miRNA pathway, potentially through regulating the spatial dynamics of the miRNA silencing complex.

To better understand the various small RNA pathways from an evolutionary perspective, I deep sequenced small RNA from several nematode species and examined the conservation and evolution of each class of small RNAs. This reveals an extraordinary sequence fluidity of piRNAs and endo-siRNAs. However, many features such as their genomic distribution and expression patterns are highly conserved. I found that nematodes produce two distinct sex-specific classes of piRNAs, suggesting different roles for piRNAs in male and female germlines.

To reveal the kinetics of miRNA-mediated silencing during the animal cell fate transition, we performed a quantitative analysis of *lin-4* miRNA-mediated silencing of its target gene, *lin-14*. Our results point to two phases of regulation: a fast *lin-14* mRNA

destabilization phase, and long-term translational inhibition that is important in maintaining the silencing of *lin-14* by the *lin-4* miRNA.

Lastly, I performed a candidate-based RNAi screen for genes involved in miRNA activity. This study leads to the finding that the mevalonate pathway regulates miRNA activity. Dolichol phosphate, synthesized from the mevalonate pathway, functions as a lipid carrier of the oligosaccharide moiety destined for protein N-linked glycosylation. Inhibition of the dolichol pathway of protein N-glycosylation also causes derepression of miRNA target mRNAs, suggesting proteins that mediate miRNA repression could be regulated by N-glycosylation.

Together, these studies highlight the importance of studying both the temporal and spatial dynamics of small RNA pathways in creating new insights.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENT

I would like to thank my dissertation exam committee, Danesh Moazed, Mike Blower, Shobha Vasudevan and Marian Walhout for taking the time to evaluate my work and share their thoughts and insights. I owe great gratitude to my dissertation advisory committee, Danesh Moazed, Fred Winston and Craig Hunter, for their kind support. Their thoughtful suggestions and warm encouragements are way beyond my projects have always made me feel so fortunate to have them as my mentors.

Probably like many other graduate students, I have much more failed projects than the few ones that worked out and are presented here. What really encourages me to keep exploring, is the huge support and freedom that I enjoyed in Gary's lab. Gary is an endless vivid source of creativity and wisdom that he constantly encourages us to think about not only interesting projects (which he refers to as 'T-pass'), but also big and harder questions (a 'T station' in his words). On top of this, Gary is the most optimistic and adventurous person I have ever met. As we self mocked (with pride): "We don't listen to orthodoxy, and that's why we are constantly wrong." If Gary led and challenged me to think about big ideas, it is Gary again who taught and showed me to do serious science. To vigorously test and clarify some controversy in the miRNA field, Gary not only asked us to repeat some experiments in a much more refined way, but also digged out his own results. It was really a mixed funny and sacred feeling to hold those gigantic films of Northern Blot that Gary did - at the year when I was born - and to make it a figure on our paper. It was through this experience that I become to be respectful to every paper and every piece of work, no matter it is big or small. On a personal level, Gary is a caring mentor, generous supporter, a backseat driver as I became comfortable with the

independence but a keen observer as always who watches every of my little progress. I thank Gary for his mentor, his encouragements and for inspiring my love of science.

I also want to thank all folks in the Ruvkun lab, in particular, my collaborators on various projects: Xiaoyun for sharing the SynMuv project from where I started my graduate work; Gabe on the *lin-14* project and it would have been an impossible task to collect samples every 3 hours for 24 hours without him; and Tai and Yan on the small RNA evolution study. Besides my collaborators, every Ruvkun lab member and many people in the department have helped me and I would like to thank. Jihong and Meng are like my big brother and sister and were always the first people I turned into when I had troubles. Jihong poked me every now and then, like poking his worms and kept me motivated. Tai, Gabe, Marty, Sylvia, Carolyn, Robbie, Chi, Xiaoyun, Yuval, Susana, Buck and Peter - make up to our awesome small RNA team that I want to thank most. It has been a great fun to work with these talented and motivated people on the most-cutting edge questions. I also cannot forget to thank Eyleen, Justine, Natasha, Christian, Sean, Meng, Alex, Yan, Ying, Amaranath, Nic, Annie, Jonah and Chris - our equally awesome metabolism and Mars team. It's really a riveting eye-opening experience to learn and think about such diverse scientific questions in this lab, and I will carry on all the inspirations along with me to the next step of my career.

I would also like to thank my summer undergraduate student, Yun Jee Kang and high school student, Evan Green. To be honest, it took me quite some time to lay back and just let them do the experiment. Although this inevitably led to more failures and roundabout compared to my teaching every details hand-by-hand, the moments when they finally got the work done made me realize that it should just be like this in training a student. As a mentor, allowing students to try and make mistakes is perhaps even more important than providing guidance. As a student myself, it's a truly invaluable experience

to ‘mentor’ my students and came to realize how generous supports I have been received in this lab. I sincerely hope some day when I have a lab of my own, I can mentor my students in a way that is close to what Gary has done for me.

I want to thank my program advisor, Dr. Bjorn Olsen, who offered me this great opportunity to explore all these excitements at Harvard and provides us a free environment to pursue whatever we love to study. Our program is small (perhaps the smallest in Harvard, and is barely heard of by most students), but Dr. Olsen with other faculty and staff have always made the program a warm family for all of us.

Most of all, I want to thank my parents. I am so blessed to be growing in this loving family that is most respectful to knowledge: my parents are both engineers and my grandparents are university teachers and government officers. I always love to listen to stories of my parents’ and even grandparents’ generations, when the new China just began her strive for a communist society. For my parents, little is more critically fate changing than that College Entrance Examination in 1977, the first one after a ten-year-long suspension of college education during the ‘Cultural Revolution’. Lucky, they were among the 270 thousand students who got admitted, out of more than 5.7 million examinees aged from 15 to 36. Years later, my father went on pursue a Ph.D. degree, after several years’ of working experience. This was when he was 37 and I was in primary school. At that time in China, Ph.D. was a rare species that when my father came back to work, he was the only doctor in the entire company. Even today, I still remember so clearly how I proud I was and that I dreamed to become a doctor myself. Now this ambition is close to be accomplished. Sincerely, I would like to dedicate this doctoral thesis to my parents, for all the inspiration and supports for pursuing what I love to do.

CHAPTER ONE

An introduction to the small RNA pathways

History of discovery

How worm geneticists discovered the new world of small RNAs

The new world of small RNA silencing was first landed by worm geneticists during the study of developmental timing in the early 1990s. The nematode *Caenorhabditis elegans* was first introduced into the scientific community by Sydney Brenner in 1960s and has the remarkable feature of having an invariable cell lineage between individuals (Sulston and Horvitz, 1977; Sulston et al., 1983). Worm geneticists took this unique opportunity to be able to trace individual cells and isolated large number of mutant strains with abnormal cell lineage, to study how the signaling and cell-cell interaction guide cell fate determination.

In particular, a hierarchy of interesting mutants stood out that bear defects in the timing of cell fate specification. These so-called 'heterochronic' mutant animals have defects in the synchrony of developmental timing between tissues, leading to either precocious or delayed switching of cell division to differentiation in the hypodermal and vulval cells (Ambros, 1989; Ambros and Horvitz, 1984; Chalfie et al., 1981). Two groups of mutants, represented by the *lin-4* retarded mutant (Chalfie et al., 1981) and the *lin-14* precocious mutant (Ambros and Horvitz, 1987), suggest that the two opposing hierarchies of genes coordinate to regulate the cell fate transition. Further genetic analysis showed that *lin-4* acts upstream of *lin-14* through down-regulation of *lin-14* activity mediated on the *lin-14* 3' untranslated region (3'UTR) (Ruvkun et al., 1989; Ruvkun et al., 1991; Wightman et al., 1991). However, the biggest surprise came when the Ambros lab found that *lin-4* does not encode a protein; instead, it gives rise to some small RNAs (Lee et al., 1993). Through a legendary collaboration between the Ambros and Ruvkun lab, it was found by sequence comparison that the 3'UTR of *lin-14* contains

several conserved regions with partial sequence complementarity to *lin-4* (Wightman et al., 1993), immediately led to a beautiful model that the *lin-4* small RNAs bind to the *lin-14* 3'UTR and in this way down-regulate the translation of *lin-14*. Although illuminating, this model was out of most people's imagination and kept being considered to be something unique to worms, rather than a general principle of gene regulation.

Fortunately, more supportive evidence about the small RNA new world emerged in the following years. In one study, it was showed that 25 nucleotide (nt)- antisense RNAs were detected in plants undergoing posttranscriptional gene silencing (PTGS) (Hamilton and Baulcombe, 1999). Meanwhile, when worm researchers designed antisense RNAs to shut down specific mRNAs (Fire et al., 1991), a surprise came when they discovered that double-stranded RNAs (dsRNAs) to be much more potent than the antisense RNAs (Fire et al., 1998). A model that dsRNAs might function in a catalytic mechanism to target homologous mRNAs for degradation was then proposed (Montgomery et al., 1998). This phenomenon, termed 'RNA interference (RNAi)', is intimately related to the small RNA mediated PTGS in plants. Lastly, a second *C. elegans* small RNA, *let-7*, was identified (Reinhart et al., 2000). Like *lin-4*, *let-7* also regulates the developmental timing. More interestingly, *let-7* is highly conserved both at the sequence and temporal expression level implying an important conserved role in regulating temporal transition in animal development (Pasquinelli et al., 2000). Together, these series of works invoked the realization of small interfering RNAs as general mechanisms in regulating gene expression and opened up an entirely new field.

The myriad facets of small RNAs

Small non-coding RNAs, including microRNAs (miRNAs), piwi-interacting RNAs (piRNAs) and endogenous small interfering RNAs (endo-siRNAs), regulate developmental and defense pathways in animals. Each class of small RNAs has unique roles and genetic requirements but invariably bind to Argonaute proteins to form effector complexes that target nucleic acids containing partial or complete complementarity to the small RNA guide.

miRNAs are important nodes in the gene regulatory network

miRNAs are ~22 nt and repress gene expression through mRNA decay and translational repression (Bartel, 2004; Djuranovic et al., 2011). In animals, miRNAs cause repression by base pairing to the 3′ UTR of their target mRNAs, which contain perfect or near-perfect sequence complementarity to nucleotides 2-7 (the 'seed region') of miRNAs and mismatches and bulges in other parts of the miRNA-mRNA duplex (Lewis et al., 2003).

In general, miRNA is first transcribed as a primary miRNA (pri-miRNA) which is several hundreds nucleotides long and folds into a hairpin structure. Pri-miRNAs are then processed by the Drosha/DGCR8 nuclear microprocessor complex. In this complex, DGCR8 (also known as Pasha) orients Drosha to cut pri-miRNAs ~11 nt from the hairpin base, generating a shortened hairpin (named precursor miRNA, or pre-miRNA) that has a ~22 nt stem with 2 nt overhang at its 3' end (Han et al., 2006; Yeom et al., 2006). Pre-miRNAs are exported from the nucleus through Exportin-5, which recognizes the 2 nt 3' overhang of pre-miRNAs (Bohnsack et al., 2004; Yi et al., 2003). In the cytoplasm, the RNase III enzyme Dicer cuts away the loop of pre-miRNAs and yields a ~22 nt miRNA:miRNA* duplex (Hutvagner et al., 2001; Ketting et al., 2001). One strand is predominantly incorporated into the Argonaute protein to form the core miRNA-induced

silencing complex (miRISC) (Bartel, 2004); the other stand, usually quickly degraded and was called the miR star strand.

The biogenesis, mechanisms of miRNAs-mediated silencing, as well as miRNAs themselves are highly conserved during evolution. For example, the *let-7* miRNA is ultra-conserved both at the sequence and temporal expression level across the animal phylogeny (Pasquinelli et al., 2000). Although new miRNAs were continuously evolving, once integrated into gene regulatory network, they become highly conserved in sequence and are rarely lost (Meunier et al., 2013; Wheeler et al., 2009). This further supports a conserved important role of miRNA in the gene regulatory network.

But what's special about miRNAs, in compared to many other regulatory mechanisms in the gene regulatory network? In other words, why the evolution chose miRNAs for particular nodes in the gene network?

We think that one big virtue for a ~22 nt non-coding RNA, is that it can be produced much faster than a protein. This advantage becomes in particular crucial when the organisms need to make fast responses, for example, when dealing with cellular damages or making cell fate transitions. Indeed, among the few miRNAs in *C. elegans* with known functions, *lin-4*, *let-7* and *lsy-6* miRNAs all function in cell fate specification and are produced just before the fate transition (Cochella and Hobert, 2012; Feinbaum and Ambros, 1999; Reinhart et al., 2000). Second, miRNAs have the potential virtue of being reversible. It has been shown that miRNAs can target mRNAs enter into processing bodies (P-bodies) (Liu et al., 2005), which are distinct foci in the cytoplasm of eukaryotic cells containing enzymes for RNA decapping and degradation. On the other hand, not all mRNAs are degraded there: some mRNAs can be stored in P-bodies, and may even shuttle back to polysomes (Brengues et al., 2005). It was found that at least one miR-122 target, the cationic amino acid transporter 1 (CAT-1) mRNA, can re-locate

from P-bodies to polysomes in response to stress (Bhattacharyya et al., 2006). Although still lacking more convincing evidence at this stage, these studies opened up an appealing possibility that the action of miRNAs could be reversible under certain conditions. Lastly, miRNAs regulation is ideal to confer robustness and precision of gene expression. By looking at overrepresented motifs in the gene regulatory network, it was observed that miRNAs often function in the incoherent feedforward loop (FFL). The FFL consists of a master transcription factor regulating a miRNA and together with it, a group of target genes. Through math modeling and computational simulation, it was shown that this circuit is in particular robust to fluctuations in the upstream regulators, thus conferring stability and precision to the downstream gene expression program. Specifically, the optimal robustness requires a modest repression of target gene expression (Ebert and Sharp, 2012; Osella et al., 2011), which is exactly the feature of miRNA-mediated silencing.

In summary, miRNAs have the virtue of fast production, potential reversibility and conferring precision to the gene expression program. These unique features together make miRNAs integral component in the gene regulatory network and highly conserved during evolution.

piRNAs and genome defense

Piwi-interacting RNAs (piRNAs) are usually 26-31 nt in size, which associate with the PIWI clade Argonautes and are most highly expressed in the germline (Seto et al., 2007). piRNAs generally function in silencing of transposable elements (TEs) and other selfish DNA sequences and in this way protect the germline integrity (Siomi et al., 2011). However, growing evidence also suggest piRNAs have versatile functions even outside

of the germline, for example, in regulating synaptic plasticity (Rajasethupathy et al., 2012).

In contrast to miRNAs, the biogenesis pathway, sequences, and modes of action for piRNAs are much more diverse across animals. The production of piRNAs is still not fully understood, but it has been suggested that it involves a primary biogenesis and secondary amplification phase. In the primary phase, a long piRNA precursor transcript is made and then processed into the piRNA (Aravin et al., 2008; Brennecke et al., 2007) usually with a 5' uracil and a size that fits into the PIWI Argonaute (Kawaoka et al., 2011). Primary piRNA can bind and induce cleavage of target transcript at a position that is 10 nt from 5' end of the primary piRNA. This triggers the target transcript being routed into the piRNA pathway and generates a secondary piRNA with its 5' end produced by the cleavage event and its 3' end processed by some unknown nuclease(s). This secondary piRNA, in turn, targets the complementary transcripts and further routes it into the piRNA pathway. As a consequence, massive piRNAs are produced along with silencing of target transcripts. This amplification phase is therefore named 'Ping-Pong' cycle (Brennecke et al., 2007).

In contrast to several other animals, *C. elegans* piRNAs are 21 nt long (also called 21U-RNAs) (Batista et al., 2008; Das et al., 2008; Ruby et al., 2006). *C. elegans* piRNAs are first transcribed as individual precursors and processed by some unknown nucleases to give rise to mature piRNAs, similar to the primary phase in other species. However, nematodes lack the 'Ping-Pong' amplification phase. Instead, *C. elegans* piRNAs bind through imperfect complementarity to target transcripts and trigger secondary endo-siRNA production to further silence the target genes (Bagijn et al., 2012; Lee et al., 2012; Shirayama et al., 2012).

Because many features of piRNA pathways are not conserved in nematodes, *Drosophila* and mouse, it is interesting to learn whether piRNA sequences, biogenesis and mechanisms of action could be conserved within much shorter evolutionary distances, for example, among closely related nematode species. This is the focus of our study discussed in Chapter II.

The dramatic expansion of endo-siRNA pathways in nematodes

Endogenous small interfering RNAs (endo-siRNAs) are produced from double-stranded RNAs (dsRNAs) and mediate gene silencing post-transcriptionally and/or via chromatin modification. They share many features and protein cofactors with the exogenous RNAi pathway. Historically, endo-siRNAs are largely thought to be restricted to *C. elegans*. Indeed, in *C. elegans*, endo-siRNA is the predominant small RNA species judged by either abundance or sequence diversity. This is likely due to the presence of several RNA-dependent RNA polymerases (RdRPs) in *C. elegans*. RdRP catalyzes the replication of RNA from an RNA template and in this way produces dsRNAs, which are then made into endo-siRNAs. However, it is becoming recognized that many more species, such as *Drosophila* (Czech et al., 2008; Ghildiyal et al., 2008; Kawamura et al., 2008; Okamura et al., 2008) and mouse (Tam et al., 2008; Watanabe et al., 2008), also possess the endo-siRNA pathways even in the absence of clear RdRP homologs. In these species, endo-siRNAs are produced from dsRNAs resulting from convergent transcription, or transcripts that form stem-loop structures.

Besides RdRPs, the *C. elegans* genome also encodes an expanding family of worm-specific Argonaute proteins (WAGO) (Figure 1.1). Out of the 25 Argonautes in *C. elegans*, 16 belong to the WAGO clade (Yigit et al., 2006) and associate with various classes of endo-siRNAs.

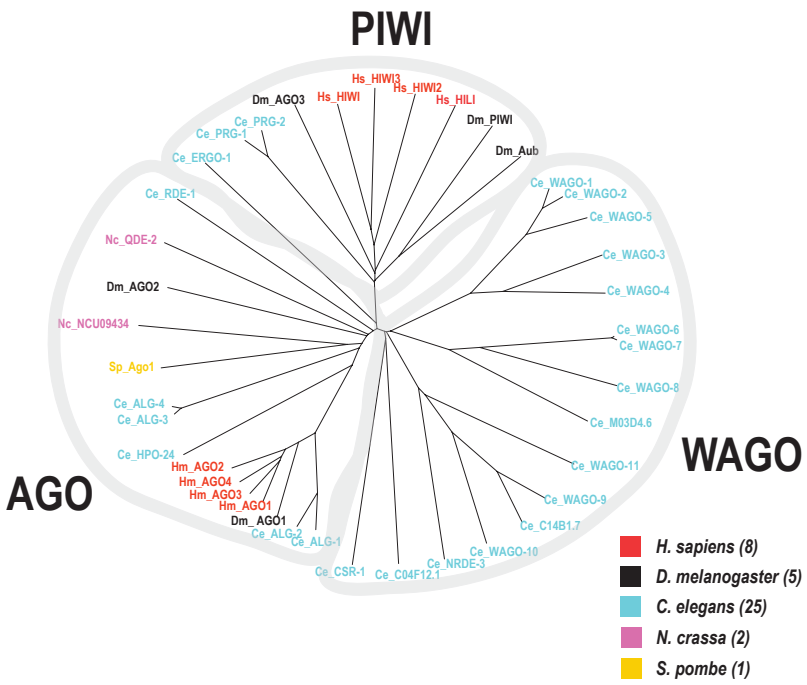


Figure 1.1 Phylogenetic tree of Argonaute proteins. Argonautes from different species are indicated in different colors, and the number of Argonautes in each genome is shown in the parenthesis. From a *C. elegans* centric point of view, the Argonaute family can be divided into three clades: the AGO, PIWI and WAGO clade.

The majority of *C. elegans* endo-siRNAs are either 22 or 26 nt long and start with 5' guanine, and are thus referred to as 22G siRNAs (22G-RNAs) or 26G siRNAs (26G-RNAs), respectively. 22G siRNAs are produced by either of the two RdRPs, RRF-1 and EGO-1, and bind to the WAGO clade Argonautes (WAGO-1-12) to silence certain protein-coding genes, transposons, pseudogenes and cryptic loci (Gu et al., 2009). A subset of 22G siRNAs produced by EGO-1 associate with the Argonaute CSR-1 (CSR-1 class siRNAs) to guide chromosome segregation (Claycomb et al., 2009; van Wolfswinkel et al., 2009). It was proposed that CSR-1 class siRNAs may also provide a

memory of self to protect endogenous genes from being routed into the piRNA and WAGO class siRNA pathways (Lee et al., 2012; Shirayama et al., 2012).

26G siRNAs fall into two classes: a spermatogenesis-enriched class which associate with the AGO clade Argonautes, ALG-3 and ALG-4 (Conine et al., 2010; Han et al., 2009); and an oocyte and embryo-enriched class which associate with the divergent PIWI-clade Argonaute, ERGO-1 (Fischer et al., 2011; Han et al., 2009; Vasale et al., 2010). Both classes of 26G siRNAs are produced by the RdRP RRF-3 and are thought to trigger secondary 22G siRNA production. However, the majority of 22G siRNAs are produced independent of a 26G siRNA trigger. How these siRNAs are produced remains unknown.

Dissertation overview

After ~20 years of research, many small RNA silencing protein cofactors were identified and many details of each small RNA pathway have been worked out. On the other hand, much more is to be learned from a dynamic and quantitative perspective to reveal the underlying mechanisms and designing principles of the pathway. In this dissertation, I present studies that examine the temporal dynamics of small RNA pathways – one from an evolutionary time scale among nematode species (Chapter II and III), which reveals an extraordinary sequence fluidity of piRNAs and endo-siRNAs during evolution. Another study of finely staged *C. elegans* during the first larval stage suggests a two-phase model of *lin-4* mediated silencing of *lin-14* (Chapter V). I also describe works identifying new cofactors functions in the miRNA pathway, potentially through regulating the spatial dynamics of the miRNA silencing complex (Chapter IV).

Reference

Ambros, V. (1989). A hierarchy of regulatory genes controls a larva-to-adult developmental switch in C. elegans. Cell *57*, 49-57.

Ambros, V., and Horvitz, H.R. (1984). Heterochronic mutants of the nematode Caenorhabditis elegans. Science *226*, 409-416.

Ambros, V., and Horvitz, H.R. (1987). The lin-14 locus of Caenorhabditis elegans controls the time of expression of specific postembryonic developmental events. Genes Dev *1*, 398-414.

Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T., and Hannon, G.J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Mol Cell *31*, 785-799.

Bagijn, M.P., Goldstein, L.D., Sapetschnig, A., Weick, E.M., Bouasker, S., Lehrbach, N.J., Simard, M.J., and Miska, E.A. (2012). Function, targets, and evolution of Caenorhabditis elegans piRNAs. Science *337*, 574-578.

Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell *116*, 281-297.

Batista, P.J., Ruby, J.G., Claycomb, J.M., Chiang, R., Fahlgren, N., Kasschau, K.D., Chaves, D.A., Gu, W., Vasale, J.J., Duan, S.*, et al.* (2008). PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans. Mol Cell *31*, 67-78.

Bhattacharyya, S.N., Habermacher, R., Martine, U., Closs, E.I., and Filipowicz, W. (2006). Relief of microRNA-mediated translational repression in human cells subjected to stress. Cell *125*, 1111-1124.

Bohnsack, M.T., Czaplinski, K., and Gorlich, D. (2004). Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. RNA *10*, 185-191.

Brengues, M., Teixeira, D., and Parker, R. (2005). Movement of eukaryotic mRNAs between polysomes and cytoplasmic processing bodies. Science *310*, 486-489.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell *128*, 1089-1103.

Chalfie, M., Horvitz, H.R., and Sulston, J.E. (1981). Mutations that lead to reiterations in the cell lineages of C. elegans. Cell *24*, 59-69.

Claycomb, J.M., Batista, P.J., Pang, K.M., Gu, W., Vasale, J.J., van Wolfswinkel, J.C., Chaves, D.A., Shirayama, M., Mitani, S., Ketting, R.F.*, et al.* (2009). The Argonaute CSR-1 and its 22G-RNA cofactors are required for holocentric chromosome segregation. Cell *139*, 123-134.

Cochella, L., and Hobert, O. (2012). Embryonic priming of a miRNA locus predetermines postmitotic neuronal left/right asymmetry in C. elegans. Cell *151*, 1229-1242.

Conine, C.C., Batista, P.J., Gu, W., Claycomb, J.M., Chaves, D.A., Shirayama, M., and Mello, C.C. (2010). Argonautes ALG-3 and ALG-4 are required for spermatogenesis-specific 26G-RNAs and thermotolerant sperm in Caenorhabditis elegans. Proc Natl Acad Sci U S A.

Czech, B., Malone, C.D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J.A., Sachidanandam, R.*, et al.* (2008). An endogenous small interfering RNA pathway in Drosophila. Nature *453*, 798-802.

Das, P.P., Bagijn, M.P., Goldstein, L.D., Woolford, J.R., Lehrbach, N.J., Sapetschnig, A., Buhecha, H.R., Gilchrist, M.J., Howe, K.L., Stark, R.*, et al.* (2008). Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the Caenorhabditis elegans germline. Mol Cell *31*, 79-90.

Djuranovic, S., Nahvi, A., and Green, R. (2011). A parsimonious model for gene regulation by miRNAs. Science *331*, 550-553.

Ebert, M.S., and Sharp, P.A. (2012). Roles for microRNAs in conferring robustness to biological processes. Cell *149*, 515-524.

Feinbaum, R., and Ambros, V. (1999). The timing of lin-4 RNA accumulation controls the timing of postembryonic developmental events in Caenorhabditis elegans. Dev Biol *210*, 87-95.

Fire, A., Albertson, D., Harrison, S.W., and Moerman, D.G. (1991). Production of antisense RNA leads to effective and specific inhibition of gene expression in C. elegans muscle. Development *113*, 503-514.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806-811.

Fischer, S.E., Montgomery, T.A., Zhang, C., Fahlgren, N., Breen, P.C., Hwang, A., Sullivan, C.M., Carrington, J.C., and Ruvkun, G. (2011). The ERI-6/7 helicase acts at the first stage of an siRNA amplification pathway that targets recent gene duplications. PLoS Genet *7*, e1002369.

Ghildiyal, M., Seitz, H., Horwich, M.D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E.L., Zapp, M.L., Weng, Z.*, et al.* (2008). Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science *320*, 1077-1081.

Gu, W., Shirayama, M., Conte, D., Jr., Vasale, J., Batista, P.J., Claycomb, J.M., Moresco, J.J., Youngman, E.M., Keys, J., Stoltz, M.J.*, et al.* (2009). Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the C. elegans germline. Mol Cell *36*, 231-244.

Hamilton, A.J., and Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science *286*, 950-952.

Han, J., Lee, Y., Yeom, K.H., Nam, J.W., Heo, I., Rhee, J.K., Sohn, S.Y., Cho, Y., Zhang, B.T., and Kim, V.N. (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell *125*, 887-901.

Han, T., Manoharan, A.P., Harkins, T.T., Bouffard, P., Fitzpatrick, C., Chu, D.S., Thierry-Mieg, D., Thierry-Mieg, J., and Kim, J.K. (2009). 26G endo-siRNAs regulate spermatogenic and zygotic gene expression in Caenorhabditis elegans. Proc Natl Acad Sci U S A *106*, 18674-18679.

Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T., and Zamore, P.D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science *293*, 834-838.

Kawamura, Y., Saito, K., Kin, T., Ono, Y., Asai, K., Sunohara, T., Okada, T.N., Siomi, M.C., and Siomi, H. (2008). Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. Nature *453*, 793-797.

Kawaoka, S., Izumi, N., Katsuma, S., and Tomari, Y. (2011). 3' End Formation of PIWI-Interacting RNAs In Vitro. Mol Cell *43*, 1015-1022.

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.

Lee, H.C., Gu, W., Shirayama, M., Youngman, E., Conte, D., Jr., and Mello, C.C. (2012). C. elegans piRNAs Mediate the Genome-wide Surveillance of Germline Transcripts. Cell.

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.

Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. Cell *115*, 787-798.

Liu, J., Valencia-Sanchez, M.A., Hannon, G.J., and Parker, R. (2005). MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. Nat Cell Biol *7*, 719-723.

Meunier, J., Lemoine, F., Soumillon, M., Liechti, A., Weier, M., Guschanski, K., Hu, H., Khaitovich, P., and Kaessmann, H. (2013). Birth and expression evolution of mammalian microRNA genes. Genome Res *23*, 34-45.

Montgomery, M.K., Xu, S., and Fire, A. (1998). RNA as a target of double-stranded RNA-mediated genetic interference in Caenorhabditis elegans. Proc Natl Acad Sci U S A *95*, 15502-15507.

Okamura, K., Chung, W.J., Ruby, J.G., Guo, H., Bartel, D.P., and Lai, E.C. (2008). The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. Nature *453*, 803-806.

Osella, M., Bosia, C., Cora, D., and Caselle, M. (2011). The role of incoherent microRNA-mediated feedforward loops in noise buffering. PLoS Comput Biol *7*, e1001101.

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

Rajasethupathy, P., Antonov, I., Sheridan, R., Frey, S., Sander, C., Tuschl, T., and Kandel, E.R. (2012). A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity. Cell *149*, 693-707.

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.

Ruby, J.G., Jan, C., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H., and Bartel, D.P. (2006). Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. Cell *127*, 1193-1207.

Ruvkun, G., Ambros, V., Coulson, A., Waterston, R., Sulston, J., and Horvitz, H.R. (1989). Molecular genetics of the Caenorhabditis elegans heterochronic gene lin-14. Genetics *121*, 501-516.

Ruvkun, G., Wightman, B., Burglin, T., and Arasu, P. (1991). Dominant gain-of-function mutations that lead to misregulation of the C. elegans heterochronic gene lin-14, and the evolutionary implications of dominant mutations in pattern-formation genes. Dev Suppl *1*, 47-54.

Seto, A.G., Kingston, R.E., and Lau, N.C. (2007). The coming of age for Piwi proteins. Mol Cell *26*, 603-609.

Shirayama, M., Seth, M., Lee, H.C., Gu, W., Ishidate, T., Conte, D., Jr., and Mello, C.C. (2012). piRNAs Initiate an Epigenetic Memory of Nonself RNA in the C. elegans Germline. Cell.

Siomi, M.C., Sato, K., Pezic, D., and Aravin, A.A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. Nat Rev Mol Cell Biol *12*, 246-258.

Sulston, J.E., and Horvitz, H.R. (1977). Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. Dev Biol *56*, 110-156.

Sulston, J.E., Schierenberg, E., White, J.G., and Thomson, J.N. (1983). The embryonic cell lineage of the nematode Caenorhabditis elegans. Dev Biol *100*, 64-119.

Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M.*, et al.* (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature *453*, 534-538.

van Wolfswinkel, J.C., Claycomb, J.M., Batista, P.J., Mello, C.C., Berezikov, E., and Ketting, R.F. (2009). CDE-1 affects chromosome segregation through uridylation of CSR-1-bound siRNAs. Cell *139*, 135-148.

Vasale, J.J., Gu, W., Thivierge, C., Batista, P.J., Claycomb, J.M., Youngman, E.M., Duchaine, T.F., Mello, C.C., and Conte, D., Jr. (2010). Sequential rounds of RNA-dependent RNA transcription drive endogenous small-RNA biogenesis in the ERGO-1/Argonaute pathway. Proc Natl Acad Sci U S A.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T.*, et al.* (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature *453*, 539-543.

Wheeler, B.M., Heimberg, A.M., Moy, V.N., Sperling, E.A., Holstein, T.W., Heber, S., and Peterson, K.J. (2009). The deep evolution of metazoan microRNAs. Evol Dev *11*, 50-68.

Wightman, B., Burglin, T.R., Gatto, J., Arasu, P., and Ruvkun, G. (1991). Negative regulatory sequences in the lin-14 3'-untranslated region are necessary to generate a temporal switch during Caenorhabditis elegans development. Genes Dev *5*, 1813-1824.

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.

Yeom, K.H., Lee, Y., Han, J., Suh, M.R., and Kim, V.N. (2006). Characterization of DGCR8/Pasha, the essential cofactor for Drosha in primary miRNA processing. Nucleic Acids Res *34*, 4622-4629.

Yi, R., Qin, Y., Macara, I.G., and Cullen, B.R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev *17*, 3011-3016.

Yigit, E., Batista, P.J., Bei, Y., Pang, K.M., Chen, C.C., Tolia, N.H., Joshua-Tor, L., Mitani, S., Simard, M.J., and Mello, C.C. (2006). Analysis of the C. elegans Argonaute family reveals that distinct Argonautes act sequentially during RNAi. Cell *127*, 747-757.

CHAPTER TWO

Evolution of the piRNA pathway in nematodes

AUTHOR CONTRIBUTIONS

This study was initially motivated by an interspecies small RNA surveillance hypothesis proposed by Gary Ruvkun. I performed all experiments and analyzed the data. Tai Montgomery provided suggestions on the small RNA analysis pipeline and data interpretation. The construction of the small RNA analysis pipeline was a collaborative effort by myself and Yan Qi. Part of this chapter is adapted from our paper “High-throughput sequencing reveals extraordinary fluidity of miRNA, piRNA and siRNA pathways in nematodes” (Shi et al., 2013).

Summary

Nematodes contain each of the broad classes of eukaryotic small RNAs, including miRNAs, endo-siRNAs and piRNAs. To better understand the evolution of these regulatory RNAs, I deep sequenced small RNA from *C. elegans* and three other closely related nematode species. Using a comparative genomics approach, I examined the conservation and evolution of each class of the small RNAs. There is no conservation of individual piRNA sequences. However, many features such as their genomic distribution, expression patterns, and tendency for piRNAs to trigger secondary siRNA production are highly conserved. We show that nematodes produce two distinct sex-specific classes of piRNAs, suggesting different roles for piRNAs in male and female germlines.

Motivating Questions

The Piwi-interacting RNA (piRNA) pathway maintains silencing of cryptic DNA sequences and in this way protects the animal germline from invading viruses and transposable elements (TEs). The establishment of immunity to new invasive TEs occurs by the incorporating of their sequences into the piRNA pool that get inherited primarily maternally, and further propagated in the following generations. In particular, works in plants and *Drosophila* have suggested that in the interspecies hybrid progeny, TEs from the maternal genome can be properly silenced by the maternal-inherited piRNAs targeting the TEs. However, some paternally inherited TEs that are absent from the maternal genome and piRNA pool becomes derepressed, consistent with the role of maternally inherited piRNAs in silencing TEs. Importantly, the derepression of paternal TEs could contribute to hybrid sterility or inviability (Blumenstiel and Hartl, 2005; Ha et al., 2009; Ish-Horowicz, 1982; Josefsson et al., 2006; Kelleher et al., 2012; Rozhkov et al., 2010).

On the other hand, small RNAs not only target TEs. We hypothesized that in the interspecies hybrid progeny, many other paternal transcripts may also be recognized as foreign and become targeted by maternal-inherited small RNAs. These paternal transcripts then could be routed into the small RNA pathways. If this hypothesis is correct, we were expecting to detect new small RNAs in the interspecies hybrid progeny, which are absent in either of its parents.

Deep sequencing small RNA in interspecies hybrid nematodes

In the *Caenorhabditis* genus, *C. elegans*, *C. briggsae*, *C. remanei* and *C. brenneri* are four species that are most commonly studied. Each of the four species is morphologically similar; however, their genomic sequences are highly divergent, with common ancestry ~110 million generations ago (Cutter et al., 2009) (Figure 2.1). The reproductive isolation between these four species primarily occur either before fertilization or during embryogenesis (Baird and Yen, 2000) (Table 2.1).

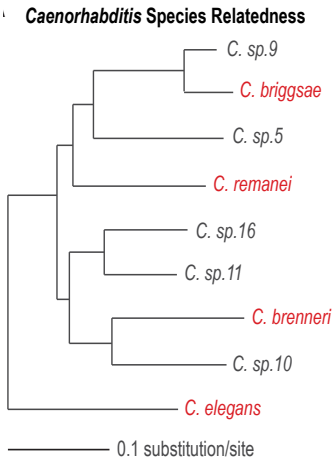


Figure 2.1 Phylogenetic relationship of the *elegans* group of *Caenorhabditis*, adapted from (Kiontke et al., 2011). *C. briggsae*, *C. remanei*, *C. brenneri*, and *C. elegans* (highlighted in red), having had their genomes sequenced and partially annotated, were chosen for this study.

Table 2.1 Reproductive isolation in several *Caenorhabditis* species

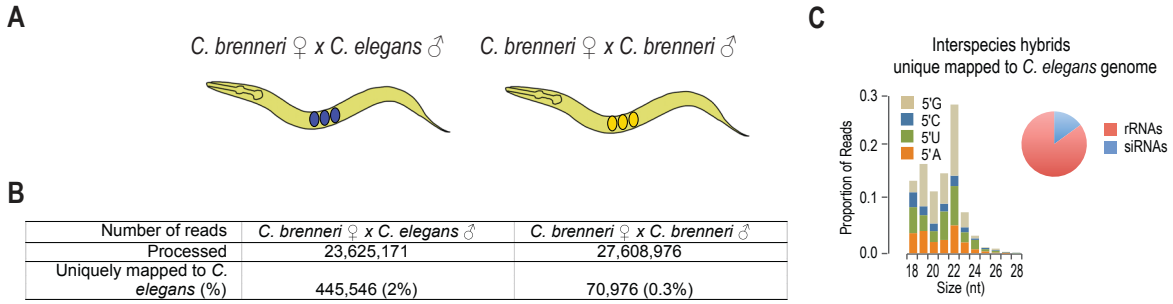
Mating	*C. elegans* ♂	*C. briggsae* ♂	*C. remanei* ♂	*C. brenneri* ♂
C. elegans ♀		No fertilization	No fertilization	No fertilization
C. briggsae ♀	No fertilization		Embryonic arrest	No fertilization
C. remanei ♀	Embryonic arrest	Embryonic arrest		No fertilization
C. brenneri ♀	Embryonic arrest	Embryonic arrest	Embryonic arrest	

We chose to deep sequence small RNAs from *C. brenneri* females crossed to *C. elegans* males. Because hand-picking worms for mating cannot be easily performed in large scale, I was not able to harvest embryos by the regular bleaching method. Instead, I collected ~100 gravid *C. brenneri* females each bearing several *C. brenneri - C. elegans* interspecies hybrid embryos. Then I extracted small RNAs from them and made a deep sequencing library, which will be referred to as 'hybrid library' for short. In parallel, *C. brenneri* females were crossed to *C. brenneri* males as a control (Figure 2.2A). This small RNA library will thus be referred to as 'control library'.

I obtained over 20 million small RNA reads from each library, 60-70% of which can be perfectly mapped to the *C. brenneri* (maternal) genome. Every small RNA that is derived from the *C. brenneri* (maternal) genome has almost the same relative abundance comparing the hybrid and control library. This suggests that mating to a different species does not change the small RNA profiles of *C. brenneri* females. I then aligned the rest of small RNA reads to the *C. elegans* (paternal) genome: ~2% of total small RNA from the hybrid library, compared to 0.3% from the control library are uniquely mapped to *C. elegans* genome (Figure 2.2B). The 70,976 small RNA reads in the control

library which can be uniquely mapped to the *C. elegans* genome could be due to contamination during library preparation or sequencing errors.

In contrast to a typical *C. elegans* embryos small RNA profile where most small RNAs are 21-23 nt starting with a 5' G or 5' U, the *C. elegans* unique small RNA reads from the interspecies hybrid library have a broad size distribution between 18-23 nt and a higher occurrence of 5' A. Surprisingly, 85% of these small RNAs are derived from ribosomal RNAs (rRNAs). This suggests that paternal ribosomes may become cleaved and degraded after fertilization. Indeed, a small RNA profiling study in carefully staged *C. elegans* early embryos have revealed a significant proportion of rRNA-derived small RNAs in 1 or 2-cell stage embryos, but not in later embryonic stages (Stoeckius et al., 2009). This turned out to be consistent with the fact that *C. brenneri* - *C. elegans* hybrid embryos are usually arrested at the 2 or 4-cell stage.



Number of reads	*C. brenneri* ♀ x *C. elegans* ♂	*C. brenneri* ♀ x *C. brenneri* ♂
Processed	23,625,171	27,608,976
Uniquely mapped to *C. elegans* (%)	445,546 (2%)	70,976 (0.3%)

Figure 2.2 Deep sequencing small RNAs in interspecies hybrid nematodes revealed paternal rRNA fragments in arrested hybrid embryos. (A) Diagram of *C. brenneri* females crossed to *C. elegans* males and bear interspecies hybrid embryos (shown in blue, left). *C. brenneri* females crossed to *C. brenneri* males (right) is the control experiment. (B) Overall statistics of small RNA libraries. About 2% of total small RNAs from the hybrid library are uniquely mapped to the *C. elegans* genome. (C) Shown is the size and 5' first nucleotide distribution of small RNAs that are uniquely mapped to the *C. elegans* genome in the hybrid library. Pie chart

(Figure 2.2 continued) shows that most small RNAs derived from the *C. elegans* genome are rRNA fragments.

Although I found a plausible explanation for our deep sequencing results, nothing particular interesting stood out. Since the hybrid embryos become immediately arrested, it hardly has any chance of making new small RNAs. As a result, this study did not provide any conclusive results in regard to our initial hypothesis. However, since I had all the small RNAs made from each of the four nematode species - as my control groups, I decided to study the small RNA pathways from an evolutionary perspective using a comparative genomics approach. The study was therefore transformed into this new direction, as presented below.

Deep sequencing small RNAs from four nematode species

For each species, I constructed 18- to 28-nt small RNA libraries from synchronized populations of young gravid adult hermaphrodites for the androdioecious (male/hermaphrodite) species *C. elegans* and *C. briggsae* and from mixed populations of adult males and females for the gonochoristic (male/female) species *C. remanei* and *C. brenneri*. I also sequenced small RNAs from synchronized early embryos and young adult males.

Annotation of piRNAs /21U-RNAs

C. elegans piRNAs (21U-RNAs) are 21 nt long and contain a 5' U. Previous studies have shown that a GTTTC core motif was strongly overrepresented and present in a limited region upstream of piRNAs in the *C. elegans* genome as well as several other nematode genomes (de Wit et al., 2009; Ruby et al., 2006). Therefore, only sequences that contain

a ‘GTTTC’ motif (allowing one mismatch at maximum) starting between -47 to -41 nt upstream of 21-nt small RNAs with a 5’ U were retained. Then, all these sequences were aligned based on position of the GTTTC core motif. To further detect nucleotide composition bias in these sequences, a position weight matrix (PWM) was calculated. I found that in each species, a large and small motif with a ~26-nt spacer which all highly resemble the *C. elegans* annotated piRNA upstream motifs were indeed strongly overrepresented. I then filtered the entire dataset of 21-nt small RNA starting with 5’ U, looking for sequences that have the motifs and define them as piRNAs/21U-RNAs. To this end, a score matrix for large motif, one for small motif and one for the distance between the two motifs were derived based on the PWM.

For each nucleotide N at each position i, f_N was the foreground (observed) frequency of N, b_N was the background frequency of N, and P was the total number of counts. b_N was estimated to be: b_A=0.34, b_T=0.34, b_G=0.16, b_C=0.16

$$score_i = \log_2\left(\frac{f_N \cdot P + b_N \cdot \sqrt{P}}{b_N \cdot P + b_N \cdot \sqrt{P}}\right)$$

$$score_{motif} = \sum_{i=1}^{N} score_i$$

Similarly,

$$score_{spacer} = \sum_{1}^{7} \log_2\left(\frac{f_N \cdot P + b_N \cdot \sqrt{P}}{b_N \cdot P + b_N \cdot \sqrt{P}}\right)$$

f_N was the foreground frequency of certain length, b_N=1/7 assuming even probabilities of spacer length ranging from 17-23 nt.

Total score = $score_{large\_motif}$ + $score_{small\_motif}$ + $score_{spacer}$

Small RNAs having a total score ≥ 15.5 are defined as piRNAs/21U-RNAs. The nucleotide composition and length of spacer sequence between the large and small motifs upstream of piRNAs are highly conserved and are shown in Figure 2.3.

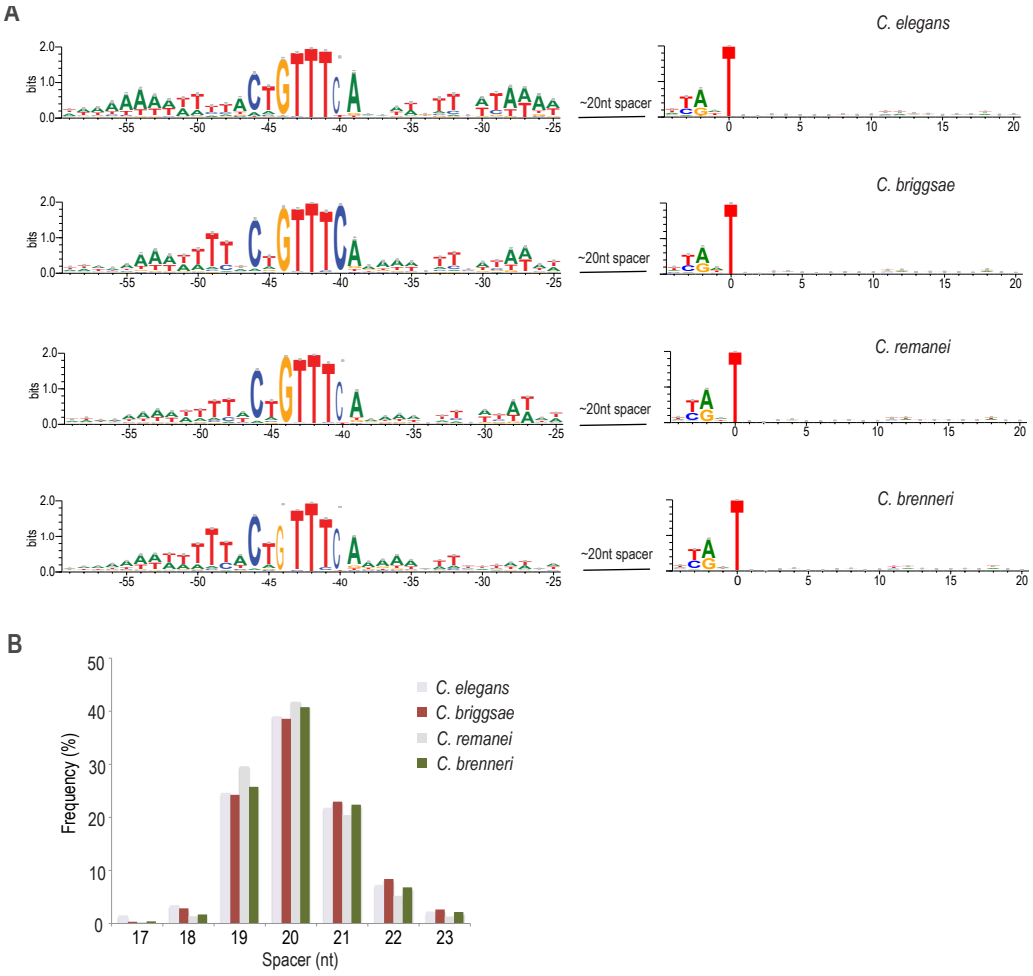


Figure 2.3 The upstream large and small motifs and distance between them are highly conserved in all four nematode species. (A) The large and small motifs upstream of piRNA loci, plotted as a sequence logo (Crooks et al., 2004), are nearly identical in all four species. (B) The spacer between large and small motifs is predominately 19-21 nt in all four species.

piRNA loci as reservoir to initiate genome-wide surveillance

The numbers of both predicted and detected piRNAs in our deep-sequencing library

were similar between *C. elegans*, with 17.6K predicted and 9.9K detected, and *C. briggsae*, with 14.8K predicted and 7.5K detected. On the other hand, *C. remanei* and *C. brenneri* possess far more piRNAs than *C. elegans* and *C. briggsae*. In *C. remanei*, 33.8K piRNAs were predicted and 23.7K were detected, and in *C. brenneri*, 54.8K piRNAs were predicted and 29.8K were detected (Table 2.2).

Table 2.2 piRNA pool sizes correlate with levels of nucleotide polymorphism

	C. elegans	*C. briggsae*	*C. remanei*	*C. brenneri*
Predicted piRNAs	17.6K	14.8K	33.8K	54.8K
Detected piRNAs	9.9K	7.5K	23.7K	29.8K
Genome Size (Mb)	100	~100	~135	~135
Neutral nucleotide site diversity	Low	Low	0.05	0.14

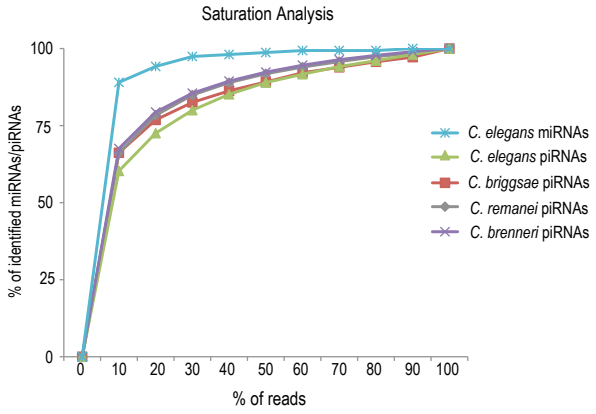


Figure 2.4 piRNA saturation analysis. The percentage of the total number of piRNAs identified in each library using a random subset of 10-100% of all sequencing reads. *C. elegans* miRNAs are shown for comparison.

A saturation analysis indicates that we are approaching saturation at similar speeds for piRNAs in each species. At the current sequencing depth, we have captured the majority of piRNAs produced at these particular developmental stages (Figure 2.4).

The genomes of *C. remanei* and *C. brenneri* are ~35% larger than those of *C. elegans* and *C. briggsae*; however, this alone does not account for the 2-3 times more piRNAs identified in *C. remanei* and *C. brenneri*.

Interestingly, *C. elegans* and *C. briggsae* are both androdioecious (male/hermaphrodite) species that reproduce primarily by selfing, and possess little natural heterozygosity. On the other hand, *C. remanei* and *C. brenneri* are both gonochoristic (male/female) species that mate at every generation and possess much higher natural heterozygosity (Barriere et al., 2009). *C. remanei* has an estimated neutral nucleotide site diversity of 0.05; which is much higher than human, mouse and *A. thaliana*, and slightly higher than *D. melanogaster*. Strikingly, polymorphism averages a 0.14 at synonymous sites in *C. brenneri*, which is the highest among all organisms tested (Cutter et al., 2013).

The interpretations are two folds. Because we are sequencing piRNAs in a population of animals, the sequence diversity can be attributed to either allelic polymorphism or the number of piRNA loci in the genome. To distinguish these two possibilities, I surveyed the nucleotide diversity of piRNA locus. If two piRNAs differ by only one or two nucleotides, they are more likely to be allelic. Indeed, ~40% piRNA in *C. remanei* differ by 1-2 nucleotides. On the other hand, 17%, 0% and 0% of piRNA differ by 1-2 nucleotide in *C. brenneri*, *C. elegans* and *C. briggsae* respectively, strongly suggests that the piRNA diversity predominantly reflects the number of piRNA loci in the genome in these other species. Currently, it is thought that piRNAs function to initiate genome-wide surveillance in the germline. In the light of this model, I therefore hypothesize that gonochoristic (male/female) species that mate at every generation are selected to keep a larger repertoire of piRNAs in order to defend against the greater diversity of paternal genome encountered during mating.

piRNAs share common features in their mechanism of action

To identify potential piRNA targets, we aligned piRNA sequences with all annotated protein-coding genes from each species. piRNA-target recognition is thought to be permissive to around three mismatches (Bagijn et al., 2012; Lee et al., 2012); thus we did the analysis allowing for 0, 1, 2, or 3 mismatches. When up to 3 mismatches are allowed, ~30% of *C. elegans* and ~20% of *C. briggsae* genes are potential targets (Figure 2.5A). Although *C. remanei* and *C. brenneri* contain a substantially larger repertoire of piRNAs, the proportions of genes with potential piRNA targets are similar to that of *C. elegans* (Figure 2.5A).

C. elegans piRNAs can trigger the production of RdRP-dependent secondary siRNAs centered on and antisense to piRNA target sites (Bagijn et al., 2012; Lee et al., 2012). To determine if piRNAs trigger siRNA formation in the other three nematodes, we assessed both sense and antisense siRNA abundance at candidate piRNA target sites. When all observed piRNAs were included in the analysis, there was only a slight enrichment of siRNAs at predicted piRNA target sites (data not shown). However, when only the top 20% most abundant piRNAs were considered, we observed a substantial enrichment of siRNAs antisense, but not sense, to the predicted piRNA target sites in each species (Figure 2.5B). Of these siRNAs, 70-80% are 22G siRNAs. These results suggest that, although individual piRNAs are not conserved, the mechanism in which they are formed and their propensity to trigger secondary siRNA formation are conserved (Bagijn et al., 2012; Lee et al., 2012).

C. elegans piRNAs are primarily derived from two broad clusters on chromosome IV (Ruby et al., 2006) (Figure 2.5C). We asked whether piRNA loci are also clustered in other species. We restricted our analysis to *C. briggsae* because *C. remanei* and *C.*

brenneri DNA sequences have not yet been assembled into chromosomes. There is extensive conservation of chromosome organization and synteny between *C. elegans* and *C. briggsae* (Hillier et al., 2007; Stein et al., 2003). In C. *briggsae*, the syntenic regions of the two major *C. elegans* piRNA clusters on chromosome IV also produced high levels of piRNAs (de Wit et al., 2009; Ruby et al., 2006) (Figure 2.5C). The two regions that give rise to the *C. elegans* piRNA clusters on chromosome IV are rearranged in *C. briggsae* such that they are separated from one another by only ~1 Mb. Interestingly, the ~1 Mb region separating the two clusters also contains a high abundance of piRNA loci. Together the region forms a continuous 6.9 Mb piRNA cluster. We also identified a second piRNA cluster on chromosome IV (13.1-15.1 Mb) and another on chromosome I (9.9-11.3 Mb) specific to *C. briggsae* (Figure 2.5C). The regions that give rise to these two *C. briggsae*-specific piRNA clusters lack continuous synteny with *C. elegans*, as determined by pairwise alignments. Two *C. briggsae* piRNA clusters previously identified on chromosomes I (7.8-9.5 Mb) and III (0-0.3 Mb) were represented by only average numbers of reads in our libraries and may have been artifacts of low sequencing depth (de Wit et al., 2009). That piRNAs in both *C. elegans* and *C. briggsae* tend to cluster and that conservation of these clusters appears to depend on long regions of continuous synteny suggests that the clusters are important for the birth of new piRNAs.

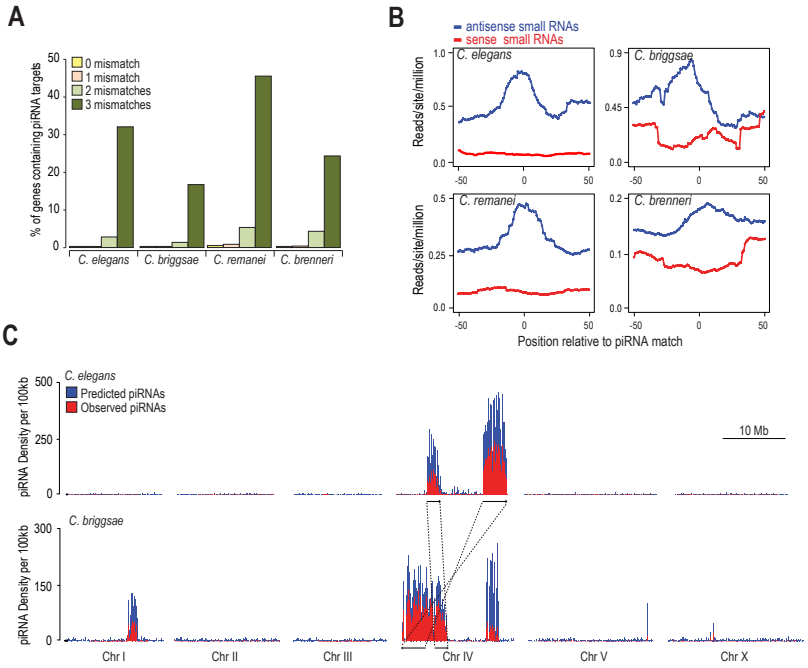


Figure 2.5 Functional and genomic features of piRNAs are conserved. (A) The percentage of protein-coding genes in each genome that could be targeted by piRNAs if 0, 1, 2 or 3 mismatches are allowed. (B) Density of small RNAs within a 100 nt window centered on the predicted target sites of the top 20% most abundant piRNAs. Small RNAs that are antisense to the predicted targets are shown in blue, and those that are sense to the targets are in red. (C) Distribution of observed (red) and predicted (blue) piRNA loci per 100 kb window in *C. elegans* (top) and *C. briggsae* (bottom). There are two piRNA clusters on *C. briggsae* chromosome IV: the 0-6.9 Mb region largely in synteny with the two *C. elegans* piRNA clusters (highlighted in lines with arrows); and the 13.1-15.1 Mb cluster. In addition, *C. briggsae* has a third piRNA cluster on chromosome I at position 9.9-11.3 Mb.

Two distinct classes of piRNAs in each species

In each of the four *Caenorhabditis* species we analyzed, we observed a highly significant positive correlation between piRNA populations in early embryos and adult

hermaphrodites or adult females+males (Figure 2.6A). In contrast, piRNA levels between adult males and adult hermaphrodites or adult females+males are only modestly correlated and show a biphasic pattern of distribution indicative of two distinct classes (Figure 2.6B). In *C. elegans*, we identified 9,307 distinct piRNAs in hermaphrodites and 6,065 distinct piRNAs in males. Of these, 5,044 were enriched >3 fold in hermaphrodites and 3,336 were enriched >3 fold in males. Only 1,493 piRNAs had similar expression levels in hermaphrodites and males (Figure 2.6C left). Each of the other species also had distinct sets of piRNAs that were enriched in either males or hermaphrodites/females+males (Figure 2.6C), suggesting that the production of distinct classes of male and female piRNAs is conserved in nematodes.

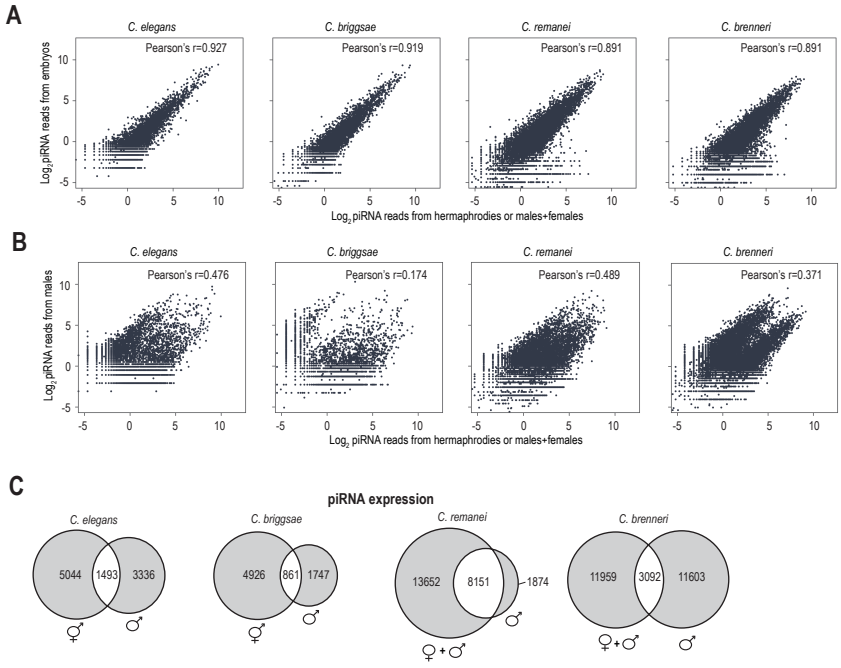


Figure 2.6 Two distinct classes of piRNAs in each species. (A) Scatter plots display the levels of piRNAs in adult hermaphrodites or adult females+males (x-axis) and early embryos (y-axis). (B) Scatter plots display the levels of piRNAs in adult hermaphrodites or adult

(Figure 2.6 continued) females+males (x-axis) and adult males (y-axis). (C) Venn diagrams show the number of piRNAs that are >three-fold enriched in hermaphrodites/females+males (left) or males (right). piRNAs shown in the overlapping section have similar expression levels in hermaphrodites/females and males.

Distinct trajectories of sex-specific piRNA evolutions in nematodes

In *C. briggsae*, *C. remanei* and *C. brenneri*, the female/hermaphrodite- or male-enriched piRNAs are localized in distinct clusters in the genome. The result of *C. briggsae* was shown (Figure 2.7A, right) because *C. remanei* and *C. brenneri* DNA sequences have not yet been assembled into chromosomes. In contrast, the genomic distributions of *C. elegans* hermaphrodite- and male-enriched piRNAs are similar, and these clusters are syntenic to the *C. briggsae* hermaphrodite-enriched piRNA cluster (Figure 2.7A, left). This suggests that *C. elegans* is likely to have lost the male-enriched piRNA clusters present in their common ancestor (Figure 2.7B).

To further identify features that could distinguish the two classes of piRNAs I analyzed the upstream motifs of each class. The upstream motifs and length of spacer between the large and small motifs are similar between the two classes of piRNAs, although, we did observe several positions in the large motif and surrounding sequence that show a stronger bias for a particular nucleotide in one class relative to the other (Figure 2.7C and D). The *C. elegans* male-enriched piRNAs have a near perfect conservation of the core 'GTTTC' motif, whereas the hermaphrodite-enriched piRNAs have a much more degenerate nucleotide preferences at the first G and last C position of the 'GTTTC' motif (Figure 2.7C). An independent work from John Kim lab showed that these differences in the large motif are sufficient to confer sex-specific expression of piRNAs (Billi et al., 2013). Interestingly, this difference does not distinguish the male

versus female/hermaphrodite-enriched piRNAs in the other three species, in which the biggest differences being the 'T' and 'A' immediate upstream and downstream of the core motif (Figure 2.7D). Importantly, both the male and female/hermaphrodite-enriched piRNAs in other three species have a near perfect conservation of the core 'GTTTC' motif, as the *C. elegans* male class of piRNAs.

Taken together these two pieces of data, I hypothesize that back in time, the common ancestor of these nematodes possessed distinct male and female piRNA clusters. As *C. elegans* evolved to a hermaphroditic species, it lost the male-enriched piRNA cluster. On the other hand, its ancient female-enriched piRNA clusters evolved to a mixed male- and hermaphrodite-enriched piRNAs through nucleotide degeneration.

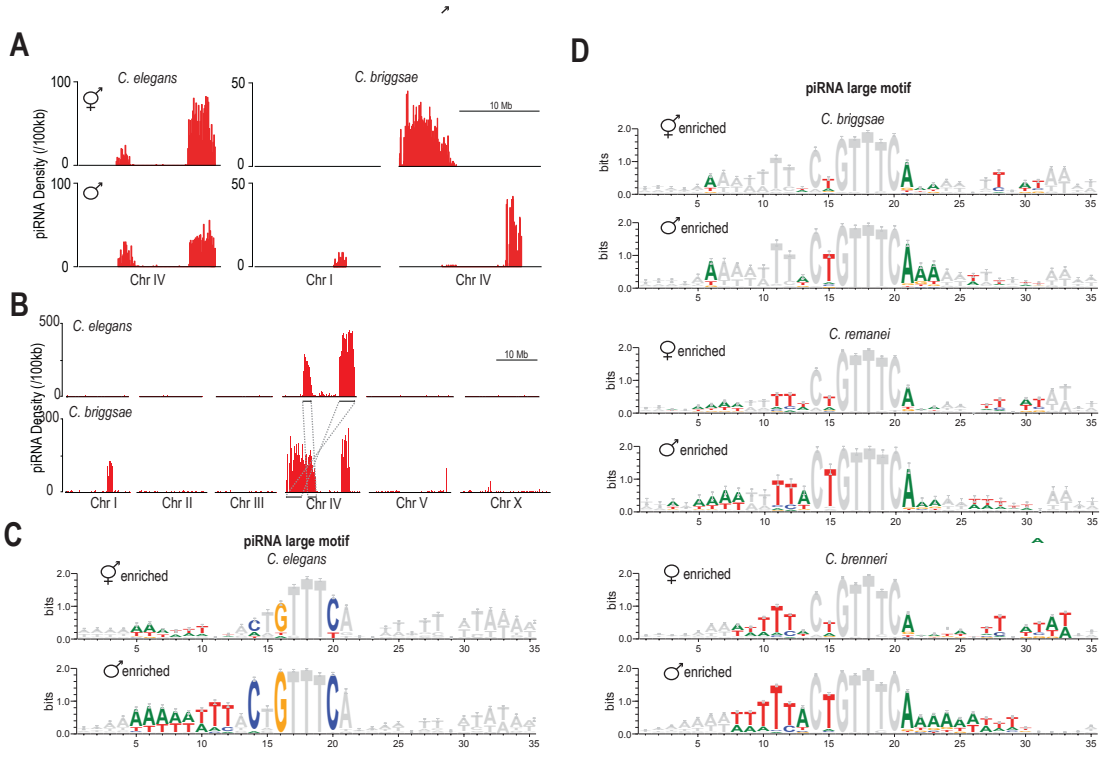


Figure 2.7 Distinct evolutionary trajectories of female/hermaphrodite- or male-enriched piRNAs. (A) Distribution of hermaphrodite or male-enriched piRNA loci per 100 kb in *C. elegans* and *C. briggsae*. (B) Distribution of piRNA loci per 100 kb window in *C. elegans* (top) and

(Figure 2.7 continued) *C. briggsae* (bottom). The two male-specific piRNA clusters in *C. briggsae* and are not present in *C. elegans*. (C, D) Sequence logos (Crooks et al., 2004) display nucleotide conservation in the large motifs upstream of hermaphrodite/female- or male-enriched piRNAs loci in *C. elegans* (C) and the other three species (D). Positions that have a substantially different weight matrix between hermaphrodites/females and males are colored.

Evolution of piRNA sequences

Although we identified nearly 70,000 piRNAs in our deep sequencing data sets in total, not a single piRNA sequence was present in more than one species. We also assessed piRNA sequence conservation when one, two, or three mismatches were allowed. Even with this less stringent criterion, only 0%, 0.01%, and 0.1% of *C. elegans* piRNAs have potential homologs in *C. briggsae* allowing for one, two, and three mismatches, respectively. It is possible that the sheer number of piRNAs (likely >15,000 in each species) and tolerance for multiple mismatches relax the sequence conservation of individual piRNA.

On the other hand, since many piRNAs can target TEs, I asked whether piRNAs were even under the positive selection, due to a potential arms race between piRNAs and TEs. To address this possibility, I surveyed the conservation of piRNA sequences among the *C. elegans* wild isolates between which there was a much shorter evolutionary distance. Specifically, I surveyed over 182,000 single-nucleotide polymorphisms (SNPs), 45596 small indels, 1116 medium size indels and 166 long indels between the common lab strain *C. elegans* Bristol N2 and the Hawaiian mapping strain CB4856 (Swan et al., 2002; Wicks et al., 2001) (R Waterston, personal communication). In the *C. elegans* N2 strain, there are 15,366 piRNAs. If the occurrence of SNPs within piRNA sequences were similar to what is expected by chance, this would

support neutral selection. On the other hand, if the occurrence of SNPs were significantly higher/lower than expected number, it would support a positive/negative selection model.

$$piRNA\_SNPs_{\exp ected\_by\_chance} = 182{,}000 \times \frac{15{,}366 \times 21}{100{,}000{,}000} = 587.3$$

In reality, there are 720 SNPs in piRNA sequences, which is not significantly higher than 587, the expected number. In addition, the position of SNPs within a piRNA is not biased toward any position. Further, the types of nucleotide substitution are mostly A-G and C-T; and the frequency of each type of substitution in piRNA sequences is not significantly different from other regions of the genome. Finally, the piRNAs bearing SNPs have a medium expression level among all piRNAs. Together, this does not support the model of positive selection of piRNAs during the *C. elegans* evolution. Rather, piRNA sequences seem to have drifted neutrally in *C. elegans*.

Future works

Stepping back to where we have come from: we are still intrigued in the potential involvement of small RNAs in speciation. As new *Caenorhabditis* species are continuously being discovered, it opens up the door for us to re-visit this question. Two *Caenorhabditis* species pairs capable of producing fertile hybrid progeny were recently described (Dey et al., 2012; Woodruff et al., 2010) and have made the *Caenorhabditis* a model system to study speciation. Furthermore, as many more species with much shorter evolutionary distances are sequenced, a survey of piRNA evolution now can be done with much higher resolution - again, at an evolutionary time scale. Specifically, a comparison between the genochoristic *C. sp. 9* and androdioecious *C. briggsae* would be very informative to investigate the evolution of piRNAs and hermaphroditism. In

collaboration with Asher Cutter lab (University of Toronto), we will deep sequence small RNA from *C. sp. 9*, *C. briggsae*, and the F1 interspecies hybrid animals (Figure 2.8).

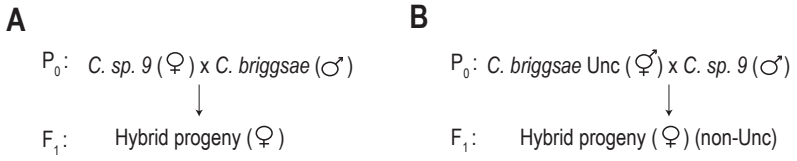


Figure 2.8 Experimental design of deep sequencing small RNAs from *C. briggsae* and *C. sp. 9* interspecies hybrid progeny. (A, B) The parental strains and F1 interspecies hybrid female animals will be collected and subject to small RNA deep sequencing. (B) In the reciprocal cross, *C. briggsae unc-119(nm67)* strain will be used, to distinguish *C. briggsae* self-progeny to the interspecies hybrid progeny (non-Unc).

Specifically, I propose to ask the following questions: (1) How piRNAs were lost and new piRNAs were born by the alignment of two genomes. (2) Whether piRNA pathways plays a role in speciation. (3) Whether there are any maternal effect or imprinting on piRNA expression.

Methods

Nematode strains

Nematode strains used in this study: *C. elegans* N2, *C. briggsae* AF16, *C. remanei* PB4641, and *C. brenneri* PB2801. Worms were cultured with bacterial strain OP50 on modified nematode growth medium (Andersen et al., 2012) containing 1% agar and 0.7% agarose to prevent burrowing of *C. brenneri*. All strains were grown at 20°C.

High-throughput sequencing and data analysis

For the construction of adult hermaphrodite/male+female, male or embryo smallRNA libraries, animals were grown at 20°C for 72–74 h post-L1 synchronization and harvested as day one gravid adult hermaphrodites for *C. elegans* and *C. briggsae* and mixed populations of adult males and females for *C. remanei* and *C. brenneri*. For male isolation, 150-200 day one adult males were handpicked from the plate. Embryos were harvested by bleach treatment of ~15,000 gravid adults. Total RNA was isolated by dounce homogenization of worms in TRI Reagent, followed by chloroform extraction and isopropanol precipitation. Small RNA high-throughput sequencing libraries were prepared as described (Montgomery et al., 2012). Briefly, 18- to 28-nt small RNAs were size-selected and treated with 20 U Tobacco Acid Phosphatase (Epicenter) at 37°C for 2 h to digest 5' tri- and diphosphates to monophosphates. Small RNAs were then ligated to the 3' adapter using T4 RNA ligase 2 truncated (NEB) for 16 h at 16°C. 5' ligations were done with T4 RNA ligase 1 (NEB) for 16 h at 16°C. Adapterligated RNAs were then reverse-transcribed and PCR-amplified using Illumina's TruSeq RNA Indexing PCR primers. Small RNA amplicons were size-selected by gel purification and subjected to Illumina HiSeq sequencing. Small RNA sequences were parsed using a custom Python program to remove adapter sequences and then mapped to the corresponding nematode reference genome (WormBase release WS230) allowing for 0 mismatches using Bowtie software (Langmead et al., 2009). For sequences mapping to multiple genomic loci, the total number of reads was divided by the number of genomic loci. Small RNA reads were then normalized to the total number of millions of mapped reads (i.e., reads per million)

piRNA/21U-RNAs annotation

piRNAs/21U-RNAs were predicted as described (Ruby et al., 2006), using a scoring matrix based on the consensus motif, spacer sequence length, and 5' U features of piRNAs. piRNA saturation analysis was performed by taking a random subset of sequencing reads with increasing size and calculating the percentage of piRNAs identified from this sublibrary.

Data access

All high-throughput sequencing data have been submitted to the NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE41461.

Reference

Andersen, E.C., Gerke, J.P., Shapiro, J.A., Crissman, J.R., Ghosh, R., Bloom, J.S., Felix, M.A., and Kruglyak, L. (2012). Chromosome-scale selective sweeps shape Caenorhabditis elegans genomic diversity. Nat Genet *44*, 285-290.

Bagijn, M.P., Goldstein, L.D., Sapetschnig, A., Weick, E.M., Bouasker, S., Lehrbach, N.J., Simard, M.J., and Miska, E.A. (2012). Function, targets, and evolution of Caenorhabditis elegans piRNAs. Science *337*, 574-578.

Baird, S.E., and Yen, W.C. (2000). Reproductive isolation in Caenorhabditis: terminal phenotypes of hybrid embryos. Evol Dev *2*, 9-15.

Barriere, A., Yang, S.P., Pekarek, E., Thomas, C.G., Haag, E.S., and Ruvinsky, I. (2009). Detecting heterozygosity in shotgun genome assemblies: Lessons from obligately outcrossing nematodes. Genome Res *19*, 470-480.

Billi, A.C., Freeberg, M.A., Day, A.M., Chun, S.Y., Khivansara, V., and Kim, J.K. (2013). A conserved upstream motif orchestrates autonomous, germline-enriched expression of Caenorhabditis elegans piRNAs. PLoS Genet *9*, e1003392.

Blumenstiel, J.P., and Hartl, D.L. (2005). Evidence for maternally transmitted small interfering RNA in the repression of transposition in Drosophila virilis. Proc Natl Acad Sci U S A *102*, 15965-15970.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res *14*, 1188-1190.

Cutter, A.D., Dey, A., and Murray, R.L. (2009). Evolution of the Caenorhabditis elegans genome. Mol Biol Evol *26*, 1199-1234.

Cutter, A.D., Jovelin, R., and Dey, A. (2013). Molecular hyperdiversity and evolution in very large populations. Mol Ecol.

de Wit, E., Linsen, S.E., Cuppen, E., and Berezikov, E. (2009). Repertoire and evolution of miRNA genes in four divergent nematode species. Genome Res *19*, 2064-2074.

Dey, A., Jeon, Y., Wang, G.X., and Cutter, A.D. (2012). Global population genetic structure of Caenorhabditis remanei reveals incipient speciation. Genetics *191*, 1257-1269.

Ha, M., Lu, J., Tian, L., Ramachandran, V., Kasschau, K.D., Chapman, E.J., Carrington, J.C., Chen, X., Wang, X.J., and Chen, Z.J. (2009). Small RNAs serve as a genetic buffer against genomic shock in Arabidopsis interspecific hybrids and allopolyploids. Proc Natl Acad Sci U S A *106*, 17835-17840.

Hillier, L.W., Miller, R.D., Baird, S.E., Chinwalla, A., Fulton, L.A., Koboldt, D.C., and Waterston, R.H. (2007). Comparison of C. elegans and C. briggsae genome sequences reveals extensive conservation of chromosome organization and synteny. PLoS Biol *5*, e167.

Ish-Horowicz, D. (1982). Transposable elements, hybrid incompatibility and speciation. Nature *299*, 676-677.

Josefsson, C., Dilkes, B., and Comai, L. (2006). Parent-dependent loss of gene silencing during interspecies hybridization. Curr Biol *16*, 1322-1328.

Kelleher, E.S., Edelman, N.B., and Barbash, D.A. (2012). Drosophila interspecific hybrids phenocopy piRNA-pathway mutants. PLoS Biol *10*, e1001428.

Kiontke, K.C., Felix, M.A., Ailion, M., Rockman, M.V., Braendle, C., Penigault, J.B., and Fitch, D.H. (2011). A phylogeny and molecular barcodes for Caenorhabditis, with numerous new species from rotting fruits. BMC Evol Biol *11*, 339.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Lee, H.C., Gu, W., Shirayama, M., Youngman, E., Conte, D., Jr., and Mello, C.C. (2012). C. elegans piRNAs Mediate the Genome-wide Surveillance of Germline Transcripts. Cell.

Montgomery, T.A., Rim, Y.S., Zhang, C., Dowen, R.H., Phillips, C.M., Fischer, S.E., and Ruvkun, G. (2012). PIWI associated siRNAs and piRNAs specifically require the Caenorhabditis elegans HEN1 ortholog henn-1. PLoS Genet *8*, e1002616.

Rozhkov, N.V., Aravin, A.A., Zelentsova, E.S., Schostak, N.G., Sachidanandam, R., McCombie, W.R., Hannon, G.J., and Evgen'ev, M.B. (2010). Small RNA-based silencing strategies for transposons in the process of invading Drosophila species. RNA.

Ruby, J.G., Jan, C., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H., and Bartel, D.P. (2006). Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. Cell *127*, 1193-1207.

Shi, Z., Montgomery, T.A., Qi, Y., and Ruvkun, G. (2013). High-throughput sequencing reveals extraordinary fluidity of miRNA, piRNA, and siRNA pathways in nematodes. Genome Res *23*, 497-508.

Stein, L.D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M.R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A.*, et al.* (2003). The genome sequence of Caenorhabditis briggsae: a platform for comparative genomics. PLoS Biol *1*, E45.

Stoeckius, M., Maaskola, J., Colombo, T., Rahn, H.P., Friedlander, M.R., Li, N., Chen, W., Piano, F., and Rajewsky, N. (2009). Large-scale sorting of C. elegans embryos reveals the dynamics of small RNA expression. Nat Methods *6*, 745-751.

Swan, K.A., Curtis, D.E., McKusick, K.B., Voinov, A.V., Mapa, F.A., and Cancilla, M.R. (2002). High-throughput gene mapping in Caenorhabditis elegans. Genome Res *12*, 1100-1105.

Wicks, S.R., Yeh, R.T., Gish, W.R., Waterston, R.H., and Plasterk, R.H. (2001). Rapid gene mapping in Caenorhabditis elegans using a high density polymorphism map. Nat Genet *28*, 160-164.

Woodruff, G.C., Eke, O., Baird, S.E., Felix, M.A., and Haag, E.S. (2010). Insights into species divergence and the evolution of hermaphroditism from fertile interspecies hybrids of Caenorhabditis nematodes. Genetics *186*, 997-1012.

CHAPTER THREE

Evolution of the miRNA pathway in nematodes

AUTHOR CONTRIBUTIONS

I performed all experiments, analyzed the results, and discussed future directions presented here, with advising from Gary Ruvkun.

Summary

Using the miRDeep2 miRNA prediction program, we identified 37 new miRNAs in *C. briggsae*, 48 new miRNAs in *C. remanei*, and 215 new miRNAs in *C. brenneri* from our small RNA deep sequencing library (for the library construction, see Chapter II). Majority of miRNA families are present in all the four species, suggesting a conserved role of miRNAs. Through an alignment of miRNA orthologs, I found that besides the nucleotide 2-7 (the 'seed region'), the second most conserved region is nucleotide 13-15. This supports that miRNA nucleotide 13-15 has a supplementary role in binding to the mRNA targets. New miRNAs can be evolved from existing ones, through nucleotide substitution, arm switch or hairpin shifting. New miRNAs can also be born de novo, and more than half of new miRNAs are in the introns. New miRNAs are generally expressed at a lower level, and have a higher tendency to locate on the X chromosome.

This chapter contains the following sections:

- Prediction of miRNAs by miRDeep2
- Sequence conservation of miRNA orthologs
- Birth of new miRNAs
- miRNA gene duplication and divergence

Prediction of miRNAs by miRDeep2

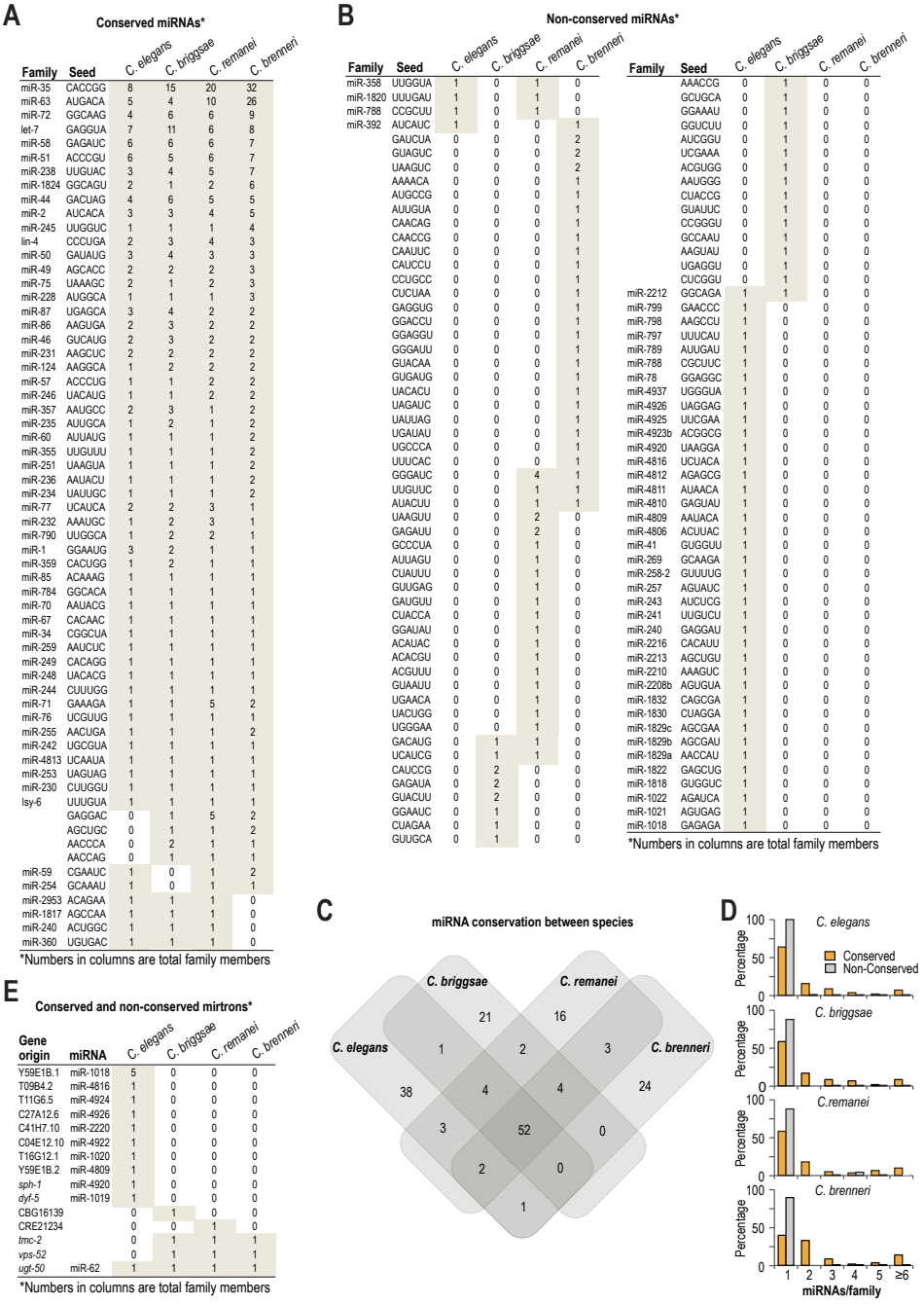
miRNAs in *C. elegans* are well-characterized (Gerstein et al., 2010; Grad et al., 2003; Kato et al., 2009; Lau et al., 2001; Lee and Ambros, 2001; Lim et al., 2003; Ruby et al., 2006). However, miRNAs in *C. briggsae* and *C. remanei* have been only partially characterized (de Wit et al., 2009), and miRNAs in *C. brenneri* have not been examined. To study the evolution of miRNAs, I first set out to obtain a more comprehensive list of miRNAs in each species.

There are several miRNA prediction software, including miRDeep2 (Friedlander et al., 2008; Friedlander et al., 2012), miRanalyzer (Hackenberg et al., 2009), miRExpress (Wang et al., 2009), miRTRAP (Hendrix et al., 2010), DSAP (Huang et al., 2010), mirTools (Zhu et al., 2010), MIReNA (Mathelier and Carbone, 2010), miRNAkey (Ronen et al., 2010) and etc. Among these software, miRDeep2 and MIReNA can predict miRNAs from small RNA deep-sequencing data sets. To test the software performance, I first asked miRDeep2 and MIReNA to predict miRNAs from the *C. elegans* small RNA deep sequencing library. miRDeep2 has a sensitivity of 73% and low false positive predictions. MIReNA has a sensitivity less than 50% and therefore was not used for de novo miRNA prediction. To conclusively call a miRNA, we required a candidate miRNA predicted by miRDeep2 to also be predicted by MIReNA and/or contain a seed sequence (positions 2–7) conserved among the *Caenorhabditis* species. By this standard, I identified 37 new miRNAs in *C. briggsae*, 48 new miRNAs in *C. remanei*, and 215 new miRNAs in *C. brenneri* but did not identify any new miRNAs in *C. elegans*. All newly identified miRNA from this study has been deposited into miRBase, the central repository for miRNA sequence information.

Conservation of miRNA families

To date, there are 101 miRNA families annotated in *C. elegans*, 84 in *C. briggsae*, 86 in *C. remanei* and 86 in *C. brenneri*, which constitute 171 distinct miRNA families (Figure 3.1A and B). Fifty-two miRNA families are conserved in all four species, sixty-two families are present in three or more species and most miRNA families within each species have homologs in at least two other species (Figure 3.1C). Few (<6%) miRNA families are conserved between only two or three nematode species (Figure 3.1C).

However, in each species at least 20% of miRNA families are unique, suggesting that miRNAs are born at relatively high rates (Figure 3.1C).



A Conserved miRNAs*

Family	Seed	C. elegans	C. briggsae	C. remanei	C. brenneri
miR-35	CACCGG	8	15	20	32
miR-63	AUGACA	5	4	10	26
miR-72	GGCAAG	4	6	6	9
let-7	GAGGUA	7	11	6	8
miR-58	GAGAUC	6	6	6	7
miR-51	ACCCGU	6	5	6	7
miR-238	UUGUAC	3	4	5	7
miR-1824	GGCAGU	2	1	2	6
miR-44	GACUAG	4	6	5	5
miR-2	AUCACA	3	3	4	5
miR-245	UUGGUC	1	1	1	4
lin-4	CCCUGA	2	3	4	3
miR-50	GAUAUG	3	4	3	3
miR-49	AGCACC	2	2	2	3
miR-75	UAAAGC	2	1	2	3
miR-228	AUGGCA	1	1	1	3
miR-87	UGAGCA	3	4	2	2
miR-86	AAGUGA	2	3	2	2
miR-46	GUCAUG	2	3	2	2
miR-231	AAGCUC	2	2	2	2
miR-124	AAGGCA	1	2	2	2
miR-57	ACCCUG	1	1	2	2
miR-246	UACAUG	1	1	2	2
miR-357	AAUGCC	2	3	1	2
miR-235	AUUGCA	1	2	1	2
miR-60	AUUAUG	1	1	1	2
miR-355	UUGUUU	1	1	1	2
miR-251	UAAGUA	1	1	1	2
miR-236	AAUACU	1	1	1	2
miR-234	UAUUGC	1	1	1	2
miR-77	UCAUCA	2	2	3	1
miR-232	AAAUGC	1	2	3	1
miR-790	UUGGCA	1	2	2	1
miR-1	GGAAUG	3	2	1	1
miR-359	CACUGG	1	2	1	1
miR-85	ACAAAG	1	1	1	1
miR-784	GGCACA	1	1	1	1
miR-70	AAUACG	1	1	1	1
miR-67	CACAAC	1	1	1	1
miR-34	CGGCUA	1	1	1	1
miR-259	AAUCUC	1	1	1	1
miR-249	CACAGG	1	1	1	1
miR-248	UACACG	1	1	1	1
miR-244	CUUUGG	1	1	1	1
miR-71	GAAAGA	1	1	5	2
miR-76	UCGUUG	1	1	1	1
miR-255	AACUGA	1	1	1	2
miR-242	UGCGUA	1	1	1	1
miR-4813	UCAAUA	1	1	1	1
miR-253	UAGUAG	1	1	1	1
miR-230	CUUGGU	1	1	1	1
lsy-6	UUUGUA	1	1	1	1
	GAGGAC	0	1	5	2
	AGCUGC	0	1	1	2
	AACCCA	0	2	1	1
	AACCAG	0	1	1	1
miR-59	CGAAUC	1	0	1	2
miR-254	GCAAAU	1	0	1	1
miR-2953	ACAGAA	1	1	1	0
miR-1817	AGCCAA	1	1	1	0
miR-240	ACUGGC	1	1	1	0
miR-360	UGUGAC	1	1	1	0

*Numbers in columns are total family members

B Non-conserved miRNAs*

Family	Seed	C. elegans	C. briggsae	C. remanei	C. brenneri
miR-358	UUGGUA	1	0	1	0
miR-1820	UUUGAU	1	0	1	0
miR-788	CCGCUU	1	0	1	0
miR-392	AUCAUC	1	0	0	1
	GAUCUA	0	0	0	2
	GUAGUC	0	0	0	2
	UAAGUC	0	0	0	2
	AAAACA	0	0	0	1
	AUGCCG	0	0	0	1
	AUUGUA	0	0	0	1
	CAACAG	0	0	0	1
	CAACCG	0	0	0	1
	CAAUUC	0	0	0	1
	CAUCCU	0	0	0	1
	CCUGCC	0	0	0	1
	CUCUAA	0	0	0	1
	GAGGUG	0	0	0	1
	GGACCU	0	0	0	1
	GGAGGU	0	0	0	1
	GGGAUU	0	0	0	1
	GUACAA	0	0	0	1
	GUGAUG	0	0	0	1
	UACACU	0	0	0	1
	UAGAUC	0	0	0	1
	UAUUAG	0	0	0	1
	UGAUAU	0	0	0	1
	UGCCCA	0	0	0	1
	UUUCAC	0	0	0	1
	GGGAUC	0	0	4	1
	UUGUUC	0	0	1	1
	AUACUU	0	0	1	1
	UAAGUU	0	0	2	0
	GAGAUU	0	0	2	0
	GCCCUA	0	0	1	0
	AUUAGU	0	0	1	0
	CUAUUU	0	0	1	0
	GUUGAG	0	0	1	0
	GAUGUU	0	0	1	0
	CUACCA	0	0	1	0
	GGAUAU	0	0	1	0
	ACAUAC	0	0	1	0
	ACACGU	0	0	1	0
	ACGUUU	0	0	1	0
	GUAAUU	0	0	1	0
	UGAACA	0	0	1	0
	UACUGG	0	0	1	0
	UGGGAA	0	0	1	0
	GACAUG	0	1	1	0
	UCAUCG	0	1	1	0
	CAUCCG	0	2	0	0
	GAGAUA	0	2	0	0
	GUACUU	0	2	0	0
	GGAAUC	0	1	0	0
	CUAGAA	0	1	0	0
	GUUGCA	0	1	0	0

Family	Seed	C. elegans	C. briggsae	C. remanei	C. brenneri
	AAACCG	0	1	0	0
	GCUGCA	0	1	0	0
	GGAAAU	0	1	0	0
	GGUCUU	0	1	0	0
	AUCGGU	0	1	0	0
	UCGAAA	0	1	0	0
	ACGUGG	0	1	0	0
	AAUGGG	0	1	0	0
	CUACCG	0	1	0	0
	GUAUUC	0	1	0	0
	CCGGGU	0	1	0	0
	GCCAAU	0	1	0	0
	AAGUAU	0	1	0	0
	UGAGGU	0	1	0	0
	CUCGGU	0	1	0	0
miR-2212	GGCAGA	1	1	0	0
miR-799	GAACCC	1	0	0	0
miR-798	AAGCCU	1	0	0	0
miR-797	UUUCAU	1	0	0	0
miR-789	AUUGAU	1	0	0	0
miR-788	CGCUUC	1	0	0	0
miR-78	GGAGGC	1	0	0	0
miR-4937	UGGGUA	1	0	0	0
miR-4926	UAGGAG	1	0	0	0
miR-4925	UUCGAA	1	0	0	0
miR-4923b	ACGGCG	1	0	0	0
miR-4920	UAAGGA	1	0	0	0
miR-4816	UCUACA	1	0	0	0
miR-4812	AGAGCG	1	0	0	0
miR-4811	AUAACA	1	0	0	0
miR-4810	GAGUAU	1	0	0	0
miR-4809	AAUACA	1	0	0	0
miR-4806	ACUUAC	1	0	0	0
miR-41	GUGGUU	1	0	0	0
miR-269	GCAAGA	1	0	0	0
miR-258-2	GUUUUG	1	0	0	0
miR-257	AGUAUC	1	0	0	0
miR-243	AUCUCG	1	0	0	0
miR-241	UUGUCU	1	0	0	0
miR-240	GAGGAU	1	0	0	0
miR-2216	CACAUU	1	0	0	0
miR-2213	AGCUGU	1	0	0	0
miR-2210	AAAGUC	1	0	0	0
miR-2208b	AGUGUA	1	0	0	0
miR-1832	CAGCGA	1	0	0	0
miR-1830	CUAGGA	1	0	0	0
miR-1829c	AGCGAA	1	0	0	0
miR-1829b	AGCGAU	1	0	0	0
miR-1829a	AACCAU	1	0	0	0
miR-1822	GAGCUG	1	0	0	0
miR-1818	GUGGUC	1	0	0	0
miR-1022	AGAUCA	1	0	0	0
miR-1021	AGUGAG	1	0	0	0
miR-1018	GAGAGA	1	0	0	0

*Numbers in columns are total family members

C miRNA conservation between species

D

E Conserved and non-conserved mirtrons*

Gene origin	miRNA	C. elegans	C. briggsae	C. remanei	C. brenneri
Y59E1B.1	miR-1018	5	0	0	0
T09B4.2	miR-4816	1	0	0	0
T11G6.5	miR-4924	1	0	0	0
C27A12.6	miR-4926	1	0	0	0
C41H7.10	miR-2220	1	0	0	0
C04E12.10	miR-4922	1	0	0	0
T16G12.1	miR-1020	1	0	0	0
Y59E1B.2	miR-4809	1	0	0	0
sph-1	miR-4920	1	0	0	0
dyf-5	miR-1019	1	0	0	0
CBG16139		0	1	0	0
CRE21234		0	0	1	0
tmc-2		0	1	1	1
vps-52		0	1	1	1
ugt-50	miR-62	1	1	1	1

*Numbers in columns are total family members

Figure 3.1 Conservation of miRNAs in nematodes. (A) Table of conserved miRNAs classified by family. Seed sequences are positions 2-7, relative to the 5' end of the miRNA. The number in each row represents the number of miRNAs in each family in each species.

(Figure 3.1 continued) Grey shading indicates presence of at least one member of a family. (B) As in A, but non-conserved miRNAs. (C) Venn diagram shows the number of miRNA families and their overlap in each of the four nematode species. (D) The percentage of families having the indicated number of members is shown for conserved and non-conserved miRNAs in each species. (E) Table of mirtrons classified by the gene hosting the mirtron.

Among conserved miRNA families, ~40-50% have multiple members within a species, whereas <13% of non-conserved families contain multiple members, suggesting that ancient miRNA families expand to confer robustness in gene regulatory networks (Figure 3.1D). A striking example of this is the miR-35 family which has expanded to contain at least eight members in each species and as many as 32 members in *C. brenneri* (Figure 3.1A). The miR-35 family is one of the few miRNA families essential for development (Alvarez-Saavedra and Horvitz 2010). Each of the other families essential for development, including miR-51, miR-58 and let-7, also contain multiple (≥5) members in each species (Figure 3.1A).

Most miRNAs are processed from primary transcripts in sequential steps involving the ribonucleases Drosha and Dicer. However, some miRNAs are instead derived from short intronic hairpins called mirtrons, during splicing, thereby bypassing Drosha cleavage (Okamura et al., 2007; Ruby et al., 2007). We found that out of the 15 *C. elegans* annotated mirtrons (Chung et al., 2011), only miR-62, embedded in the third intron of *ugt-50*, is conserved in the other three nematodes (Figure 3.1E). miR-62 has 100% sequence conservation in all four nematodes, and the conservation of the intron sequence itself is much higher than that of other *ugt-50* introns (Figure 3.2). Although the role of miR-62 is unknown, the strong selective pressure to maintain it hints at an important function. In addition to miR-62, we identified one non-conserved mirtron in *C.*

briggsae, one in *C. remanei* and two present in *C. briggsae*, *C. remanei* and *C. brenneri*, but not in *C. elegans* (Figure 3.1E).

Figure 3.2 The miR-62 mirtron is highly conserved. (A) The miR-62 mature sequences (in purple) are identical in all four nematode species. Nucleotides at other positions of miR-62 mirtron that are different between species are shown in grey. (B) The level of conservation of miR-62 mirtron is substantially higher than other introns in *ugt-50*.

Sequence conservation of miRNA orthologs

In animals, miRNAs cause repression by base pairing to the 3′ untranslated region of their target mRNAs, which contain perfect or near-perfect sequence complementarity to nucleotides 2-7 (the 'seed region') of miRNAs and mismatches and bulges in other parts of the miRNA-mRNA duplex. However, there is no consensus on where and how many mismatches/bulges can be tolerated between a functional miRNA-target mRNA interaction. As a result, although many miRNA prediction algorithms have been developed, they share few predicted targets in common. Prediction of miRNAs with high confidence is still a significantly challenging problem in the field.

miRNAs present in all the four nematode species suggest that they were present

in the common ancestor and are likely to have important functional roles. In particular, the region in the mature miRNA important for target recognition would be most conserved. Therefore, this allowed me to study which region of the mature miRNA might be important for binding with target mRNAs, by looking at the level of nucleotide conservation at each position of orthologous miRNAs.

By the alignment of 56 miRNAs that have a 1-1 ortholog in each of the four species, I found that 30 mature miRNAs have 100% sequence conservation, 14 miRNAs have sequence divergence in only 1-2 positions and 12 miRNAs bear divergence in more than 2 positions. Then, I counted for each position of the mature miRNA, how many times there is divergence among orthologous miRNAs (Figure 3.3). I found that besides the 'seed region', nucleotides 13-15 are second most conserved region.

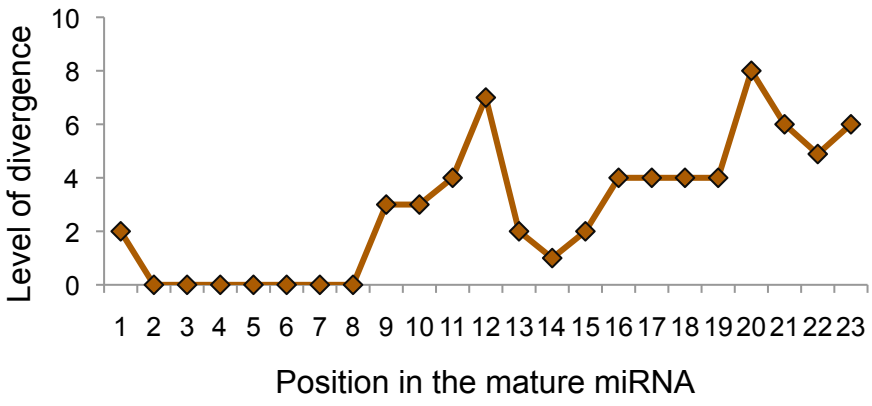


Figure 3.3 Nucleotides 13-15 in the mature miRNA are second most conserved, beside nucleotide 2-8, the extended seed region. Shown are the levels of divergence between orthologous miRNAs, in the four nematode species.

Importantly, this result echoes many previous studies (Brennecke et al., 2005; Friedman et al., 2009; Lai and Posakony, 1998; Wightman et al., 1993) and a recent one

from the Zamore lab (Wee et al., 2012). Together, these results further consolidate that nucleotide 13-16 play a supplementary role for miRNA binding to target transcripts. The incorporation of this principle might improve the miRNA target prediction algorithm.

Birth of new miRNAs

Among the over 200 miRNAs present in *C. elegans*, 47 are specific in *C. elegans* suggesting these miRNAs were born after the divergence of *C. elegans* with other nematode species. It was estimated that the four nematode species diverged from their last common ancestor approximately ~110 million generations ago (Cutter et al., 2009). Therefore, as *C. elegans* evolved, ~0.43 new miRNAs were gained every million generation. However, since miRNAs were constantly dying as well, the miRNA birth rate would be higher than 0.43 per million generation.

A comparison between conserved/old versus new *C. elegans* miRNAs indicates that new miRNAs are generally weakly expressed (Figure 3.4). This is consistent with the hypothesis that new miRNAs play a less significant role in the gene regulatory network, partially due to their low expression levels.

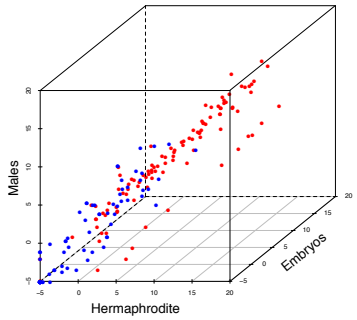


Figure 3.4 New miRNAs are generally weakly expressed. Shown are expression levels of conserved/old (red) and new (blue) *C. elegans* miRNAs at the three developmental stages, in a log scale.

52% of new miRNAs (n=39), whereas only 18% of conserved/old miRNAs (n=32) in *C. elegans* are in the intron of protein-coding genes ($p < 0.01$, Chi-square test). This suggests that new miRNAs are more likely to be evolved from introns in nematodes. Intriguingly, I found that new miRNAs are also more likely to be located on the X chromosome (Figure 3.5). It is possible that the X-inactivation in the germline prevents the expression of these miRNAs during the germline development and in this way reduces any potential deleterious effects of these new miRNAs on embryos. However, it was showed that miRNAs can escape the X-inactivation in mammals (Meunier et al., 2013). The reason for the enrichment of new miRNAs on the X chromosome is unclear.

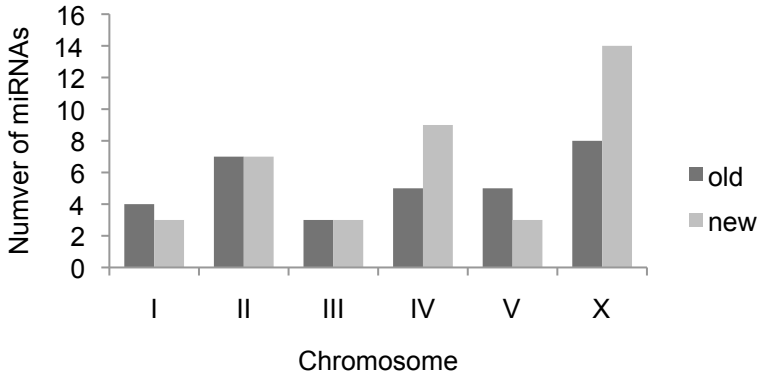


Figure 3.5 The genomic locations of conserved/old and new miRNAs in *C. elegans*.

In *Arabidopsis*, the 22-nt miRNAs primarily arise from foldback precursors containing asymmetric bulges (Cuperus et al., 2010). However, it is less clear how new miRNAs were evolved in animals. To track the evolutionary trajectory of newly evolved miRNAs in *C. elegans*, I aligned the newly evolved *C. elegans* miRNAs with genomes of other nematode species. Since most new miRNAs in the intragenic regions cannot be

aligned to other genomes, I focused the following analysis on several *C. elegans* new miRNAs located in the introns. For example, by an alignment of *C. elegans mir-4826* mirtron with the corresponding *C. briggsae* intron, I found that this intron in *C. briggsae* does not have a miRNA-like secondary structure (Figure 3.6B, right). In *C. elegans*, this intron has undergone multiple deletions, substitutions and small insertions (Figure 3.6A) and eventually evolves a perfect hairpin secondary structure (Figure 3.6B, left). Several other *C. elegans* new miRNAs were also evolved like *mir-4826* and are not shown here one by one. Together, it suggests that in nematodes, miRNAs can evolve de novo through multiple rounds of mutations.

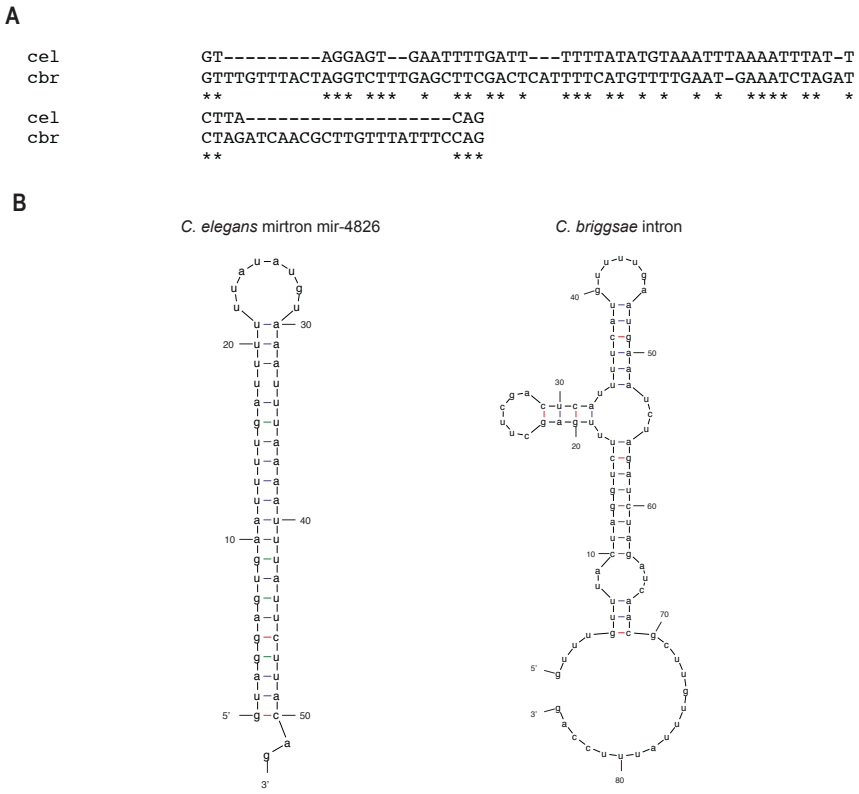


Figure 3.6 Born of the *C. elegans mir-4826* mirtron. (A) Sequence alignment of the *C. elegans mir-4826* mirtron against the *C. briggsae* corresponding intron. (B) Secondary structure prediction of the *C. elegans mir-4826* mirtron and the *C. briggsae* corresponding intron, using the mfold algorithm (Zuker, 2003).

miRNA gene duplication and divergence

Besides de novo birth of new miRNAs, miRNA genes also evolve from existing ones. In particular, miRNA gene duplications lessen the selective pressure on one miRNA paralog and this provides a reservoir for the evolution of new miRNA. Taking the most duplicated *mir-35* family of miRNAs as an example, they underwent three basic types of changes. The most common divergence among miRNA paralogs is nucleotide substitution (Figure 3.7A), suggesting point mutations during evolution. All of nucleotide substitutions in the *mir-35* family occur outside of the seed sequence, mainly at the 3' end of the mature miRNA, indicating they might still possess overall similar target specificity. Indeed, the *C. elegans mir-35* family members are redundantly required for embryonic development (Alvarez-Saavedra and Horvitz, 2010; Massirer et al., 2012). The second type of evolution is through arm switching. Originally, the mature *mir-41* is located at the 3' arm of the hairpin: as for all the *C. elegans mir-41* paralogs and its orthologs in the other three species. The *C. elegans mir-41* hairpin bears several mutations in the stem region and this causes the 5' arm of the hairpin to become the predominant miRNA (Figure 3.7B). Since this miRNA has different seed sequence, a new miRNA with entirely different set of target genes was evolved. Lastly, miRNA genes can evolve through hairpin shifting. The mature *mir-37* was originally located at the 3' arm of the hairpin, as shown on the top of Figure 3.7C. In *C. elegans*, the locus evolved a second hairpin immediately downstream of the ancient hairpin, and the mature *mir-37* is made from its 5' arm (Figure 3.7C). This can give rise to a different new miRNA, if arm switch occurs after the hairpin shifting event.

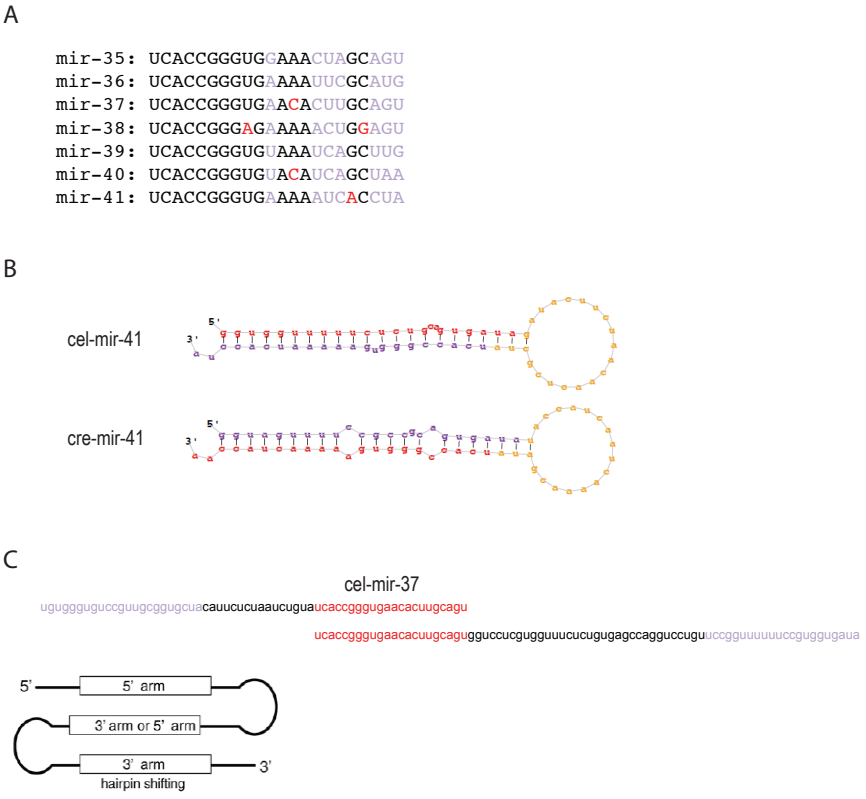


Figure 3.7 Evolution of the *mir-35* miRNA gene family. (A) Nucleotide divergence in the *C. elegans mir-35* family. Positions highlighted in purple were mutated before some of the *mir-35* family gene duplications. Positions highlighted in red bear divergence in only one miRNA and therefore the mutation was likely to happen after all gene duplications. (B) Arm switching. In *C. remanei*, the miRNA on the 3' arm of *mir-41* hairpin is much more abundant whereas in *C. elegans*, the miRNAs on the 5' arm is more abundant. (C) Hairpin shifting, the diagram is adapted from (de Wit et al., 2009). For (B) and (C), the mature miRNA is shown in red and miRNA star strand in purple.

Together, our analyses revealed that majority of miRNAs are highly conserved across the nematode species. Meanwhile, many new miRNAs were born in each species either de novo or through evolution of existing ones via nucleotide substitution, arm switching or hairpin shifting.

Reference

Alvarez-Saavedra, E., and Horvitz, H.R. (2010). Many families of C. elegans microRNAs are not essential for development or viability. Curr Biol *20*, 367-373.

Brennecke, J., Stark, A., Russell, R.B., and Cohen, S.M. (2005). Principles of microRNA-target recognition. PLoS Biol *3*, e85.

Cuperus, J.T., Carbonell, A., Fahlgren, N., Garcia-Ruiz, H., Burke, R.T., Takeda, A., Sullivan, C.M., Gilbert, S.D., Montgomery, T.A., and Carrington, J.C. (2010). Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in Arabidopsis. Nat Struct Mol Biol *17*, 997-1003.

Cutter, A.D., Dey, A., and Murray, R.L. (2009). Evolution of the Caenorhabditis elegans genome. Mol Biol Evol *26*, 1199-1234.

de Wit, E., Linsen, S.E., Cuppen, E., and Berezikov, E. (2009). Repertoire and evolution of miRNA genes in four divergent nematode species. Genome Res *19*, 2064-2074.

Friedlander, M.R., Chen, W., Adamidi, C., Maaskola, J., Einspanier, R., Knespel, S., and Rajewsky, N. (2008). Discovering microRNAs from deep sequencing data using miRDeep. Nat Biotechnol *26*, 407-415.

Friedlander, M.R., Mackowiak, S.D., Li, N., Chen, W., and Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res *40*, 37-52.

Friedman, R.C., Farh, K.K., Burge, C.B., and Bartel, D.P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. Genome Res *19*, 92-105.

Gerstein, M.B., Lu, Z.J., Van Nostrand, E.L., Cheng, C., Arshinoff, B.I., Liu, T., Yip, K.Y., Robilotto, R., Rechtsteiner, A., Ikegami, K*., et al.* (2010). Integrative Analysis of the Caenorhabditis elegans Genome by the modENCODE Project. Science.

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.

Hackenberg, M., Sturm, M., Langenberger, D., Falcon-Perez, J.M., and Aransay, A.M. (2009). miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments. Nucleic Acids Res *37*, W68-76.

Hendrix, D., Levine, M., and Shi, W. (2010). miRTRAP, a computational method for the systematic identification of miRNAs from high throughput sequencing data. Genome Biol *11*, R39.

Huang, P.J., Liu, Y.C., Lee, C.C., Lin, W.C., Gan, R.R., Lyu, P.C., and Tang, P. (2010). DSAP: deep-sequencing small RNA analysis pipeline. Nucleic Acids Res *38*, W385-391.

Kato, M., de Lencastre, A., Pincus, Z., and Slack, F.J. (2009). Dynamic expression of small non-coding RNAs, including novel microRNAs and piRNAs/21U-RNAs, during Caenorhabditis elegans development. Genome Biol *10*, R54.

Lai, E.C., and Posakony, J.W. (1998). Regulation of Drosophila neurogenesis by RNA:RNA duplexes? Cell *93*, 1103-1104.

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.

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

Massirer, K.B., Perez, S.G., Mondol, V., and Pasquinelli, A.E. (2012). The miR-35-41 Family of MicroRNAs Regulates RNAi Sensitivity in Caenorhabditis elegans. PLoS Genet *8*, e1002536.

Mathelier, A., and Carbone, A. (2010). MIReNA: finding microRNAs with high accuracy and no learning at genome scale and from deep sequencing data. Bioinformatics *26*, 2226-2234.

Meunier, J., Lemoine, F., Soumillon, M., Liechti, A., Weier, M., Guschanski, K., Hu, H., Khaitovich, P., and Kaessmann, H. (2013). Birth and expression evolution of mammalian microRNA genes. Genome Res *23*, 34-45.

Ronen, R., Gan, I., Modai, S., Sukacheov, A., Dror, G., Halperin, E., and Shomron, N. (2010). miRNAkey: a software for microRNA deep sequencing analysis. Bioinformatics *26*, 2615-2616.

Ruby, J.G., Jan, C., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H., and Bartel, D.P. (2006). Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. Cell *127*, 1193-1207.

Wang, W.C., Lin, F.M., Chang, W.C., Lin, K.Y., Huang, H.D., and Lin, N.S. (2009). miRExpress: analyzing high-throughput sequencing data for profiling microRNA expression. BMC Bioinformatics *10*, 328.

Wee, L.M., Flores-Jasso, C.F., Salomon, W.E., and Zamore, P.D. (2012). Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. Cell *151*, 1055-1067.

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.

Zhu, E., Zhao, F., Xu, G., Hou, H., Zhou, L., Li, X., Sun, Z., and Wu, J. (2010). mirTools: microRNA profiling and discovery based on high-throughput sequencing. Nucleic Acids Res *38*, W392-397.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res *31*, 3406-3415.

CHAPTER FOUR

The mevalonate pathway regulates miRNA activity in

Caenorhabditis elegans

AUTHOR CONTRIBUTIONS

I performed all experiments, analyzed the results, and discussed future directions presented here, with advising from Gary Ruvkun. Part of this chapter is adapted from our paper “The mevaloante pathway regulates microRNA activity in *Caenorhabditis elegans*” (Shi and Ruvkun, 2012).

Summary

The mevalonate pathway is highly conserved and mediates the production of isoprenoids, which feed into biosynthetic pathways for sterols, dolichol, ubiquinone, heme, isopentenyl adenine, and prenylated proteins. We found that in *Caenorhabditis elegans*, the non-sterol biosynthetic outputs of the mevalonate pathway are required for the activity of miRNAs in silencing their target mRNAs. Inactivation of genes that mediate multiple steps of the mevalonate pathway causes derepression of several miRNA target genes, with no disruption of the miRNA levels, suggesting a role in miRNA-induced silencing complex (miRISC) activity. Dolichol phosphate, synthesized from the mevalonate pathway, functions as a lipid carrier of the oligosaccharide moiety destined for protein N-linked glycosylation. Inhibition of the dolichol pathway of protein N-glycosylation also causes derepression of miRNA target mRNAs. The proteins that mediate miRNA repression are therefore likely to be regulated by N-glycosylation. Conversely, drugs such as statins, which inhibit the mevalonate pathway, may compromise miRNA repression as well as the more commonly considered cholesterol biosynthesis.

Motivating Questions

This project was initially motivated by a hypothesis that some small RNAs are covalently linked to proteins, forming RNA bar codes that could mediate interaction with complementary coded proteins and nucleic acids, proposed by Gary Ruvkun. There are a few pieces of evidence that bacterial small RNAs can be coupled to modified nucleotides such as Coenzyme A (CoA) (Kowtoniuk et al., 2009) and Nicotinamide adenine dinucleotide (NAD) (Chen et al., 2009). These modifications would allow small

RNAs to covalently interact with proteins, for example via disulfide bond or thioester bond. Although the potential functions of these small RNA-protein complexes were completely speculative, we hypothesized that the chemical modifications of small RNAs could be important for their functions. To test this hypothesis, we assembled a cherry-picked RNAi library, targeting genes function in various biosynthetic pathways (for example, CoA metabolisms), genes predicted to interact with RNAs, as well as proteins enriched in cysteine that potentially can be covalently linked to modified RNAs. The genes targeted by this cherry-picked RNAi library are listed in Table 4.1.

I specifically asked whether the presumed RNA modifications are important for miRNA activity. If our hypotheses were true, the RNAi library should be enriched for genes important for miRNA function, compared to a random assembled RNAi library. In *C. elegans*, the *let-7* miRNA regulates developmental timing events during the fourth larval stage (L4)-to-adult transition. Loss of *let-7* activity either by mutation of the *let-7* gene or inactivation of *dcr-1/Dicer* or *alg-1/Argonaute*, the core components in miRNA maturation and function, cause retarded heterochronic phenotypes in which larval developmental patterns are reiterated and adult-specific specializations do not occur (Grishok et al., 2001; Reinhart et al., 2000). To identify genes that act in the miRNA pathway, I screened the sulfome RNAi library for gene inactivations that enhance a weak *let-7(mg279)* reduction-of-function mutation. However, there was only one strong hit from this screen, *hmgs-1*, which encodes the *C. elegans* ortholog of HMG-CoA synthase. Although our cherry-picked RNAi library was not enriched for genes important for miRNA function, I decided to continue studying the *hmgs-1* gene anyway.

Table 4.1 Cherry-picked RNAi library

Gene	Gene Public Name	Gene Description
B0024.9	trx-2	
B0228.5	trx-1	trx-1 encodes a thioredoxin, a small redox protein that functions as a protein-disulfide reductase
B0416.5a		
C01B7.1	C01B7.1	
C02A12.1	gst-33	
C02F5.1	knl-1	KNL-1 is an essential kinetochore component that is required for proper spindle elongation and chromosome separation, and in the kinetochore assembly pathway, plays a key role in linking the initiation of kinetochore formation with the construction of a functional microtubule-binding interface; in the assembly pathway, KNL-1 functions downstream of the DNA-proximal kinetochore components CeCENP-A/HCP-3 and CeCENP-B/HCP-4 and upstream of the outer kinetochore components HIM-10/Nuf2p, NDC-80/HEC1, CeBUB-1, HCP-1, and CeCLASP2/CLS-2
C03D6.3	cel-1	cel-1 encodes a mRNA capping enzyme, with a N-terminal region with RNA triphosphatase activity and a C-terminal region containing motifs found in yeast and vaccinia virus capping enzyme guanylytransferases; cel-1 is required for embryonic viability, body morphology, and vulval development.
C06E7.3	C06E7.3	
C07A12.4	pdi-2	
C12C8.2	C12C8.2	
C14B1.1	pdi-1	
C14B1.4	tag-125	tag-125/C14B1.4 encodes an ortholog of the histone methyltransferase subunit WDR5 (OMIM:609012) that antagonizes SynMuv transcriptional repressors.
C14B1.5	C14B1.5	C14B1.5 encodes an ortholog of S. cerevisiae YIL103 and human DPH2L1/OVCA1 (OMIM:603527, deleted or downregulated in ovarian tumors); C14B1.5 is paralogous to S. cerevisiae DPH2/YKL191W, a protein component of diphtamide synthesis.
C14B9.2	C14B9.2	
C14B9.7	rpl-21	rpl-21 encodes a large ribosomal subunit L21 protein; by homology, RPL-21 is predicted to function in protein biosynthesis; in C. elegans, RPL-21 activity is required for embryonic and germline development.
C16A3.3	C16A3.3	
C16C10.12	C16C10.12	
C23H3.3	C23H3.3	
C23H3.5	C23H3.5	
C24F3.5	abt-1	abt-1 encodes a predicted ATP-binding cassette (ABC) transporter that is a member of the ABCA subfamily of transport proteins; ABT-1 is predicted to function as a transmembrane protein that couples energy to transport of various molecules across membranes, but as loss of abt-1 activity via RNAi results in no obvious defects, the precise role of abt-1 in C. elegans development and/or behavior is not yet known.
C24F3.5	abt-1	abt-1 encodes a predicted ATP-binding cassette (ABC) transporter that is a member of the ABCA subfamily of transport proteins; ABT-1 is predicted to function as a transmembrane protein that couples energy to transport of various molecules across membranes, but as loss of abt-1 activity via RNAi results in no obvious defects, the precise role of abt-1 in C. elegans development and/or behavior is not yet known.
C25A1.6	C25A1.6	
C29E4.2	kle-2	
C29E4.7	C29E4.7	
C30H7.2	C30H7.2	
C34B2.2	kbp-5	
C34B7.4	C34B7.4	
C44B12.3	C44B12.3	
C44B7.10	C44B7.10	
C49F5.1	sams-1	
C52E12.3	sqv-7	SQV-7 promotes glycosaminoglycan biosynthesis by translocating UDP-glucuronic acid, UDP-N-acetylgalactosamine, and UDP-galactose into the lumen of the Golgi apparatus
C53D5.5	C53D5.5	The C53D5.5 gene encodes an ortholog of the human gene GAMMA-GLUTAMYLTRANSFERASE 1 (GGT1), which when mutated leads to glutathionuria (OMIM:231950).

Table 4.1 (continued)

Gene	Gene Public Name	Gene Description
C53D5.6	imb-3	imb-3 encodes an importin-beta-like protein orthologous to Drosophila, vertebrate, and yeast importin/karyopherin-beta3; IMB-3 is predicted to function as a nuclear transport factor that, with the RAN-1 GTPase, regulates nuclear import of ribosomal proteins
C54D2.4	sul-3	sul-3 is orthologous to the human gene ARYLSULFATASE B (ARSB; OMIM:253200), which when mutated leads to mucopolysaccharidosis type VI.
C54G10.2	rfc-1	
C55B7.6	sulp-1	sulfate permease family of anion transporters; by homology, SULP-1 is predicted to function as an anion transporter that regulates cellular pH and volume via transmembrane movement of electrolytes and fluids
D1053.1	gst-42	gst-42 is orthologous to the human gene GLUTATHIONE TRANSFERASE ZETA-1 (also known as MALEYLACETOACETATE ISOMERASE; GSTZ1; OMIM:603758), which when mutated is thought to lead to a variety of type I tyrosinemia.
D2005.5	drh-3	
D2023.2	pyc-1	pyruvate carboxylase ortholog
D2096.4	sqv-1	sqv-1 encodes a UDP-glucuronic acid decarboxylase, biochemically active in vitro, that is required for cytokinesis of one-cell embryos and for vulval morphogenesis
DH11.3	pgp-11	pgp-11 encodes an ATP-binding protein that is a member of the P-glycoprotein subclass of the ATP-binding cassette (ABC) transporter superfamily; pgp-11 is predicted to function as a transmembrane protein that couples energy to transport of various molecules across membranes
E02H1.1	E02H1.1	
E02H1.4	pme-2	
EEED8.5	mog-5	The mog-5 gene encodes a DEAH helicase orthologous to the Drosophila CG8241, the human HRH1, and the S. cerevisiae PRP22 proteins.
F01G4.3	F01G4.3	
F08B4.6	hst-1	
F08B4.7	F08B4.7	
F08C6.2	F08C6.2	F08C6.2 encodes a lipid-activated CTP:phosphocholine cytidylyltransferase (CCT), with CCT activity in vitro; recombinant F08C6.2 enzyme is most activated by a 1:1 mixture of phosphatidylcholine:oleate vesicles
F11G11.2	gst-7	glutathione S-transferase.
F11G11.3	gst-6	glutathione S-transferase.
F12F6.3	rib-1	The rib-1 gene encodes an ortholog of human EXT1, which when mutated leads to hereditary multiple exostoses, type I (OMIM:133700).
F13A7.10	gst-44	
F14D12.5	sulp-2	sulp-2 encodes one of eight C. elegans members of the sulfate permease family of anion transporters
F17A9.1	F17A9.1	F17A9.1 encodes a divergent ONECUT class CUT homeobox protein with a single N-terminal cut domain; F17A9.1 has an atypical tyrosine residue at position 48 of its homeodomain rather than a phenylalanine or tryptophan residue; the cut domain may be a compact DNA-binding domain composed of alpha helices; phylogenetically, F17A9.1 is (somewhat distantly) affiliated with C17H12.9, Drosophila ONECUT, and mammalian HNF6 proteins; F17A9.1 has no obvious function in mass RNAi assays.
F21H7.1	gst-22	
F22B8.6	F22B8.6	
F25B4.6		HMG-CoA synthase
F26E4.10	drsh-1	Drosha; by homology, DRSH-1 is predicted to function as an endoribonuclease that, in the nucleus, initiates cleavage of primary miRNA transcripts (pri-mRNAs) into pre-miRNAs that are then exported to the cytoplasm for further processing
F26E4.12	F26E4.12	
F26E4.8	tba-1	
F26H11.1	kbp-3	
F26H9.4	F26H9.4	
F26H9.6	rab-5	rab-5 encodes a rab related protein of the Ras GTPase superfamily that affects both the localization of P-granules and of PAR-2, and also affects embryonic and larval viability and the cytoplasmic appearance of cells in the early embryo.
F26H9.8	F26H9.8	

Table 4.1 (continued)

Gene	Gene Public Name	Gene Description
F28D1.10	gex-3	The gex-3 gene encodes a homolog of NAP1/NCKAP1, a mammalian protein ligand of the small GTPase Rac1, and of Drosophila HEM2/NAP1/KETTE; gex-3 is required for tissue morphogenesis and cell migrations; in gex-3 mutants, cells differentiate properly but fail to become organized.
F29F11.1	sqv-4	sqv-4 encodes a UDP-glucose 6-dehydrogenase, biochemically active in vitro, that is required for cytokinesis of one-cell embryos and for vulval morphogenesis; SQV-4 is orthologous to Drosophila SUGARLESS, human UGDH (OMIM:603370), and zebrafish JEKYLL
F35E8.8	gst-38	
F35G12.8	smc-4	The smc-4 gene encodes a homolog of the SMC4 subunit of mitotic condensin; SMC-4 acts with MIX-1 to enable chromosome segregation.
F35G2.4	phy-2	
F37B1.1	gst-24	
F37B1.2	gst-12	
F37B1.3	gst-14	
F37B1.4	gst-15	
F37B1.5	gst-16	
F37B1.7	gst-18	
F37B1.8	gst-19	
F37B12.1	F37B12.1	
F37B12.2	gcs-1	gcs-1 encodes the C. elegans ortholog of gamma-glutamine cysteine synthetase heavy chain (GCS(h)); GCS-1 is predicted to function, in a conserved oxidative stress response pathway, as a phase II detoxification enzyme that catalyzes the rate-limiting first step in glutathione biosynthesis
F37D6.1	mus-101	
F37F2.3	gst-25	
F41D9.5	sulp-3	sulp-3 encodes one of eight C. elegans members of the sulfate permease family of anion transporters; by homology, SULP-3 is predicted to function as an anion transporter that regulates cellular pH and volume via transmembrane movement of electrolytes and fluids; a sulp-3::GFP transcriptional fusion is expressed exclusively in the pharyngeal muscles.
F42E11.1	pgp-4	pgp-4 encodes an ATP-binding protein that is a member of the P-glycoprotein subclass of the ATP-binding cassette (ABC) transporter superfamily
F42G8.6	moc-3	Molybdopterin synthase sulfurylase, orgholog to human Adenylyltransferase and sulfurtransferase MOCS3, to generate MPT
F43D2.1	F43D2.1	
F43E2.4	haf-2	haf-2 encodes a predicted transmembrane protein of the ATP-binding cassette (ABC) transporter superfamily; by homology, HAF-2 is proposed to function in ATP-dependent transport of molecules across plasma and intracellular membranes; however, as loss of HAF-2 function via RNA-mediated interference (RNAi) does not result in any abnormalities
F46E10.9	dpy-11	dpy-11 encodes a membrane- associated thioredoxin- like (TRX) protein that affects body shape and ray morphology; the TRX domain displays catalytic activity in vitro, and dpy-11 is expressed in cytoplasm of hypodermis.
F49E10.5	ctbp-1	tag-45 encodes a D-isomer specific 2-hydroxyacid dehydrogenase.
F49E10.5	ctbp-1	tag-45 encodes a D-isomer specific 2-hydroxyacid dehydrogenase.
F49E2.1	F49E2.1	Description: F49E2.1 is orthologous to the human gene MOLYBDENUM COFACTOR SYNTHESIS-STEP 1 PROTEIN A-B SPLICE TYPE III
F49H6.5	F49H6.5	The F49H6.5 gene encodes a homolog of the human gene MOCS1A, which when mutated leads to molybdenum cofactor deficiency
F54C8.1	F54C8.1	
F54C8.2	cpar-1	
F54C8.3	emb-30	emb-30 encodes an anaphase-promoting complex/cyclosome (APC/C) component orthologous to mammalian APC-4 and Schizosaccharomyces pombe Lid1; EMB-30 is required for the metaphase-to-anaphase transition during meiosis and mitosis, for establishing anterior-posterior polarity in the early embryo, and for proper localization of germline granules and the maternally provided PAR-2 and PAR-3 proteins.
F54C8.4	F54C8.4	
F54D5.1	pcs-1	
F54G8.3	ina-1	

Table 4.1 (continued)

Gene	Gene Public Name	Gene Description
F56B3.10	gst-40	
F58A4.3	hcp-3	The hcp-3 gene encodes a centromere protein (CENP)-A homolog required for kinetochore function; inactivation of hcp-3 in one-cell embryos by RNAi causes a complete loss of kinetochores, with total failure of chromosomes to segregate properly during mitosis, to recruit components to the kinetochore other than HCP-3, or to assemble a stable mitotic spindle; in addition, HCP-4 fails to localize properly to the kinetochore in hcp-3(RNAi) embryos.
F59C6.4	F59C6.4	
H06O01.1	pdi-3	pdi-3 encodes a protein disulfide isomerase (Updike and Strome) required for normal cuticle collagen deposition and, subliminally, for maintenance of normal body shape; PDI-3 has both PDI and calcium-dependent transglutaminase activity in vitro, crosslinking proteins through a gamma-glutamyl epsilon-lysine dual residue
K01G5.6	rib-2	The rib-2 gene encodes an ortholog of human EXT2, which when mutated leads to hereditary multiple exostoses, type II (OMIM:133701).
K02F2.2	K02F2.2	
K02F2.3	tag-203	
K03D10.3	K03D10.3	The K03D10.3 gene encodes a MYST acetyltransferase orthologous to the Drosophila MALES-ABSENT-ON-THE-FIRST (MOF) and CG1894 proteins, the human MOF protein, and the S. cerevisiae SAS2 protein.
K08F4.11	gst-3	gst-3 encodes a predicted glutathione S-transferase.
K08F4.6	gst-2	
K08F4.7	gst-4	gst-4 encodes a predicted glutathione S-transferase; mRNA is expressed in adults and accumulation increases in response to paraquat.
K10B3.6	K10B3.6	
K10B3.7	gpd-3	gpd-3 encodes a predicted glyceraldehyde 3-phosphate dehydrogenase that affects embryonic viability; GPD-3 interacts with LIN-2 in two-hybrid assays.
K10B3.8	gpd-2	gpd-2 encodes one of four C. elegans glyceraldehyde-3-phosphate dehydrogenases (GAPDHs
K10B3.9	mai-1	mai-1 is homologous to mitochondrial intrinsic ATPase inhibitor protein (IF(1)), which blocks reverse action (ATP hydrolysis) by F(0)F(1)-ATPase when its (normally required) proton gradient is lost
K12G11.1	sulp-4	members of the sulfate permease family of anion transporters
K12G11.2	sulp-5	sulp-5 encodes one of eight C. elegans members of the sulfate permease family of anion transporters; by homology, SULP-5 is predicted to function as an anion transporter that regulates cellular pH and volume via transmembrane movement of electrolytes and fluids
M03F8.2	pst-1	
M03F8.3	M03F8.3	
M04F3.1	M04F3.1	
R03A10.3	R03A10.3	Molybdenum cofactor sulfurase, in Moco activation, ortholog to Arabidopsis thaliana Aba3
R03D7.6	gst-5	
R03G5.2	sek-1	SEK-1 has MAPKK activity and belongs to the MAPKK family; SEK-1 can activate both JNK-1 and PMK-1 in the yeast Hog pathway.
R05H10.5	R05H10.5	
R06A4.7	mes-2	mes-2 encodes a SET domain-containing protein that is orthologous to the Drosophila Polycomb group protein Enhancer of zeste [E(Z)]
R07B1.4	gst-36	
R07B5.8	R07B5.8	The R07B5.8 gene encodes a MYST acetyltransferase homologous to the S. cerevisiae SAS3 protein; it is also more distantly homologous to the human MOZ and MORF proteins, which share a similar domain organization C-terminal to the MYST domain.
R107.6	cls-2	cls-2 encodes one of three predicted orthologs of mammalian CLASPs and of Drosophila ORBIT/MAST, microtubule-binding proteins required for fibroblast polarization and mitosis; cls-2 is required in mass RNAi assays for embryonic development and normal mitotic spindles; it has been claimed that, in an RNAi screen of potential microtubule tip-binding proteins, only cls-2(RNAi) yielded embryonic lethality and meiotic defects.
R107.7	gst-1	gst-1 encodes a putative glutathione S-transferase with highest similarity to the pi class.
R10E11.4	sqv-3	sqv-3 encodes a beta(1,4)-galactosyltransferase
R11G1.3	gst-11	

Table 4.1 (continued)

Gene	Gene Public Name	Gene Description
R11G1.3	gst-11	
R12B2.4	him-10	related to the Nuf2 kinetochore proteins. in C. elegans, him-10 activity is essential for the proper structure and function of mitotic and meiotic kinetochores and thus, for proper attachment and segregation of chromosomes during mitosis and meiosis
R12E2.1	R12E2.1	
R12E2.2	R12E2.2	
R13D7.7	gst-41	
R13F6.1	kbp-1	
R186.3	R186.3	
R186.4	lin-46	one of two C. elegans paralogs of bacterial MoeA proteins and mammalian gephyrins (E domain, only). Last step in molybdenum cofactor synthesis, incorportation of Mo
R186.7	R186.7	
R53.1	R53.1	
T03F1.8	T03F1.8	
T03F1.9	hcp-4	The hcp-4 gene encodes a centromere protein (CENP)-C homolog, holocentric protein (HCP)-4.
T04H1.4	rad-50	
T05G5.4	T05G5.4	
T05G5.5	T05G5.5	
T06H11.4	moc-1	an ortholog of human GEPHYRIN which when mutated leads to molybdenum cofactor (MoCo) deficiency; MOC-1 is also paralogous to LIN-46
T07A9.6	daf-18	daf-18 encodes a lipid phosphatase homologous to the human PTEN tumor suppresor
T10B5.5	T10B5.5	
T10B5.6	knl-3	knl-3 encodes a novel protein; KNL-3 activity is essential for formation of a functional kinetochore and thus, for proper chromosome segregation and spindle pole separation
T13A10.11	tag-32	
T14G10.1	pps-1	pps-1 is orthologous to human PAPSS1 (OMIM:603262) and human PAPSS2 (OMIM:603005, mutated in spondyloepimetaphyseal dysplasia).
T20G5.11	rde-4	
T20G5.2	cts-1	cts-1 encodes a citrate synthase, predicted to be mitochondrial, that is required for embryonic development.
T21G5.5	star-2	
T23G11.3	gld-1	gld-1 encodes a protein containing a K homology RNA binding domain that is required for meiotic cell cycle progression during oogenesis in parallel with gld-2, and also affects spermatogenesis
T23G7.5	pir-1	
T24D1.1	sqv-5	chondroitin synthase that both initiates and elongates chondroitin chains
T26C5.1	gst-13	
T27A3.6	T27A3.6	T27A3.6 is orthologous to the human gene MOLYBDENUM COFACTOR BIOSYSTHESIS PROTEIN E, function in MPT synthesis
T28A11.11	gst-23	
VC5.4	mys-1	The VC5.4 gene encodes a MYST acetyltransferase orthologous to the Drosophila EG0007.7, human TIP60 and S. cerevisiae ESA1 proteins
W01A11.6	moc-2	Molybdopterin biosynthesis protein, still, ortholog of human gephyrins
W01B11.2	sulp-6	sulp-6 encodes one of eight C. elegans members of the sulfate permease family of anion transporters; by homology, SULP-6 is predicted to function as an anion transporter that regulates cellular pH and volume via transmembrane movement of electrolytes and fluids
W01B6.9	ndc-80	ndc-80 (cogc-7) encodes an ortholog of mammalian COG-7 (OMIM:606978, mutated in congenital disorder of glycosylation), a subunit of lobe B of the conserved oligomeric Golgi complex (COGC)
W04G3.6	sulp-7	sulp-7 encodes one of eight C. elegans members of the sulfate permease family of anion transporters
W06B3.1	W06B3.1	

Table 4.1 (continued)

Gene	Gene Public Name	Gene Description
W08D2.7	W08D2.7	
W09D6.6	hmt-1	haf-5 encodes a predicted transmembrane half-molecule ATP-binding cassette (ABC) transporter
Y110A2AL.13	Y110A2AL.13	
Y110A2AL.14	sqv-2	sqv-2 encodes a glycosaminoglycan galactosyltransferase II, biochemically active in vitro, that is required for cytokinesis of one-cell embryos and for vulval morphogenesis; the common requirement for SQV-2 in both cytokinesis and morphogenesis may be to promote filling an extracellular space with hygroscopic proteoglycans (either in the eggshell, or underneath the L4 cuticle), which in turn may cause the space to fill with fluid.
Y110A7A.18	ppw-2	
Y110A7A.19	Y110A7A.19	
Y113G7A.11	ssu-1	sulfotransferase
Y32G9A.1	gst-37	
Y43F4B.6	klp-19	klp-19 encodes a plus-end-directed microtubule motor protein that is most closely related to motors of the kinesin-4 family
Y45G12C.13	Y45G12C.13	
Y47D3B.1	Y47D3B.1	
Y47D3B.10	dpy-18	An alpha subunit of prolyl-4-hydroxylase which is a procollagen modifying enzyme required for exoskeleton formation, morphogenesis and maintenance of body shape; it is expressed throughout the hypodermis and certain head and posterior neurons.
Y47G6A.24	mis-12	mis-12 encodes the C. elegans homolog of human and Schizosaccharomyces pombe Mis12 protein; mis-12 activity is required for proper attachment of chromosomes to the mitotic spindle
Y48E1B.10	gst-20	
Y48G1A.5	imb-5	imb-5 encodes an importin-beta-like protein orthologous to mammalian CAS proteins (cellular apoptosis susceptibility) and Saccharomyces cerevisiae CSE1 (chromosome segregation 1); IMB-5 is predicted to function in nuclear transport of proteins required for mitotic progression or apoptosis as well as in re-export of importin-alpha, a nuclear import protein; in C. elegans, IMB-5 is essential for embryogenesis and required for normal pronuclear envelope dynamics, and may also play a role in vulval morphogenesis.
Y48G1BL.2	atm-1	atm-1 encodes an ortholog of human ATM (OMIM:208900) that is required for the checkpoint response to DNA damage; human ATM encodes a phosphatidylinositol-3 kinase homolog that is biochemically activated by cellular irradiation, and mutation of ATM leads to ataxia-telengiectasia.
Y50D4C.4	sqv-6	sqv-6 encodes a xylosyltransferase, active in cell culture, that is required for cytokinesis of one-cell embryos and for vulval morphogenesis
Y53F4B.29	gst-26	
Y53F4B.30	gst-27	
Y53F4B.31	gst-28	
Y53F4B.32	gst-29	
Y53F4B.33	gst-39	
Y53F4B.35	gst-31	
Y53F4B.37	gst-32	
Y54E10A.2	cogc-1	COG-1/ldlBp
Y54E10A.3	Y54E10A.3	
Y54E10BR.4	Y54E10BR.4	
Y55F3AM.15	csn-4	csn-4 encodes a protein with similarity to human COP9 proteasome subunit 4.
Y55F3AR.3	Y55F3AR.3	
Y65B4A.1	Y65B4A.1	
Y67D8C.5	Y67D8C.5	
Y71F9AL.1	Y71F9AL.1	
Y75B7B.2	Y75B7B.2	
Y80D3A.2	emb-4	Uncloned locus that is required maternally for embryogenesis.

Table 4.1 (continued)

Gene	Gene Public Name	Gene Description
Y80D3A.2	emb-4	Uncloned locus that is required maternally for embryogenesis.
ZK1005.1	pme-5	pme-5 encodes a poly (ADP-ribose) polymerase (PARP) that is a member of a conserved family of enzymes that catalyze: 1) the synthesis of poly (ADP-ribose), and 2) the covalent attachment of this polymer to glutamic acid residues of acceptor proteins such as histones and topoisomerases in order to regulate cellular processes such as maintenance of chromtin structure, programmed cell death, and DNA replication and repair
ZK1127.10	ZK1127.10	ZK1127.10 is orthologous to the human gene UNKNOWN (PROTEIN FOR MGC:9471) (CTH; OMIM:219500), which when mutated leads to disease.
ZK1127.4	ZK1127.4	
ZK1127.5	ZK1127.5	
ZK1127.7	ZK1127.7	
ZK1307.5	sqv-8	SQV-8 is homologous to three distinct glucuronyl transferases (GlcAT-I, GlcAT-P and GlcAT-D) that play a role in the synthesis of different glycoconjugates; the common requirement for SQV-8 in both cytokinesis and morphogenesis may be to promote filling an extracellular space with hygroscopic proteoglycans (either in the eggshell, or underneath the L4 cuticle), which in turn may cause the space to fill with fluid.
ZK1320.9	ZK1320.9	
ZK287.2	sulp-8	sulp-8 encodes one of eight C. elegans members of the sulfate permease family of anion transporters; by homology, SULP-8 is predicted to function as an anion transporter that regulates cellular pH and volume via transmembrane movement of electrolytes and fluids; a SULP-8::GFP fusion is expressed in the basolateral membrane of the excretory cell, intestine, and rectal gland cells.
ZK546.11	gst-30	
ZK637.10	trxr-2	
ZK697.6	gst-21	

Introduction of the mevalonate pathway

The mevalonate pathway is present in all higher eukaryotes and many bacteria and mediates the production of isoprenoids. The isoprenoids feed into a wide range of biosynthetic pathways: sterols, primarily cholesterol; dolichol, which serves as the lipid carrier of the oligosaccharide moiety destined for protein N-linked glycosylation; ubiquinone and heme A which function in the electron transport chain; prenylated proteins; and isopentenyl adenine, which is present in position 37 of tRNAs that read codons starting with U (Goldstein and Brown, 1990) (Figure 4.1). Cholesterol, the bulk product of the mevalonate pathway in humans and many other organisms, is important for membrane structure and steroid hormone synthesis. *C. elegans* possesses a functional mevalonate pathway but lacks all enzymes for the synthesis of sterols, suggesting that mevalonate in *C. elegans* is an important precursor for other biosynthetic

pathways (Morck et al., 2009). Although it does not synthesize cholesterol itself, *C. elegans* requires exogenously supplied cholesterol for growth and development (Gerisch et al., 2001; Yochem et al., 1999).

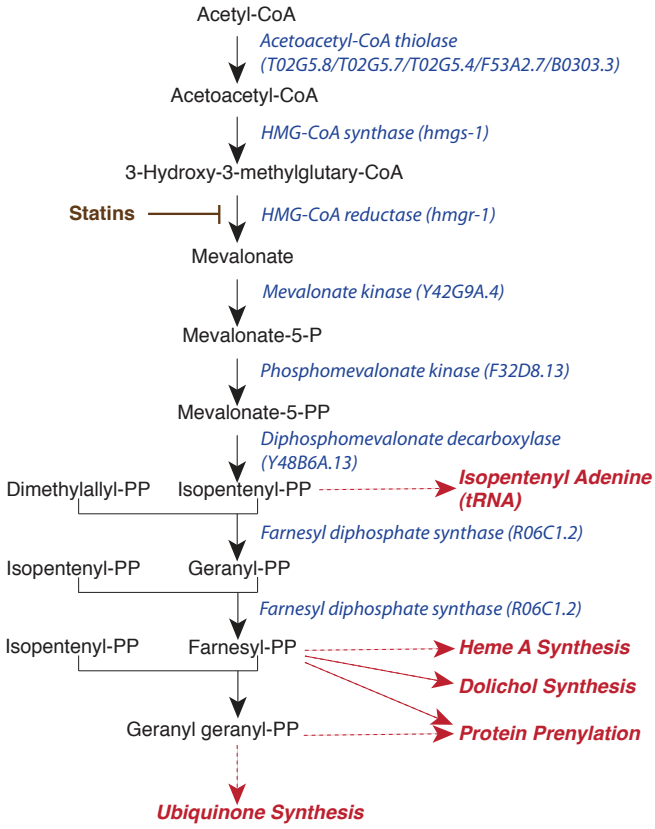


Figure 4.1 Diagram of the *C. elegans* mevalonate pathway.

hmgs-1/HMG-CoA synthase functions in the *let-7* miRNA pathway

Inactivation of *hmgs-1* by RNAi causes *let-7*-like phenotypes in three independent assays: first, the burst-through-vulva phenotype characteristic of *let-7* strong loss-of-function mutations (Figure 4.2A); second, the failure to express a reporter gene for an adult-specific collagen, *col-19::gfp* (Figure 4.2B); and third, the failure to produce alae, an adult-specific cuticle structure, at the nominal adult stage (Table 4.2).

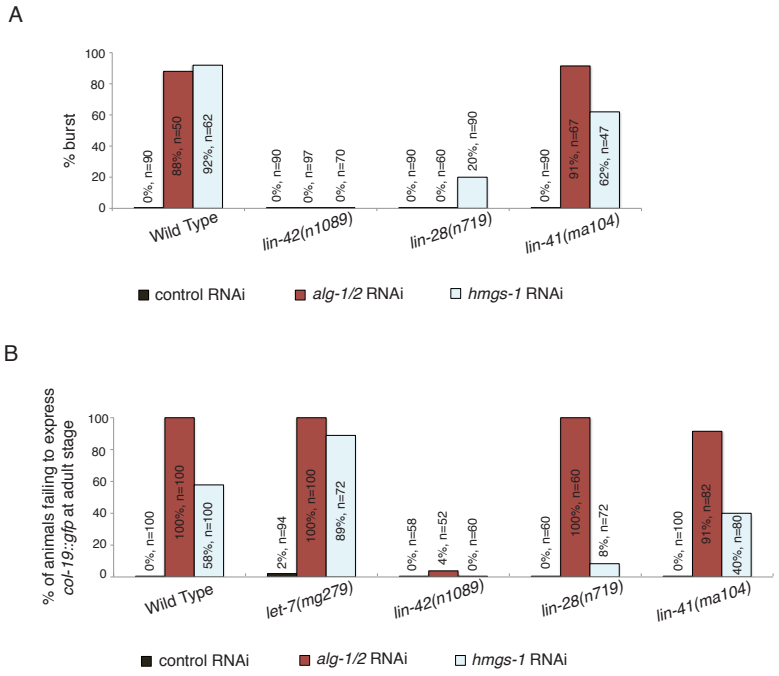


Figure 4.2 Inactivation of *hmgs-1* causes *let-7*-like phenotypes. (A) Shown are the percentage of animals that burst after the L4-to-adult molt upon treatment with control, *alg-1/2* or *hmgs-1* RNAi in the indicated genetic background. Inactivation of *alg-1/2* or *hmgs-1* causes bursting with high penetrance in the wild-type animals, but lower penetrance in the *lin-42(n1089)* or *lin-28(n719)* loss-of-function mutants. (B) Inactivation of *hmgs-1* causes animals to not express *col-19::gfp* in hyp7 cells at adult stage. The penetrance of this phenotype is elevated in the *let-7(mg279)* mutant and decreased in the *lin-42(n1089)*, *lin-28(n719)* or *lin-41(ma104)* loss-of-function mutants. Inactivation of *alg-1/2* also causes animals to not express *col-19::gfp* in hyp7 cells; this phenotype is suppressed in *lin-42(n1089)*.

Two lines of evidence further support that *hmgs-1* functions in the *let-7*-regulated heterochronic pathway. First, although *hmgs-1* inactivation causes relatively weak or incompletely penetrant retarded phenotypes on its own, these phenotypes are strongly enhanced in sensitized genetic backgrounds with compromised *let-7* activity (Table 4.2

and Figure 4.3). For example, upon *hmgs-1* inactivation, 9% of wild-type animals, but 100% of *alg-1(gk214)* and 67% of *ain-1(ku322)* mutants fail to produce adult-specific lateral alae (Table 4.2). These mutations in genes that encode the ALG-1/Argonaute protein or the AIN-1/*A*LG-1 *IN*teracting protein compromise miRNA function, and inactivation of *hmgs-1* is strongly synergistic with these mutations. Second, the retarded phenotypes caused by *hmgs-1* inactivation are suppressed by the loss-of-function of validated *let-7* target genes. For example, the retarded phenotypes of *hmgs-1* inactivation are completely suppressed by *lin-42(n1089)*, are partially suppressed by *lin-28(n719)*, and are weakly suppressed by *lin-41(ma104)*, a hypomorphic mutation that causes only weak precocious phenotypes (Figure 4.2). These data suggest that *hmgs-1* functions in the *let-7* pathway via the regulation of the activity of validated *let-7* target genes *lin-28*, *lin-41* and *lin-42*.

Table 4.2 The effect of *hmgs-1* inactivation in hypodermal cell fate speciation

			% of animals having the adult alae[a]			
	Strain	RNAi	No alae	Gapped	Complete	(n)
1	Wild Type	control	0	0	100	30
2	Wild Type	*hmgs-1*	9	3	88	32
3	*alg-1(gk214)*	control	45	55	0	20
4	*alg-1(gk214)*	*hmgs-1*	100	0	0	20
5	*alg-2(ok304)*	control	0	0	100	15
6	*alg-2(ok304)*	*hmgs-1*	21	29	50	24
7	*ain-1(ku322)*	control	10	38	52	21
8	*ain-1(ku322)*	*hmgs-1*	67	28	5	21

[a]The percentage of animals having no/gapped/complete alae structures were assessed after the L4-adult molt, only one side of each animal was assayed.

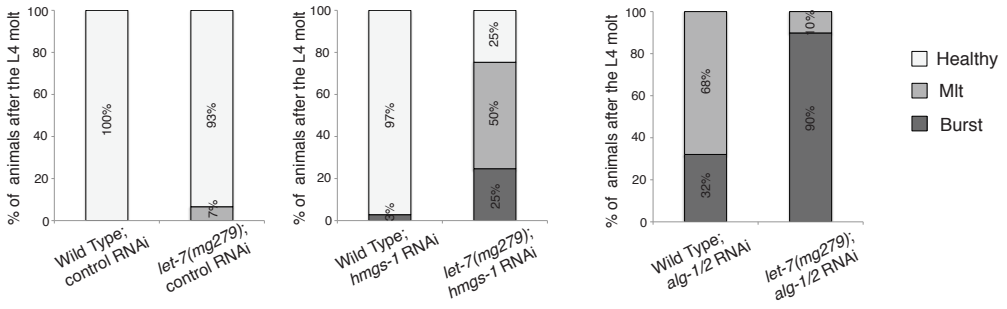


Figure 4.3 The percentage of Mlt and burst animals upon a mild knock-down of *hmgs-1* or *alg-1/2* by diluted RNAi is enhanced by the *let-7(mg279)* mutation. Shown are the percentage of healthy, molting defective (Mlt) and burst animals after the L4-to-adult molt in wild type and *let-7(mg279)* mutants. Mlt animals retain eggs following a defective molt and are eventually consumed by their progeny. Animals were fed starting at the larval stage one (L1) with *E. coli* expressing the *hmgs-1* or *alg-1/2* dsRNA diluted with control *E. coli* expressing a benign dsRNA and scored after the L4-adult molt. The percentage of burst and Mlt animals is enhanced in the *let-7(mg279)* mutant compared to wild type when fed with diluted *hmgs-1* or *alg-1/2* RNAi.

hmgs-1 is required for *let-7* family and *lin-4* miRNA silencing of target genes

To ask more directly whether *hmgs-1* functions in the miRNA pathway, we assayed whether miRNA target mRNAs become derepressed upon inactivation of this gene. We focused on the genetically verified targets of the *let-7* family and *lin-4* miRNAs. The hunchback factor *hbl-1 (H*unch*b*ack *L*ike*)* is silenced synergistically by the *let-7* family of miRNAs (*mir-48*, *mir-241* and *mir-84*) during the L2 to L3 stage transition (Abbott et al., 2005). Knocking down *hmgs-1* by RNAi prevents down-regulation of *hbl-1::gfp* at the L3 stage. This resembles the phenotype caused by mutations in the *let-7* family of miRNAs (Figure 4.4A). We assayed whether the silencing of *lin-14* by the *lin-4* miRNA during the late first larval stage (L1) is dependent on *hmgs-1*. LIN-14 protein levels become

derepressed at the late L1 stage by approximately two fold comparing *hmgs-1* RNAi-treated to stage-matched control animals, and the derepression is still apparent at the L2 stage (Figure 4.4B). To ask if the desilencing of *lin-14* upon inactivation of *hmgs-1* is due to reduced *lin-4* miRNA repression of *lin-14* via its 3' untranslated region (3'UTR), we analyzed the down-regulation of *lin-14* in the *lin-14(n355)* gain-of-function mutant, which lacks all the sites in the *lin-14* 3'UTR that are complementary to *lin-4* and *let-7* and its paralogs (Hayes and Ruvkun, 2005; Wightman et al., 1991). *lin-14* is not further desilenced when *hmgs-1* is inactivated in the *lin-14(n355)* mutant background (Figure 4.4C, top). *lin-14* is also not further desilenced in the *lin-4(e912)* null mutant upon inactivation of *hmgs-1* (Figure 4.4C, bottom). This indicates that the derepression of *lin-14* after *hmgs-1* inactivation is dependent on the *lin-14* 3'UTR and *lin-4* miRNA. Together, these results show that *hmgs-1* is required for silencing of *lin-14* by the *lin-4* miRNA.

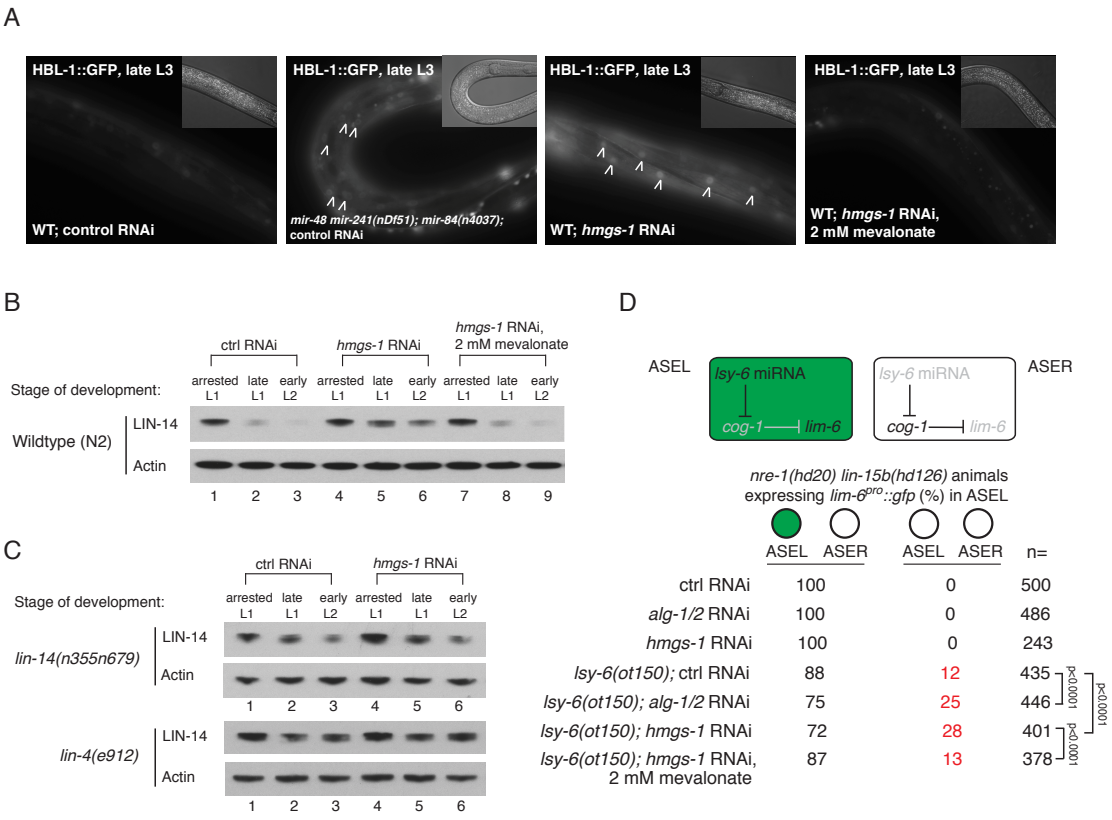


Figure 4.4 Inactivation of *hmgs-1* causes desilencing of miRNA target genes. (A) Inactivation of *hmgs-1* causes defects in the down-regulation of *hbl-1::gfp* at the L3 stage. This resembles the phenotype caused by mutations in the *let-7* family of miRNAs: *mir-48*, *mir-241* and *mir-84*. The desilencing of *hbl-1::gfp* upon *hmgs-1* RNAi is rescued by supplementing with 2 mM mevalonate. Images were captured using the same exposure settings and processed identically. Arrowheads point to the desilenced *hbl-1::gfp* in the nuclei of hyp7 cells. Insets are Nomarski images. (B-C) Immunoblots. Actin probed as a control for even loading. (B) Inactivation of *hmgs-1* causes defects in the down-regulation of *lin-14*, the target of the *lin-4* miRNA, at the late L1 stage (lane 5) and early L2 stage (lane 6). Mevalonate supplementation rescues this phenotype (lanes 8 and 9). (C) Inactivation of *hmgs-1* does not further desilence *lin-14* in the *lin-14(n355n679)* mutant lacking the *lin-14* 3'UTR or the *lin-4(e912)* null mutant (comparing lanes 5 and 6 to lanes 2 and 3). (D) The *lsy-6* miRNA is expressed in the ASEL but not ASER neuron in wild type. It is required for ASEL specification, judged by the expression of *lim-6^{pro}::gfp* in ASEL,

(Figure 4.4 continued) which is promoted by down-regulation of *cog-1* by *lsy-6*. In a sensitized genetic background with a weak allele of *lsy-6*, *ot150*, inactivation of *alg-1/2* or *hmgs-1* significantly enhanced the ASEL specification defect. Supplementing 2 mM mevalonate rescued the phenotype of *hmgs-1* inactivation. Brackets indicate statistically significant difference judged by two-tailed chi-square test.

hmgs-1 acts in miRNA pathways in other cell types as well

hmgs-1 also modulates the activity of miRNAs whose functions are unrelated to developmental timing. *lsy-6* is a miRNA that regulates the specification of the taste neurons, ASE left (ASEL) and ASE right (ASER), which even though they are bilaterally symmetric express distinct patterns of receptor genes based on the asymmetric activity of the *lsy-6* miRNA (Chang et al., 2003; Johnston and Hobert, 2003). Specifically expressed in less than ten neurons including ASEL but not ASER, *lsy-6* down-regulates the *cog-1* transcription factor only in ASEL, thus distinguishing the gene expression profile of ASEL from ASER (Chang et al., 2003; Johnston and Hobert, 2003). The ASEL neuron of *lsy-6(ot71)* null mutants fails to down-regulate *cog-1* and adopts the ASER pattern of gene expression as a result. On the other hand, animals bearing a hypomorphic allele of *lsy-6*, *ot150*, display the ASEL specification defect with incomplete penetrance. Inactivation of genes that are key for miRNA activity, e.g. *nhl-2*, significantly enhances the ASEL fate specification defect in the *lsy-6(ot150)*, but not in the wild-type background (Hammell et al., 2009). We asked whether knocking down *hmgs-1* causes an ASEL specification defect by scoring *lim-6^{pro}::gfp*, an ASEL-specific reporter. To enhance the efficiency of RNAi in neurons, we crossed the *lim-6^{pro}::gfp* reporter into the RNAi-hypersensitive *nre-1(hd20) lin-15b(hd126)* mutant background (Schmitz et al., 2007). RNAi was initiated at the L3 stage of *lsy-6(ot150)*; *nre-1(hd20) lin-15b(hd126)*

parental (P_0) animals, and *lim-6^{pro}::gfp* was scored in the progeny. 28% of *hmgs-1* RNAi-treated animals (n=401), compared to 12% of control RNAi-treated animals (n=435) showed the ASEL specification defect (Figure 4.4D). However, the progeny of *nre-1(hd20) lin-15b(hd126)* animals with the wild-type *lsy-6* gene did not show any ASEL specification defect upon inactivation of *hmgs-1*. Similar results were obtained by knocking down *alg-1/2* by RNAi. This result supports a requirement for *hmgs-1* for the efficient down-regulation of *cog-1* by the *lsy-6* miRNA.

Taken together, these observations suggest that *hmgs-1* modulates the function of many and perhaps all miRNAs in multiple tissues, and at multiple stages during development.

***hmgs-1* acts downstream of miRNA biogenesis and loading of ALG-1/Argonaute**

We asked whether *hmgs-1* is required for miRNA biogenesis/accumulation or activity. To distinguish between these possibilities, we first assayed the mature miRNA levels by real-time PCR. The levels of *let-7, lin-4* and *mir-55* all remained unchanged upon knocking down *hmgs-1* (Figure 4.5A) despite the fact that the targets of *let-7* and *lin-4* became derepressed. We also assayed the *let-7* level in the *let-7(mg279)* mutant, which has a reduced level of mature *let-7* miRNA resulting from defects in the splicing and processing of the *let-7* transcript (Bracht et al., 2004). Inactivation of *hmgs-1* did not reduce the level of mature *let-7* even in this sensitized genetic background (data not shown). In contrast, knocking down *alg-1/2* caused a significant reduction of miRNA levels, consistent with its role in miRNA biogenesis and stability.

To assay whether *hmgs-1* regulates the competence of ALG-1/Argonaute in loading miRNAs, we purified miRISC from synchronized L4-stage *alg-1(gk214)* mutants rescued with an HA-ALG-1 single-copy construct. The HA-ALG-1-bound *let-7*, *lin-4* and

mir-55 levels remained unchanged upon inactivation of *hmgs-1* (Figure 4.5B and C), indicating ALG-1 loading is unaltered. We also found that the guide/passenger-strand ratio of these miRNAs remained unchanged when *hmgs-1* was inactivated (Figure 4.5D). Together, these results position *hmgs-1* downstream of miRISC loading and duplex unwinding. It suggests that one or multiple downstream steps, for example the competence of miRISC in finding and silencing its target mRNAs, are dependent on *hmgs-1*.

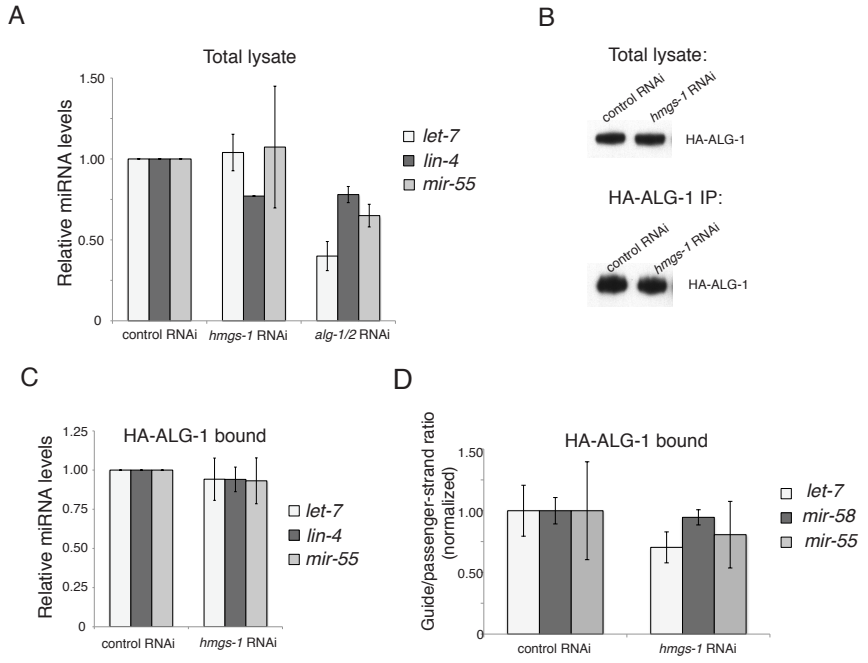


Figure 4.5 *hmgs-1* acts downstream of miRNA biogenesis/accumulation and loading of ALG-1. (A) Shown are the mature miRNA levels in total worm lysate, determined by real-time PCR. The miRNA levels are reduced upon *alg-1/2* inactivation, but remain unchanged upon *hmgs-1* inactivation. (B-D) HA-ALG-1 was immunoprecipitated from animals treated with control or *hmgs-1* RNAi and the level of HA-ALG-1-bound miRNAs was determined. (B) Equal amount of HA-ALG-1 was purified from the control and *hmgs-1* RNAi-treated animals. Shown is the Western blot of HA-ALG-1 in total lysate (top) and from HA-ALG-1 immunoprecipitation (IP) (bottom). (C) Relative levels of *let-7*, *lin-4* and *mir-55* bound by HA-ALG-1: they remain unchanged upon

(Figure 4.5 continued) *hmgs-1* inactivation. In (A) and (C), for each miRNA, the result is shown relative to its level in animals treated with control RNAi. (D) The guide/passenger-strand ratio of *let-7*, *mir-58* and *mir-55* was unaltered after inactivation of *hmgs-1*. The results were normalized to the mean guide/passenger-strand ratio in control RNAi-treated animals. The mean and standard deviation was calculated from three biological replicates. Error bars represent SEM.

We also surveyed whether *hmgs-1* regulates the protein level and/or cellular localization of the core miRNA cofactors. We monitored ALG-1/Argonaute and AIN-1/*A*LG-1 *I*N*t*eracting protein. Neither the overall expression level nor the subcellular localization of GFP::ALG-1 or AIN-1::GFP was altered upon knocking down *hmgs-1* (Figure 4.6).

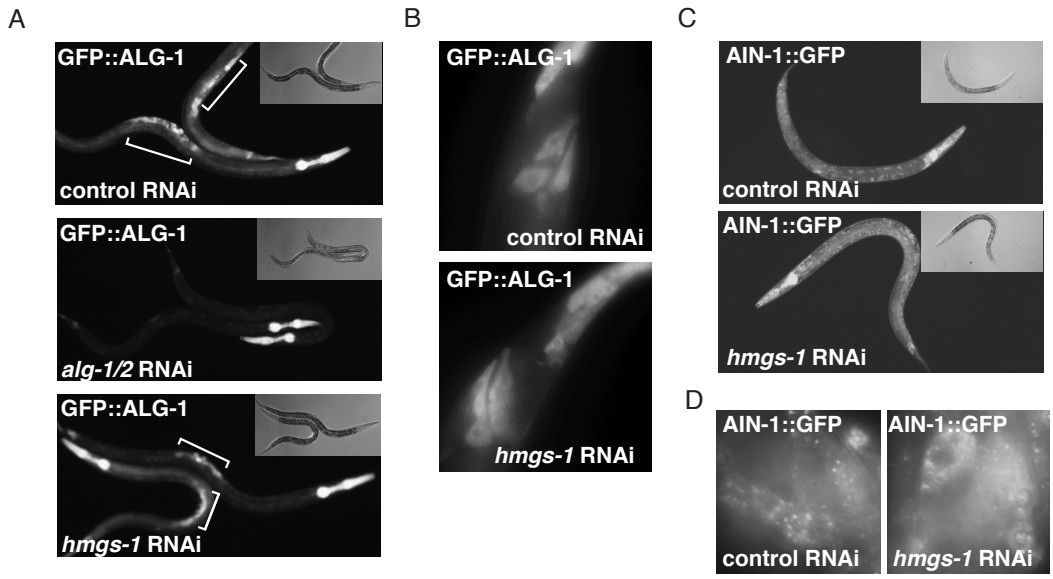


Figure 4.6 *hmgs-1* does not regulate the overall expression pattern or subcellular localization of ALG-1/Argonaute and AIN-1/*A*LG-1*I*N*t*eracting proteins. (A) Global GFP::ALG-1 expression in control, *alg-1/2* RNAi or *hmgs-1* RNAi-treated L4 animals. Brackets indicate the vulval and somatic gonadal expression of GFP::ALG-1 in control and *hmgs-1* RNAi but not *alg-1/2* RNAi-treated animals. (B) Shown are several cells in the tail region. GFP::ALG-1 is largely diffuse

(Figure 4.6 continued) in the cytoplasm. This pattern is not affected upon RNAi depletion of *hmgs-1*. (C) AIN-1::GFP is ubiquitously expressed in L4 animals, with highest expression in the head neurons. This pattern is not affected upon RNAi depletion of *hmgs-1*. (D) Shown are several cells in the tail region. AIN-1::GFP exhibits punctate cellular localization, and this pattern is not affected upon RNAi depletion of *hmgs-1*.

The non-cholesterol output of the mevalonate pathway modulates miRNA activity

Humans and some other organisms have two forms of HMG-CoA synthase: the cytosolic form, which acts in the mevalonate pathway, and the mitochondrial form, which functions in the production of ketone bodies during starvation. *C. elegans* bears just the *hmgs-1* ortholog of HMG-CoA synthase, which is predicted to be cytosolic. Therefore, we hypothesized that the isoprenoid output of the mevalonate pathway has a role in miRNA activity. Three strands of evidence support this hypothesis:

First, we reasoned that if the retarded phenotypes caused by *hmgs-1* inactivation are due to reduced biosynthetic outputs of the mevalonate pathway, then supplementing mevalonate, the downstream product of HMG-CoA synthase, should rescue these phenotypes. Indeed, mevalonate supplementation completely rescued all retarded phenotypes caused by inactivation of *hmgs-1* (Figure 4.7A, Table 4.3), but did not rescue the retarded phenotypes induced, for example, by inactivation of the Argonaute gene, *alg-1/2* (Table 4.3). Mevalonate supplementation also rescued the desilencing of *hbl-1::gfp* and *lin-14* in *hmgs-1* RNAi-treated animals (Figure 4.4A and B). Furthermore, the ASEL specification defect upon knocking down *hmgs-1* was also rescued by mevalonate (Figure 4.4D). In contrast, supplementation to even 50 μg/ml cholesterol (the standard Nematode Growth Medium contains 5 μg/ml cholesterol) did not rescue any of the retarded phenotypes caused by knocking down *hmgs-1* (Table 4.3). This indicates

that instead of cholesterol, other biosynthetic products of the mevalonate pathway modulate miRNA activity.

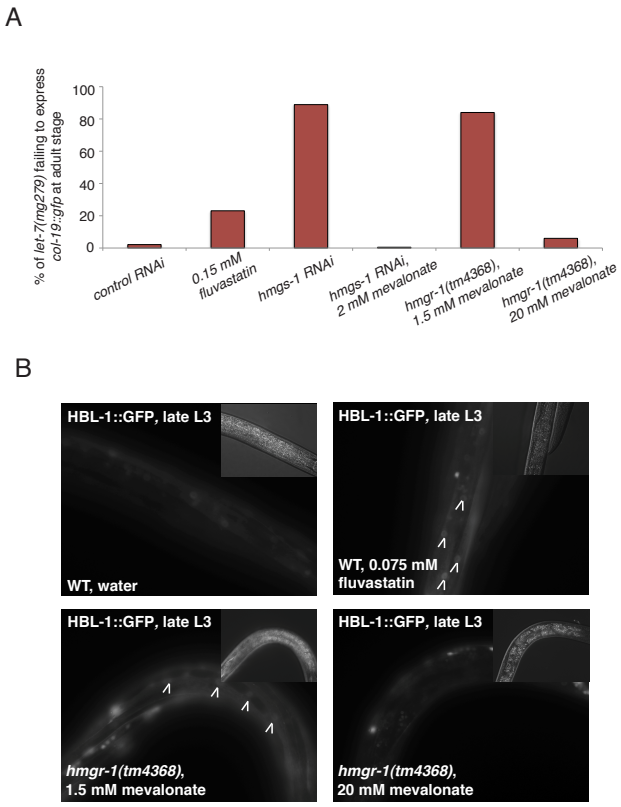


Figure 4.7 The mevalonate pathway modulates miRNA activity. (A) Inactivation of the mevalonate pathway by either application of fluvastatin, inactivation of *hmgs-1* by RNAi or mutation of *hmgr-1,* with a low level of mevalonate (1.5 mM) supplied in the medium, causes *let-7(mg279)* mutants to fail to express *col-19::gfp* in hyp7 cells. This phenotype can be rescued by supplementing mevalonate. (B) Inactivation of the mevalonate pathway by application of fluvastatin or mutation of *hmgr-1,* with a low amount of mevalonate (1.5 mM) supplied in the medium, causes defects in the down-regulation of *hbl-1::gfp* at the L3 stage. However, *hmgr-1(tm4368)* animals growing on high mevalonate (20 mM) show wild-type down-regulation of *hbl-1::gfp*. Images were captured using the same exposure settings and processed identically. Arrowheads point to the desilenced *hbl-1::gfp* in the nuclei of hyp7 cells. Insets are Nomarski images.

Table 4.3 Mevalonate supplementation rescues gene inactivation of *hmgs-1*

	No supplement	55 μg/ml cholesterol	2 mM mevalonate
control RNAi	healthy gravid adult	healthy gravid adult	healthy gravid adult
alg-1/2 RNAi	burst	burst	burst
hmgs-1 RNAi	burst	burst	healthy gravid adult (complete rescue)

Second, HMG-CoA reductase (encoded by *hmgr-1*) also acts in the miRNA pathway. HMG-CoA reductase is the rate-limiting enzyme that acts immediately downstream of HMG-CoA synthase in the production of mevalonate. The *C. elegans hmgr-1(tm4368)* mutant strain bears a 620-bp deletion that spans the first three exons, causing a likely null mutation in this gene. The homozygous *hmgr-1(tm4368)* mutants that segregate from a heterozygote arrest at the L1 stage. However, if mevalonate is added to the growth media to 20 mM final concentration, the homozygous *hmgr-1(tm4368)* mutants are viable and fertile. We found that *hmgr-1(tm4368); let-7(mg279)* mutants grown with low (1.5 mM) mevalonate failed to express *col-19::gfp*, a defect that was not observed when mevalonate was increased to 20 mM (Figure 4.7A). In addition, *hmgr-1(tm4368)* mutants grown with 1.5 mM mevalonate failed to properly down-regulate *hbl-1::gfp* at the L3 stage, a phenotype that was also rescued by a higher concentration of mevalonate (Figure 4.7B).

Third, we found that statins, cholesterol-lowering drugs that inhibit HMG-CoA reductase activity, can compromise *let-7* activity. When fluvastatin was added to the growth medium, it caused *let-7(mg279)* animals to fail to express *col-19::gfp* at the adult stage (Figure 4.7A). Fluvastatin also prevented the proper down-regulation of *hbl-1::gfp* at the L3 stage (Figure 4.7B), similar to RNAi depletion of *hmgs-1* or the *hmgr-1(tm4368)* mutation.

Together, the above results strongly indicate that in *C. elegans*, the non-sterol outputs of the mevalonate pathway modulate miRNA activity.

miRNAs in *C. elegans* are unlikely to bear abundant isopentenyl modification

Isoprenoids, the end products of the mevalonate pathway, feed into a wide range of downstream pathways in addition to the better known synthesis of sterols (Goldstein and Brown, 1990). This includes protein prenylation; tRNA isopentenylation; and biosynthesis of ubiquinone, heme A and dolichol. Inspired by our initial motivating ideas and supported by the fact that tRNA A37 bears an isopentenyl modification, I hypothesized that miRNAs could be modified by isopentenyl moiety synthesized from the mevalonate pathway.

The isopentenyl modification of tRNA at A37 has been suggested to stabilize the adjacent A36 base pairing with the codon starting with U; therefore reduce the fatal first position misreading by inhibiting the wobble capacity (Persson et al., 1994; Robins et al., 1967). We therefore hypothesized that the mevalonate pathway modulates miRNA activity via a similar isopentenyl modification of miRNAs, which might stabilize the base pairing between the miRNA and its target mRNA, or facilitate its sorting to appropriate cellular compartments. We tested this hypothesis by purifying *C. elegans* 18-28nt small RNA, then digesting the RNA with nuclease P1 to single nucleotide for mass spectrometry analysis. Although we can clearly detect the N6-(Δ2-isopentenyl)adenosine from nuclease P1 treated tRNAs as a positive control (Figure 4.8A and C), we did not detect this species from nuclease P1 treated *C. elegans* small RNA (Figure 4.8B). Because miRNAs represent ~30% of 18-28 nt small RNAs, we concluded that the majority of *C. elegans* miRNAs are not isopentenyl modified.

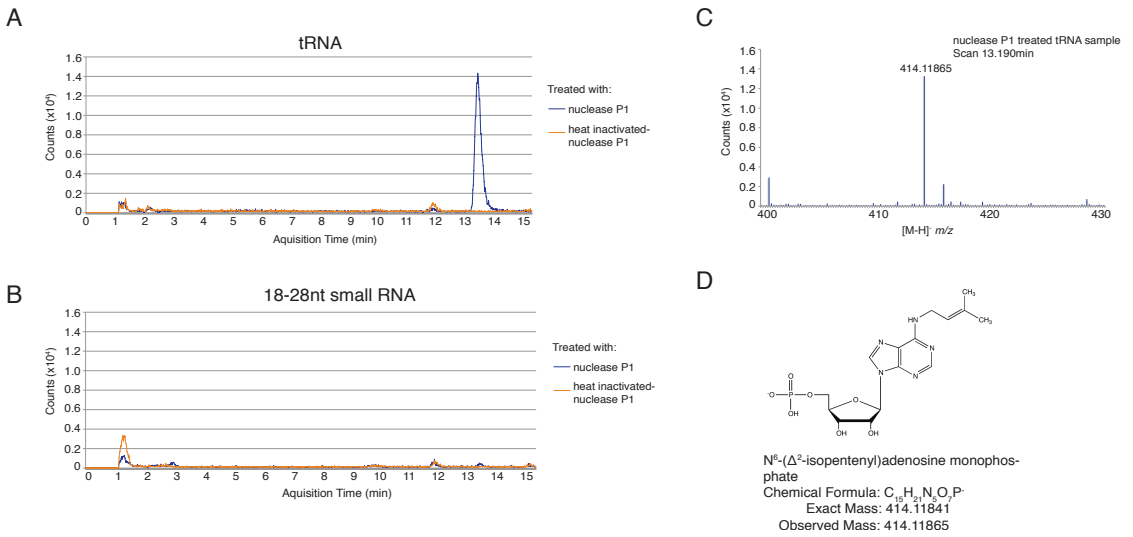


Figure 4.8 No isopentenyl-modified miRNAs were detected in *C. elegans*. (A) The extracted ion chromatogram (EIC) of yeast tRNA sample at $[M-H]^-$ *m/z*=414.11841, the expected mass of N^6-(Δ^2-isopentenyl)adenosine. This species is only detected in the nuclease P1 treated, but not heat inactivated nuclease P1 treated tRNA sample. (B) The EIC of *C. elegans* 18-28nt small RNA sample at $[M-H]^-$ *m/z*=414.11841. This species cannot be detected. (C) The $[M-H]^-$ *m/z*=414.11865 species is clearly and only detected in the nuclease P1 treated tRNA sample. It elutes at 13.190 min. (D) The chemical structure, formula, exact mass and observed mass of N^6-(Δ^2-isopentenyl)adenosine.

The dolichol pathway for protein N-linked glycosylation is required for miRNA activity

To further delineate which downstream biosynthetic output of the mevalonate pathway modulates miRNA activity, we inactivated genes corresponding to each step in the production and usage of isoprenoids and surveyed the phenotypes. We screened for which of these gene inactivations caused *let-7(mg279)* animals to fail to express *col-19::gfp* (Table 4.4).

Table 4.4 Phenotypes of gene inactivations in the *let-7(mg279); [col-19::gfp]* background

Gene Targeted[a]	Locus	Description	Phenotypes in P0 animals[b]	Phenotypes in F1 animals[c]
Mevalonate pathway				
T02G5.8	*kat-1*	Acetyl-CoA acetyltransferase		
T02G5.7		Acetyl-CoA acetyltransferase		
T02G5.4		Acetyl-CoA acetyltransferase		
F53A2.7		Acetyl-CoA acetyltransferase		
B0303.3		Acetyl-CoA acetyltransferase		
F25B4.6	*hmgs-1*	HMG-CoA synthase	Burst and sterile. Fail to express *col-19::gfp*	nd
F08F8.2	*hmgr-1*	HMG-CoA reductase		Weak *col-19::gfp*
Y42G9A.4	*mvk-1*	mevalonate kinase		
Y48B6A.13		Mevalonate pyrophosphate decarboxylase		
R06C1.2	*fdps-1*	Polyprenyl synthetase		Weak *col-19::gfp*
Protein prenylation				
R02D3.5	*fnta-1*	farnesyltransferase, alpha subunit		Burst
F23B12.6	*fntb-1*	beta subunit of farnesyltransferase		
Y48E1B.3		geranylgeranyltransferase Type I, beta subunit		
M57.2		geranylgeranyltransferase type II, alpha subunit		Arrested at L3
B0280.1	*ggtb-1*	geranylgeranyltransferase type II, beta subunit		
Coenzyme Q biosynthesis				
C24A11.9	*coq-1*	trans-prenyltranferases		Arrested at L2
F57B9.4	*coq-2*			
Y57G11C.11	*coq-3*			
K07B1.2	*coq-6*			
ZC395.2	*clk-1*			
Dolichol synthesis and N-glycosylation				
Y60A3A.14		ALG7 homolog		Pale-looking, weak *col-19::gfp*
R10D12.12		UDP-N-acetylglucosamine transferase subunit ALG13 homolog		
M02B7.4		UDP-N-acetylglucosamine transferase subunit ALG14 homolog		
T26A5.4		ALG1 homolog		
F09E5.2		ALG2 homolog		
B0361.8		ALG11 homolog		
K09E4.2		Dolichyl-P-Man:Man(5)GlcNAc(Cvetkovic et al.)-PP-dolichyl mannosyltransferase		
C14A4.3		Mannosyltransferase ALG9 homolog		
ZC513.5		Mannosyltransferase ALG12 homolog		
C08B11.8		Glucosyltransferase ALG6 homolog		
C08H9.3		Glucosyltransferase ALG8 homoolog		
T24D1.4	*tag-179*	Alpha-1,2 glucosyltransferase ALG10 homolog		
T22D1.4		Oligosaccharyltransferase, alpha subunit (ribophorin I)	Fail to express *col-19::gfp*	nd
M01A10.3	*ostd-1*	Oligosaccharyltransferase subunit. Ortholog of yeast SWP1, human Ribophorin II	Fail to express *col-19::gfp*	nd
T09A5.11	*ostb-1*	Oligosaccharyltransferase subunit. Ortholog of yeast WBP1, human OST48	Fail to express *col-19::gfp*	nd
F57B10.10	*dad-1*	Oligosaccharyltransferase subunit. Ortholog of yeast OST2, human DAD-1	Fail to express *col-19::gfp*	nd
T12A2.2		Oligosaccharyltransferase, STT3 subunit	Fail to express *col-19::gfp*	nd
Heme A and Heme O biosynthesis				
Y46G5A.2		protoheme IX farnesyltransferase		
T06D8.5		Cytochrome oxidase assembly factor COX15		
tRNA isopentenylation				
ZC395.6	*gro-1*	tRNA -isopentenylpyrophosphate transferase		

Table 4.4 (continued)

(nd) Not Determined

[a]The gene targeted by the RNAi clone was confirmed by sequencing.

[b]RNAi was initiated started at the L1 stage, and the phenotypes were scored after the L4-adult molt.

[c]RNAi was initiated started at the L1 stage of parental (P_0) animals, and the phenotypes were scored in progeny. However, the phenotypes were not determined if the RNAi causes lethality/sterility in P_0 animals.

If animals appeared WT after RNAi, cell is left empty.

The gene inactivations that caused a *let-7*-like phenotype, the failure to up-regulate *col-19::gfp* expression at the adult stage, all correspond to proteins that act in the dolichol pathway for protein N-linked glycosylation (Table 4.4. For a diagram of the dolichol pathway, see Figure 4.9).

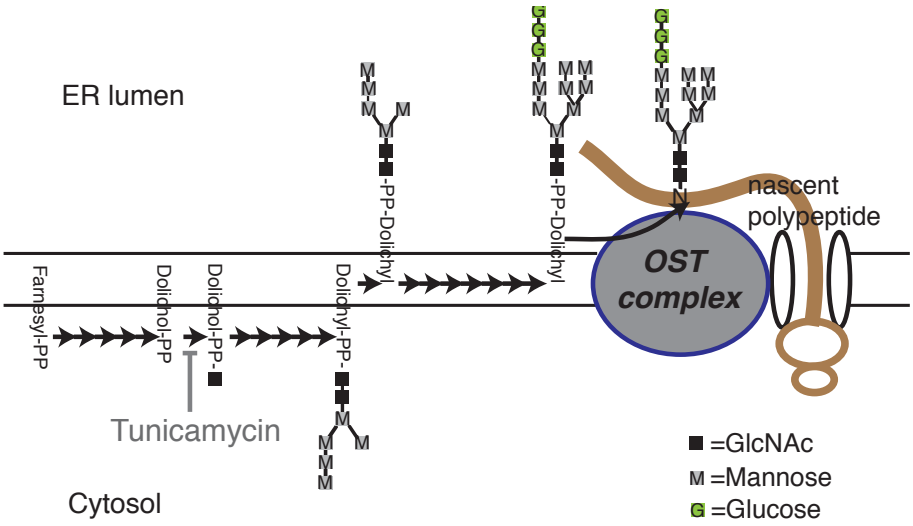


Figure 4.9 Dolichol phosphate is synthesized from the mevalonate pathway and has a role in protein N-linked glycosylation. Dolichol phosphate synthesized from the mevalonate pathway serves as the lipid carrier of the oligosaccharide moiety destined for protein N-linked glycosylation. The transfer of oligosaccharide to an asparagine residue on a nascent polypeptide is catalyzed by oligosaccharyltransferase (OST) complex on the ER membrane. Tunicamycin blocks the reaction of UDP-GlcNAc and dolichol-phosphate in the first step of the dolichol pathway.

The *C. elegans* oligosaccharyltransferase (OST) complex, which carries out protein N-glycosylation, has five subunits: the catalytic subunit T12A2.2/STT3, and four accessory subunits: T22D1.4/ribophorin I, OSTB-1, OSTD-1 and DAD-1. Depleting any of these subunits by RNAi caused a defect in *col-19::gfp* expression in the *let-7(mg279)* mutant, but not the wild-type background (Figure 4.10A), suggesting a strong genetic interaction with *let-7*. To further validate this result, we performed the same assay using tunicamycin. Tunicamycin is an antibiotic that blocks the reaction of UDP-GlcNAc and dolichol phosphate in the first step of the dolichol pathway and thus inhibits the synthesis of N-linked glycoproteins. When tunicamycin was added to the worm growth medium, it caused *let-7(mg279)* adult animals to fail to express *col-19::gfp*, in a dose-dependent manner (Figure 4.10B). More directly, we found that inactivation of *T12A2.2/STT3* disrupted down-regulation of the *hbl-1::gfp* reporter at the L3 stage (Figure 4.10C). Furthermore, knocking down *T12A2.2/STT3* also exacerbated the ASEL neuron specification defect in the *lsy-6(ot150)* mutant (Figure 4.10D), similar to the inactivation of the mevalonate pathway.

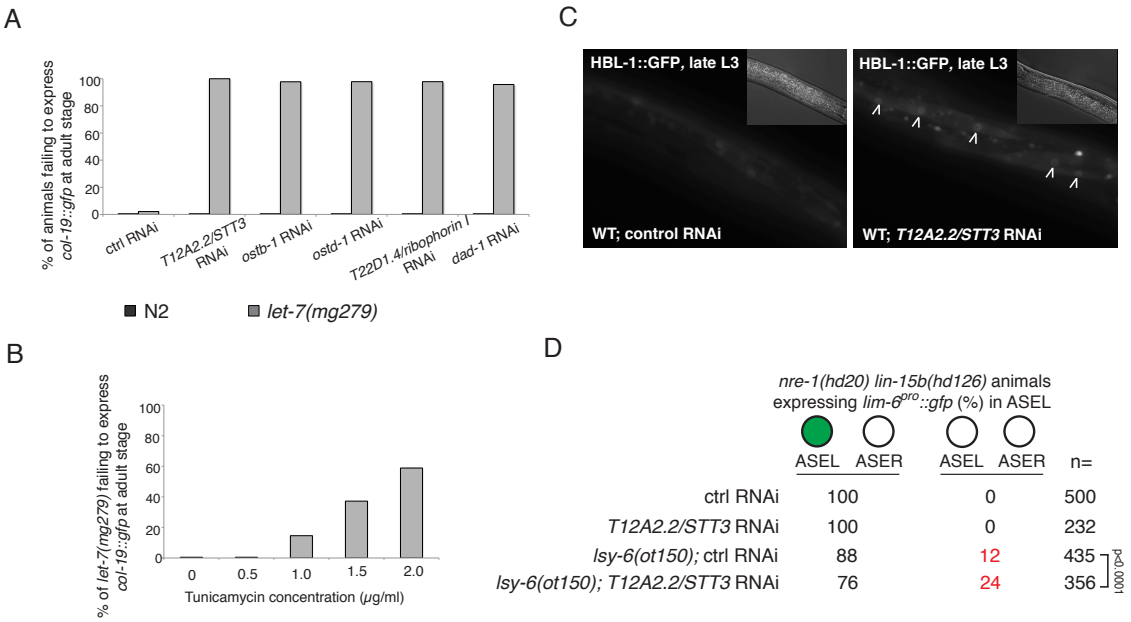


Figure 4.10 The dolichol pathway for protein N-glycosylation is required for miRNA activity. (A) Inhibiting the oligosaccharyltransferase (OST) activity by RNAi depletion of any of its five subunits causes defects in the nominal adult-stage up-regulation of *col-19::gfp* in hyp7 cells, in a *let-7(mg279)* but not wild-type background. (B) Tunicamycin causes defects in the nominal adult-stage up-regulation of *col-19::gfp* in the *let-7(mg279)* mutant in a dose-dependent manner. (C) Inhibiting the OST activity by RNAi depletion of its catalytic subunit, *T12A2.2/STT3*, causes defects in the down-regulation of *hbl-1::gfp* at the L3 stage. Images were captured using the same exposure settings and processed identically. Arrowheads point to the desilenced *hbl-1::gfp* in the nuclei of hyp7 cells. Insets are Nomarski images. (D) RNAi depletion of *T12A2.2/STT3* enhances the ASEL specification defect in the *lsy-6*(*ot150)* mutant but not wild-type genetic background. Brackets indicate a statistically significant difference judged by two-tailed chi-square test.

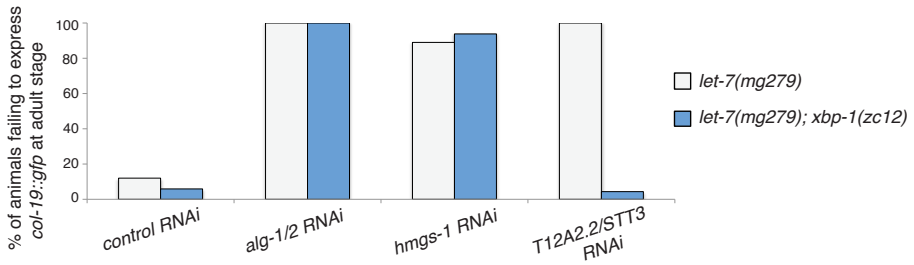
Because N-glycosylation facilitates protein folding in the endoplasmic reticulum (ER), blockage of N-glycosylation by RNAi, mutation or drug causes protein misfolding, which induces ER stress and the unfolded protein response (Shiu et al.). We asked if the

desilencing of miRNA target genes is an element of the unfolded protein response. We found that several gene inactivations that strongly induce ER UPR did not prevent proper down-regulation of *hbl-1::gfp* at the L3 stage (Figure 4.11A). However, several gene inactivations that induce ER UPR did enhance the failure to express *col-19::gfp* in the *let-7(mg279)* mutant, and a mutation in the unfolded protein response factor *xbp-1 (Calfon et al., 2002)* suppressed this defect (Figure 4.11B). The *xbp-1(zc12)* mutation also suppressed the failure to express *col-19::gfp* induced by inactivation of *T12A2.2/STT3*, but not the failure to express *col-19::gfp* induced by inactivation of *alg-1/2* or *hmgs-1* (Figure 4.11B). This suggests that the mevalonate pathway is required for miRNA activity for more than one reason: first, a relatively direct role of N-glycosylation in miRISC function possibly via regulating the sorting of miRISC to appropriate cellular membrane compartment; and second, a relaying signaling cascade downstream of ER UPR in opposing miRNA activity (Figure 4.11C).

A

Gene Inactivation	Discription	*hsp-4::gfp* expression	Defect in down-regulation of *hbl-1::gfp* [% (n)][a]	Defect in *col-19::gfp* expression [% (n)][b]	
				let-7(mg279)	*xbp-1(zc12); let-7(mg279)*
control		very low	0% (12)	13% (195)	0% (110)
Y54E10BR.5	signal peptidase	high	0% (12)	83% (86)	0% (88)
arf-3	ADP-ribosylation factor related protein	high	0% (12)	80% (85)	72% (100)
fat-6	delta-9 fatty acid desaturase	high	0% (8)	10% (80)	65% (82)
fah-1	putative fumarylacetoacetate hydrolase	medium	0% (10)	38% (77)	0% (87)
pdi-3	disulfide isomerase	medium	0% (12)	87% (83)	0% (90)
sams-1	S-adenosylmethionine synthetase	high	nd	6% (73)	0% (82)
tkt-1	Transketolase	high	nd	1% (100)	0% (87)
ero-1	ER-associated oxidoreductin	high	nd	10% (90)	0% (83)

B



C

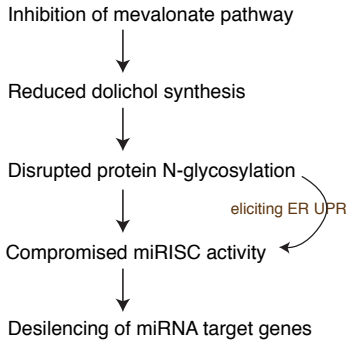


Figure 4.11 Induction of ER stress mildly compromises *let-7* activity. (A) Inactivation of genes required for ER homeostasis causes up-regulation of *hsp-4/BiP::gfp*, the hallmark of ER stress. These gene inactivations do not cause a defect in the down-regulation of *hbl-1::gfp* at the L3 stage. However, a subset of gene inactivations cause *let-7(mg279)* mutant animals to fail to express *col-19::gfp* in hyp7 cells at the adult stage. In some but not all cases, this defect of *col-19::gfp* expression can be suppressed by the *xbp-1(zc12)* mutation. (nd) Not Determined. [a]The expression of *hbl-1::gfp* in the hyp was scored at the late L3 stage, and the precentage of animals having derepressed *hbl-1::gfp* is indicated. [b]The expression of *col-19::gfp* was scored at

(Figure 4.11 continued) the adult stage, and the precentage of animals failing to express *col-19::gfp* in hyp7 cells is indicated. (B) Shown are the percentage of animals failing to express *col-19::gfp* in hyp7 cells at the adult stage. The *xbp-1(zc12)* mutation suppresses the failure of *col-19::gfp* expression in *T12A2.2/STT3* RNAi-treated *let-7(mg279)* mutants. However *xbp-1(zc12)* does not suppress the failure of *col-19::gfp* expression caused by the inactivation of *alg-1/2* or *hmgs-1*. (C) The dolichol phosphate/protein N-glycosylation output of the mevalonate pathway is required in the miRNA repression of target mRNAs in *C. elegans*.

Discussion and future directions

We showed that inactivation of the mevalonate pathway does not decrease the biogenesis nor the loading of miRNAs into ALG-1/Argonaute. The most likely hypothesis is that mevalonate pathway might be important for the membrane-association of ALG-1/Argonaute. This hypothesis gained support from earlier studies of the Argonaute protein, which was named as GERp95 (Golgi ER protein 95kDa) (Cikaluk et al., 1999; Tahbaz et al., 2001) and was found to be a peripheral membrane protein located on the Golgi and/or ER in several mammalian cell lines. In remains mysterious how Argonaute associates with the membrane and how it is specifically targeted to certain membrane compartments. We therefore hypothesized that some membrane proteins recruit Argonaute, and such membrane proteins may depend on N-glycosylation for its proper localization and function. To this end, we purified *C. elegans* ALG-1/Argonuate by immunoprecipitation and asked if there is any co-immunoprecipitated glycosylated protein(s). However, we were unable to obtain conclusive results, due to the low sensitivity of the detection method (data not shown). On the other hand, it is possible that other biosynthetic products of the mevalonate pathway mediate the membrane

association of Argonaute. As a matter of fact, the Voinnet group showed that cholesterol and some other isoprenoids are required for the membrane association of Argonaute 1 in *Arabidopsis* (Brodersen et al., 2012). Unlike many other organisms, *C. elegans* lack the cholesterol biosynthesis downstream of mavalonate pathway, and thus excludes the possibility of cholesterol being involved here.

What might be the biological function of the membrane-association of Argonaute? Recent studies revealed that the RNA-Induced Silencing Complex (RISC) co-localizes with the Multi-Vesicular Bodies (MVB) (Gibbings et al., 2009). Indeed, the Endosomal Sorting Complex Required for Transport (ESCRT) complex responsible for the MVB assembly was found to be required for the miRNA-mediated silencing (Gibbings et al., 2009; Lee et al., 2009). Therefore, it is an attractive hypothesis that the membrane association of RISC may facilitate their recycling, or the trafficking of small RNAs between cells through exosomes (Dunoyer et al., 2010; Feinberg and Hunter, 2003; Gibbings and Voinnet, 2010; Molnar et al., 2010). Indeed, secretory miRNAs have been detected in human peripheral blood, the signature of which emerges as a novel biomarker for the diagnosis of myocardial infarction (Meder et al., 2010). There are several experiments to further investigate this hypothesis. (1) To assay the fine subcellular localization of *C. elegans* ALG-1/Argonaute using super resolution microscopy, with and without the statin treatment. (Cvetkovic et al.) To obtain the dynamics of *lin-4* mediated silencing of *lin-14* and calculate whether one molecule of *lin-4* miRNA can silence several *lin-14* mRNA targets by multiple turnover. If this is the case, we can then ask whether the multiple turnovers of the *lin-4* miRNA are regulated by the mevalonate pathway.

METHODS

Feeding RNAi

In addition to the RNAi clones used in Table 4.1 and Table 4.4, the following gene was knocked down by feeding RNAi using the Ahringer RNAi library (1): *F48F7.1* (*alg-1/Argonaute*), which probably also targets *alg-2* due to the high sequence similarity, and therefore is referred to as *alg-1/2* RNAi. HT115 bacteria carrying the empty vector L4440, which expresses dsRNA homologous to no worm gene, were used as a control. Bacterial clones were cultured at 37 °C for 15 h before seeding the RNAi plates. After induction of dsRNA for 24 h at room temperature, worms were placed on RNAi plates.

LIN-14 Western Blots

Because *hmgs-1* is an essential gene for fertility, we applied a mild gene knockdown by feeding the parental (P0) animals with Escherichia coli expressing *hmgs-1* dsRNA diluted with control E. coli expressing dsRNA homologous to no worm gene, starting at the fourth (L4) larval stage. Embryos were isolated from the P0 animals to synchronize their progeny by hatching in the absence of food. A fraction of synchronized L1s were flash-frozen in liquid nitrogen and others were fed on plates seeded with undiluted RNAi bacteria at 20 °C and collected after 20 and 23 h. At 24 h, worms were visually inspected to ensure they were all at the early L2 stage by counting the number of germ cells, the divided intestinal nuclei, and disappearance of L1 alae. Worm lysate preparation and LIN-14 Western blotting were performed as previously described (Cvetkovic et al.). Blots were reprobed with actin antibody (Abcam; ab3280) as loading control.

Quantification of microRNA by real-time PCR

The following procedure was adapted from (Shi and Chiang, 2005). RNA was isolated,

DNase treated, and polyadenylated by poly(A) polymerase. An adapter primer containing a unique 5′ sequence and 12 Ts and ending in VN-3′ was used to make cDNA. This anchors the adapter to the beginning of the poly(A) tail by virtue of the VN-3′ nucleotides (V = A, C, or G; n = A, T, C, or G). The cDNA is then amplified with a forward primer based on the entire tested miRNA sequence and a reverse primer complementary to the adapter. The PCR amplification was monitored by SYBR Green incorporation, and a corresponding threshold cycle (CT) was obtained. The quantity of miRNA, relative to two internal reference genes, U6 and 18s rRNA, was calculated using the formula 2–ΔCT, where ΔCT = (CT miRNA – CT reference). For each miRNA, the result was shown relative to its level in wild-type animals treated with control RNAi. The mean and SD were calculated from three biological replicates.

ALG-1 immunoprecipitation

ALG-1 was purified from synchronized L4-stage *alg-1(gk214)* mutants rescued with an HA-ALG-1 single copy construct. About 0.5 mL of worms was flash-frozen in liquid nitrogen, followed by grinding with a mortar and pestle. An equal volume of cell lysis buffer [50 mM Tris·HCl (pH 7.4), 100 mM KCl, 2.5 mM MgCl2, 0.1% Nonidet P-40, 0.5 mM PMSF, 1 Complete proteinase inhibitor tablet (Roche)/15 mL, 40 U/mL RNaseOUT (Invitrogen)] was added, and the lysate was homogenized on a head-to-tail rotor for 15 min. Debris was spun down in a tabletop centrifuge at 12,000 x g for 5 min at 4 °C. The cell lysate was precleared by adding 20 µL protein A agarose bead slurry (Roche) and rotating for 10 min. The cleared lysate was incubated with 3 µL HA antibody (clone 12CA5; Roche) for 20 min and then 100 µL protein A agarose bead slurry for 20 min at 4 °C. The beads were then washed eight times for 40 min in total. Ten percent of the immunoprecipitation (IP) sample was used for Western blot analysis. Ninety percent of

the IP sample was treated with proteinase K (1.0 µg/µL; Ambion) at 65 °C for 15 min. RNA was extracted with phenol-chloroform and subjected to miRNA real-time PCR analysis.

Reference

Abbott, A.L., Alvarez-Saavedra, E., Miska, E.A., Lau, N.C., Bartel, D.P., Horvitz, H.R., and Ambros, V. (2005). The let-7 MicroRNA family members mir-48, mir-84, and mir-241 function together to regulate developmental timing in Caenorhabditis elegans. Dev Cell *9*, 403-414.

Bracht, J., Hunter, S., Eachus, R., Weeks, P., and Pasquinelli, A.E. (2004). Trans-splicing and polyadenylation of let-7 microRNA primary transcripts. RNA *10*, 1586-1594.

Brodersen, P., Sakvarelidze-Achard, L., Schaller, H., Khafif, M., Schott, G., Bendahmane, A., and Voinnet, O. (2012). Isoprenoid biosynthesis is required for miRNA function and affects membrane association of ARGONAUTE 1 in Arabidopsis. Proc Natl Acad Sci U S A.

Calfon, M., Zeng, H., Urano, F., Till, J.H., Hubbard, S.R., Harding, H.P., Clark, S.G., and Ron, D. (2002). IRE1 couples endoplasmic reticulum load to secretory capacity by processing the XBP-1 mRNA. Nature *415*, 92-96.

Chang, S., Johnston, R.J., Jr., and Hobert, O. (2003). A transcriptional regulatory cascade that controls left/right asymmetry in chemosensory neurons of C. elegans. Genes Dev *17*, 2123-2137.

Chen, Y.G., Kowtoniuk, W.E., Agarwal, I., Shen, Y., and Liu, D.R. (2009). LC/MS analysis of cellular RNA reveals NAD-linked RNA. Nat Chem Biol *5*, 879-881.

Cikaluk, D.E., Tahbaz, N., Hendricks, L.C., DiMattia, G.E., Hansen, D., Pilgrim, D., and Hobman, T.C. (1999). GERp95, a membrane-associated protein that belongs to a family of proteins involved in stem cell differentiation. Mol Biol Cell *10*, 3357-3372.

Cvetkovic, A., Menon, A.L., Thorgersen, M.P., Scott, J.W., Poole, F.L., 2nd, Jenney, F.E., Jr., Lancaster, W.A., Praissman, J.L., Shanmukh, S., Vaccaro, B.J.*, et al.* (2010). Microbial metalloproteomes are largely uncharacterized. Nature *466*, 779-782.

Dunoyer, P., Schott, G., Himber, C., Meyer, D., Takeda, A., Carrington, J.C., and Voinnet, O. (2010). Small RNA Duplexes Function as Mobile Silencing Signals Between Plant Cells. Science.

Feinberg, E.H., and Hunter, C.P. (2003). Transport of dsRNA into cells by the transmembrane protein SID-1. Science *301*, 1545-1547.

Gerisch, B., Weitzel, C., Kober-Eisermann, C., Rottiers, V., and Antebi, A. (2001). A hormonal signaling pathway influencing C. elegans metabolism, reproductive development, and life span. Dev Cell *1*, 841-851.

Gibbings, D., and Voinnet, O. (2010). Control of RNA silencing and localization by endolysosomes. Trends Cell Biol *20*, 491-501.

Gibbings, D.J., Ciaudo, C., Erhardt, M., and Voinnet, O. (2009). Multivesicular bodies associate with components of miRNA effector complexes and modulate miRNA activity. Nat Cell Biol *11*, 1143-1149.

Goldstein, J.L., and Brown, M.S. (1990). Regulation of the mevalonate pathway. Nature *343*, 425-430.

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.

Hammell, C.M., Lubin, I., Boag, P.R., Blackwell, T.K., and Ambros, V. (2009). nhl-2 Modulates microRNA activity in Caenorhabditis elegans. Cell *136*, 926-938.

Hayes, G.D., and Ruvkun, G. (2005). "Control of developmental timing by microRNAs in C. elegans." Ph.D. thesis (Cambridge, MA, Harvard University).

Johnston, R.J., and Hobert, O. (2003). A microRNA controlling left/right neuronal asymmetry in Caenorhabditis elegans. Nature *426*, 845-849.

Kowtoniuk, W.E., Shen, Y., Heemstra, J.M., Agarwal, I., and Liu, D.R. (2009). A chemical screen for biological small molecule-RNA conjugates reveals CoA-linked RNA. Proc Natl Acad Sci U S A *106*, 7768-7773.

Lee, Y.S., Pressman, S., Andress, A.P., Kim, K., White, J.L., Cassidy, J.J., Li, X., Lubell, K., Lim do, H., Cho, I.S.*, et al.* (2009). Silencing by small RNAs is linked to endosomal trafficking. Nat Cell Biol *11*, 1150-1156.

Meder, B., Keller, A., Vogel, B., Haas, J., Sedaghat-Hamedani, F., Kayvanpour, E., Just, S., Borries, A., Rudloff, J., Leidinger, P.*, et al.* (2010). MicroRNA signatures in total peripheral blood as novel biomarkers for acute myocardial infarction. Basic Res Cardiol.

Molnar, A., Melnyk, C.W., Bassett, A., Hardcastle, T.J., Dunn, R., and Baulcombe, D.C. (2010). Small Silencing RNAs in Plants Are Mobile and Direct Epigenetic Modification in Recipient Cells. Science.

Morck, C., Olsen, L., Kurth, C., Persson, A., Storm, N.J., Svensson, E., Jansson, J.O., Hellqvist, M., Enejder, A., Faergeman, N.J.*, et al.* (2009). Statins inhibit protein lipidation and induce the unfolded protein response in the non-sterol producing nematode Caenorhabditis elegans. Proc Natl Acad Sci U S A *106*, 18285-18290.

Persson, B.C., Esberg, B., Olafsson, O., and Bjork, G.R. (1994). Synthesis and function of isopentenyl adenosine derivatives in tRNA. Biochimie *76*, 1152-1160.

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.

Robins, M.J., Hall, R.H., and Thedford, R. (1967). N-6-(delta-3-isopentenyl) adenosine. A component of the transfer ribonucleic acid of yeast and of mammalian tissue, methods of isolation, and characterization. Biochemistry *6*, 1837-1848.

Schmitz, C., Kinge, P., and Hutter, H. (2007). Axon guidance genes identified in a large-scale RNAi screen using the RNAi-hypersensitive Caenorhabditis elegans strain nre-1(hd20) lin-15b(hd126). Proc Natl Acad Sci U S A *104*, 834-839.

Shi, R., and Chiang, V.L. (2005). Facile means for quantifying microRNA expression by real-time PCR. Biotechniques *39*, 519-525.

Shi, Z., and Ruvkun, G. (2012). The mevalonate pathway regulates microRNA activity in Caenorhabditis elegans. Proc Natl Acad Sci U S A *109*, 4568-4573.

Shiu, P.K., Zickler, D., Raju, N.B., Ruprich-Robert, G., and Metzenberg, R.L. (2006). SAD-2 is required for meiotic silencing by unpaired DNA and perinuclear localization of SAD-1 RNA-directed RNA polymerase. Proc Natl Acad Sci U S A *103*, 2243-2248.

Tahbaz, N., Carmichael, J.B., and Hobman, T.C. (2001). GERp95 belongs to a family of signal-transducing proteins and requires Hsp90 activity for stability and Golgi localization. J Biol Chem *276*, 43294-43299.

Wightman, B., Burglin, T.R., Gatto, J., Arasu, P., and Ruvkun, G. (1991). Negative regulatory sequences in the lin-14 3'-untranslated region are necessary to generate a temporal switch during Caenorhabditis elegans development. Genes Dev *5*, 1813-1824.

Yochem, J., Tuck, S., Greenwald, I., and Han, M. (1999). A gp330/megalin-related protein is required in the major epidermis of Caenorhabditis elegans for completion of molting. Development *126*, 597-606.

CHAPTER FIVE

Dual regulation of the *lin-14* target mRNA by the *lin-4* miRNA

AUTHOR CONTRIBUTIONS

This work was greatly supported by Gary Ruvkun. In collaboration with Gabe Hayes, we monitored the dynamics of *lin-14* mRNA and protein as well as *lin-4* miRNA levels in finely staged animals during early larval development. Gabe Hayes characterized the *lin-14(n355)* mutant. Gary Ruvkun performed the northern blot experiment shown in Figure 5.3. This chapter was adapted from our manuscript “Dual regulation of the *lin-14* target mRNA by the *lin-4* miRNA”.

Summary

In animals, miRNAs typically bind with partial complementarity to sequences in the 3' untranslated (UTR) regions of target mRNAs, to induce a decrease in the production of the encoded protein. The relative contributions of translational inhibition of intact mRNAs and degradation of mRNAs caused by binding of the miRNA vary; for many genetically validated miRNA targets, translational repression has been implicated whereas some other analyses of other miRNA targets have revealed only modest translational repression and more significant mRNA destabilization. In *C. elegans,* the *lin-4* miRNA accumulates during early larval development, binds to target elements in the *lin-14* mRNA and causes a sharp decrease in the abundance of LIN-14 protein. Here, we monitor the dynamics of *lin-14* mRNA and protein as well as *lin-4* miRNA levels in finely staged animals during early larval development. We find complex regulation of *lin-14*, with an initial modest decline in *lin-14* mRNA abundance followed by fluctuation but little further decline of *lin-14* mRNA levels yet continuing and more dramatic decline in LIN-14 protein abundance. We show that the translational inhibition of *lin-14* is dependent on binding of the *lin-4* miRNA to multiple *lin-4* complementary sites in the *lin-14* 3'UTR. Our results point to the importance of translational inhibition in silencing of *lin-14* by *lin-4* miRNA.

Characterization of the *lin-14(n355gf)* mutant

The *lin-14(n355)* gain-of-function mutation was initially described as a possible translocation that separates the *lin-14* promoter and coding region from most of its 3'UTR, relieving it from repression by the *lin-4* miRNA (Wightman et al., 1991). To characterize the structure of the 3'UTR in the *n355* mutant we performed 3' RACE. We found that *n355* is an inversion that causes a break at nucleotide 254 of the *lin-14* 3'UTR

and a fusion to an intergenic sequence of the X chromosome corresponding to cosmid ZC373, downstream of the gene *col-176* (the sequences flanking the fusion were *lin-14* to ZC373: *attatcccca*TCATTTCGAG). A series of adenosine residues that did not correspond to the genomic sequence, presumably indicating the polyadenosine tail, began at position 82 of the sequence derived from ZC373, suggesting that the *lin-14(n355)* 3'UTR is ~335 nt, far shorter than the normally 1.6 kb wild type *lin-14* 3'UTR (data not shown). The *n355* inversion removes all of the characterized complementary sites for *lin-4* (Wightman et al., 1993) from the *lin-14* 3'UTR (Figure 5.1). The *n355* mutant causes retarded heterochronic phenotypes and continued expression of the LIN-14 protein at larval stages later than the first larval stage (L1), similar to a *lin-4* null mutant, further indicating that the *n355* mutation prevents the *lin-14* 3'UTR from responding to *lin-4* (Ambros and Horvitz, 1987; Lee et al., 1993).

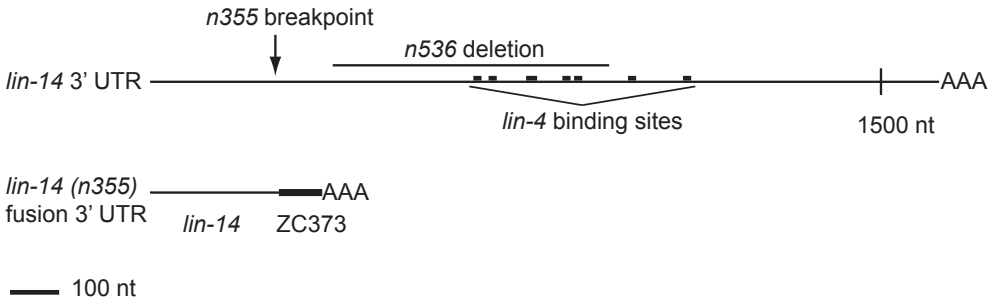


Figure 5.1 Diagram of the *lin-14* 3'UTR. Shown are the positions of the seven *lin-4* binding sites, which are all absent in the *n355* mutant allele. In the *n355* mutant, the 3' end of the *lin-14* gene is fused to intergenic sequence corresponding to cosmid ZC373. n536 allele bears a 607 bp deletion in the *lin-14* 3'UTR, which deletes five of the seven *lin-4* binding sites.

Temporal analyses revealed two phases of regulation of *lin-14* by *lin-4* miRNA

To resolve the dynamics of *lin-4* mediated down-regulation of *lin-14*, we performed temporal analyses of *lin-14* mRNA and protein as well as *lin-4* miRNA levels in finely staged wild-type and *lin-14(n355n679)* mutant animals collected at 3-h intervals from the early first larval (L1) to early second larval (L2) stage at 20°C. The *n355* allele decouples *lin-14* from *lin-4* repression resulting in gain-of-function retarded phenotypes, and slower development rate, but the *lin-14*(*n679)* V299D missense mutation confers a temperature-sensitive suppression of the *lin-14(n355)* retarded phenotype, due to a reduction-in-function mutation in the encoded LIN-14 protein; thus while the LIN-14 protein production continues after the normal mid-larval stage one downregulation by the *lin-4* miRNA, that temporally misexpressed LIN-14 protein is non-functional for specification of L1 cell fates (Reinhart and Ruvkun, 2001). At 20°C, the *lin-14(n355n679)* mutant animals grow and develop at similar rates as wild-type animals, allowing us to compare side-by-side the *lin-4* and *lin-14* dynamics in wild-type and *lin-14(n355n679)* mutant animals in which *lin-14* is relieved from repression by *lin-4*. More importantly, the misexpression of LIN-14 protein is decoupled from the mis-specification of L1 stage cell fates, so that molecular phenotypes can be interpreted without the complication of indirect developmental fate phenotypes.

To quantitatively assay the full-length polyadenylated *lin-14* mRNA, we used oligo$(dT)_{20}$ primer to reverse-transcribe mRNA isolated from synchronized animals and performed quantitative PCR (qPCR) analysis. Relative *lin-14* levels were obtained by normalization to *rpl-32*, a large ribosomal subunit that is abundant and stably expressed throughout development. The results are shown relative to the *lin-14* level when releasing from L1 diapause, or 0 hours of larval development. In wild-type animals, *lin-14* mRNA decreased to 49% and then to 38% of the 0 hour sample at 9 hours and 12 hours of larval development, respectively (Figure 5.2A). From 12 hours to 24 hours of larval

development, *lin-14* mRNA levels fluctuate with little evidence of further decline comparing 24 hour to 12 hour time point (p=0.23, one-tailed t-test) (Figure 5.2A). On the other hand, *lin-14* mRNA levels in the *lin-14(n355n679)* mutants fluctuate modestly , and show little evidence of the monotonic reduction throughout the L1 stage of wild type (Figure 5.2C).

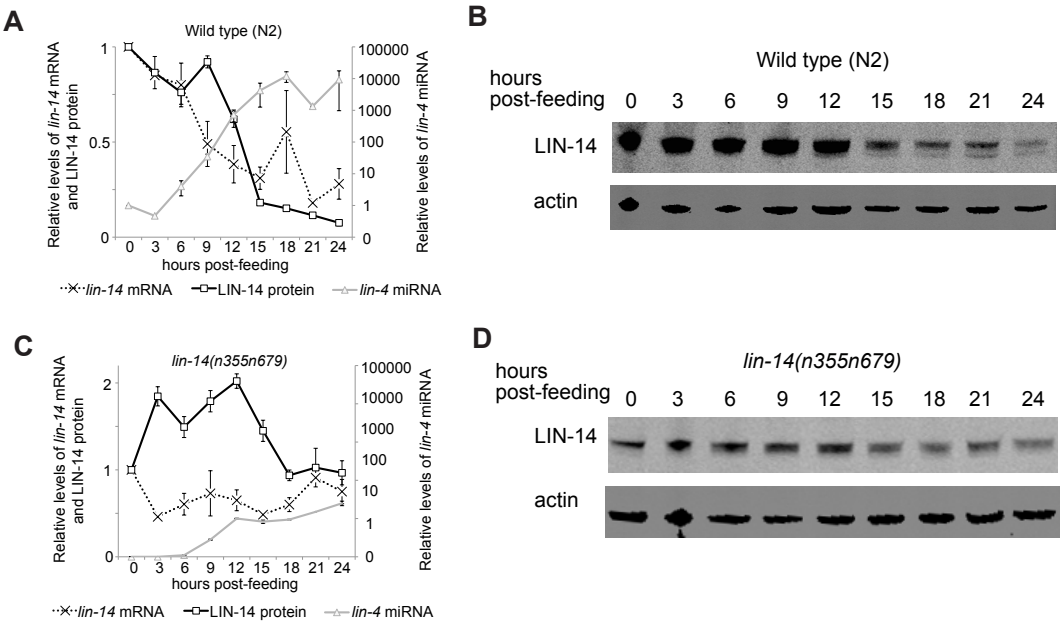


Figure 5.2 Temporal analyses of *lin-14* mRNA, protein and *lin-4* miRNA levels in wild-type and *lin-14(n355n679)* mutant animals. (A, C) Quantification of *lin-14* mRNA, LIN-14 protein and *lin-4* miRNA from early L1 (0 hours post-feeding) to early L2 (24 hours post-feeding) in wild-type (A) and *lin-14(n355n679)* mutant animals (C). All results were shown relative to the level at the 0 h time point. Error bars represent SEM for two independent experiments. (B, D) Representative immunoblots showing the abundance of LIN-14 protein in wild-type (B) and *lin-14(n355n679)* mutant animals (D) over 24 h of development. Actin serves as a control for the normalization of LIN-14.

We monitored LIN-14 protein levels in the same samples that were analyzed for *lin-14* mRNA abundance. LIN-14 protein levels were analyzed by Odyssey CLx Infrared Fluorescent Western Blot and quantified by normalization to actin. The results are shown relative to the LIN-14 protein level when releasing from L1 diapause, or 0 hours of larval development. In wild-type animals, LIN-14 protein levels are stable during the first 9 hours of larval development (Figure 5.2A and B). Subsequently, it declined to 62% and further to 18% at 12 hours and 15 hours (Figure 5.2A and B), which follows the decline of *lin-14* mRNA levels from 6-12 hours (Figure 5.2A). Since LIN-14 protein levels decline by ~5 fold and *lin-14* mRNA levels decline by ~2.5 fold, it suggests that both mRNA decay and inhibition of protein translation contribute to silencing of *lin-14* at early stages. From 15 hours to 24 hours of larval development, LIN-14 protein levels decrease from 18% to 8% of the 0 hour levels (Figure 5.2A and B). In contrast, LIN-14 protein levels in the *lin-14(n355n679)* mutants do not show any significant down-regulation throughout the first larval stage (Figure 5.2C and D), suggesting the ~10-fold down-regulation of LIN-14 protein in wild-type animals is indeed mediated by *lin-4* miRNA binding to *lin-14* 3'UTR.

To track the mature *lin-4* miRNA levels, we performed Taqman miRNA assays in the same samples as above. In wild-type animals, *lin-4* miRNA was present at very low level in embryos, which only can be detected by deep-sequencing (Stoeckius et al., 2009) but not by Northern Blot, until 9-12 hours of larval development at 20°C (Feinbaum and Ambros, 1999; Holtz and Pasquinelli, 2009). Consistent with previous studies, our analyses showed that *lin-4* miRNA levels begin to rise substantially at 9 hours, and its level becomes ~5000 fold higher at 24 hours compared to 0 hours. Curiously, the initial decline of *lin-14* mRNA levels during 6-12 hours happens prior to the significant accumulation of *lin-4* miRNA. When *lin-4* miRNA becomes abundant at later stages, *lin-*

14 mRNA levels fluctuate yet show little evidence of significant down-regulation; however, LIN-14 protein levels continue to decrease. Together, our analyses point to two phases of regulation: a fast *lin-14* mRNA destabilization as soon as *lin-4* miRNA emerges, and long-term translational inhibition that initiates and is in particularly important for maintaining the silencing of *lin-14* mRNA by the *lin-4* miRNA.

Strikingly, we found that *lin-4* miRNA levels in the *lin-14(n355n679)* mutants were ~50-200-fold lower compared to stage-matched wild-type animals (Figure 5.2A and C). We note that the mutant LIN-14 V299D protein encoded by the *lin-14(n355n679)* locus fails to specify larval stage one cell fates and instead specifies larval stage two cell fates at the nominal L1 stage. Thus the failure to up-regulate the *lin-4* miRNA in this *lin-14* mutant suggests that LIN-14 gene activity may normally act upstream of *lin-4* miRNA expression which then feeds back on production of LIN-14 protein. The defect in LIN-14 protein activity in a *lin-14(n355n679)* mutant may break this autoregulatory loop. Alternatively, in *C. elegans*, miRNAs are protected from degradation by their target mRNAs (Chatterjee et al., 2011; Chatterjee and Grosshans, 2009). Therefore, it is possible that *lin-4* miRNA loses this protection in *lin-14(n355n679)* mutants due to the absence of *lin-4* binding sites in the *lin-14* 3'UTR. This would suggest that in contrast to models of hundreds of mRNA targets of miRNAs, in the case of *lin-4*, there may just be the *lin-14* mRNA target that is key for accumulation of the *lin-4* miRNA.

The *lin-14* mRNA levels of wild-type and mutant bearing a *lin-14* 3'UTR deletion are equal both at early and late stages of animal development

Another approach to assay whether *lin-4* repression acts via mRNA destabilization or translational inhibition is to compare the levels of wild-type *lin-14* mRNA to mutant *lin-14* mRNA missing *lin-4* complementary sites in a heterozygous animal. To this end, we

assayed *lin-14* mRNA levels in the *lin-14(n536n540)/szT1* heterozygous strain with one wild-type and one *n536n540* allele of *lin-14*. *lin-14(n536)* bears a 607 bp deletion that removes five of the seven *lin-4* complementary elements in the *lin-14* 3'UTR (Figure 5.1) and is a gain-of-function mutation causing retarded heterochronic development due to derepression of *lin-14* (Wightman et al., 1991; Wightman et al., 1993). The *n540* allele, an amber mutation at position 280 (Lys to amber), was isolated as a recessive suppressor of the *lin-14(n536gf)* mutant (Ambros and Horvitz, 1987; Ruvkun et al., 1989). We reasoned that if *lin-4* miRNA silences *lin-14* via enhancing mRNA decay, then *lin-14(n536n540)* mRNA should be stabilized relative to the wild-type *lin-14* mRNA due to reduction of *lin-4*-mediated regulation. However, Northern blot of RNA isolated from *lin-14(n536n540)/szT1* heterozygotes showed that the 3.5 kb wild-type *lin-14* and 2.9 kb *lin-14(n536n540)* mRNA levels are nearly equal at both embryonic stage before *lin-4* miRNA is expressed and at mid-L4 stage, after *lin-4* repression has occurred (Figure 5.3). This experiment indicates that there is little down-regulation of *lin-14* mRNA between the pre-*lin-4* expression stage and the L4 stage from either the wild type allele or the *lin-14* allele missing many *lin-4*-complementary regions. However, the two residual *lin-4* complementary sites in the *lin-14(n536)* 3'UTR are insufficient to mediate the silencing effect by *lin-4* through translational repression because the n536 mutation derepresses the L2 and later stage expression of the LIN-14 protein. Our result is also largely consistent with a recent study (Stadler et al., 2012) and further stresses the importance of translational inhibition in *lin-4* miRNA-mediated silencing of *lin-14*.

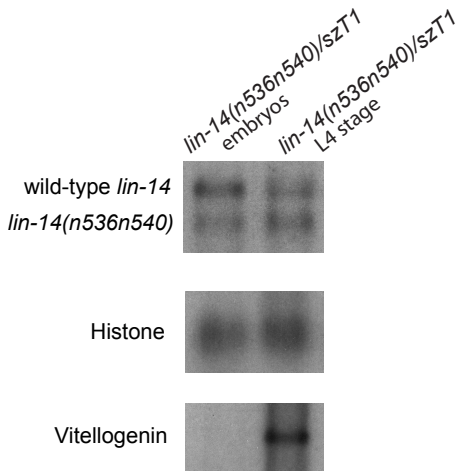


Figure 5.3 The levels of *lin-14* mRNA derived from the wild-type and a *lin-14* mutant allele bearing a 3'UTR deletion are equal at the embryonic and L4 stages. Shown is a Northern blot of *lin-14*. The levels of wild-type *lin-14* (3.5 kb) and *lin-*14*(n536n540)* mRNA (2.9 kb) in the *lin-14(n536n540)/szT1* heterozygotes are almost equal at both the embryonic and L4 stages. Histone mRNA is blotted as a control for even loading. Vitellogenin mRNA encoding yolk polypeptides that is most abundant during oogenesis is shown to indicate animal stages.

Discussion

The results we present here consolidate the earlier studies where *lin-14* was found to be silenced by *lin-4* without being significantly destabilized at the mRNA level (Olsen and Ambros, 1999; Wightman et al., 1993), as well as a recent ribosome profiling study showing translational control of *lin-14* by *lin-4* (Stadler et al., 2012). Although cases have been reported where regulation by miRNAs can cause destabilization of target mRNAs in *C. elegans* (Bagga et al., 2005), the relative contributions of mRNA degradation and translational repression in miRNA-mediated repression are variable (Ding and Grosshans, 2009). For example, *lin-41* mRNA levels regulated by *let-7* miRNA showed

the strongest reduction, whereas *lin-14* mRNA showed modest reduction (Bagga et al., 2005). However, the *lin-41* mRNA degradation noted by Bagga et al was not observed by Stadler et al, who instead suggested translational control of *lin-41* by the *let-7* miRNA (Stadler et al., 2012).

Several recent studies have suggested a two-phase model for miRNA-mediated silencing that begins with translational repression, followed by mRNA deadenylation and decay which consolidates the silencing (Bazzini et al., 2012; Djuranovic et al., 2011, 2012; Fabian et al., 2009; Selbach et al., 2008; Zdanowicz et al., 2009). Interestingly, we observed a distinct dynamics for *lin-4* silencing of *lin-14*. First, *lin-14* mRNA level declines to ~40% as soon as *lin-4* miRNA is first expressed during 6-9 hours of L1 development, prior to the decrease of LIN-14 protein level. Although translational inhibition may also contribute to this initial silencing of *lin-14*, currently this hypothesis is not supported by solid evidence. On the other hand, between 12-24 hours of larval development, *lin-14* mRNA levels fluctuate with no obvious further decline whereas LIN-14 protein levels continue to decrease. Furthermore, by early L2 stage LIN-14 protein level has declined by ~10 fold and *lin-14* mRNA levels decline by ~3 fold, suggesting that long-term translational inhibition maintains silencing of *lin-14* mRNA by the *lin-4* miRNA.

It remains unclear why the modes of action of miRNA-mediated silencing are different depending on the biological and experimental context. In particular, it is intriguing why and how the predominant mechanism of silencing could change even for the same miRNA and its target mRNA in the same organism. Since the levels of *lin-4* miRNA increase more than 500 fold from early L1 stage when *lin-14* mRNA just begins to decay to early L2 stage when *lin-14* has been stably silenced, we speculate that the number of *lin-4* molecules binding to *lin-14* mRNA 3'UTR might determine the mode of silencing. Specifically, the seven *lin-4* complementary sites in *lin-14* 3'UTR could allow a

sensitive response to *lin-4*, if a single or few *lin-4* miRNA molecules bind to *lin-14* 3'UTR causes mRNA decay, whereas more miRNAs binding causes translational control. Such an architecture would trigger a fast decline in LIN-14 protein synthesis even when the level of *lin-4* is low, to achieve a sharp transition in development. To maintain silencing, translation of *lin-14* mRNA is inhibited probably through binding of multiple *lin-4* molecules to *lin-14* mRNA 3'UTR as *lin-4* massively accumulates. Supporting this hypothesis, we found that the *lin-14(n536n540)* mRNA missing five out of the seven *lin-4* complementary elements in its 3'UTR is destabilized to a similar level as the wild-type *lin-14* mRNA. However, *lin-14(n536)* is not properly silenced and causes retarded heterochronic phenotype (Wightman et al., 1991; Wightman et al., 1993). This suggests that the two remaining *lin-4* complementary sites are sufficient to cause the normal *lin-4*-triggered decay of *lin-14* mRNA, but insufficient to induce *lin-4* triggered translational repression.

Finally, our fine-stage analysis showed that *lin-14* mRNA drops initially during the first larval intermolt stage, but returns to a higher level during the L1-L2 molt. It is an emerging trend that mRNA levels could oscillate in animals entering and exiting the molt. Therefore, an important lesson is that looking at a single time point could be misleading, which may contribute to the differences in results and conclusions obtained from different labs.

Methods

Characterization of the *lin-14(n355)* inversion

RNA samples derived from wild type and *n355* were analyzed by 3' RACE (Roche). The most prominent band that was detected in *n355* and not wild type corresponded to the

length of the *lin-14* 3'UTR in the *n355* mutant that was deduced by Wightman et al. (1991). Sequencing of this band showed that at position 254 of the 3'UTR the *lin-14* sequence was fused to an intergenic region of the X chromosome corresponding to cosmid ZC373. PCR using primers homologous to *lin-14* and ZC373 and DNA from the *n355* mutant as template yielded bands that confirmed the chromosomal structure in the *n355* mutant that was predicted by 3' RACE (data not shown).

Quantitative Western blot

Odyssey CLx Infrared Fluorescent Western Blot was performed following the vendor's protocol. Briefly, NuPAGE® LDS sample buffer (2x) equal to the total volume of the worms was added and samples were boiled for 5 minutes. The lysates were pelleted in a microfuge and the supernatant loaded to a 4-12% Bis-Tris gel (Invitrogen) for electrophoresis. The proteins were transferred to nitrocellulose membrane (Bio-Rad). The membrane was blocked in blocking buffer (5% nonfat milk in PBS+0.1% Tween® 20) for 1.5 h at room temperature. Subsequently, the membrane was incubated in rabbit anti-LIN-14 (1:1000) and mouse anti-actin (1:5000, Abcam) primary antibodies at 4°C overnight. After washing four times with PBS+0.1% Tween® 20, the membrane was incubated in RDye 800CW Goat anti-Rabbit IgG (1:10000) and IRDye 680RD Goat anti-Mouse IgG (1:5000) at room temperature for 1 hour. The membrane was washed four times with PBS+0.1% Tween® 20, briefly rinsed with PBS and imaged with the Odyssey CLx infrared imaging system. LIN-14 and actin blots were scanned using the 800 nm and 700 nm channels, respectively. The abundance of LIN-14 was quantified by normalization to actin.

Quantitative RT-PCR analysis of *lin-14* mRNA

Total RNA was purified using TRIzol Reagent (Molecular Research Center), DNase treated (TURBO™ DNase, Ambion), and reverse-transcribed using oligo$(dT)_{20}$ primer (Invitrogen). The PCR amplification was monitored by SYBR green incorporation, and a corresponding threshold cycle (C_T) was obtained. The quantity of *lin-14* mRNA, relative to *rpl-32*, was calculated using the formula $2^{-\Delta CT}$, where $\Delta CT = (C_{T\ lin-14} - C_{T\ rpl-32})$. The *lin-14* level was then normalized to its level at 0 h after release from L1 diapause. Primers used: *lin-14* forward primer spanning the last exon-exon junction (caaaaactgagagcgaaacg) and *lin-14* reverse primer in the last exon (tggaccttgaagaggaggag).

Taqman miRNA Assays

Taqman miRNA Assays (Applied Biosystems) were performed following the vendor's protocol. Briefly, 100 ng total RNA was reverse transcribed using a miRNA-specific RT primer. The real-time PCR amplification was performed and a corresponding threshold cycle (C_T) was obtained. The quantity of *lin-4* miRNA, relative to U6 as an internal reference gene, was calculated using the formula $2^{-\Delta CT}$, where $\Delta CT = (C_{T\ lin-4} - C_{T\ U6})$. The *lin-4* level was then normalized to its level at 0 h after release from L1 diapause.

Northern Blot

We collected synchronized embryos and L4 staged animals of *lin-14(n536n540)/szT1* heterozygotes. *szT1* is a (I;X) translocation balancer chromosome that balances the slow growing and very small brood size of the *lin-14(n536n540)* homozygotes. While the *lin-14(n536n540)/szT1* mutants segregate homozygous larval lethal *szT1* homozygotes and homozygous *lin-14(n536n540)* animals, they constitute a minor component of the population because of the much larger brood and faster growth of the *lin-*

14(n536n540)/szT1 heterozygotes. Total RNA was purified using guanidium isothiocyanate disruption and purification through a CsCl cushion via ultracentrifugation, polyA selected, and separated on a 1.2% formaldehyde agarose gel. RNA was transferred to Nylon membrane, UV crossed linked, then hybridized to radiolabeled DNA probes that were generated from the 3.8 kb EcoRI fragment bearing the last 7 exons of *lin-14* and its 3'UTR. A DNA probe for the histone gene was used for normalization of mRNA content per lane, and a probe for the vitellogenin gene was used to indicate animal stages. X-Omat films exposed to the Northern blot were scanned on an optical scanner.

Reference

Ambros, V., and Horvitz, H.R. (1987). The lin-14 locus of Caenorhabditis elegans controls the time of expression of specific postembryonic developmental events. Genes Dev *1*, 398-414.

Bagga, S., Bracht, J., Hunter, S., Massirer, K., Holtz, J., Eachus, R., and Pasquinelli, A.E. (2005). Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. Cell *122*, 553-563.

Bazzini, A.A., Lee, M.T., and Giraldez, A.J. (2012). Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. Science *336*, 233-237.

Chatterjee, S., Fasler, M., Bussing, I., and Grosshans, H. (2011). Target-Mediated Protection of Endogenous MicroRNAs in C. elegans. Dev Cell *20*, 388-396.

Chatterjee, S., and Grosshans, H. (2009). Active turnover modulates mature microRNA activity in Caenorhabditis elegans. Nature *461*, 546-549.

Ding, X.C., and Grosshans, H. (2009). Repression of C. elegans microRNA targets at the initiation level of translation requires GW182 proteins. EMBO J *28*, 213-222.
Djuranovic, S., Nahvi, A., and Green, R. (2011). A parsimonious model for gene regulation by miRNAs. Science *331*, 550-553.

Djuranovic, S., Nahvi, A., and Green, R. (2012). miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. Science *336*, 237-240.

Fabian, M.R., Mathonnet, G., Sundermeier, T., Mathys, H., Zipprich, J.T., Svitkin, Y.V., Rivas, F., Jinek, M., Wohlschlegel, J., Doudna, J.A.*, et al.* (2009). Mammalian miRNA RISC recruits CAF1 and PABP to affect PABP-dependent deadenylation. Mol Cell *35*, 868-880.

Feinbaum, R., and Ambros, V. (1999). The timing of lin-4 RNA accumulation controls the timing of postembryonic developmental events in Caenorhabditis elegans. Dev Biol *210*, 87-95.

Holtz, J., and Pasquinelli, A.E. (2009). Uncoupling of lin-14 mRNA and protein repression by nutrient deprivation in Caenorhabditis elegans. RNA *15*, 400-405.

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.

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.

Reinhart, B.J., and Ruvkun, G. (2001). Isoform-specific mutations in the Caenorhabditis elegans heterochronic gene lin-14 affect stage-specific patterning. Genetics *157*, 199-209.

Ruvkun, G., Ambros, V., Coulson, A., Waterston, R., Sulston, J., and Horvitz, H.R. (1989). Molecular genetics of the Caenorhabditis elegans heterochronic gene lin-14. Genetics *121*, 501-516.

Selbach, M., Schwanhausser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. Nature *455*, 58-63.

Stadler, M., Artiles, K., Pak, J., and Fire, A. (2012). Contributions of mRNA abundance, ribosome loading, and post- or peri-translational effects to temporal repression of C. elegans heterochronic miRNA targets. Genome Res *22*, 2418-2426.

Stoeckius, M., Maaskola, J., Colombo, T., Rahn, H.P., Friedlander, M.R., Li, N., Chen, W., Piano, F., and Rajewsky, N. (2009). Large-scale sorting of C. elegans embryos reveals the dynamics of small RNA expression. Nat Methods *6*, 745-751.

Wightman, B., Burglin, T.R., Gatto, J., Arasu, P., and Ruvkun, G. (1991). Negative regulatory sequences in the lin-14 3'-untranslated region are necessary to generate a temporal switch during Caenorhabditis elegans development. Genes Dev *5*, 1813-1824.

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.

Zdanowicz, A., Thermann, R., Kowalska, J., Jemielity, J., Duncan, K., Preiss, T., Darzynkiewicz, E., and Hentze, M.W. (2009). Drosophila miR2 primarily targets the m7GpppN cap structure for translational repression. Mol Cell *35*, 881-888.

APPENDIX I

Summary of Publications

Studies presented in this dissertation:

Chapter II and III

Shi, Z., Montgomery, T.A., Qi, Y., and Ruvkun, G. (2013). High-throughput sequencing reveals extraordinary fluidity of miRNA, piRNA, and siRNA pathways in nematodes. Genome Res 23, 497-508.

Chapter IV

Shi, Z., and Ruvkun, G. (2012). The mevalonate pathway regulates microRNA activity in Caenorhabditis elegans. Proc Natl Acad Sci U S A 109, 4568-4573.

Chapter V

Shi, Z., Hayes, G., and Ruvkun, G. Dual regulation of the *lin-14* target mRNA by the *lin-4* microRNA: in revision

Additional small RNA-related publications not presented in this dissertation:

Wu, X., **Shi, Z.**, Cui, M., Han, M., and Ruvkun, G. (2012). Repression of germline RNAi pathways in somatic cells by retinoblastoma pathway chromatin complexes. PLoS Genet 8, e1002542.

Abstract: The retinoblastoma (Rb) tumor suppressor acts with a number of chromatin cofactors in a wide range of species to suppress cell proliferation. *The Caenorhabditis elegans* retinoblastoma gene and many of these cofactors, called synMuv B genes, were identified in genetic screens for cell lineage defects caused by growth factor

misexpression. Mutations in many synMuv B genes, including *lin-35/Rb*, also cause somatic misexpression of the germline RNA processing P granules and enhanced RNAi. We show here that multiple small RNA components, including a set of germline-specific Argonaute genes, are misexpressed in the soma of many synMuv B mutant animals, revealing one node for enhanced RNAi. Distinct classes of synMuv B mutants differ in the subcellular architecture of their misexpressed P granules, their profile of misexpressed small RNA and P granule genes, as well as their enhancement of RNAi and the related silencing of transgenes. These differences define three classes of synMuv B genes, representing three chromatin complexes: a LIN-35/Rb-containing DRM core complex, a SUMO-recruited Mec complex, and a synMuv B heterochromatin complex, suggesting that intersecting chromatin pathways regulate the repression of small RNA and P granule genes in the soma and the potency of RNAi. Consistent with this, the DRM complex and the synMuv B heterochromatin complex were genetically additive and displayed distinct antagonistic interactions with the MES-4 histone methyltransferase and the MRG-1 chromodomain protein, two germline chromatin regulators required for the synMuv phenotype and the somatic misexpression of P granule components. Thus intersecting synMuv B chromatin pathways conspire with synMuv B suppressor chromatin factors to regulate the expression of small RNA pathway genes, which enables heightened RNAi response. Regulation of small RNA pathway genes by human retinoblastoma may also underlie its role as a tumor suppressor gene.

Contribution: I performed some of the enhanced RNAi assays in double mutants of synMuvB genes. I also contributed to the identification of some RNAi factors as

synMuvB target genes by querying microarray datasets, and helped the validation of a set of these genes via RT-qPCR.