Published in final edited form as:

J Mol Evol. 2013 December ; 77(0): 206–220. doi:10.1007/s00239-013-9588-6.

Pleiotropic constraints, expression level and the evolution of miRNA sequences

Richard Jovelin

Department of Ecology and Evolutionary Biology, University of Toronto, Ontario M5S 3B2, Canada

Abstract

Post-transcriptional gene regulation mediated by microRNAs (miRNAs) plays critical roles during development by modulating gene expression and conferring robustness to stochastic errors. Phylogenetic analyses suggest that miRNA acquisition could play a role in phenotypic innovation. Moreover, miRNA-induced regulation strongly impacts genome evolution, increasing selective constraints on 3'UTRs, protein sequences and expression level divergence. Thus, it is essential to understand the factors governing sequence evolution for this important class of regulatory molecules. Investigation of the patterns of molecular evolution at miRNA loci have been limited in *Caenorhabditis elegans* because of the lack of a close outgroup. Instead, I used *Caenorhabditis briggsae* as the focus point of this study because of its close relationship to *Caenorhabditis* sp. 9. I also corroborated the patterns of sequence evolution in *Caenorhabditis* using published orthologous relationships among miRNAs in *Drosophila*. In nematodes and in flies, miRNA sequence divergence is not influenced by the genomic neighbourhood (i.e. intronic or intergenic) but is nevertheless affected by the genomic context because X-linked miRNAs evolve faster than autosomal miRNAs. However, this effect of chromosomal linkage can be explained by differential expression levels rather than a fast-X effect. The results presented here support a universal negative relationship between rates of molecular evolution and expression level, and suggest that mutations in highly expressed miRNAs are more likely to be deleterious because they potentially affect a larger number of target genes. Finally, I show that many single-family member miRNAs evolve faster than miRNAs from multigene families and have limited functional scope, suggesting that they are not strongly integrated in gene regulatory networks.

Keywords

miRNA; expression level; molecular evolution; *Caenorhabditis*; *Drosophila*

Introduction

Animal genomes encode a set of shared toolkit genes with function in cell signaling, cell adhesion and transcriptional regulation that diversified early during metazoan evolution (Guder et al. 2006; Kasbauer et al. 2007; Matus et al. 2007; Nichols et al. 2006; Technau et al. 2005). Lineage-specific duplications and losses of developmental regulatory genes contribute to their functional diversification (Canestro et al. 2009; Jovelin et al. 2010), and can generate genetic incompatibilities leading to speciation (Lynch and Force 2000). The redeployment of this shared toolkit within regulatory networks, through the evolution of transcription factor binding sites, in various tissues during development is thought to have played a major role in the evolution of animal form (Carroll 2008). Nevertheless, post-

transcriptional control of gene activity mediated by microRNAs (miRNAs) is now recognized as a fundamental layer of gene regulation (Bartel and Chen 2004). MiRNAs encode ~22 nucleotide-long mature RNAs processed from a stereotypical hairpin precursor transcript (Bartel 2004) that bind to the 3' untranslated region (UTR) of messenger RNAs (mRNA), primarily through a 5' 7-bp complementary seed motif, to repress translation and less frequently to cause mRNA degradation (Bartel 2009). MiRNAs play critical roles in development and during homeostasis by modulating gene expression level in a switch-like function but also by buffering expression variation and fine-tuning expression to optimum levels (Bartel and Chen 2004; Felekakis et al. 2011; Herranz and Cohen 2010; Li et al. 2009; Mukherji et al. 2011; Wu et al. 2009).

The origin of miRNAs is ancient, dating back to the early metazoans (Grimson et al. 2008). However, if miRNAs are widespread regulatory molecules, differences in miRNA origin, biogenesis and target recognition suggest an independent origin of plant and animal miRNAs, perhaps from a common short interfering RNA system (Axtell et al. 2011). In contrast to developmental regulatory proteins, novel miRNA families are continuously added to animal genomes (Wheeler et al. 2009) because of a net positive gain despite high birth and death rates (Lu et al. 2008b), making them attractive markers for the resolution of deep phylogenetic branching (Campbell et al. 2011; Heimberg et al. 2010). Once integrated in genomes, miRNAs could have profound effects in re-shaping regulatory interactions by controlling dosage of their target genes. Remarkably, a burst of miRNA acquisition in basal vertebrates is coincidental with an increase in morphological complexity (Heimberg et al. 2008), and the rate of miRNA acquisition also accelerated during the early diversification of placental mammals and later in the hominoid lineage (Iwama et al. 2013), suggesting a role for miRNAs in phenotypic innovation.

With as many as hundreds of genes being under the control of a single miRNA, and with a typical combinatorial mode of action (Brennecke et al. 2005; Lewis et al. 2005), miRNA-mediated regulation significantly impacts genome evolution. Post-transcriptional regulation mediated by miRNAs strongly influences 3'UTR sequence evolution because of selection for the conservation of miRNA binding sites and for the avoidance of miRNA targeting (Farh et al. 2005; Jan et al. 2011; Stark et al. 2005). Moreover, the number of miRNA regulators is negatively correlated with protein sequence evolution (Chen et al. 2011; Cheng et al. 2009), and target genes may evolve slower than non-targets because of selection for optimal codon usage (Takuno and Innan 2011). In addition, miRNAs play a role in the evolution of gene expression variation between species and within populations, by constraining expression divergence for some targets but also increasing variation in expression level at other targets (Cui et al. 2007; Lu and Clark 2012).

Sequence variation at miRNA loci and their binding sites could result in substantial rewiring of gene regulatory networks. Population genetic analyses of miRNAs and their binding sites show that miRNA-induced regulation is generally under strong purifying selection but nevertheless detected instances of non-neutral patterns of sequence variation and strong population differentiation at miRNAs and/or binding sites (Chen and Rajewsky 2006; Loh et al. 2011; Lu et al. 2008a; Quach et al. 2009; Saunders et al. 2007), further implicating miRNAs as a significant source of phenotypic variation. Exact matching of nucleotides 2-8, the 5' seed motif, is required for interaction with most miRNA targets, although some other sites require matching in the 3' end or in the center of the mature sequence (Bartel 2009; Brennecke et al. 2005; Shin et al. 2010). Nucleotide changes at these sites are likely to impede miRNA regulation and may shift the target spectrum. However, any mutations in the miRNA gene can potentially affect miRNA-mediated regulation by altering miRNA processing and/or miRNA-target interaction (de Meaux et al. 2008; Duan et al. 2007; Jazdzewski et al. 2009; Jazdzewski et al. 2008; Lee et al. 2011). It is therefore essential to

determine the factors that influence sequence variation in the entire hairpin as well as in the mature region for a class of genes that plays crucial roles in gene regulatory networks and that contributes substantially in shaping nucleotide variation across genomes.

First identified in the nematode *Caenorhabditis elegans* 20 years ago (Lee et al. 1993), miRNAs in this species have been extensively characterized, with 223 miRNA genes annotated in miRBase release 19 (Kozomara and Griffiths-Jones 2011). Additional sequencing of small RNAs expressed in the related species *Caenorhabditis briggsae* and *Caenorhabditis remanei* has revealed several modes of miRNA evolution through seed shifting, hairpin shifting (formation of a new hairpin with sequence upstream or downstream of the miRNA) and arm-switching of the miRNA within the precursor hairpin (de Wit et al. 2009). However, the high divergence times with related species, resulting in the lack of a suitable species for sequence comparison, has prevented thorough investigations of miRNA sequence evolution in *C. elegans*. Although only 140 miRNAs have thus far been annotated in *C. briggsae*, this species is particularly well suited for investigating patterns of molecular evolution in miRNAs because of its close relative *Caenorhabditis* sp. 9 (Kiontke et al. 2011). *C. briggsae* and *C. sp. 9* diverged in the recent past as interspecific crosses still can produce fertile hybrid progeny (Kozłowska et al. 2012; Woodruff et al. 2010), and this species pair has proved useful for investigating protein sequence evolution (Jovelin and Phillips 2011). In addition, *Caenorhabditis* sp. 5 is a moderately divergent outgroup to the *C. briggsae* – *C. sp. 9* species pair (Kiontke et al. 2011), having split from their last common ancestor ~ 18 million years ago (Cutter 2008). Moreover, analysis of a subset of the *C. briggsae* miRNAs uncovered a relatively high level of nucleotide polymorphism and identified alleles predicted to alter function based on principles of target recognition and miRNA processing (Jovelin and Cutter 2011), prompting further investigation of nucleotide variation in *Caenorhabditis* miRNAs.

Here, I performed a homology search of the *C. briggsae* miRNAs in the genome assemblies of *C. sp. 9* and *C. sp. 5* in order to examine patterns of sequence evolution in nematode miRNAs. I also used published orthologous relationships among miRNAs in *Drosophila* (Nozawa et al. 2010) in order to evaluate the generality of these findings and to increase our understanding of miRNA sequence evolution. I show that rates of nucleotide variation at miRNA loci and mature sequences strongly depend on miRNA expression level, supporting the view that gene expression plays an important role in molecular evolution. By examining nucleotide variation in the mature sequence and the remaining of the hairpin separately, I show that selective constraints in highly expressed miRNAs are associated with the fitness cost of deleterious mutations with pleiotropic effects, affecting a larger number of target genes.

Materials and Methods

Nematode and fly miRNAs

The list of 140 *C. briggsae* miRNAs annotated in miRBase 19 (Kozomara and Griffiths-Jones 2011) was used as query in a BLASTN search (Altschul et al. 1990) for investigating the miRNA content in the two most closely related species *Caenorhabditis* sp. 9 and *Caenorhabditis* sp. 5 (Kiontke et al. 2011) (Genome Sequencing Center, Washington University, St. Louis, unpublished data). The entire hairpin was used in the BLAST search against the genome assembly of *C. sp. 5* and the entire hairpin with ~600 bp of flanking sequence was used in the BLAST search against the genome assembly of *C. sp. 9* to ensure identification of the *C. sp. 9* orthologs and to avoid retrieving non-orthologous miRNAs with similar sequences. In addition, it was possible for *C. sp. 9* to identify orthologs of clustered miRNAs based on synteny mapping and sequence conservation. BLAST hits were not considered as miRNAs if no hairpin structure with a minimum folding energy lower than

–15 Kcal/mol, as determined by RNAfold (Gruber et al. 2008), could be identified. The mature sequences of *Cbr-mir-72*, *Cbr-mir-83*, *Cbr-mir-246* and *Cbr-mir-360* in miRBase are predicted from comparison with *Caenorhabditis elegans* miRNAs but differ from the mature sequences identified in *C. briggsae* using small RNA sequencing (de Wit et al. 2009). Mature sequences with experimental support from (de Wit et al. 2009) were used instead. The list of hairpin and mature sequences is available as Supplementary Data. Sequences of miRNAs in *Drosophila* and their orthologous relationships were obtained from the literature, defining 151 orthologous groups (Nozawa et al. 2010). Gain of function mutation phenotypes of *D. melanogaster* miRNAs corresponding to the constitutively active act5C-Gal4 driver are from (Schertel et al. 2012).

Sequence analyses

Sequences were first automatically aligned with CLUSTAL W (Thompson et al. 1994) and each alignment was then manually curated using BioEdit (Hall 1999). Sequence divergence was measured between *C. briggsae* and *C. sp. 9* orthologs with a Jukes-Cantor distance in MEGA 5, excluding indels (Tamura et al. 2011). The mean nucleotide divergence in each miRNA orthologous group, estimated with a Jukes-Cantor distance in MEGA 5, was used as the estimate of nucleotide divergence for *Drosophila* miRNAs. Pairwise sequence divergence between *Drosophila melanogaster* and *Drosophila simulans* (or the next closest species if the miRNA was absent from *D. simulans*) gave qualitatively similar results, although not always statistically significant because of the low level of nucleotide divergence at miRNA loci between these two species (not shown). For each miRNA in *Caenorhabditis* and in *Drosophila*, the rate of nucleotide substitutions was computed separately for the mature sequence and the backbone of the hairpin (hairpin minus the mature miRNA), in addition to the entire hairpin. In order to investigate selective constraints operating on miRNA loci in *Caenorhabditis*, I compared the level of sequence divergence between *C. briggsae* and *C. sp. 9* orthologs for the mature sequence, the backbone of the hairpin (excluding the mature miRNA) and sequences immediately flanking miRNAs.

Hairpin and mature sequences of members of the *mir-35* family in *C. elegans* and in *Caenorhabditis remanei* were downloaded from miRBase 19 (Kozomara and Griffiths-Jones 2011), aligned with the *mir-35* family members in *C. briggsae*, *C. sp. 9* and *C. sp. 5* with CLUSTAL W (Thompson et al. 1994) in an attempt to examine further their evolutionary relationships using phylogenetic analysis. Neighbor-Joining trees were built with the entire hairpin and mature sequences separately in MEGA 5 (Tamura et al. 2011), using distances computed with the Maximum Composite Likelihood method and pairwise gap deletions. Branch support was evaluated with the bootstrapping procedure with 500 replicates and branches with less than 50% support were collapsed.

Expression level and target genes

The number of reads from RNA-seq was used as a measure of expression level. miRNA expression levels in mixed-stage worms of the reference *C. briggsae* strain AF16 grown under standard conditions were obtained from (de Wit et al. 2009). First, expression level was binned into three groups similar to the classification of (Liang and Li 2009): 1) $n \geq 100$, the high expression group which contains 39 miRNAs, 2) $100 > n \geq 15$, the medium expression group which contains 33 miRNAs, 3) $15 > n \geq 0$, the low expression groups which contains 68 miRNAs. Second, I examined the relationship between expression level and nucleotide divergence independently from bin grouping using Spearman's rank correlation. Expression levels of miRNAs in *D. melanogaster* were obtained from (Berezikov et al. 2011).

Predicted targets of *C. briggsae* miRNAs were retrieved from TargetScanWorm 5.1 (Ruby et al. 2006). Similar results were obtained with targets predicted from TargetScanWorm 6.2 (Jan et al. 2011), without 3' compensatory sites (not shown). Predicted targets of *D. melanogaster* miRNAs were obtained from TargetScanFly 6.2 (Lewis et al. 2005).

Orthologs of the *C. briggsae* miRNA targets were identified in *C. sp.9* and *C. sp. 5*, respectively using a BLASTN search with the entire gene and a TBLASTN search with the *C. briggsae* protein sequence, in order to investigate the possibility of co-evolution between divergent miRNAs with a substitution in the seed motif and their targets. Respectively 95 and 96 orthologs could be unambiguously identified in *C. sp. 5* and in *C. sp. 9* out of a total of 111 targets for the *mir-35* and *mir-790* miRNAs in *C. briggsae*. Orthologous protein sequences were aligned and used to derive codon-based nucleotide sequence alignments in order to identify the last exon with the terminal stop codon. Up to 500 bp of sequence immediately downstream the stop codon of each orthologous target in *C. briggsae*, *C. sp. 9* and *C. sp. 5* was searched for complementary motifs to the ancestral and derived seeds, using definition of target sites from TargetScanWorm. Similar results were obtained when using 300 bp or 1 Kb of downstream sequence (Supplementary Figure 3).

The program DAVID (Huang et al. 2009) was used to examine the enrichment of Gene Ontology (GO) terms in targets of X-linked and autosomal miRNAs using the GO terms related to biological processes only (GOterm_BP_fat). Only significant ($P < 0.05$) GO terms after Benjamini correction were considered. Clusters of GO terms were retained only if they contained at least two significant GO terms.

Statistical analyses were performed using JMP 4.0.4 (SAS Institute Inc., Cary, NC, USA). Significance of pairwise comparisons between the means of two samples was examined using Wilcoxon two-sample tests. Differences among the rate of sequence divergence and among the numbers of target genes in the three expression groups were determined using an ANOVA. Spearman's rank correlation was used to investigate the dependency between variables.

Results

Mature miRNAs are subject to strong purifying selection

Orthologs of nearly all *C. briggsae* miRNAs were identified with confidence in the genome of *C. sp. 9*. Homology search of the 140 *C. briggsae* miRNAs in the two related species *C. sp. 9* and *C. sp. 5* identified 124 and 111 orthologs respectively (Figure 1B). Some of the disparities in miRNA numbers between the three species could result from the quality of the genome assemblies for *C. sp. 5* and *C. sp. 9* (see also Supplementary Text), or from high sequence divergence limiting the homology search. However, there is no bias for miRNAs missing in *C. sp. 5* to be more rapidly evolving between *C. briggsae* and *C. sp. 9* than for miRNAs present in all three species (mature: $P = 0.29$; backbone: $P = 0.10$). Nevertheless, some of these disparities do reflect differences in miRNA complement between the three species. Five miRNAs – *mir-45*, *mir-77*, *mir-232*, *mir-357* and *mir-2222* – with two copies in the *C. briggsae* genome have a single copy in *C. sp. 9* and *C. sp. 5* (Figure 1A), indicating that they specifically duplicated in the lineage leading to *C. briggsae*. With the exception of the *Cbr-mir-232* paralogs, separated by 970 Kb, and the *Cbr-mir-357* copies found on different chromosomes, the paralogs of the three other duplicated miRNAs are less than 3 Kb distant in *C. briggsae*. In addition, lineage-specific duplications of *mir-35* genes suggest rapid evolution of an X-linked miRNA cluster (Supplementary Text and Supplementary Figure 1). It cannot be excluded that some *mir-35* genes in *C. sp. 5* are in fact paralogs of *C. briggsae mir-35* genes that duplicated after the split between the two species. However, all *mir-35* genes identified in *C. sp. 9*, and used to compute nucleotide divergence (see below),

are unambiguous orthologs of *C. briggsae* *mir-35* genes because they were identified using conserved synteny and sequence conservation extending beyond the miRNAs genes. The mature sequences of the miRNAs in *C. sp. 9* and *C. sp. 5* belong respectively to 75 and 67 families defined by a distinct seed motif (Figure 1C). Although deep sequencing of RNA samples is needed to fully annotate miRNAs in *C. sp. 9* and *C. sp. 5*, the homology search suggests that the miRNA content of the two most closely related species *C. briggsae* and *C. sp. 9* greatly overlaps. Hairpin and mature sequences of *C. sp. 5* and *C. sp. 9* miRNAs are available online as Supplementary Data.

The number of nucleotide substitutions is on average 1.6 times lower in the backbone (i.e. hairpin minus mature miRNA) relative to flanking sequences, revealing the action of purifying selection expected for this segment of the miRNA (Figure 2A). This is consistent with miRNA stars (miRNA*: complementary mature sequences) having some regulatory activity (Guo and Lu 2010; Okamura et al. 2008; Yang et al. 2011), and with the structural and sequence requirements for miRNA processing and activity (Bartel 2004; Cullen 2004; Liu et al. 2008; Yue et al. 2011). However, in agreement with their regulatory function, mature miRNAs experience the strongest purifying selection and have on average 6.2-fold less nucleotide divergence than the sequences flanking the miRNAs and 3.8-fold less than the backbone (Figure 2A).

A disproportional increase in the substitution rate in one lineage may reveal a shift in the regime of selection among lineages, causing a greater number of mutations to become fixed either because of adaptive or relaxed purifying selection. I used *C. sp. 5* as an outgroup to test if variation in selective constraints could be detected between *C. briggsae* and *C. sp. 9* miRNAs by polarizing substitutions between the two lineages and comparing branch lengths for each ortholog group. On average, the level of nucleotide divergence does not differ significantly between *C. briggsae* and *C. sp. 9* both for the mature miRNA ($K_{Cbr} = 0.0066$, $K_{Csp9} = 0.0110$; $P = 0.24$) and for the backbone ($K_{Cbr} = 0.0339$, $K_{Csp9} = 0.0367$; $P = 0.62$). This result is unaffected by miRNAs not found in *C. sp. 5* because faster evolving miRNAs between *C. briggsae* and *C. sp. 9* are no more likely to be lacking orthologs in *C. sp. 5* (mature: $P = 0.29$; backbone: $P = 0.10$). Thus, variation in evolutionary rates may reflect properties of individual miRNAs instead of a global shift in selective constraints in *C. briggsae* or in *C. sp. 9*.

Substitutions in the seed predict different sets of target genes

Despite overall strong conservation of the mature miRNA, 8% (9/106) of the miRNAs common to all three species and 3% (4/124) between the two most closely related species *C. briggsae* and *C. sp. 9* have substitutions in the seed sequence (Figure 2B). Exact matching of the seed sequence with target sites in the 3'UTR of messenger RNAs is essential for proper interaction (Bartel 2009; Brennecke et al. 2005), suggesting that these divergent miRNAs have acquired different functions by regulating distinct sets of target genes. Alternatively, stabilizing selection at the level of the interaction between miRNAs and their targets may promote sequence divergence between orthologous miRNAs if substitutions are compensated in the target sites, a situation akin to the co-evolution between transcription factors and their binding sites (Baker et al. 2011).

If miRNAs with seed substitutions in one species co-evolved with their targets to maintain similar function and target gene repertoires, then target sites corresponding to the new seed may be found disproportionately in the 3'UTR of the same target genes relative to the other species. Four *C. briggsae* miRNAs – *mir-35b-2*, *mir-35b-6*, *mir-35c-2* and *mir-790-2* – with divergent seed in their orthologs (or paralogs) in *C. sp.9* or *C. sp.5* (Figure 2B) belong to families conserved in *C. elegans* and have predicted targets. As mentioned above, *mir-35* genes tentatively annotated as orthologs between *C. briggsae* and *C. sp. 5* could in fact be

paralogs. However, the seeds of the *C. sp. 9* orthologs and *C. sp. 5* homologs define novel miRNA families not present in *C. briggsae* and *C. elegans*. Moreover, these miRNAs are paralogs of miRNAs that have conserved seeds in all three species. *Cbr-mir-790-1* and its orthologs in *C. sp. 5* and *C. sp. 9* all share the same seed with *Cbr-mir-790-2*. All *Cbr-mir-35* paralogs have the same seed that is also shared with 13 out of 14 *mir-35* genes in *C. sp. 9* and 14 out of 16 *mir-35* genes in *C. sp. 5*. This implicates that the seed substitutions occurred in *C. sp. 9* for *mir-790-2* and *mir-35c-2* and in the two *C. sp. 5* *mir-35* genes even if these are paralogs and not orthologs of *Cbr-mir-35b-2* and *Cbr-mir-35b-6* (Figure 2B). Thus, it can be inferred that the set of targets in *C. briggsae* represents the ancestral set of target genes for the *mir-35* and *mir-790* families in *C. sp. 5* and *C. sp. 9*.

The frequency of target sites corresponding to the novel seed is increased, but only slightly, in species having a substitution in the miRNA seed. Moreover, orthologs of *mir-790-2* targets in *C. sp. 5* and *C. briggsae*, and orthologs of *mir-35c-2* targets in *C. sp. 5* also have sites complementary to the novel seed present only in *C. sp. 9* (Figure 3). Note, however, that these four divergent miRNAs are paralogs of miRNAs with conserved seeds in all three species (Figure 1A). Because selection presumably maintains target sites corresponding to the ancestral seed in all three species, we may expect the frequencies of the sites corresponding to the ancestral and derived seeds to be close to 0.5 if the divergent miRNAs co-evolved with their target genes. Overall, then, there is no strong pattern of co-evolution, and the low frequency of the target sites of the derived seeds suggests that orthologs of these four miRNAs regulate largely distinct sets of target genes, if they retained any functionality. Similar results were found when using different lengths of downstream sequence (Supplementary Figure 3).

Genomic neighborhood has no effect on miRNA sequence divergence

A large proportion (28%) of the miRNAs in *C. briggsae* is located in introns. In chordates, novel miRNAs tend to be overrepresented in introns of protein-coding genes (Campo-Paysaa et al. 2011; Meunier et al. 2013; Wang et al. 2010). In addition, there is increasing evidence that young miRNAs tend to evolve faster than more phylogenetically conserved miRNAs (Fahlgren et al. 2010; Lu et al. 2008a; Meunier et al. 2013; Nozawa et al. 2010). However, intronic miRNAs are no more divergent than intergenic miRNAs in *Caenorhabditis* (mature: $K_{inter} = 0.0245$, $K_{intron} = 0.0138$, Wilcoxon-two sample $P = 0.43$; backbone: $K_{inter} = 0.0863$, $K_{intron} = 0.0656$, Wilcoxon-two sample $P = 0.13$). This result suggests that novel and rapidly diverging miRNAs are not overrepresented in *C. briggsae* introns. This pattern could be explained if novel miRNAs in *Caenorhabditis* emerge *de novo* as frequently in introns as in intergenic regions. Interestingly, the ratios of *C. briggsae* miRNAs missing in *C. sp. 5* to miRNAs shared between the two species that are intronic (7/32) or intergenic (22/76) are not significantly different (Fisher's exact test $P = 0.68$). However, caution should be taken in interpreting this result as evidence for equal rates of origination of miRNAs in introns and intergenic regions. Missing miRNAs in *C. sp. 5* do not necessarily indicate recent origin in the lineage leading to *C. briggsae* because they could have been secondarily lost or could be missing because of the nature of the *C. sp. 5* genome assembly. Nevertheless, among the miRNA families missing from *C. sp. 5*, the ratios of conserved families (i.e. present in *C. elegans*, *C. remanei*, *C. briggsae* and *C. sp. 9*) to novel families (i.e. not present in *C. elegans* and *C. remanei*) are not significantly different (Fisher's exact test $P = 0.14$) between intronic miRNAs (0/3) and intergenic miRNAs (3/1), suggesting that novel miRNAs may not originate preferentially in introns in *Caenorhabditis*.

In *Drosophila*, novel miRNAs arising from random hairpin structures originated equally from introns and intergenic sequences (Nozawa et al. 2010). Thus, I asked if the pattern of divergence in intronic miRNAs is similar for fruitflies and nematodes. Indeed, intronic

miRNAs in *Drosophila* do not diverge faster than intergenic miRNAs (mature: $K_{inter} = 0.0270$, $K_{intron} = 0.0202$, $P = 0.58$; backbone: $K_{inter} = 0.1325$, $K_{intron} = 0.1421$, $P = 0.94$), as expected if novel and rapidly diverging miRNAs originate at similar rates in introns and intergenic regions.

X-linked miRNAs evolve faster than autosomal miRNAs

In *C. briggsae*, 39% (54 out of 140) of the miRNAs are located on the X chromosome and, as a group, X-linked miRNAs are more divergent than autosomal miRNAs (Figure 4A). The backbone and the mature sequences of X-linked miRNAs evolve respectively 1.6 and 2 times faster than the corresponding regions in autosomal miRNAs. This difference is highly significant for the backbone ($P = 0.008$), although it is not significant for the mature sequence ($P = 0.2$) (Figure 4A). Importantly, the faster rate of sequence divergence is localized to the hairpin of X-linked miRNAs and does not extend to their flanking sequence (X-linked: $K_{flanking} = 0.131$, autosomal: $K_{flanking} = 0.125$; $P = 0.24$). I used the program DAVID (Huang et al. 2009) to examine the enrichment of Gene Ontology terms of miRNA targets in order to investigate possible functions of X-linked miRNAs. It is unlikely that differences in substitution rates between X-linked and autosomal miRNAs in *Caenorhabditis* are caused predominantly by differential function, because the targets of autosomal miRNAs have Gene Ontologies that greatly overlap with those of X-linked miRNA targets (Supplementary Table 1, Supplementary Table 2). Similar to nematodes, X-linked miRNAs in *Drosophila* accumulate more substitutions in the backbone and in the mature sequence than autosomal miRNAs, and in flies this difference is highly significant for both regions of the hairpin (Figure 4B). Thus, chromosomal linkage has an effect on miRNA sequence evolution in both nematodes and flies.

The level of miRNA expression strongly influences miRNA sequence evolution

The expression level of X-linked miRNAs is 2.3-fold lower ($P = 0.0007$) than for autosomal miRNAs in *Caenorhabditis* and 5.4-fold lower ($P = 0.0029$) in *Drosophila*. This is consistent with the finding that highly expressed human miRNAs have fewer substitutions and polymorphisms than miRNAs expressed at low levels (Liang and Li 2009), and with the negative correlation between sequence divergence in the backbone and miRNA expression level in flies (Shen et al. 2011).

The sequences of miRNAs expressed at high level are more constrained than the sequences of miRNAs expressed at intermediate and low levels (Figure 5). Expression level exerts a negative influence on sequence divergence across the entire hairpin, but also separately in the backbone and in the mature sequence (hairpin: $F(2, 121) = 4.89$, $P = 0.010$; backbone: $F(2, 121) = 4.88$, $P = 0.009$; mature: $F(2, 121) = 3.85$, $P = 0.024$). Similarly, nucleotide divergence in the mature miRNA, in the backbone and in the entire hairpin is negatively correlated with the level of miRNA expression in *Caenorhabditis*, independently of bin grouping (Table 1). In *Drosophila*, stronger constraints in the backbone of miRNAs expressed at higher levels is consistent with previous results (Shen et al. 2011), and this pattern is extended here to the mature sequence and the entire hairpin (Table 1). These results, together with those previously published for human and fly miRNAs (Liang and Li 2009; Shen et al. 2011), demonstrate that the impact of expression level on miRNA nucleotide variation is not restricted to specific taxa and strongly suggest that it may be a general phenomenon.

A possible explanation for the relationship between expression level and miRNA sequence evolution is that mutations in highly expressed miRNAs are more likely to be pleiotropic, with larger consequences on fitness. In support of this prediction, miRNAs expressed at low levels regulate, on average, significantly fewer target genes ($F(2, 112) = 3.91$, $P = 0.023$).

than highly expressed miRNAs in *Caenorhabditis* (Figure 6). Moreover, expression level correlates positively with the number of predicted target genes in *Caenorhabditis* (Spearman's $\rho = 0.2919$, $P = 0.0015$) and in *Drosophila* (Spearman's $\rho = 0.3046$, $P = 0.0002$). Importantly, the number of target genes correlates negatively and significantly with nucleotide divergence in the mature miRNA in *Caenorhabditis* and in *Drosophila*, but not with nucleotide divergence in the hairpin or in the backbone (Table 1). These results suggest that mutations in the mature sequence of highly expressed miRNAs are thus likely to have broad functional consequences and are disproportionately selected against. In contrast, lowly expressed miRNAs may experience only weak purifying selection, enabling mutations to drift to fixation.

Many miRNAs with unique seed sequence may have limited function

Half (4 out of 9) of the miRNAs with a substitution in the seed sequence belong to multi-member families and 3 out of 9 are members of the *mir-35* family (Figure 2B). Functional redundancy provided by paralogs buffers the deleterious effects of mutations, allowing substitutions to accumulate in protein-coding genes (Gu et al. 2003; Hsiao and Vitkup 2008; Ramani et al. 2012). Can the presence of paralogs lower selective constraints on miRNA sequence? Individual deletions of most *C. elegans* miRNAs do not cause large phenotypic abnormalities, suggesting functional redundancy among miRNAs (Miska et al. 2007) and/or subtle phenotypic effects that require fine-grained analyses (Abbott 2010; Ramani et al. 2012). This is even more striking for the *C. elegans mir-35* family, because each member of the *mir-35-41* cluster can rescue the lethal phenotype caused by the deletion of the entire cluster (Alvarez-Saavedra and Horvitz 2010).

However, miRNAs from families with unique seed do not evolve slower than multi-family member miRNAs (mature: $K_{single} = 0.0297$, $K_{multiple} = 0.0148$, $P = 0.07$; backbone: $K_{single} = 0.0821$, $K_{multiple} = 0.0752$, $P = 0.33$), indicating that they are not subject to stronger purifying selection (Figure 7A). One possibility is that many miRNAs with unique seed sequences are not well integrated in regulatory networks and have limited function. This is plausible because, in *Drosophila*, the birth and death rates of miRNAs are high, suggesting that if miRNAs primarily originate *de novo* they may have limited function at birth so that their loss would have little consequences on the organism's fitness (Lu et al. 2008b; Nozawa et al. 2010).

To further test the hypothesis that miRNAs with unique seed have limited function, I used a recently published dataset of gain-of-function mutations for 180 miRNAs in *Drosophila melanogaster* (Schertel et al. 2012). Gain-of-function (GOF) mutations obtained by overexpression have been useful for investigating miRNA function because they circumvent functional redundancy (Schertel et al. 2012). In *Drosophila*, the level of sequence divergence differs significantly for the mature sequence between miRNAs belonging to multi-member families and those with unique seeds (Figure 7A). However, single family member miRNAs evolve faster than multi family member miRNAs, opposite to the expected pattern if purifying selection would be relaxed because of the presence of paralogs (mature: $K_{single} = 0.0310$, $K_{multiple} = 0.0087$, $P = 0.014$; backbone: $K_{single} = 0.1362$, $K_{multiple} = 0.1305$, $P = 0.91$). First, the expression level of multi-family member miRNAs is 1.12-fold higher than for single family member miRNAs ($P < 0.0001$). Second, single family member miRNAs regulate on average fewer target genes than multi family member miRNAs ($T_{single} = 112.37$, $T_{multiple} = 190.48$, $P = 0.001$). Third, the ratio of miRNAs without a GOF phenotype to miRNAs having a GOF phenotype is higher among single family member miRNAs (87/41) than among multi-family members (25/27) ($\chi^2 = 6.225$, $P = 0.013$) (Figure 7B). Finally, the single family miRNAs without a GOF phenotype have a 2.3-fold lower expression level ($P = 0.002$) and have more divergent mature sequences ($K_{no-pheno} = 0.0361$,

$K_{pheno} = 0.0234$, $P = 0.041$) that single family miRNAs with a GOF phenotype (Figure 7C). Together, these results support the idea that many miRNAs with unique seeds have limited functional scope and are not strongly integrated in regulatory networks.

Discussion

In this study, I investigated nucleotide divergence at miRNA loci in *Caenorhabditis* and corroborated the findings by comparison with *Drosophila*. The pattern of nucleotide variation among miRNAs in *Caenorhabditis* is consistent with a previous report examining nucleotide polymorphism and divergence in a small subset of the *C. briggsae* miRNAs (Jovelin and Cutter 2011). Moreover, the low level of nucleotide variation localizing to mature miRNAs agrees with the pattern observed in animals, including humans, and plants (Chen and Rajewsky 2006; Ehrenreich and Purugganan 2008; Fahlgren et al. 2010; Loh et al. 2011; Quach et al. 2009; Saunders et al. 2007). Negative selection is clearly the dominant mode of selection experienced by miRNAs. However, the number of substitutions per site among all mature sequences ranges from 0 to 41%, indicating that selective constraints vary among nematode miRNAs. Moreover, the variation in evolution rates among miRNAs is not due to a disproportional increase in the substitution rate in one lineage but reflects properties of the miRNAs. The purpose of this study is to determine the factors contributing to the disparity in evolutionary rates among miRNAs.

Intronic miRNAs

It has been suggested that introns provide a convenient environment for the *de novo* emergence of miRNAs because mutations in unstructured transcripts could give rise to hairpin structures processed by the miRNA biogenesis pathway (Campo-Paysaa et al. 2011). Given the propensity of new miRNAs to originate in introns, at least in chordates (Campo-Paysaa et al. 2011; Meunier et al. 2013; Wang et al. 2010), and the rapid sequence evolution of young miRNAs in several species (Fahlgren et al. 2010; Lu et al. 2008a; Meunier et al. 2013; Nozawa et al. 2010), one may expect the evolutionary rate of miRNAs to be influenced by their genomic neighborhood.

However, intronic miRNAs do not diverge faster than intergenic miRNAs in *Caenorhabditis*, suggesting that novel miRNAs may not originate preferentially in introns. Full consideration of the origin of miRNAs in *Caenorhabditis* will require their analysis in a phylogenetic context. However, this is complicated without independent and experimentally derived catalogs of miRNAs. High sequence divergence in the hairpin backbone of *Caenorhabditis* miRNAs (de Wit et al. 2009) makes orthology inferences difficult, perhaps because of the high substitution rate observed in nematodes (Aguinaldo et al. 1997; Mushegian et al. 1998) and because of the presumably large divergence times among species with a sequenced genome (Cutter 2008). As an example, homology search of *C. briggsae* miRNAs in *C. sp. 11*, a more distantly related species (Kiontke et al. 2011), retrieved only 88 miRNAs (data not shown). Nevertheless, novel miRNA families present in *C. briggsae* and *C. sp. 9* do not seem to be disproportionally found in introns when using *C. elegans* and *C. remanei* as more distant outgroups. Thus, novel miRNAs in *Caenorhabditis* could emerge as frequently from hairpins in introns as from hairpins in intergenic regions, and/or older miRNAs may not move away from the intron they originate from. For instance, it is noteworthy that intronic miRNAs tend to remain associated with the same host gene over evolutionary time in mammals (Hoepfner et al. 2009). Indirect evidence that novel miRNAs are not overrepresented in introns in *Caenorhabditis* also comes from comparison with *Drosophila*. Phylogenetic analysis of miRNA origin in *Drosophila* shows equal rates of origination in introns and intergenic regions (Nozawa et al. 2010). As we may expect, intronic miRNAs do not evolve faster than intergenic miRNAs in *Drosophila*, as in

Caenorhabditis, despite higher substitution rates in young fly miRNAs (Lu et al. 2008a; Nozawa et al. 2010).

Faster sequence evolution of X-linked miRNAs

Chromosomal linkage has an effect on miRNA sequence evolution both in nematodes and flies because X-linked miRNAs evolve faster than autosomal miRNAs (Figure 4). For instance, the X-linked *mir-35* cluster in *C. briggsae* is characterized both by rapid gene duplication and/or gene loss (Figure 1A and Supplementary Figure 1) and by extensive nucleotide variation (Figure 2B). Although the domain of expression of this cluster is unknown in *C. briggsae*, it is noteworthy that the only *mir-35* cluster in *C. elegans*, the *mir-35-41* cluster located on chromosome II, is expressed exclusively in the gonad and predominantly during oogenesis (Alvarez-Saavedra and Horvitz 2010). Two X-linked miRNA clusters show rapid sequence evolution and complex patterns of duplications in primates (Li et al. 2010; Zhang et al. 2007). The pattern for these two miRNA clusters, expressed predominantly in testes, is consistent with the findings that X-linked miRNAs expressed in the testis show higher sequence divergence than autosomal miRNAs in mammals (Guo et al. 2009; Meunier et al. 2013). Spatio-temporal analysis of miRNA expression and functional enrichment of target genes suggest that the evolution of X-linked miRNAs expressed in testes may be driven by selection for their function in spermatogenesis and testis development (Guo et al. 2009; Li et al. 2010; Meunier et al. 2013; Zhang et al. 2007). Target genes of *C. briggsae* X-linked miRNAs are enriched in Gene Ontology terms related to reproduction, but autosomal miRNAs have Gene Ontologies that greatly overlap with those of X-linked miRNAs (Supplementary Table 1, Supplementary Table 2).

Theory predicts that recessive beneficial mutations will be more effectively selected on the X chromosome because they are readily expressed in the heterogametic individuals, resulting in faster rate of sequence divergence for X-linked loci relative to autosomal loci. And yet, more effective purifying selection at X-linked loci will result in disproportionately higher accumulation of slightly deleterious mutations on autosomal loci and faster rate of sequence evolution (Vicoso and Charlesworth 2006). However, higher male mutation rate, adaptive evolution from standing genetic variation, sexually antagonistic fitness effects, female-biased ratio and low effective population size can affect the predicted pattern of faster-X evolution in different directions (Vicoso and Charlesworth 2006). In *Drosophila*, the faster-X effect is weak for protein-coding genes because the fixation of adaptive mutations is countered by the purging of deleterious mutations (Mank et al. 2010), and seems to be more pronounced for rapidly evolving genes (Vicoso et al. 2008). It is unclear how the mating system of *C. briggsae* might affect the fast-X effect. Low male frequency in natural populations of primarily selfing XX hermaphrodites (Félix and Dureau 2012; Félix et al. 2013) could restrict the fast-X effect to male-biased X-linked genes. However, the selfing mode of reproduction could enhance the selection of new mutations by reducing heterozygosity. Nevertheless, X-linked genes in the selfing species *C. elegans* evolve more slowly than autosomal loci, and this pattern holds for X-linked male-bias genes and for rapidly evolving sperm genes located on the X chromosome (Cutter and Ward 2005). Higher codon bias for X-linked loci in *Drosophila* and *Caenorhabditis* is nevertheless compatible with stronger purifying selection on the X chromosome (Singh et al. 2005). Thus, the pattern of sequence evolution for X-linked and autosomal miRNAs is unlikely to be predominantly due to functional differences or because of a fast-X effect. Instead, differences in rates of sequence divergence between X-linked and autosomal miRNAs may be explained by differential expression levels.

Expression level and pleiotropic constraints

Here I showed that expression level strongly influences miRNA evolutionary rates in *Caenorhabditis* and in *Drosophila* and that it affects substitution rates across the entire hairpin but also in the backbone and in the mature miRNA separately (Figure 5, Table 1). These findings agree with reports for human and fly miRNAs (Liang and Li 2009; Shen et al. 2011), suggesting that the relationship between miRNA expression level and sequence evolution is not taxon-specific. Moreover, the inverse relationship between sequence variation and expression level of miRNAs emphasizes the role of gene expression level in molecular evolution and is consistent with the negative correlation between expression level and sequence divergence for long intergenic noncoding RNAs (Managadze et al. 2011) and for protein coding genes (Drummond et al. 2005).

The relationship between expression level and rates of protein sequence evolution is best explained by a fitness cost associated with protein misfolding, imposing selection for translational accuracy and translational robustness (Drummond and Wilke 2008; Drummond and Wilke 2009; Yang et al. 2010). Proper folding also is crucial to the proper processing and functionality of miRNAs. The secondary structure of the pre-miRNA is essential for cleavage of the miRNA duplex (Bartel 2004; Cullen 2004). Also, there is evidence that substitutions in the miRNA hairpin alter miRNA processing, reducing the expression level of functional molecules and/or generating alternative mature sequences, in addition to reducing miRNA-induced regulation (de Meaux et al. 2008; Duan et al. 2007; Jazdzewski et al. 2009; Jazdzewski et al. 2008; Lee et al. 2011). Consistent with the finding that the loop sequence controls the fidelity of mature miRNA formation in the 5' end (Yue et al. 2011), Shen et al. (2011) demonstrated that constraints in the backbone of miRNAs in *Drosophila* increase to minimize the error rate of miRNA processing as miRNA expression level increases. However, the authors focused on fly miRNAs for which the mature sequence is invariant (Shen et al. 2011). Moreover, because variation in the mature sequence can both affect miRNA processing and regulation (Duan et al. 2007), the negative relationship between expression level and nucleotide variation in the mature sequence (Figure 5, Table 1) (Liang and Li 2009) requires further explanation.

One hypothesis is that miRNAs expressed at low levels evolve faster because changes in these miRNAs would have limited fitness consequences. This prediction is supported by a positive correlation between the expression level of a miRNA and the number of target genes it regulates (Figure 6), suggesting that mutations in highly expressed miRNAs are more likely to be pleiotropic. Regulation of many targets may require miRNAs to be expressed at higher levels and/or miRNAs expressed in more tissues and cell types may have more targets than miRNAs expressed in a few cells. If expression level lowers miRNA sequence evolution because of functional constraints associated with the pleiotropic effect of deleterious mutations, then we would expect this effect to manifest itself on the part of the miRNA carrying the regulatory activity. Indeed, the number of target genes is negatively and significantly correlated with sequence divergence in the mature miRNA only, both for *Caenorhabditis* and for *Drosophila* (Table 1).

The results presented here support a universal negative relationship between rates of nucleotide substitution and expression level. However, for miRNAs, selective constraints on nucleotide variation in highly expressed genes are not only related to a cost associated with misfolding and processing (Shen et al. 2011) but also to fitness costs associated with target interaction.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

I thank Asher Cutter for comments on the manuscript. I also thank the Associate Editor and one anonymous reviewer for valuable comments and suggestions. This work was supported by grants from the Ontario Ministry of Research and Innovation and the National Health Institutes (R01-GM096008).

References

- Abbott AL. Uncovering new functions for microRNAs in *Caenorhabditis elegans*. *Curr Biol*. 2010; 21:R668–71. [PubMed: 21920301]
- Aguinaldo AM, Turbeville JM, Linford LS, Rivera MC, Garey JR, Raff RA, Lake JA. Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature*. 1997; 387:489–93. [PubMed: 9168109]
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990; 215:403–10. [PubMed: 2231712]
- Alvarez-Saavedra E. microRNAs are not essential for development or viability. *Curr Biol*. 2010; 20:367–73. [PubMed: 20096582]
- Axtell MJ, Westholm JO, Lai EC. Vive la difference: biogenesis and evolution of microRNAs in plants and animals. *Genome Biol*. 2011; 12:221. [PubMed: 21554756]
- Baker CR, Tuch BB, Johnson AD. Extensive DNA-binding specificity divergence of a conserved transcription regulator. *Proc Natl Acad Sci U S A*. 2011; 108:7493–8. [PubMed: 21498688]
- Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*. 2004; 116:281–97. [PubMed: 14744438]
- Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell*. 2009; 136:215–33. [PubMed: 19167326]
- Bartel DP, Chen CZ. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nat Rev Genet*. 2004; 5:396–400. [PubMed: 15143321]
- Berezikov E, Robine N, Samsonova A, Westholm JO, Naqvi A, Hung JH, Okamura K, Dai Q, Bortolamiol-Becet D, Martin R, Zhao Y, Zamore PD, Hannon GJ, Marra MA, Weng Z, Perrimon N, Lai EC. Deep annotation of *Drosophila melanogaster* microRNAs yields insights into their processing, modification, and emergence. *Genome Res*. 2011; 21:203–15. [PubMed: 21177969]
- Brennecke J, Stark A, Russell RB, Cohen SM. Principles of microRNA-target recognition. *PLoS Biol*. 2005; 3:e85. [PubMed: 15723116]
- Campbell LI, Rota-Stabelli O, Edgecombe GD, Marchioro T, Longhorn SJ, Telford MJ, Philippe H, Rebecchi L, Peterson KJ, Pisani D. MicroRNAs and phylogenomics resolve the relationships of Tardigrada and suggest that velvet worms are the sister group of Arthropoda. *Proc Natl Acad Sci U S A*. 2011; 108:15920–4. [PubMed: 21896763]
- Campo-Paysaa F, Semon M, Cameron RA, Peterson KJ, Schubert M. microRNA complements in deuterostomes: origin and evolution of microRNAs. *Evol Dev*. 2011; 13:15–27. [PubMed: 21210939]
- Canestro C, Catchen JM, Rodriguez-Mari A, Yokoi H, Postlethwait JH. Consequences of lineage-specific gene loss on functional evolution of surviving paralogs: ALDH1A and retinoic acid signaling in vertebrate genomes. *PLoS Genet*. 2009; 5:e1000496. [PubMed: 19478994]
- Carroll SB. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell*. 2008; 134:25–36. [PubMed: 18614008]
- Chen K, Rajewsky N. Natural selection on human microRNA binding sites inferred from SNP data. *Nat Genet*. 2006; 38:1452–6. [PubMed: 17072316]
- Chen SC, Chuang TJ, Li WH. The relationships among microRNA regulation, intrinsically disordered regions, and other indicators of protein evolutionary rate. *Mol Biol Evol*. 2011
- Cheng C, Bhardwaj N, Gerstein M. The relationship between the evolution of microRNA targets and the length of their UTRs. *BMC Genomics*. 2009; 10:431. [PubMed: 19751524]
- Cui Q, Yu Z, Purisima EO, Wang E. MicroRNA regulation and interspecific variation of gene expression. *Trends Genet*. 2007; 23:372–5. [PubMed: 17482307]

- Cullen BR. Transcription and processing of human microRNA precursors. *Mol Cell*. 2004; 16:861–5. [PubMed: 15610730]
- Cutter AD. Divergence times in *Caenorhabditis* and *Drosophila* inferred from direct estimates of the neutral mutation rate. *Mol Biol Evol*. 2008; 25:778–86. [PubMed: 18234705]
- Cutter AD, Ward S. Sexual and temporal dynamics of molecular evolution in *C. elegans* development. *Mol Biol Evol*. 2005; 22:178–88. [PubMed: 15371532]
- de Meaux J, Hu JY, Tartler U, Goebel U. Structurally different alleles of the ath-*MIR824* microRNA precursor are maintained at high frequency in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A*. 2008; 105:8994–9. [PubMed: 18579782]
- de Wit E, Linsen SE, Cuppen E, Berezikov E. Repertoire and evolution of miRNA genes in four divergent nematode species. *Genome Res*. 2009; 19:2064–74. [PubMed: 19755563]
- Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. Why highly expressed proteins evolve slowly. *Proc Natl Acad Sci U S A*. 2005; 102:14338–43. [PubMed: 16176987]
- Drummond DA, Wilke CO. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell*. 2008; 134:341–52. [PubMed: 18662548]
- Drummond DA, Wilke CO. The evolutionary consequences of erroneous protein synthesis. *Nat Rev Genet*. 2009; 10:715–24. [PubMed: 19763154]
- Duan R, Pak C, Jin P. Single nucleotide polymorphism associated with mature *miR-125a* alters the processing of pri-miRNA. *Hum Mol Genet*. 2007; 16:1124–31. [PubMed: 17400653]
- Ehrenreich IM, Purugganan MD. Sequence variation of MicroRNAs and their binding sites in *Arabidopsis*. *Plant Physiol*. 2008; 146:1974–82. [PubMed: 18305205]
- Fahlgren N, Jogdeo S, Kasschau KD, Sullivan CM, Chapman EJ, Laubinger S, Smith LM, Dasenko M, Givan SA, Weigel D, Carrington JC. MicroRNA gene evolution in *Arabidopsis lyrata* and *Arabidopsis thaliana*. *Plant Cell*. 2010; 22:1074–89. [PubMed: 20407027]
- Farh KK, Grimson A, Jan C, Lewis BP, Johnston WK, Lim LP, Burge CB, Bartel DP. The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. *Science*. 2005; 310:1817–21. [PubMed: 16308420]
- Felekis K, Voskarides K, Dweep H, Sticht C, Gretz N, Deltas C. Increased number of microRNA target sites in genes encoded in CNV regions. Evidence for an evolutionary genomic interaction? *Mol Biol Evol*. 2011; 28:2421–4.
- Félix MA, Duveau F. Population dynamics and habitat sharing of natural populations of *Caenorhabditis elegans* and *C. briggsae*. *BMC Biol*. 2012; 10:59. [PubMed: 22731941]
- Félix MA, Jovelin R, Ferrari C, Han S, Cho YR, Andersen EC, Cutter AD, Braendle C. Species richness, distribution and genetic diversity of *Caenorhabditis* nematodes in a remote tropical forest. *BMC Evol Biol*. 2013; 13:10. [PubMed: 23311925]
- Grimson A, Srivastava M, Fahey B, Woodcroft BJ, Chiang HR, King N, Degnan BM, Rokhsar DS, Bartel DP. Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature*. 2008; 455:1193–7. [PubMed: 18830242]
- Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL. The Vienna RNA websuite. *Nucleic Acids Res*. 2008; 36:W70–4. [PubMed: 18424795]
- Gu Z, Steinmetz LM, Gu X, Scharfe C, Davis RW, Li WH. Role of duplicate genes in genetic robustness against null mutations. *Nature*. 2003; 421:63–6. [PubMed: 12511954]
- Guder C, Philipp I, Lengfeld T, Watanabe H, Hobmayer B, Holstein TW. The Wnt code: cnidarians signal the way. *Oncogene*. 2006; 25:7450–60. [PubMed: 17143289]
- Guo L, Lu Z. The fate of miRNA* strand through evolutionary analysis: implication for degradation as merely carrier strand or potential regulatory molecule? *PLoS One*. 2010; 5:e11387. [PubMed: 20613982]
- Guo X, Su B, Zhou Z, Sha J. Rapid evolution of mammalian X-linked testis microRNAs. *BMC Genomics*. 2009; 10:97. [PubMed: 19257908]
- Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*. 1999; 41:95–98.

- Heimberg AM, Cowper-Sal-lari R, Semon M, Donoghue PC, Peterson KJ. microRNAs reveal the interrelationships of hagfish, lampreys, and gnathostomes and the nature of the ancestral vertebrate. *Proc Natl Acad Sci U S A*. 2010; 107:19379–83. [PubMed: 20959416]
- Heimberg AM, Sempere LF, Moy VN, Donoghue PC, Peterson KJ. MicroRNAs and the advent of vertebrate morphological complexity. *Proc Natl Acad Sci U S A*. 2008; 105:2946–50. [PubMed: 18287013]
- Herranz H, Cohen SM. MicroRNAs and gene regulatory networks: managing the impact of noise in biological systems. *Genes Dev*. 2010; 24:1339–44. [PubMed: 20595229]
- Hoepfner MP, White S, Jeffares DC, Poole AM. Evolutionarily stable association of intronic snoRNAs and microRNAs with their host genes. *Genome Biol Evol*. 2009; 1:420–8. [PubMed: 20333211]
- Hsiao TL, Vitkup D. Role of duplicate genes in robustness against deleterious human mutations. *PLoS Genet*. 2008; 4:e1000014. [PubMed: 18369440]
- Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009; 4:44–57. [PubMed: 19131956]
- Iwama H, Kato K, Imachi H, Murao K, Masaki T. Human MicroRNAs Originated from Two Periods at Accelerated Rates in Mammalian Evolution. *Mol Biol Evol*. 2013; 30:613–26. [PubMed: 23171859]
- Jan CH, Friedman RC, Ruby JG, Bartel DP. Formation, regulation and evolution of *Caenorhabditis elegans* 3'UTRs. *Nature*. 2011; 469:97–101. [PubMed: 21085120]
- Jazdzewski K, Liyanarachchi S, Swierniak M, Pachucki J, Ringel MD, Jarzab B, de la Chapelle A. Polymorphic mature microRNAs from passenger strand of pre-miR-146a contribute to thyroid cancer. *Proc Natl Acad Sci U S A*. 2009; 106:1502–5. [PubMed: 19164563]
- Jazdzewski K, Murray EL, Franssila K, Jarzab B, Schoenberg DR, de la Chapelle A. Common SNP in pre-miR-146a decreases mature miR expression and predisposes to papillary thyroid carcinoma. *Proc Natl Acad Sci U S A*. 2008; 105:7269–74. [PubMed: 18474871]
- Jovelin R, Cutter AD. MicroRNA sequence variation potentially contributes to within-species functional divergence in the nematode *Caenorhabditis briggsae*. *Genetics*. 2011; 189:967–76. [PubMed: 21890738]
- Jovelin R, Phillips PC. Expression level drives the pattern of selective constraints along the insulin/Tor signal transduction pathway in *Caenorhabditis*. *Genome Biol Evol*. 2011; 3:715–22. [PubMed: 21849326]
- Jovelin R, Yan YL, He X, Catchen J, Amores A, Canestro C, Yokoi H, Postlethwait JH. Evolution of developmental regulation in the vertebrate FgfD subfamily. *J Exp Zool B Mol Dev Evol*. 2010; 314:33–56. [PubMed: 19562753]
- Kasbauer T, Towb P, Alexandrova O, David CN, Dall'armi E, Staudigl A, Stiening B, Bottger A. The Notch signaling pathway in the cnidarian *Hydra*. *Developmental Biology*. 2007; 303:376–90. [PubMed: 17184766]
- Kiontke KC, Felix MA, Ailion M, Rockman MV, Braendle C, Penigault JB, Fitch DH. A phylogeny and molecular barcodes for *Caenorhabditis*, with numerous new species from rotting fruits. *BMC Evol Biol*. 2011; 11:339. [PubMed: 22103856]
- Kozłowska JL, Ahmad AR, Jahesh E, Cutter AD. Genetic variation for postzygotic reproductive isolation between *Caenorhabditis briggsae* and *Caenorhabditis* sp. 9. *Evolution*. 2012; 66:1180–95. [PubMed: 22486697]
- Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res*. 2011; 39:D152–7. [PubMed: 21037258]
- Lee HC, Yang CW, Chen CY, Au LC. Single point mutation of microRNA may cause butterfly effect on alteration of global gene expression. *Biochem Biophys Res Commun*. 2011; 404:1065–9. [PubMed: 21195055]
- Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*. 1993; 75:843–54. [PubMed: 8252621]
- Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*. 2005; 120:15–20. [PubMed: 15652477]

- Li J, Liu Y, Dong D, Zhang Z. Evolution of an X-linked primate-specific micro RNA cluster. *Mol Biol Evol.* 2010; 27:671–83. [PubMed: 19933172]
- Li X, Cassidy JJ, Reinke CA, Fischboeck S, Carthew RW. A microRNA imparts robustness against environmental fluctuation during development. *Cell.* 2009; 137:273–82. [PubMed: 19379693]
- Liang H, Li WH. Lowly expressed human microRNA genes evolve rapidly. *Mol Biol Evol.* 2009; 26:1195–8. [PubMed: 19299536]
- Liu G, Min H, Yue S, Chen CZ. Pre-miRNA loop nucleotides control the distinct activities of *mir-181a-1* and *mir-181c* in early T cell development. *PLoS One.* 2008; 3:e3592. [PubMed: 18974849]
- Loh YH, Yi SV, Streelman JT. Evolution of microRNAs and the diversification of species. *Genome Biol Evol.* 2011; 3:55–65. [PubMed: 21169229]
- Lu J, Clark AG. Impact of microRNA regulation on variation in human gene expression. *Genome Res.* 2012; 22:1243–54. [PubMed: 22456605]
- Lu J, Fu Y, Kumar S, Shen Y, Zeng K, Xu A, Carthew R, Wu CI. Adaptive evolution of newly emerged micro-RNA genes in *Drosophila*. *Mol Biol Evol.* 2008a; 25:929–38. [PubMed: 18296702]
- Lu J, Shen Y, Wu Q, Kumar S, He B, Shi S, Carthew RW, Wang SM, Wu CI. The birth and death of microRNA genes in *Drosophila*. *Nat Genet.* 2008b; 40:351–5. [PubMed: 18278047]
- Lynch M, Force AG. The origin of interspecific genomic incompatibility via gene duplication. *Am Nat.* 2000; 156:590–605.
- Managadze D, Rogozin IB, Chernikova D, Shabalina SA, Koonin EV. Negative correlation between expression level and evolutionary rate of long intergenic noncoding RNAs. *Genome Biol Evol.* 2011; 3:1390–404. [PubMed: 22071789]
- Mank JE, Vicoso B, Berlin S, Charlesworth B. Effective population size and the Faster-X effect: empirical results and their interpretation. *Evolution.* 2010; 64:663–74. [PubMed: 19796145]
- Matus DQ, Thomsen GH, Martindale MQ. FGF signaling in gastrulation and neural development in *Nematostella vectensis*, an anthozoan cnidarian. *Development Genes and Evolution.* 2007; 217:137–48. [PubMed: 17237944]
- Meunier J, Lemoine F, Soumillon M, Liechti A, Weier M, Guschanski K, Hu H, Khaitovich P, Kaessmann H. Birth and expression evolution of mammalian microRNA genes. *Genome Res.* 2013; 23:34–45. [PubMed: 23034410]
- Miska EA, Alvarez-Saavedra E, Abbott AL, Lau NC, Hellman AB, McGonagle SM, Bartel DP, Ambros VR, Horvitz HR. Most *Caenorhabditis elegans* microRNAs are individually not essential for development or viability. *PLoS Genet.* 2007; 3:e215. [PubMed: 18085825]
- Mukherji S, Ebert MS, Zheng GX, Tsang JS, Sharp PA, van Oudenaarden A. MicroRNAs can generate thresholds in target gene expression. *Nat Genet.* 2011; 43:854–9. [PubMed: 21857679]
- Mushegian AR, Garey JR, Martin J, Liu LX. Large-scale taxonomic profiling of eukaryotic model organisms: a comparison of orthologous proteins encoded by the human, fly, nematode, and yeast genomes. *Genome Res.* 1998; 8:590–8. [PubMed: 9647634]
- Nichols SA, Dirks W, Pearce JS, King N. Early evolution of animal cell signaling and adhesion genes. *Proceedings of the National Academy of Sciences of the United States of America.* 2006; 103:12451–6. [PubMed: 16891419]
- Nozawa M, Miura S, Nei M. Origins and evolution of microRNA genes in *Drosophila* species. *Genome Biol Evol.* 2010; 2:180–9. [PubMed: 20624724]
- Okamura K, Phillips MD, Tyler DM, Duan H, Chou YT, Lai EC. The regulatory activity of microRNA* species has substantial influence on microRNA and 3' UTR evolution. *Nat Struct Mol Biol.* 2008; 15:354–63. [PubMed: 18376413]
- Quach H, Barreiro LB, Laval G, Zidane N, Patin E, Kidd KK, Kidd JR, Bouchier C, Veuille M, Antoniewski C, Quintana-Murci L. Signatures of purifying and local positive selection in human miRNAs. *Am J Hum Genet.* 2009; 84:316–27. [PubMed: 19232555]
- Ramani AK, Chuluunbaatar T, Verster AJ, Na H, Vu V, Pelte N, Wannissorn N, Jiao A, Fraser AG. The majority of animal genes are required for wild-type fitness. *Cell.* 2012; 148:792–802. [PubMed: 22341449]

- Ruby JG, Jan C, Player C, Axtell MJ, Lee W, Nusbaum C, Ge H, Bartel DP. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell*. 2006; 127:1193–207. [PubMed: 17174894]
- Saunders MA, Liang H, Li WH. Human polymorphism at microRNAs and microRNA target sites. *Proc Natl Acad Sci U S A*. 2007; 104:3300–5. [PubMed: 17360642]
- Schertel C, Rutishauser T, Forstemann K, Basler K. Functional characterization of *Drosophila* microRNAs by a novel in vivo library. *Genetics*. 2012; 192:1543–52. [PubMed: 23051640]
- Shen Y, Lv Y, Huang L, Liu W, Wen M, Tang T, Zhang R, Hungate E, Shi S, Wu CI. Testing hypotheses on the rate of molecular evolution in relation to gene expression using microRNAs. *Proc Natl Acad Sci U S A*. 2011; 108:15942–7. [PubMed: 21911382]
- Shin C, Nam JW, Farh KK, Chiang HR, Shkumatava A, Bartel DP. Expanding the microRNA targeting code: functional sites with centered pairing. *Mol Cell*. 2010; 38:789–802. [PubMed: 20620952]
- Singh ND, Davis JC, Petrov DA. X-linked genes evolve higher codon bias in *Drosophila* and *Caenorhabditis*. *Genetics*. 2005; 171:145–55. [PubMed: 15965246]
- Stark A, Brennecke J, Bushati N, Russell RB, Cohen SM. Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. *Cell*. 2005; 123:1133–46. [PubMed: 16337999]
- Takuno S, Innan H. Selection fine tunes the expression of microRNA target genes in *Arabidopsis thaliana*. *Mol Biol Evol*. 2011; 28:2429–34. [PubMed: 21478371]
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011; 28:2731–9. [PubMed: 21546353]
- Technau U, Rudd S, Maxwell P, Gordon PM, Saina M, Grasso LC, Hayward DC, Sensen CW, Saint R, Holstein TW, Ball EE, Miller DJ. Maintenance of ancestral complexity and non-metazoan genes in two basal cnidarians. *Trends Genet*. 2005; 21:633–9. [PubMed: 16226338]
- Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 1994; 22:4673–80. [PubMed: 7984417]
- Vicoso B, Charlesworth B. Evolution on the X chromosome: unusual patterns and processes. *Nat Rev Genet*. 2006; 7:645–53. [PubMed: 16847464]
- Vicoso B, Haddrill PR, Charlesworth B. A multispecies approach for comparing sequence evolution of X-linked and autosomal sites in *Drosophila*. *Genet Res (Camb)*. 2008; 90:421–31. [PubMed: 19061532]
- Wang QH, Zhou M, Sun J, Ning SW, Li Y, Chen L, Zheng Y, Li X, Lv SL, Li X. Systematic analysis of human microRNA divergence based on evolutionary emergence. *FEBS Lett*. 2010; 585:240–8. [PubMed: 21130764]
- Wheeler BM, Heimberg AM, Moy VN, Sperling EA, Holstein TW, Heber S, Peterson KJ. The deep evolution of metazoan microRNAs. *Evol Dev*. 2009; 11:50–68. [PubMed: 19196333]
- Woodruff GC, Eke O, Baird SE, Felix MA, Haag ES. Insights into species divergence and the evolution of hermaphroditism from fertile interspecies hybrids of *Caenorhabditis* nematodes. *Genetics*. 2010; 186:997–1012. [PubMed: 20823339]
- Wu CI, Shen Y, Tang T. Evolution under canalization and the dual roles of microRNAs: a hypothesis. *Genome Res*. 2009; 19:734–43. [PubMed: 19411598]
- Yang JR, Zhuang SM, Zhang J. Impact of translational error-induced and error-free misfolding on the rate of protein evolution. *Mol Syst Biol*. 2010; 6:421. [PubMed: 20959819]
- Yang JS, Phillips MD, Betel D, Mu P, Ventura A, Siepel AC, Chen KC, Lai EC. Widespread regulatory activity of vertebrate microRNA* species. *Rna*. 2011; 17:312–26. [PubMed: 21177881]
- Yue SB, Trujillo RD, Tang Y, O'Gorman WE, Chen CZ. Loop nucleotides control primary and mature miRNA function in target recognition and repression. *RNA Biol*. 2011; 8:1115–23. [PubMed: 22142974]
- Zhang R, Peng Y, Wang W, Su B. Rapid evolution of an X-linked microRNA cluster in primates. *Genome Res*. 2007; 17:612–7. [PubMed: 17416744]

Fig. 1.

A. Conservation matrix among miRNAs in three nematode species. The name of the *Caenorhabditis briggsae* miRNA used as query against the genome assemblies of *Caenorhabditis* sp. 9 (Csp9) and *Caenorhabditis* sp. 5 (Csp5) is shown in left column. Dark grey boxes indicate orthologs with conserved mature sequences. Light grey boxes indicate orthologous miRNAs with substitutions in the mature sequence. *C. briggsae* miRNAs with no orthologs are shown with white boxes. Crosses illustrate ambiguous orthology status due to missing data in the genome assembly. Note that orthology is tentatively assigned for members of the *mir-35* family between *C. briggsae* and *C. sp. 5* but that it cannot be excluded that some of the *mir-35* genes are paralogs rather than orthologs (Supplementary Text and Supplementary Figure 1). B. The number of orthologous miRNAs in each species. C. The number of miRNA families defined by distinct seed motif

Fig. 2.

A. miRNAs experience strong purifying selection, as evident from comparisons of mean nucleotide divergence between the mature sequence, the backbone (the hairpin minus the mature miRNA) and their flanking sequence. Wilcoxon two-sample tests: *** $P < 0.001$. Error bars indicate ± 1 standard error of the mean (SEM). B. Nucleotide substitutions between orthologous miRNAs (or perhaps between paralogous miRNAs for *Csp5-mir-35* genes, see text for details) in the seed motif predict functional divergence and the regulation of distinct sets of target genes. The seed sequence is in bold and dots represent identical nucleotides relative to the top miRNA. Shaded boxes highlight substitutions between orthologs in the seed. Dashes indicate indels. Cbr: *C. briggsae*; C9: *C. sp. 9*; C5: *C. sp. 5*

Fig. 3.

No pattern of co-evolution between divergent miRNAs in *C. sp. 5* and *C. sp. 9* and the targets of their orthologs (or paralogs) in *C. briggsae*. The frequency of target sites corresponding to the novel seed is low and is only slightly increased in species having a substitution in the seed. Sequences of miRNA seeds are shown on the left. The frequency of the target sites for the ancestral seed (dark grey) and the derived divergent seed (white) are plotted on the left

Fig. 4. X-linked miRNAs evolve faster than autosomal miRNAs in *Caenorhabditis* (A) and in *Drosophila* (B). Comparison of mean nucleotide divergence between miRNAs located on the X chromosome and on the autosomes in *C. briggsae* and in *D. melanogaster*. Wilcoxon two-sample tests: ** $P < 0.01$, *** $P < 0.001$, NS: not significant. Error bars indicate ± 1 SEM

Fig. 5.

The average level of nucleotide divergence in the miRNA hairpin, the backbone and the mature sequence decreases with increasing levels of miRNA expression. Error bars indicate ± 1 SEM. ANOVA: hairpin: $F(2, 121) = 4.89$, $P = 0.010$; backbone: $F(2, 121) = 4.88$, $P = 0.009$; mature: $F(2, 121) = 3.85$, $P = 0.024$

Fig. 6.

The average number of target genes increases with higher levels of miRNA expression. Error bars indicate ± 1 SEM. ANOVA: $F(2, 112) = 3.91$, $P = 0.023$

Fig. 7.

Many miRNAs with unique seed have limited function in regulatory networks. A. Nucleotide divergence is higher, on average, for miRNAs with unique seed than for miRNAs with paralogs in *Caenorhabditis* and in *Drosophila*. B. The proportion of miRNAs without gain-of-function phenotype is significantly higher for single-family member miRNAs in *Drosophila*. C. The level of nucleotide divergence is significantly higher for miRNAs without phenotype among miRNAs with unique seed. In contrast, miRNAs with a phenotype do not evolve more slowly when they have paralogs. Unique: miRNAs with unique seed; Multi: miRNAs belonging to multi-member families. A and C: Wilcoxon two-sample tests, * $P < 0.05$. B: χ^2 test * $P < 0.05$. Error bars indicate ± 1 SEM

Table 1

Spearman's rank correlations between nucleotide divergence in miRNAs in *Caenorhabditis* and *Drosophila* and several predictor variables. Nucleotide divergence is measured for the entire miRNA hairpin and also separately for the backbone and for the mature sequence.

	<i>Caenorhabditis</i>			<i>Drosophila</i>		
	hairpin	backbone	mature	hairpin	backbone	mature
miRNA expression level	-0.3900 ***	-0.3982 ***	-0.3211 ***	-0.2570 **	-0.1621 *	-0.4960 ***
Number of targets	-0.1786	-0.1265	-0.2414 *	-0.1103	-0.0328	-0.2269 **
Minimum free energy	0.1037	0.0880	0.1345	-0.0394	-0.0404	0.1101

* $P < 0.05$,

** $P < 0.01$,

*** $P < 0.001$