Genome **Biology**

REVIEW

Vive la différence: biogenesis and evolution of microRNAs in plants and animals

Michael J Axtell*[1], Jakub O Westholm[2] and Eric C Lai*[2]

Abstract

MicroRNAs are pervasive in both plants and animals, but many aspects of their biogenesis, function and evolution differ. We reveal how these differences contribute to characteristic features of microRNA evolution in the two kingdoms.

The first microRNAs (miRNAs) to be discovered, lin-4 and let-7, were found to be regulators of *Caenorhabditis elegans* development [1-3], and they established a paradigm for eukaryotic gene regulation in which short hairpins generate RNAs of approximately 22 nucleotides (nt) that repress specific target mRNAs. miRNAs have proved to be pervasive in both animals [4-6] and plants [7,8], acting as sequence-specific guides for target recognition [9,10]. Several thousand miRNAs have now been found in dozens of plants and animals [11]. Moreover, the biogenesis and activity of miRNAs are strongly related to those of small interfering RNAs (siRNAs) that mediate RNA interference, another ancient mechanism for post-transcriptional gene silencing [12].

Although miRNAs mediate diverse aspects of development and physiology in both plants and animals [13,14]**,** there are substantial differences between them. For example, the loci that produce miRNAs have distinct genomic arrangements in each kingdom. Furthermore, miRNAs are excised from precursor transcripts by different pathways in the two kingdoms, and in different subcellular compartments. Once made, plant and animal miRNAs have vastly different suites of direct targets; the number of direct targets of a given animal miRNA generally exceeds that of a given plant miRNA by at least an order of magnitude [15]. Herein, we focus on how

these differences contribute to, and are the result of, distinct evolutionary characteristics of miRNAs in the two kingdoms. We also highlight many commonalities between the respective systems that may reflect a shared evolutionary heritage or convergent strategies for handling and metabolizing double-stranded RNAs.

Distinct characteristics of miRNA pathways in plants and animals

What is a miRNA? Answering this question is not a simple task, as no single definition clearly and specifically encompasses all miRNAs. Although practical guides for miRNA annotation in plants and animals exist [16,17], not all loci reported in the miRBase registry [18] have been annotated to the same degree of confidence. In general, miRNAs are the products of inverted repeat transcripts that are precisely cleaved by RNase III enzyme(s) in the Dicer and/or Drosha protein families to yield small RNAs of approximately 21 to 24 nucleotides that guide Argonaute (AGO) proteins to complementary targets. Analogous, but distinct, core pathways govern the biogenesis of most miRNAs in plants and animals. Although we focus on these canonical miRNA pathways, a plethora of alternative pathways exist. Indeed, the diversity and flexibility of miRNA biogenesis pathways, in concert with related mechanisms that generate siRNAs, have made a significant contribution to miRNA evolution. In addition, while a hallmark of most studied miRNAs is the precise manner in which they are excised from precursor hairpins, there are examples of imprecisely cleaved miRNAs. As we shall see, this phenomenon might have implications for miRNA evolution, but it also poses challenges for the accurate distinction of *bona fide* miRNAs from fortuitous hairpins associated with short RNAs not generated by a specific biogenesis machinery.

Canonical miRNA biogenesis in plants versus animals

The biogenesis of plant miRNAs has been documented most thoroughly in *Arabidopsis thaliana* (Figure 1a). Primary miRNA (pri-miRNA) transcripts are products of RNA polymerase II that contain a hairpin RNA secondary structure [19]. The length of plant pri-miRNA

*Correspondence: Michael J Axtell, mja18@psu.edu; Eric C Lai, laie@mskcc.org
[1]Department of Biology, The Pennsylvania State University, 208 Mueller Laboratory, University Park, PA 16802, USA
[2]Department of Developmental Biology, Sloan-Kettering Institute, 1275 York Avenue, Box 252, New York, NY 10065, USA

BioMed Central

hairpins is heterogeneous, ranging from approximately 70 to many hundreds of bases. Of the four Dicer-like enzymes in *Arabidopsis*, Dicer-like 1 (DCL1) is responsible for the bulk of miRNA biogenesis [20]. DCL1 usually cleaves from the base of the pri-miRNA hairpin to yield a precursor-miRNA (pre-miRNA) hairpin, and cleaves again to release a miRNA/miRNA* duplex [21], although 'loop-first' processing, where the first DCL1-catalyzed cut occurs proximal to the loop, can also occur [22,23]. Most plant pri-miRNA hairpins produce a single miRNA/miRNA* duplex, but some loci, including *MIR159* and *MIR319*, consistently produce multiple duplexes [22-24].

Plant miRNA/miRNA* biogenesis is completed within the nucleus [25] in specialized subnuclear regions termed D-bodies [26,27]. Several accessory factors also contribute to the efficiency and fidelity of miRNA/miRNA* excision in plants (for a recent review, see [14]). The 3'-most nucleotides of the initial miRNA/miRNA* duplex are then 2'-*O*-methylated by the nuclear HEN1 protein [28]; this modification prevents non-templated 3'-polymerization that accelerates miRNA turnover [29]. HASTY, a plant homolog of Exportin-5, is then thought to export miRNA/miRNA* duplexes for loading into cytoplasmic AGO proteins [25], of which AGO1 is the predominant carrier of plant miRNAs. AGO1 can act as a 'slicer' to direct the endonucleolytic cleavage of target RNAs [30,31]; most other plant AGOs are also likely to possess slicing capabilities [32].

The collected studies from *Drosophila*, *C. elegans* and mammalian cells [12] indicate a conserved animal mechanism that is analogous to, but distinct from, plant miRNA biogenesis (Figure 1b). Most animal miRNAs are transcribed by RNA polymerase II, although a subset of animal miRNAs are products of RNA polymerase III [33]. The major difference compared with plants is the segregated cleavage of miRNA precursors by nuclear and cytoplasmic RNase III enzymes. All animals use the Drosha RNase III enzyme, which partners with the double-stranded RNA-binding domain protein DGCR8 (known as Pasha in invertebrates), to liberate pre-miRNA hairpins from pri-miRNA transcripts. The lengths of pre-miRNAs are more consistent in animals than in plants, with most in the 55- to 70-nucleotide range; however, select *Drosophila* pre-miRNAs can approach 200 nucleotides [34].

Following nuclear export of pre-miRNAs by Exportin-5, they are cleaved into miRNA/miRNA* duplexes by cytoplasmic Dicer (a single enzyme in *C. elegans* and vertebrates, and Dicer-1 in *Drosophila*). These are loaded into miRNA effector Argonautes (*Drosophila* dAGO1, *C. elegans* ALG1/2, and vertebrate Ago1 to Ago4). Of the mammalian Ago proteins, only Ago2 has Slicer activity [35,36]; *Drosophila* dAGO1 has

Slicer activity, but appears to have poorer turnover than its paralog dAGO2, the major carrier of endogenous siRNAs (endo-siRNAs) [37,38]. Curiously, while plant miRNAs are universally methylated at their 3' ends by HEN1, most products of animal miRNA genes are not. An exception regards *Drosophila* miRNA* strands, which are preferentially loaded into dAGO2 (Figure 1b). All dAGO2 cargoes, including miRNA* strands, endo-siRNAs and exogenous siRNAs from viruses or artificial dsRNA, are methylated by HEN1 as single-stranded species [39-42]. In addition to this core machinery, several accessory factors influence the biogenesis efficiency, fidelity and sorting of animal miRNAs [43]. Notably, a growing number of these factors act in cell-specific or state-specific manners to regulate miRNA production or activity, indicating that neither process is constitutive.

Genomic arrangement of plant and animal miRNA genes

In plants, most miRNA-encoding loci comprise independent, non-protein-coding transcription units. Among the rare intronic plant miRNAs is one present within an intron of *DCL1* orthologs; this miRNA might direct feedback regulation of miRNA biogenesis [44,45]. Plant miRNA hairpins sometimes occur in genomic clusters, strongly suggesting expression of multiple hairpins from a single pri-miRNA. Around one-fifth of annotated miRNAs in *Arabidopsis*, rice and poplar occur in tandem clusters at distances less than 10 kb [46]. Most clusters in these species (61% to 90%) contain hairpins encoding identical mature miRNAs, suggesting that they were the result of local tandem duplications and serve to increase the dosage of a particular miRNA from a single promoter [46]. The minority of plant miRNA clusters that produce more than one mature miRNA family nearly all encode mature species that are not conserved outside of the genus within which they were first described.

The genomic patterns of animal miRNA genes are significantly different from those of plants. Although many derive from stand-alone non-protein-coding loci, approximately 30% are located on the sense strands of introns [47]. There are only a few cases of miRNAs transcribed antisense to introns, suggesting some evolutionary benefit to sense intronic location. The simplest notion is that this arrangement permits miRNAs to take advantage of *cis*-regulatory elements that direct the expression of the host mRNA; however, intronic miRNAs can also be controlled by independent regulatory elements [48-50]. Unlike in plants, there are also occasional examples of miRNA biogenesis from exons of animal protein-coding genes, including UTRs and coding sequences (CDSs) [51,52].

Clustering of miRNA genes is more common in animals than in plants: up to 40% of miRNAs in

nematodes, flies and mammals are clustered in their respective genomes. Curiously, while there are many cases of locally duplicated miRNA hairpins, it is common for animal miRNA clusters to encode unrelated mature miRNAs. Only 5% to 20% of operons are composed exclusively of duplicated miRNAs in these three well-studied animal clades. Therefore, amplification of specific miRNA levels is not sufficient to explain the composition of animal miRNA operons. Instead, these different genomic origins and arrangements reflect distinct evolutionary styles of canonical miRNAs in plants and animals.

Non-canonical biogenesis pathways for inverted repeat transcripts: miRNAs and siRNAs

Many non-canonical mechanisms convert precursor transcripts into miRNAs and/or siRNAs. The strategies that are most relevant to miRNA evolution are pathways that metabolize inverted repeat transcripts (for reviews, see [53,54]). In animal cells, the first major alternative to

the canonical miRNA pathway came with the recognition of mirtrons (Figure 1b), which are pre-miRNA hairpins whose ends are defined by splicing instead of Drosha cleavage [55-57]. Following their debranching into a linear form, they are diced into conventional miRNAs. Mirtron biogenesis has not been extensively documented in plants, but one short hairpin intron in rice (*MIR1429*) generates specific miRNA/miRNA* reads indicative of mirtron processing [58].

In addition to other types of Drosha-independent miRNAs in animals [59], the conserved vertebrate miR-451 matures by a Dicer-independent mechanism [60-62]. Following Drosha cleavage, the *pre-mir-451* hairpin has only 19 bp of stem, which is too short to be cleaved by Dicer. Instead, it is loaded directly into Ago2, the sole vertebrate Argonaute-class Slicer enzyme. Ago2 cleaves its 3′ hairpin arm generating a 30-nucleotide species, whose 3′ end is resected to yield the mature miRNA of approximately 23 nucleotides (Figure 1b). The miR-451 pathway is instructive in that it does not
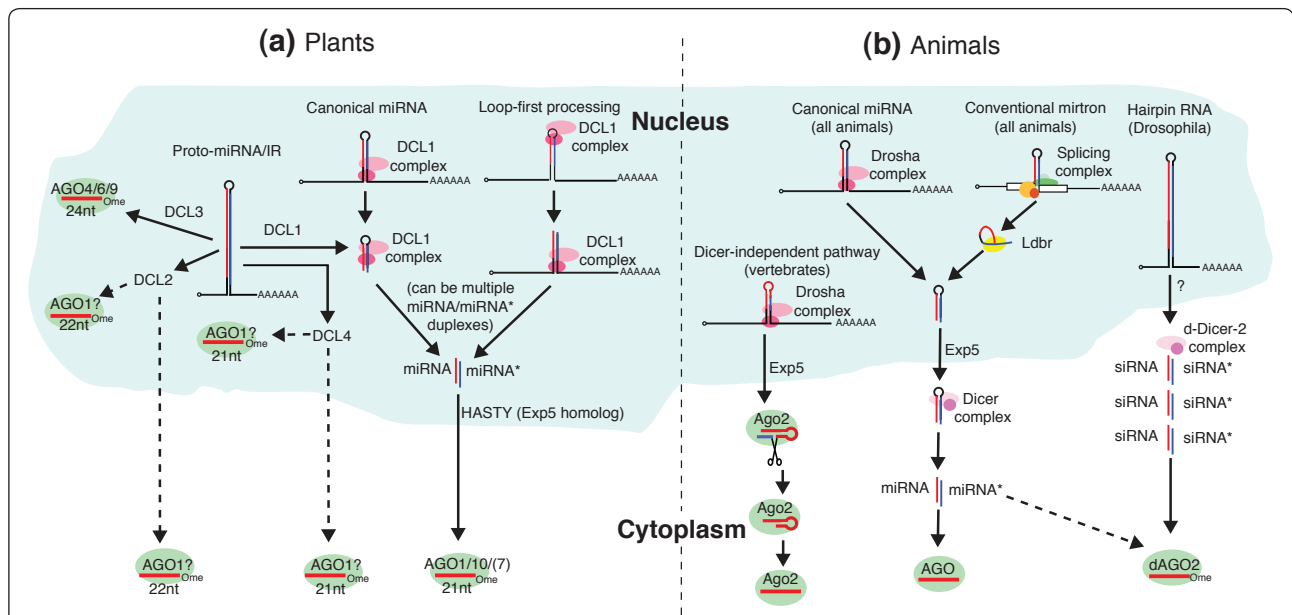

Figure 1. Major biogenesis pathways of small RNAs from inverted repeat transcripts in plants and animals. (a) In plants, canonical microRNAs (miRNAs) are produced by the nuclear RNase III Dicer-like1 (DCL1), which cuts from the base of the hairpin towards the loop; a subset of plant miRNAs are processed from the loop towards the hairpin. One miRNA/miRNA* duplex is shown, but there can be several such duplexes depending on the length of the stem. These are transported from the nucleus via HASTY, an Exportin5 (Exp5) homolog, for loading into an Argonaute (AGO) complex. The main miRNA effector in plants is AGO1, and to a lesser extent AGO10 and other AGOs; AGO7 carries the exceptional miRNA miR390. Long well-paired hairpins (proto-miR/inverted repeat (IR) transcripts) can be processed by a diversity of Dicers to generate either miRNAs or small interfering RNAs (siRNAs). The subcellular location for dicing by DCL2 and DCL4, and subsequent AGO loading of the resulting siRNAs, is not yet clear. nt, nucleotide. **(b)** In animals, canonical miRNAs are processed by the nuclear RNase III enzyme Drosha. The precursor miRNA (pre-miRNA) hairpin is exported to the cytoplasm by Exp5 to generate a single miRNA/miRNA* duplex, which is loaded into a miRNA class AGO protein (*Drosophila* dAGO1, *Caenorhabditis elegans* ALG1/2, or vertebrate Ago1 to Ago4). There are Drosha-independent non-canonical pathways, including the mirtron pathway where intron splicing and lariat debranching generate pre-miRNA hairpins. Also, vertebrate miR-451 is matured by a Dicer-independent route. Here, Drosha cleavage generates a short hairpin that is loaded into the 'Slicer' Ago2, which cleaves its 3′ arm; this is resected to yield the mature miRNA. Unlike other vertebrate miRNAs, miR-451 can only be matured in Ago2. Finally, in the *Drosophila* hairpin RNA pathway, long inverted repeats are processed by the endogenous siRNA pathway, being cleaved by d-Dicer2 to generate siRNAs that load dAGO2. Many *Drosophila* miRNA* species are also preferentially sorted into dAGO2 (dashed arrow).

proceed via a miRNA/miRNA* intermediate; typical annotation strategies require such paired duplexes for confident miRNA calls. AGO-mediated miRNA biogenesis from short hairpins has not been reported in plants to date.

In *Drosophila*, artificial long inverted repeat transcripts efficiently silence homologous transcripts, permitting transgenic RNAi [63]. These exogenous triggers prefaced the recognition of endo-siRNA substrates, including hairpin RNAs (hpRNAs). These long structured hairpins resemble long miRNA hairpins, in that they have internal mismatches and bulged positions; however, they are not processed by miRNA machinery. Instead, they traverse a siRNA pathway (Figure 1b), and are cleaved by Dicer-2 to generate duplexes that are preferentially loaded into AGO2 [64-66]. Endogenous processing of long hairpins in mammalian cells appears limited, because extensive dsRNAs trigger the antiviral interferon response. Some hpRNA-type loci are expressed and processed in embryonic stem cells and ovaries, indicating that this mode of biogenesis exists in vertebrates [67-69].

In contrast to vertebrates, plants have an extensive capacity to process long inverted repeat transcripts into small RNAs. There are currently no examples of well-conserved hairpin small RNA loci in plants, but they might conceivably play species-specific roles [70]. However, as in *Drosophila*, artificial hairpins are useful for reverse genetics in plants, indicating that perfect inverted repeat transcripts are readily accepted by small RNA biogenesis pathways. Interestingly, there exists a clear continuum of hairpin-derived small RNAs in plants, ranging from canonical miRNAs (defined by the precise production of a discrete miRNA/miRNA* duplex, or phased duplexes) to heterogeneously processed hairpins exhibiting enormous size variation [45,71]. The more imprecisely processed plant hairpins are generally processed by one or more of the Dicer-like enzymes DCL2, DCL3 or DCL4 (Figure 1a), which are also associated with production of dsRNA-derived siRNAs of various functions [45,70]. This continuum highlights the subjective nature of annotating miRNAs, the hallmark of which is the precision of their small RNA ends. It is a particular challenge to categorize ambiguous plant inverted repeats that generate an abundance of reads, of which only a subset conform to putative miRNA/miRNA* duplexes.

Target recognition by plant and animal miRNAs

The substantial differences between the biogenesis of animal and plant miRNAs are also reflected in the differences in their requirements for target recognition. It has long been known that plant miRNAs often have targets with perfect [72] or, more frequently, near-perfect [10] complementarity, facilitating relatively simple identification. Canonical plant miRNA target sites are found in 5′ UTRs, ORFs and 3′ UTRs, as well as within non-protein-coding transcripts, suggesting that all RNA contexts are equally amenable to miRNA-directed regulation in plants. Many of these plant miRNA targets succumb to AGO-catalyzed cleavage when they encounter a cognate miRNA; the characteristic remnants of these cleavage reactions enable molecular confirmation of plant miRNA target predictions *in vivo* [72-74]. However, not all plant miRNA-target interactions lead to AGO-catalyzed slicing. Some plant miRNA targets have conserved central mismatches embedded within perfectly base-paired regions at the 5′ and 3′ ends that allow AGO/miRNA binding but prevent slicing [75,76], whereas others seem to be refractory to target cleavage despite extensive complementarity [77]. In addition, even for plant miRNA targets that are sliced, slicing is often not the sole mechanism by which miRNAs repress target gene expression in plants: several experiments, involving multiple plant miRNA families, have demonstrated a pervasive contribution of translational repression to plant miRNA function [77-79].

Whether operating via slicing and/or translational repression, though, most evidence indicates that plant miRNAs require extensive pairing to their targets. By contrast, it is rare for animal miRNAs to identify targets with 'plant-like' complementarity. The initial target sites identified for the founding miRNA lin-4, within the 3′ UTR of *C. elegans lin-14*, exhibited only partial complementarity [1,2]. Subsequent studies in the *Drosophila* system elucidated arrays of approximately seven-nucleotide conserved 3′ UTR motifs termed Brd boxes, GY boxes and K boxes, which mediate critical post-transcriptional repression of Notch target genes during sensory bristle and eye development [80-82]. These motifs proved to represent binding sites for most of the initially described *Drosophila* miRNAs [5], and defined their capacity to identify targets via complementarity to their 5′ ends, preferentially at nucleotides 2 to 8 [9]. Extensive computational and experimental studies verified this as the major mode of miRNA target recognition in animals, with Watson-Crick pairing of positions 2 to 8 of the miRNA referred to as 'seed-pairing' [83,84]. Additional features, such as an adenosine following the seed match, the location within the 3′ UTR, proximity to other miRNA-binding sites, and the degree of local secondary structure, also influence target site activity [15].

Although there is also clear evidence for evolutionary selection of animal miRNA-binding sites in coding regions or even 5′ UTR sites [85,86], most of the well-studied target sites in this kingdom occur in 3′ UTRs. The efficacy of CDS or 5′ UTR sites appears to be hampered owing to competition with ribosomes [87].

Since artificial siRNAs that guide target slicing operate efficiently via coding sites in animal transcripts, similar to plant miRNAs, the paucity of CDS targeting is consistent with the view that relatively little miRNA target regulation in animals is mediated by AGO-catalyzed slicing; indeed, efforts to identify sliced remnants of AGO-catalyzed cleavage in mammalian samples yield very few targets [88,89]. The molecular mechanism(s) of miRNA targeting remain under investigation, and many have been proposed [90,91]. Although much attention has been focused on translational inhibition mechanisms, there is also evidence that bulk regulatory properties of animal miRNAs can be explained by mRNA degradation, possibly through induction of deadenylation [80,81,92,93].

In vertebrates, about 30% of transcripts contain probable miRNA-binding sites that have been conserved between mammals and chicken [94,95], and a comparable breadth of targeting has been detected in invertebrates [34,96,97]. Genome-wide transcriptome [92,98] and proteome [99,100] studies provide experimental support for the breadth of miRNA targeting in animal cells, and further indicate that many functional miRNA:target interactions are not well conserved. Moreover, at least some functional animal miRNA binding sites lack seed-pairing [2,3,101]. miRNA targeting in animals has broad potential to be combinatorial, since individual targets often bear conserved target sites for different miRNAs [94,95]. The scope of miRNA targeting appears to be drastically different in plants. Less than 1% of the transcripts in *Arabidopsis* are known or predicted miRNA targets [20,74] and there appears to be little, if any, combinatorial control. Almost all known plant miRNA targets have a single target site and are regulated by just one miRNA. These genome-wide principles may support the notion that animal miRNAs generally cast a wide net of mostly subtle regulatory effects across the transcriptome, while plant miRNAs have more focused and stronger regulatory effects on a relative handful of key targets. Of course, these generalizations should not be over-interpreted. Although quantitatively mild regulation can be of substantial importance to normal development or physiology, the loss of potent regulatory interactions is sometimes of surprisingly minimal phenotypic consequence [13].

Overall, the pairing requirements and target breadth of animal and plant miRNAs are clearly distinct, and this might be related to differences in their evolutionary emergence. Nevertheless, there is reason to believe that evolution has acted upon a backdrop of shared ancestral mechanisms. A recent study showed that *Drosophila* endo-siRNAs loaded in AGO2 require their 2′-*O*-methylation to remain stable in the presence of highly complementary targets [102]. Reciprocally, instability of

Drosophila miRNAs loaded in AGO1 was induced by providing them an artificial, perfectly complementary target, and similar findings applied to mammalian miRNAs. It was proposed that the general rarity of highly complementary targets for animal miRNAs has permitted them to dispense with a 3′ protection pathway [102]. By contrast, the fact that most plant miRNAs do have highly complementary targets may provide a pressure for obligate 3′ methylation of plant miRNAs. Moreover, the piwi-interacting RNA (piRNA) class of small RNAs carried by animal Piwi proteins are also methyated by Hen1 [42,103,104], and key functions of piRNAs include the recognition of perfectly matching transposon transcripts [105]. These data suggest that an evolutionarily ancient aspect of small regulatory RNA-mediated regulation is sensitive to the status of target pairing.

Contrasting modes of evolutionary emergence of plant and animal miRNAs

miRNA formed from intragenomic duplication

How are new miRNAs formed? An important clue came from the observation that some recently evolved miRNA hairpins in plants exhibit complementarity with their target mRNAs that extends beyond the region of the mature miRNA [106]. This observation suggested an evolutionary scenario where an inverted duplication of a gene gave rise to a 'proto-miRNA', which, when transcribed, would make a hairpin capable of producing small RNAs with perfect complementarity to the parental transcripts (Figure 2a). Over time, mutational drift obscures the extensive homology to the parental transcript and refines the precision of small RNA processing, leaving just a single region (the mature miRNA) that retains complementarity (Figure 2b). Consistent with this hypothesis, evidence for extended complementarity of plant miRNA hairpins to target mRNAs is restricted to less-conserved (and therefore presumably younger) loci [20,106,107].

'Proto-miRNA' loci in plants are likely to transit through a stage where small RNAs are imprecisely processed by one or more of the siRNA-generating DCL enzymes (Figure 2b). This hypothesis is supported by numerous examples of recently evolved plant *MIRNA* hairpins that are processed by DCL4, DCL3 or DCL2 instead of, or in addition to, the canonical miRNA Dicer DCL1 [45,108-110]. In addition to generating multiple sizes of small RNAs (for example, 21, 22 and 24 nucleotides) characteristic of different Dicers, the small RNAs of a given size may be only partially phased or unphased altogether. During this transitional period, many small RNAs from the same foldback would be complementary to the target(s), thus allowing beneficial regulatory relationships between the hairpin and the target to be selected for without an immediate

requirement for the precise processing that characterizes canonical miRNAs. This functional, yet transitional, state may be suited to plants (relative to animals) because of their requirement for a high degree of small RNA-target complementarity, which may consequently minimize off-target effects.

A strategy for miRNA genesis from pre-existing RNA structures also seems to occur with miniature inverted-repeat transposable elements (MITEs), whose eponymous inverted repeats, when transcribed, create

hairpin RNAs resembling proto-*MIRNA*s in plants [111]. A subset of animal miRNAs also derive from MITEs or other repetitive elements [112-114], and at least some of these may recognize mRNA targets bearing complementary repeat-related sequences. However, as animal miRNAs rarely exhibit 'plant-like' extensive complementarity to targets, the target duplication model does not seem to apply broadly in this kingdom. It is worth considering the vertebrate *mir-196* genes in this context. The three members of this family are located in
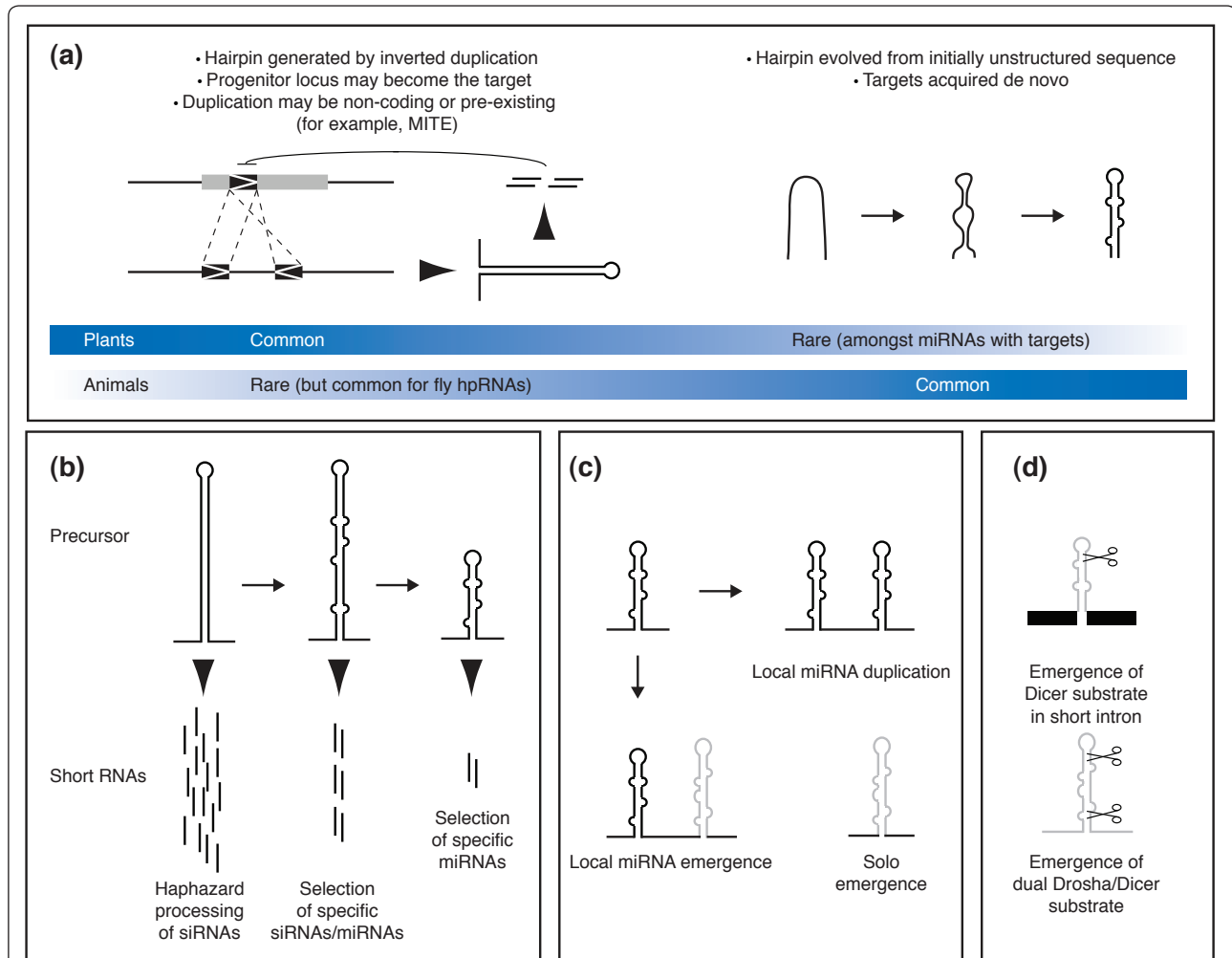

Figure 2. Modes of microRNA emergence in plants and animals. (a) Left: intragenomic duplications of protein-coding genes (or non-coding regions) can generate long foldbacks, which can be diced into small RNAs capable of targeting the progenitor transcript. This phenomenon seems common in plants, where extensive target complementarity is the rule, and ancestral relationships between microRNAs (miRNAs) and their targets can sometimes be detected; *Drosophila* hairpin RNA (hpRNA) may emerge similarly. MITE, miniature inverted-repeat transposable element. Right: inverted repeats might also emerge from initially unstructured sequences. This appears to be the dominant mode of miRNA emergence in animals. It also occurs in plants, but only rarely do such miRNAs appear to acquire functional targets. **(b)** Inferred model for plant miRNA emergence from long foldbacks; arrows indicate evolutionary relationships, arrowheads indicate small RNAs produced from a given hairpin. Long hairpins are processed haphazardly, often by different Dicers, to generate heterogeneous small interfering RNAs (siRNAs). As regulatory relationships are refined, the precision and phasing of hairpin processing may increase. Shortening of the hairpin to produce a single defined duplex may represent a mature state of plant miRNA evolution. **(c)** Expansion of miRNA clusters. In both plants and animals, local duplication may increase the dosage of a given miRNA. In animals, there may be an advantage for Drosha cleavage of hairpins emerging near extant miRNAs, leading to operons of unrelated miRNAs. **(d)** Different biogenesis mechanisms impose distinct demands on gene birth. Mirtrons need only evolve the capacity for one RNase III cleavage by Dicer, whereas canonical miRNAs need to gain the ability to be cleaved consecutively by Drosha and Dicer.

homologous positions within the four HOX genomic clusters, which encode conserved homeodomain proteins that govern anterior-posterior identities of body segments. Several Hox genes have miR-196 seed matches in their 3′ UTRs, but the *HOXB8* 3′ UTR has a highly conserved, fully complementary site to miR-196 [115,116]. The chromosomal proximity of *mir-196a-1* to *HOXB8* (separated only by *HOXB9*) is suggestively similar to the proximity of several young plant miRNA genes to their targets, and raises the possibility that *mir-196* genes evolved from a local duplication of HOX genes.

The *Drosophila* hpRNA pathway also offers an informative comparison. A number of endo-siRNAs generated by hpRNA loci have clear targets bearing nearly perfectly complementary sites that mediate their downregulation [64,65]. In the case of hp-CG18854 and its target *CG8289*, extensive target homology clearly indicates their ancestral relationship. This mimics an early state in target-derived plant miRNA emergence. In the case of hp-CG4068, the target *mus308* bears a perfectly paired antisense target site to its most abundant endo-siRNA, but lacks extended flanking homology. This may be analogous to a mature state in plant miRNA history. We emphasize that there does not appear to be an evolutionary relationship between hpRNAs and miRNAs in *Drosophila* (that is, there is no evidence that hpRNAs eventually become stabilized as miRNAs), and the hpRNAs as a class are rapidly evolving. Nevertheless, the similarities between *Drosophila* hpRNAs and plant miRNA genes are striking, apparently reflecting convergent evolutionary strategies.

Emergence of miRNAs from initially unstructured sequences

Because vanishingly few animal miRNAs seem to have derived from their target genes, it has long been assumed that a major route for miRNA birth in animals is through *de novo* emergence of RNA hairpins that gain competence for miRNA biogenesis (Figure 2a). Recall that plant miRNAs commonly have one-to-one or one-to-a-few target relations, but that animal miRNAs mediate broad regulatory networks owing to their minimal six- to seven-nucleotide target pairing requirements. If we assume that gene regulation in any extant individual is the product of substantial selective pressures for an optimal state, the introduction of a novel regulatory RNA is likely to be either neutral or detrimental, and only rarely beneficial [117,118]. In plants, the potentially detrimental influence of an emergent foldback on a miRNA sequence target might be mitigated by increasing the activity or expression of that target. However, in animals, one might imagine that emergent miRNA foldbacks might have the potential to misregulate a large cohort of target genes, from which a return to normalcy would not be easy.

Therefore, it has been posited that newborn miRNAs of animals are likely to 'creep' quietly into existence, beginning with low expression levels whose regulatory activities are tolerated by any targets encountered [117,118]. It also seems probable that *de novo* hairpins would not be fully endowed with characteristics permitting efficient miRNA biogenesis, and this would also limit their maturation. Recent annotation efforts in *Drosophila melanogaster* identified candidate hairpins that have evidence for miRNA biogenesis (that is, have many reads, have star species, and/or have reads in Argonaute immunoprecipitates), but do not show as clear evidence for precision of processing as do other more canonical miRNA loci [52]. These include loci that clearly exhibit patterns of random RNA breakdown layered on top of specific Drosha/Dicer-1/AGO1 biogenesis, suggesting that they are evolutionary intermediates that are only partially processed by the miRNA pathway (Figure 3).

Under the appropriate circumstances, then, the occasional beneficial regulation mediated by an emerging miRNA might be selected for. This would occur concomitantly with purging of target sites that mediate detrimental regulation, otherwise favoring loss of the emerging miRNA locus [117,118]. Such events might permit mutations within the hairpin that could improve its cleavage by Drosha and/or Dicer enzymes, as well as improve transcriptional capacity. These qualities are indeed mirrored in the general expression patterns of animal miRNAs. Among highly expressed miRNAs, basically all are deeply conserved within a given clade (for example, Drosophilids or vertebrates). The lowly expressed miRNAs include some conserved loci, which might be due to their tissue-restricted expression, but essentially all of the evolutionarily newborn miRNAs fall into the low-expression group [34,52,119].

Although the birth of many plant *MIRNA* loci can be explained by the target duplication model, comparisons between the closely related species *Arabidopsis thaliana* and *Arabidopsis lyrata* have strongly implied that many *MIRNA* loci arose recently from inverted repeats formed from random intergenic sequences [107,120,121]. Such *MIRNA* loci are likely to be rapidly lost due to mutational drift, as they are born without pre-existing target homology; indeed, targets are not readily detectable for many of these evolutionarily young miRNAs, and the patterns of nucleotide substitutions in many of them suggest neutral drift rather than the constrained patterns of substitutions found for older, clearly functional miRNAs [107,121]. Nevertheless, it has been inferred that targets can occasionally be captured for young plant miRNAs. This is perhaps most clear in a few cases, where plant *MIRNA* hairpins born of genic duplications have acquired targets distinct from their originating loci; for

instance, *Arabidopsis* miR447 and miR856 have validated targets that are distinct from their loci of origin [20,107].

Distinguishing *bona fide* miRNAs from RNA degradation products

The genomes of most higher eukaryotes are predicted to encode at least 10^5 to 10^6 putative hairpins with substantial similarity to validated pri-miRNA hairpins [122-124]. In theory, this constitutes an enormous reservoir of putative miRNA substrates, whose trace regulatory activities might be subject to selection and evolutionary stabilization. But how many predicted hairpins in a given genome are actually competent to be specifically processed by the miRNA biogenesis machinery? Conserved miRNA genes are amenable to computational discovery by signatures of hairpins that exhibit evolutionarily stable arms and diverging terminal loops [122,123]. However, it is not currently possible to prospectively annotate the miRNAs encoded by a genome, in the absence of comparative genomics, with any reasonable degree of specificity or sensitivity. Therefore, the mere existence of large numbers of predicted miRNA-like hairpins does not imply the existence of similar numbers of evolutionarily emergent miRNA genes.

In light of broad transcription across euchromatin of plants and animals [125,126], including substantial amounts of transcribed sequence that are removed by quality control mechanisms, it seems inevitable that short pieces of RNA will eventually be associated with most of the genome, including most predicted genomic hairpins. For example, many hundreds of candidate miRNA hairpins in various *Drosophila* species were initially annotated on the basis of singleton reads [119]. However, these were either not recovered in substantially larger datasets from the same tissue sources, or generated heterogeneously sized reads mapping through the predicted hairpins [127], indicating that few of these were genuine miRNAs produced by RNase III processing.

More recently, short RNAs were recovered from >100,000 hairpins in the *D. melanogaster* genome (from a starting set of nearly 1 billion short RNA reads from almost 200 libraries) [52]. However, confident canonical miRNA production could only be assigned to approximately 200 loci. These included many miRNA hairpins that are recently or newly evolved in *D. melanogaster*, but they did not include hundreds of other miRNA candidates predicted from comparative analysis of the 12 sequenced Drosophilids [34,128]. Even though most of these predictions were associated with at least some small RNA reads, none exhibited read patterns and/or sizes that were consistent with Drosha/Dicer-mediated processing (Figure 3). Although caution should be exercised in interpreting negative evidence, such data
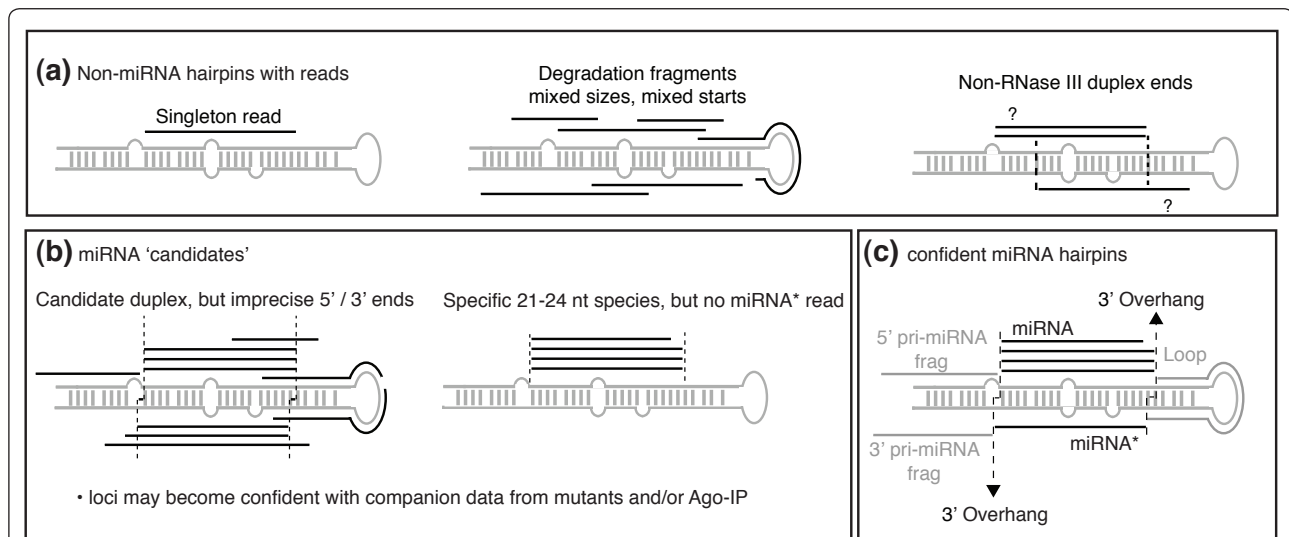

Figure 3. The complexity of annotating microRNAs from reads mapped to predicted hairpins. (a) Examples of loci that should not be annotated as microRNAs (miRNAs): hairpins with single reads, heterogeneously sized reads, and/or putatively duplexed reads lacking 3′ overhangs. **(b)** Examples of loci with some, but insufficient, evidence for miRNA biogenesis; such loci are worth segregating as candidates pending further study. For instance, depletion of reads from miRNA biogenesis mutants or enrichment in Ago complexes could elevate their status from 'candidate' to 'confident'. It is also worth considering that candidate miRNA hairpins with relatively imprecise processing patterns may represent transitional intermediates in miRNA birth. nt, nucleotide. **(c)** Confident miRNA hairpins generate relatively precise miRNA/miRNA* duplexes with 3′ overhangs. As datasets grow, it is often possible to observe cloned terminal loops or 5′/3′ fragments (frag) of the primary miRNA (pri-miRNA) base, whose phasing with miRNA/miRNA* termini provide stringent evidence for *in vivo* cleavage reactions. Note that the vertebrate locus *mir-451* matures by direct hairpin cleavage by Ago2, and not via a miRNA/miRNA* intermediate; thus, the criteria outlined in this figure are not applicable to *mir-451*-class substrates.

support the notion that a fairly limited number of genomic regions are competent for recognition by the miRNA pathway when transcribed.

These findings are echoed by studies of other animals and plants. For example, while novel mammalian miRNAs continued to be reported, a recent systematic analysis actually revised the estimates of mammalian miRNAs downward, and called into question a substantial subset of extant annotations [129]. Similarly in *Arabidopsis*, many hundreds of miRNA genes were predicted by comparing hairpins with inferred target sites [130]. However, very few of these have since been validated by deep sequencing [20,45]. Altogether, these observations suggest that relatively few of the many predicted genomic hairpins are substrates of miRNA biogenesis pathways, and hint that substantial bioinformatic progress remains to bring miRNA prediction up to a par with protein-coding gene prediction.

It is clear that more data on the efficiency and specificity of miRNA processing are desirable, especially as all entries in the miRBase registry are currently treated as equivalent [11]. Although some loci are now deemed suspect, in principle, the majority of entries are confident miRNA genes that exhibit reasonably precise processing. The reality is that with ever-increasing depths of small RNA sequencing, all miRNA loci exhibit some level of terminal heterogeneity in their cloned products, and this tends to blur the division between loci that can be confidently inferred to have transited an RNase III pathway acting on a precursor hairpin, as opposed to fortuitous hairpins that generate degradation products (Figure 3a). At present, a cautious approach to miRNA annotation seems warranted, in which miRNA 'candidates' are segregated from more confident miRNA loci [52]. These include loci that exhibit plausible miRNA/miRNA* duplexes but have substantial reads not conforming to RNase III products, as well as hairpins lacking a star species (Figure 3b). In particular, the lack of cloned miRNA* species was not previously seen as an impediment to miRNA annotation; however, the depth of next-generation sequencing now makes it reasonable to demand miRNA* species for confident annotation. (Exceptions may be made with appropriate data; that is, if there is strong evidence from loss of reads in miRNA biogenesis mutants, or enrichment of reads in Ago-IP samples.)

Perhaps more importantly, there are few data on how 'well' a miRNA hairpin is processed. If only 1% of a miRNA hairpin transits the processing machinery, and these are processed precisely, then it is equally eligible for annotation as loci for which 100% of the hairpin is converted into miRNAs. The efficiency of processing cannot be inferred from counts of small RNAs alone, since some endogenous miRNA loci might generate rare reads owing to cell-specific expression, or perhaps post-transcriptional inhibition of some aspect of their biogenesis [43]. It is only recently that efforts have been made to generate systematic data on the effectiveness of miRNA processing, by cloning small RNA libraries from pools of cells transfected with different miRNA expression constructs and quantifying the output reads [129]. Larger scale data on forced expression of predicted hairpins, especially ones that lack endogenous reads or exhibit heterogenous reads, should provide valuable insights into the potentially partial capacity of some substrates to enter miRNA biogenesis pathways.

Distinct evolutionary flux of different miRNA subclasses

In general, small RNAs that populate a miRNA-effector AGO have the functional attributes of a miRNA, regardless of whether it was produced by a canonical or alternative pathway. However, this does not mean that all miRNA substrates evolve similarly. For example, the evolution of plant miRNAs from target duplications can be compared with those from incidental hairpins. Because the plant miRNA system appears to require extensive target complementarity, it is presumed that only rarely will fortuitous small RNA-generating hairpins emerge and then subsequently acquire beneficial targets. Therefore, these classes of plant miRNAs are expected to emerge and disappear with distinct dynamics [106,107,120,121].

Does the genomic location of miRNA hairpins influence their emergence? Canonical miRNAs in animals, unlike in plants, are commonly located in introns. Because animal miRNAs mostly emerge from incidental foldbacks, the introns of *bona fide* transcription units might be a privileged location for miRNA birth already endowed with directed primary transcription. We have also mentioned that animal miRNAs, unlike their plant brethren, are frequently arranged in operons composed of dissimilar species. This suggests that the emergence of animal miRNAs is privileged by location near an extant miRNA (Figure 2c). Access to the Drosha/DGCR8 processing complex is the gatekeeper for entry into the canonical miRNA pathway, and its activity has been reported to be co-transcriptional [131-133]. Consequently, physical proximity of an emergent hairpin to an established miRNA hairpin could enhance its nuclear cleavage relative to 'solo' emergent hairpins; this might be particularly important for the processing of suboptimal hairpins. Newly evolved miRNAs have indeed been detected in proximity to much more deeply conserved animal miRNAs [52]. The lack of Drosha homologs in plants might render this characteristic less relevant in the plant kingdom.

What is the influence of the biogenesis mechanism on miRNA evolution? If we compare the canonical miRNAs with mirtrons (Figure 2d), a subset of mirtrons are well conserved in a given animal clade (that is, *Drosophila* or vertebrates), and these diverge more quickly in their terminal loop than their hairpin arms [55,134], as seen with canonical miRNAs. However, relatively few mirtrons are well conserved and no mirtrons are common between invertebrates and vertebrates, whereas many canonical miRNAs are identical over this distance. This implied that mirtrons, as a class, emerge and disappear more quickly than canonical miRNAs. This was confirmed with deep sequencing of small RNAs from three Drosophilid species; this showed that mirtrons comprise a steadily increasing fraction of confidently annotated miRNA loci as the evolutionary branch length under consideration decreases [127]. Recent analysis of mirtrons from ultradeep sequencing has extended the species-specific catalog of fly and nematode mirtrons further still [52,135].

The distinct evolution of mirtrons and canonical miRNAs might relate to the structural hurdles needed to become a substrate of the respective biogenesis machineries (Figure 2d). In the case of canonical miRNAs, a substrate must simultaneously adopt conformations that permit its cleavage by both Drosha and Dicer. In the case of mirtrons, a substrate must be spliced and be a target of Dicer. We do not know how many endogenous substrates of Drosha there are, but there are currently no more than some hundreds known in any given animal species. By contrast, the many short constitutively spliced introns (for example, there are 27,000 to 30,000 introns <120 nucleotides in flies and nematodes) comprise a large pool of loci that have passed one processing hurdle. Structure-function studies have shown that mirtrons are flexible with regard to primary sequence, provided that they retain splicing functionality and adopt substantial hairpin structure with 3′ overhangs [135]. In fact, the nature of splice sites (GU...YAG) should position many introns to pair the 5′ G with the 3′ Y, leaving the AG as a two-nucleotide 3′ overhang.

These factors might conspire to aid the evolutionary emergence of animal mirtrons, relative to canonical miRNAs. Conversely, the apparent near absence of mirtrons in plants could well be a consequence of a distinct biogenesis mechanism that does not rely upon processing by two separate RNase III enzymes. Because plant miRNA maturation relies upon a single Dicer-like protein to completely liberate a miRNA/miRNA* duplex from a pri-miRNA transcript, the pre-miRNA intermediates produced by mirtrons may be either unnecessary or perhaps even unrecognized by the plant DCL1 complex.

Concluding remarks

The major plant and animal miRNA pathways differ with respect to their biochemical mechanisms, the extent of their preferred target pairing, and numbers of functional targets. These differences have resulted in distinct characteristics of the evolution of plant and animal miRNAs. In particular, the co-evolution of target:miRNA pairs is common in plants, whereas it seems much more common for animal miRNAs to emerge and then acquire target genes. One interpretation is that this reflects independent emergence of miRNA pathways in plants and animals, on the backbone of an ancestral RNAi pathway that metabolized dsRNA into short RNAs that populate Argonaute proteins. This system might have emerged to defend against invasive nucleic acids such as viruses and transposons, and subsequently been adapted to generate miRNAs from endogenous inverted repeat transcripts. However, there are also analogies between plant and animal miRNA pathways. For example, certain vertebrate miRNA targets, as well as *Drosophila* hpRNA targets, exhibit 'plant-like' extensive complementarity. There is reciprocally a growing appreciation that plant miRNAs have emerged from incidentally emerged hairpins, akin to the presumed dominant mode for animal miRNA birth. Therefore, an alternative interpretation is that a miRNA pathway was extant in the last common ancestor of plants and animals, but became differentially deployed in these kingdoms.

In either case, it is clear that a limited set of core proteins, namely RNase III enzymes and Argonaute proteins, have been joined in remarkably diverse ways to control gene expression via small RNAs. Recent studies of fungal small RNA pathways provide additional evidence for innovation of RNase III-independent mechanisms for siRNA and miRNA production [136,137], for which we can only guess at the underlying reasons that permitted the loss of canonical pathways and invention of new pathways. Altogether, it is evident that miRNAs are not a unitary entity, but instead encompass a variety of conceptually related phenomena, whose evolutionary pressures differ according to mechanism of biogenesis and even genomic location. Understanding the principles that govern the evolutionary flux of these myriad small RNA pathways will provide a fundamental complement to understanding the flux of protein-coding genes [138].

Abbreviations

bp, base pair; CDS, coding sequence; endo-siRNa, endogenous siRNA; hpRNA, hairpin RNA; miRNA, microRNA; MITE, miniature inverted-repeat transposable element; ORF, open reading frame; piRNA, piwi-interacting RNA; pre-miRNA, precursor miRNA; pri-miRNA, primary miRNA; siRNA, small interfering RNA; UTR, untranslated region.

References

1. 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**:843-854.
2. Wightman B, Ha I, Ruvkun G: Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* 1993, **75**:855-862.
3. Reinhart BJ, Slack F, Basson M, Pasquinelli A, Bettinger J, Rougvie A, Horvitz HR, Ruvkun G: The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* 2000, **403**:901-906.
4. Lee RC, Ambros V: An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* 2001, **294**:862-864.
5. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T: Identification of novel genes coding for small expressed RNAs. *Science* 2001, **294**:853-858.
6. Lau N, Lim L, Weinstein E, Bartel DP: An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 2001, **294**:858-862.
7. Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP: MicroRNAs in plants. *Genes Dev* 2002, **16**:1616-1626.
8. Park W, Li J, Song R, Messing J, Chen X: CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Curr Biol* 2002, **12**:1484-1495.
9. Lai EC: microRNAs are complementary to 3′ UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat Genet* 2002, **30**:363-364.
10. Rhoades MW, Reinhart BJ, Lim LP, Burge CB, Bartel B, Bartel DP: Prediction of plant microRNA targets. *Cell* 2002, **110**:513-520.
11. Kozomara A, Griffiths-Jones S: miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* 2011, **39**:D152-D157.
12. Kim VN, Han J, Siomi MC: Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Biol* 2009, **10**:126-139.
13. Flynt AS, Lai EC: Biological principles of microRNA-mediated regulation: shared themes amid diversity. *Nat Rev Genet* 2008, **9**:831-842.
14. Voinnet O: Origin, biogenesis, and activity of plant microRNAs. *Cell* 2009, **136**:669-687.
15. Bartel DP: MicroRNAs: target recognition and regulatory functions. *Cell* 2009, **136**:215-233.
16. Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, Dreyfuss G, Eddy SR, Griffiths-Jones S, Marshall M, Matzke M, Ruvkun G, Tuschl T: A uniform system for microRNA annotation. *RNA* 2003, **9**:277-279.
17. Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, Bowman JL, Cao X, Carrington JC, Chen X, Green PJ, Griffiths-Jones S, Jacobsen SE, Mallory AC, Martienssen RA, Poethig RS, Qi Y, Vaucheret H, Voinnet O, Watanabe Y, Weigel D, Zhu JK: Criteria for annotation of plant MicroRNAs. *Plant Cell* 2008, **20**:3186-3190.
18. miRBase [http://www.mirbase.org/]
19. Xie Z, Allen E, Fahlgren N, Calamar A, Givan SA, Carrington JC: Expression of *Arabidopsis MIRNA* genes. *Plant Physiol* 2005, **138**:2145-2154.
20. Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Law TF, Grant SR, Dangl JL, Carrington JC: High-throughput sequencing of *Arabidopsis* microRNAs: evidence for frequent birth and death of MIRNA genes. *PLoS ONE* 2007, **2**:e219.
21. Kurihara Y, Watanabe Y: Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc Natl Acad Sci USA* 2004, **101**:12753-12758.
22. Bologna NG, Mateos JL, Bresso EG, Palatnik JF: A loop-to-base processing mechanism underlies the biogenesis of plant microRNAs miR319 and miR159. *EMBO J* 2009, **28**:3646-3656.
23. Addo-Quaye C, Snyder JA, Park YB, Li YF, Sunkar R, Axtell MJ: Sliced microRNA targets and precise loop-first processing of MIR319 hairpins revealed by analysis of the *Physcomitrella patens* degradome. *RNA* 2009, **15**:2112-2121.
24. Zhang W, Gao S, Zhou X, Xia J, Chellappan P, Zhang X, Jin H: Multiple distinct

25. small RNAs originate from the same microRNA precursors. *Genome Biol* 2010, **11**:R81.
25. Park MY, Wu G, Gonzalez-Sulser A, Vaucheret H, Poethig RS: Nuclear processing and export of microRNAs in *Arabidopsis*. *Proc Natl Acad Sci USA* 2005, **102**:3691-3696.
26. Fang Y, Spector DL: Identification of nuclear dicing bodies containing proteins for microRNA biogenesis in living Arabidopsis plants. *Curr Biol* 2007, **17**:818-823.
27. Song L, Han MH, Lesicka J, Fedoroff N: Arabidopsis primary microRNA processing proteins HYL1 and DCL1 define a nuclear body distinct from the Cajal body. *Proc Natl Acad Sci USA* 2007, **104**:5437-5442.
28. Yu B, Yang Z, Li J, Minakhina S, Yang M, Padgett RW, Steward R, Chen X: Methylation as a crucial step in plant microRNA biogenesis. *Science* 2005, **307**:932-935.
29. Li J, Yang Z, Yu B, Liu J, Chen X: Methylation protects miRNAs and siRNAs from a 3′-end uridylation activity in *Arabidopsis*. *Curr Biol* 2005, **15**:1501-1507.
30. Baumberger N, Baulcombe DC: *Arabidopsis* ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proc Natl Acad Sci USA* 2005, **102**:11928-11933.
31. Qi Y, Denli AM, Hannon GJ: Biochemical specialization within *Arabidopsis* RNA silencing pathways. *Mol Cell* 2005, **19**:421-428.
32. Vaucheret H: Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes Dev* 2006, **20**:759-771.
33. Borchert GM, Lanier W, Davidson BL: RNA polymerase III transcribes human microRNAs. *Nat Struct Mol Biol* 2006, **13**:1097-1101.
34. Ruby JG, Stark A, Johnston WK, Kellis M, Bartel DP, Lai EC: Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. *Genome Res* 2007, **17**:1850-1864.
35. Liu J, Carmell MA, Rivas FV, Marsden CG, Thomson JM, Song JJ, Hammond SM, Joshua-Tor L, Hannon GJ: Argonaute2 is the catalytic engine of mammalian RNAi. *Science* 2004, **305**:1437-1441.
36. Meister G, Landthaler M, Patkaniowska A, Dorsett Y, Teng G, Tuschl T: Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell* 2004, **15**:185-197.
37. Okamura K, Ishizuka A, Siomi H, Siomi MC: Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev* 2004, **18**:1655-1666.
38. Forstemann K, Horwich MD, Wee L, Tomari Y, Zamore PD: *Drosophila* microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. *Cell* 2007, **130**:287-297.
39. Czech B, Zhou R, Erlich Y, Brennecke J, Binari R, Villalta C, Gordon A, Perrimon N, Hannon GJ: Hierarchical rules for Argonaute loading in *Drosophila*. *Mol Cell* 2009, **36**:445-456.
40. Okamura K, Liu N, Lai EC: Distinct mechanisms for microRNA strand selection by *Drosophila* Argonautes. *Mol Cell* 2009, **36**:431-444.
41. Ghildiyal M, Xu J, Seitz H, Weng Z, Zamore PD: Sorting of *Drosophila* small silencing RNAs partitions microRNA* strands into the RNA interference pathway. *RNA* 2010, **16**:43-56.
42. Horwich MD, Li C, Matranga C, Vagin V, Farley G, Wang P, Zamore PD: The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr Biol* 2007, **17**:1265-1272.
43. Kim YK, Heo I, Kim VN: Modifications of small RNAs and their associated proteins. *Cell* 2010, **143**:703-709.
44. Axtell MJ, Snyder JA, Bartel DP: Common functions for diverse small RNAs of land plants. *Plant Cell* 2007, **19**:1750-1769.
45. Rajagopalan R, Vaucheret H, Trejo J, Bartel DP: A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes Dev* 2006, **20**:3407-3425.
46. Merchan F, Boualem A, Crespi M, Frugier F: Plant polycistronic precursors containing non-homologous microRNAs target transcripts encoding functionally related proteins. *Genome Biol* 2009, **10**:R136.
47. Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A: Identification of mammalian microRNA host genes and transcription units. *Genome Res* 2004, **14**:1902-1910.
48. Aboobaker AA, Tomancak P, Patel N, Rubin GM, Lai EC: *Drosophila* microRNAs exhibit diverse spatial expression patterns during embryonic development. *Proc Natl Acad Sci USA* 2005, **102**:18017-18022.
49. Isik M, Korswagen HC, Berezikov E: Expression patterns of intronic microRNAs in *Caenorhabditis elegans*. *Silence* 2010, **1**:5.
50. Martinez NJ, Ow MC, Reece-Hoyes JS, Barrasa MI, Ambros VR, Walhout AJ:

Genome-scale spatiotemporal analysis of *Caenorhabditis elegans* microRNA promoter activity. *Genome Res* 2008, **18**:2005-2015.

51. Han J, Pedersen JS, Kwon SC, Belair CD, Kim YK, Yeom KH, Yang WY, Haussler D, Blelloch R, Kim VN: **Posttranscriptional crossregulation between Drosha and DGCR8.** *Cell* 2009, **136**:75-84.

52. 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**:203-215.

53. Miyoshi K, Miyoshi T, Siomi H: **Many ways to generate microRNA-like small RNAs: non-canonical pathways for microRNA production.** *Mol Genet Genomics* 2010, **284**:95-103.

54. Okamura K, Chung WJ, Lai EC: **The long and short of inverted repeat genes in animals: microRNAs, mirtrons and hairpin RNAs.** *Cell Cycle* 2008, **7**:2840-2845.

55. Okamura K, Hagen JW, Duan H, Tyler DM, Lai EC: **The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*.** *Cell* 2007, **130**:89-100.

56. Ruby JG, Jan CH, Bartel DP: **Intronic microRNA precursors that bypass Drosha processing.** *Nature* 2007, **448**:83-86.

57. Flynt AS, Chung WJ, Greimann JC, Lima CD, Lai EC: **microRNA biogenesis via splicing and exosome-mediated trimming in Drosophila.** *Mol Cell* 2010, **38**:900-907.

58. Zhu QH, Spriggs A, Matthew L, Fan L, Kennedy G, Gubler F, Helliwell C: **A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains.** *Genome Res* 2008, **18**:1456-1465.

59. Czech B, Hannon GJ: **Small RNA sorting: matchmaking for Argonautes.** *Nat Rev Genet* 2010, **12**:19-31.

60. Yang JS, Maurin T, Robine N, Rasmussen KD, Jeffrey KL, Chandwani R, Papapetrou EP, Sadelain M, O'Carroll D, Lai EC: **Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis.** *Proc Natl Acad Sci USA* 2010, **107**:15163-15168.

61. Cifuentes D, Xue H, Taylor DW, Patnode H, Mishima Y, Cheloufi S, Ma E, Mane S, Hannon GJ, Lawson ND, Wolfe SA, Giraldez AJ: **A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity.** *Science* 2010, **328**:1694-1698.

62. Cheloufi S, Dos Santos CO, Chong MM, Hannon GJ: **A dicer-independent miRNA biogenesis pathway that requires Ago catalysis.** *Nature* 2010, **465**:584-589.

63. Kennerdell JR, Carthew RW: **Heritable gene silencing in Drosophila using double-stranded RNA.** *Nat Biotechnol* 2000, **18**:896-898.

64. Okamura K, Chung W-J, Ruby JG, Guo H, Bartel DP, Lai EC: **The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs.** *Nature* 2008, **453**:803-806.

65. Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, Perrimon N, Kellis M, Wohlschlegel J, Sachidanandam R, Hannon GJ, Brennecke J: **An endogenous siRNA pathway in *Drosophila*.** *Nature* 2008, **453**:798-802.

66. Kawamura Y, Saito K, Kin T, Ono Y, Asai K, Sunohara T, Okada T, Siomi MC, Siomi H: ***Drosophila* endogenous small RNAs bind to Argonaute2 in somatic cells.** *Nature* 2008, **453**:793-797.

67. Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, Cheloufi S, Hodges E, Anger M, Sachidanandam R, Schultz RM, Hannon GJ: **Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes.** *Nature* 2008, **453**:534-538.

68. Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T, Surani MA, Sakaki Y, Sasaki H: **Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes.** *Nature* 2008, **453**:539-543.

69. Babiarz JE, Ruby JG, Wang Y, Bartel DP, Blelloch R: **Mouse ES cells express endogenous shRNAs, siRNAs, and other microprocessor-independent, Dicer-dependent small RNAs.** *Genes Dev* 2008, **22**:2773-2785.

70. Dunoyer P, Brosnan CA, Schott G, Wang Y, Jay F, Alioua A, Himber C, Voinnet O: **An endogenous, systemic RNAi pathway in plants.** *EMBO J* 2010, **29**:1699-1712.

71. Zhang X, Henderson IR, Lu C, Green PJ, Jacobsen SE: **Role of RNA polymerase IV in plant small RNA metabolism.** *Proc Natl Acad Sci USA* 2007, **104**:4536-4541.

72. Llave C, Xie Z, Kasschau KD, Carrington JC: **Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA.** *Science* 2002, **297**:2053-2056.

73. German MA, Pillay M, Jeong DH, Hetawal A, Luo S, Janardhanan P, Kannan V, Rymarquis LA, Nobuta K, German R, De Paoli E, Lu C, Schroth G, Meyers BC, Green PJ: **Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends.** *Nat Biotechnol* 2008, **26**:941-946.

74. Addo-Quaye C, Eshoo TW, Bartel DP, Axtell MJ: **Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome.** *Curr Biol* 2008, **18**:758-762.

75. Axtell MJ, Jan C, Rajagopalan R, Bartel DP: **A two-hit trigger for siRNA biogenesis in plants.** *Cell* 2006, **127**:565-577.

76. Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, Rubio-Somoza I, Leyva A, Weigel D, Garcia JA, Paz-Ares J: **Target mimicry provides a new mechanism for regulation of microRNA activity.** *Nat Genet* 2007, **39**:1033-1037.

77. Brodersen P, Sakvarelidze-Achard L, Bruun-Rasmussen M, Dunoyer P, Yamamoto YY, Sieburth L, Voinnet O: **Widespread translational inhibition by plant miRNAs and siRNAs.** *Science* 2008, **320**:1185-1190.

78. Chen X: **A microRNA as a translational repressor of APETALA2 in *Arabidopsis* flower development.** *Science* 2004, **303**:2022-2025.

79. Gandikota M, Birkenbihl RP, Hohmann S, Cardon GH, Saedler H, Huijser P: **The miRNA156/157 recognition element in the 3' UTR of the *Arabidopsis* SBP box gene SPL3 prevents early flowering by translational inhibition in seedlings.** *Plant J* 2007, **49**:683-693.

80. Lai EC, Burks C, Posakony JW: **The K box, a conserved 3' UTR sequence motif, negatively regulates accumulation of *Enhancer of split* Complex transcripts.** *Development* 1998, **125**:4077-4088.

81. Lai EC, Posakony JW: **The Bearded box, a novel 3' UTR sequence motif, mediates negative post-transcriptional regulation of *Bearded* and *Enhancer of split* Complex gene expression.** *Development* 1997, **124**:4847-4856.

82. Leviten MW, Lai EC, Posakony JW: **The *Drosophila* gene *Bearded* encodes a novel small protein and shares 3' UTR sequence motifs with multiple *Enhancer of split* complex genes.** *Development* 1997, **124**:4039-4051.

83. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB: **Prediction of mammalian microRNA targets.** *Cell* 2003, **115**:787-798.

84. Brennecke J, Stark A, Russell RB, Cohen SM: **Principles of microRNA-target recognition.** *PLoS Biol* 2005, **3**:e85.

85. Lytle JR, Yario TA, Steitz JA: **Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR.** *Proc Natl Acad Sci U S A* 2007, **104**:9667-9672.

86. Schnall-Levin M, Zhao Y, Perrimon N, Berger B: **Conserved microRNA targeting in *Drosophila* is as widespread in coding regions as in 3' UTRs.** *Proc Natl Acad Sci USA* 2010, **107**:15751-15756.

87. Gu S, Jin L, Zhang F, Sarnow P, Kay MA: **Biological basis for restriction of microRNA targets to the 3' untranslated region in mammalian mRNAs.** *Nat Struct Mol Biol* 2009, **16**:144-150.

88. 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.

89. Karginov FV, Cheloufi S, Chong MM, Stark A, Smith AD, Hannon GJ: **Diverse endonucleolytic cleavage sites in the mammalian transcriptome depend upon microRNAs, Drosha, and additional nucleases.** *Mol Cell* 2010, **38**:781-788.

90. Brodersen P, Voinnet O: **Revisiting the principles of microRNA target recognition and mode of action.** *Nat Rev Mol Cell Biol* 2009, **10**:141-148.

91. Fabian MR, Sonenberg N, Filipowicz W: **Regulation of mRNA translation and stability by microRNAs.** *Annu Rev Biochem* 2010, **79**:351-379.

92. Guo H, Ingolia NT, Weissman JS, Bartel DP: **Mammalian microRNAs predominantly act to decrease target mRNA levels.** *Nature* 2010, **466**:835-840.

93. Giraldez AJ, Mishima Y, Rihel J, Grocock RJ, Van Dongen S, Inoue K, Enright AJ, Schier AF: **Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs.** *Science* 2006, **312**:75-79.

94. 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.

95. Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, Macmenamin P, da Piedade I, Gunsalus KC, Stoffel M, Rajewsky N: **Combinatorial microRNA target predictions.** *Nat Genet* 2005, **37**:495-500.

96. Grun D, Wang Y-L, Langenberger D, Gunsalus KC, Rajewsky N: **microRNA target predictions across seven *Drosophila* species and comparison to mammalian targets.** *PLoS Comput Biol* 2005, **1**:e13.

97. Jan CH, Friedman RC, Ruby JG, Bartel DP: **Formation, regulation and**

evolution of Caenorhabditis elegans 3' UTRs. *Nature* 2011, **469**:97-101.
98. Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM: Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 2005, **433**:769-773.
99. Selbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N: Widespread changes in protein synthesis induced by microRNAs. *Nature* 2008, **455**:58-63.
100. Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP: The impact of microRNAs on protein output. *Nature* 2008, **455**:64-71.
101. Ha I, Wightman B, Ruvkun G: A bulged *lin-4/lin-14* RNA duplex is sufficient for *Caenorhabditis elegans lin-14* temporal gradient formation. *Genes Dev* 1996, **10**:3041-3050.
102. Ameres SL, Horwich MD, Hung JH, Xu J, Ghildiyal M, Weng Z, Zamore PD: Target RNA-directed trimming and tailing of small silencing RNAs. *Science* 2010, **328**:1534-1539.
103. Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD: A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* 2006, **313**:320-324.
104. Saito K, Sakaguchi Y, Suzuki T, Suzuki T, Siomi H, Siomi MC: Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi-interacting RNAs at their 3' ends. *Genes Dev* 2007, **21**:1603-1608.
105. Senti KA, Brennecke J: The piRNA pathway: a fly's perspective on the guardian of the genome. *Trends Genet* 2010, **26**:499-509.
106. Allen E, Xie Z, Gustafson AM, Sung GH, Spatafora JW, Carrington JC: Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. *Nat Genet* 2004, **36**:1282-1290.
107. 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**:1074-1089.
108. Vazquez F, Blevins T, Ailhas J, Boller T, Meins F Jr: Evolution of *Arabidopsis* MIR genes generates novel microRNA classes. *Nucleic Acids Res* 2008, **36**:6429-6438.
109. Chellappan P, Xia J, Zhou X, Gao S, Zhang X, Coutino G, Vazquez F, Zhang W, Jin H: siRNAs from miRNA sites mediate DNA methylation of target genes. *Nucleic Acids Res* 2010, **38**:6883-6894.
110. Wu L, Zhou H, Zhang Q, Zhang J, Ni F, Liu C, Qi Y: DNA methylation mediated by a microRNA pathway. *Mol Cell* 2010, **38**:465-475.
111. Piriyapongsa J, Jordan IK: Dual coding of siRNAs and miRNAs by plant transposable elements. *RNA* 2008, **14**:814-821.
112. Piriyapongsa J, Marino-Ramirez L, Jordan IK: Origin and evolution of human microRNAs from transposable elements. *Genetics* 2007, **176**:1323-1337.
113. Piriyapongsa J, Jordan IK: A family of human microRNA genes from miniature inverted-repeat transposable elements. *PLoS ONE* 2007, **2**:e203.
114. Smalheiser NR, Torvik VI: Mammalian microRNAs derived from genomic repeats. *Trends Genet* 2005, **21**:322-326.
115. Mansfield JH, Harfe BD, Nissen R, Obenauer J, Srineel J, Chaudhuri A, Farzan-Kashani R, Zuker M, Pasquinelli AE, Ruvkun G, Sharp PA, Tabin CJ, McManus MT: MicroRNA-responsive 'sensor' transgenes uncover Hox-like and other developmentally regulated patterns of vertebrate microRNA expression. *Nat Genet* 2004, **36**:1079-1083.
116. Yekta S, Shih IH, Bartel DP: MicroRNA-directed cleavage of HOXB8 mRNA. *Science* 2004, **304**:594-596.
117. Chen K, Rajewsky N: The evolution of gene regulation by transcription factors and microRNAs. *Nat Rev Genet* 2007, **8**:93-103.
118. Bartel DP, Chen CZ: Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nat Rev Genet* 2004, **5**:396-400.
119. 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* 2008, **40**:351-355.

120. de Felippes FF, Schneeberger K, Dezulian T, Huson DH, Weigel D: Evolution of *Arabidopsis thaliana* microRNAs from random sequences. *RNA* 2008, **14**:2455-2459.
121. Ma Z, Coruh C, Axtell MJ: *Arabidopsis lyrata* small RNAs: transient MIRNA and small interfering RNA loci within the *Arabidopsis* genus. *Plant Cell* 2010, **22**:1090-1103.
122. Lai EC, Tomancak P, Williams RW, Rubin GM: Computational identification of *Drosophila* microRNA genes. *Genome Biol* 2003, **4**:R42.
123. Lim LP, Glasner ME, Yekta S, Burge CB, Bartel DP: Vertebrate microRNA genes. *Science* 2003, **299**:1540.
124. Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, Barad O, Barzilai A, Einat P, Einav U, Meiri E, Sharon E, Spector Y, Bentwich Z: Identification of hundreds of conserved and nonconserved human microRNAs. *Nat Genet* 2005, **37**:766-770.
125. Chekanova JA, Gregory BD, Reverdatto SV, Chen H, Kumar R, Hooker T, Yazaki J, Li P, Skiba N, Peng Q, Alonso J, Brukhin V, Grossniklaus U, Ecker JR, Belostotsky DA: Genome-wide high-resolution mapping of exosome substrates reveals hidden features in the Arabidopsis transcriptome. *Cell* 2007, **131**:1340-1353.
126. Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermuller J, Hofacker IL, Bell I, Cheung E, Drenkow J, Dumais E, Patel S, Helt G, Ganesh M, Ghosh S, Piccolboni A, Sementchenko V, Tammana H, Gingeras TR: RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* 2007, **316**:1484-1488.
127. Berezikov E, Liu N, Flynt AS, Hodges E, Rooks M, Hannon GJ, Lai EC: Evolutionary flux of canonical microRNAs and mirtrons in Drosophila. *Nat Genet* 2010, **42**:6-9.
128. Stark A, Kheradpour P, Parts L, Brennecke J, Hodges E, Hannon GJ, Kellis M: Systematic discovery and characterization of fly microRNAs using 12 *Drosophila* genomes. *Genome Res* 2007, **17**:1865-1879.
129. Chiang HR, Schoenfeld LW, Ruby JG, Auyeung VC, Spies N, Baek D, Johnston WK, Russ C, Luo S, Babiarz JE, Blelloch R, Schroth GP, Nusbaum C, Bartel DP: Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. *Genes Dev* 2010, **24**:992-1009.
130. Lindow M, Krogh A: Computational evidence for hundreds of non-conserved plant microRNAs. *BMC Genomics* 2005, **6**:119.
131. Ballarino M, Pagano F, Girardi E, Morlando M, Cacchiarelli D, Marchioni M, Proudfoot NJ, Bozzoni I: Coupled RNA processing and transcription of intergenic primary microRNAs. *Mol Cell Biol* 2009, **29**:5632-5638.
132. Morlando M, Ballarino M, Gromak N, Pagano F, Bozzoni I, Proudfoot NJ: Primary microRNA transcripts are processed co-transcriptionally. *Nat Struct Mol Biol* 2008, **15**:902-909.
133. Kim YK, Kim VN: Processing of intronic microRNAs. *EMBO J* 2007, **26**:775-783.
134. Berezikov E, Chung W-J, Willis J, Cuppen E, Lai EC: Mammalian mirtron genes. *Mol Cell* 2007, **28**:328-336.
135. Chung WJ, Agius P, Westholm JO, Chen M, Okamura K, Robine N, Leslie CS, Lai EC: Computational and experimental identification of mirtrons in *Drosophila melanogaster* and *Caenorhabditis elegans*. *Genome Res* 2011, **21**:286-300.
136. Drinnenberg IA, Weinberg DE, Xie KT, Mower JP, Wolfe KH, Fink GR, Bartel DP: RNAi in budding yeast. *Science* 2009, **326**:544-550.
137. Lee HC, Li L, Gu W, Xue Z, Crosthwaite SK, Pertsemlidis A, Lewis ZA, Freitag M, Selker EU, Mello CC, Liu Y: Diverse pathways generate microRNA-like RNAs and Dicer-independent small interfering RNAs in fungi. *Mol Cell* 2010, **38**:803-814.
138. Kaessmann H: Origins, evolution, and phenotypic impact of new genes. *Genome Res* 2010, **20**:1313-1326.