Evolution of plant microRNAs and their targets

Michael J. Axtell1 and John L. Bowman2,3

1 Department of Biology and Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA 16802, USA
2 University of California, Davis, Section of Plant Biology, 1 Shields Avenue, Davis, CA 95616, USA
3 School of Biological Sciences, Monash University, Melbourne, VIC 3800, Australia

MicroRNAs (miRNAs) are a specialized class of small silencing RNAs that regulate gene expression. They have a limited phylogenetic distribution among eukaryotes, suggestive of at least two independent origins from an ancestral small RNA-producing pathway. A set of 21 abundantly expressed miRNAs are clearly conserved among the angiosperms; many of these function to regulate transcription factors involved in developmental control. Recent experiments have uncovered a much larger set of weakly expressed, less conserved miRNAs in plants, and this group has provided insights into the origins of miRNAs and their targets. These data have provided a coherent set of hypotheses explaining the birth, selection and death of miRNAs in land plants.

Diversity and ancient origins of eukaryotic small RNAs

MicroRNAs (miRNAs) are just one of three distinct types of small RNAs that are currently understood. The short interfering RNAs (siRNAs) are typically processed by Dicer proteins from long, perfectly double-stranded RNA (dsRNA) precursors, although in some cases they are the direct polymerization products of RNA-dependent RNA polymerases [1,2]. By contrast, the Piwi-interacting RNAs (piRNAs) arise from uncharacterized precursor molecules without the need for Dicer processing (for review, see [3]). All three classes are united not just by their small sizes but also by the fact that they function as specificity determinants for the repressive activities of ribonucleoprotein complexes containing proteins in the Argonaute–Piwi superfamily. The small RNA ‘selects’ RNA targets on the basis of Watson–Crick complementarity, after which the Argonaute or Piwi protein represses the expression of the target either by hydrolyzing the backbone of the RNA target, preventing translation of the target RNA, or by directing chromatin modifications to the genomic locus that encodes the target. Alternatively, some small RNA targets might be activated by Argonaute inter-family interactions [4], whereas others have evolved as ‘mimics’ to titrate Argonaute complexes away from other targets [5].

Both the siRNAs and piRNAs can be further subdivided on the basis of species- and lineage-specific details; land plants, in particular, exhibit extraordinary diversity in siRNA biogenesis and function (reviewed by Rymarquis et al., this issue). However, both classes are fundamentally similar in that they often serve to defend against endogenous or exogenous nucleic acid-based parasites. The importance of siRNAs in plant resistance against viruses is emphatically underscored by the observation that most, if not all, plant viruses encode a suppressor of the siRNA response (for review, see [6]). Many endogenous Arabidopsis siRNAs also arise from transposons and/or gene-poor, repeat-rich pericentromeric heterochromatin, implying that they have functions analogous to those of piRNAs [7–9]. Likewise, recent results from studies of the fruit fly Drosophila melanogaster demonstrate that piRNAs are restricted to germ cells, where they are specifically required to repress the expression of a broad menagerie of transposons [10], whereas the Drosophila siRNA machinery confers antiviral resistance [11,12]. In stark contrast, miRNAs function to post-transcriptionally regulate the expression of endogenous mRNAs in trans, and thus are integral, evolved components of gene regulatory networks rather than being adaptive responders to RNA-based threats.

The existence of Argonautes and Dicers, and the expression of siRNAs and/or the ability to trigger gene silencing with dsRNA (i.e., a process known as RNA interference [RNAi]) is commonplace among disparate eukaryotes (Figure 1). Besides the major multi-cellular lineages (i.e., animals, fungi and land plants), siRNAs and/or RNAi have been reported in diverse eukaryotes, including the slime mold Dictyostelium discoideum [13], the chlorophytic green alga Chlamydomonas reinhardtii [14], and the trypanosome Trypanosoma brucei [15]. Although many eukaryotic lineages have yet to be analyzed (Figure 1), very few species are clearly ‘RNAi-negative’; these include the brewer’s yeast Saccharomyces cerevisiae [16] and certain trypanosome pathogens [17]. Canonical piRNAs are much more spotty in their distribution, having only been described in animals to date. However, the split between Piwi and Argonaute proteins is quite ancient; Piwi genes are also present in ciliates and slime molds [18]; in addition, the ~28-nt ‘scan RNAs’, which direct DNA elimination in the ciliate Tetrahymena, function in a complex with a Piwi protein [19,20] and are thus also piRNAs, albeit with a highly divergent molecular function. The pan-eukaryotic distribution of siRNAs and RNAi implies that they arose early in eukaryotic evolution, whereas the broad mechanistic similarities...
Among miRNAs, siRNAs and piRNAs strongly suggest that they descended from a common, ancestral small RNA-based silencing system. Given that miRNAs and piRNAs appear to be more phylogenetically restricted than siRNAs and RNAi, it might be that they are more recent innovations. The fact that observations of Argonaute–Piwi-associated small RNAs have been restricted to eukaryotic organisms suggests that this ancestral system arose specifically in the eukaryotic lineage. However, the observation of Argonaute proteins in archaebacteria [21–23] suggests that the advent of the Argonaute–Piwi-based small RNA system was much earlier.

The origins of miRNAs
In a similar manner to piRNAs, miRNAs are phylogenetically restricted (Figure 1). Until 2007, miRNAs were only known from land plants, animals and certain animal viruses. Furthermore, animal and plant miRNAs are quite distinct from one another in four areas: miRNA biogenesis, miRNA precursors, target regulation, and conservation of miRNAs.

**miRNA biogenesis**
Animal miRNAs are derived from a compartmentalized, two-step process in which the nuclear localized RNaseIII Drosha first defines one end of the miRNA–miRNA* duplex, and then the pre-miRNA hairpin is exported to the cytosol, where Dicer completes the processing [24]. In some cases, the Drosha function is replaced by de-branching of the 2′-3′ linked intron lariats, which are by-products of spliceosome-mediated intron processing [25,26]. However, even with such ‘miRtrons’, miRNA biogenesis remains a compartmentalized, two-step process. By contrast, plant miRNA–miRNA* duplexes are generated in the nucleus by a single RNaseIII protein, Dicer-like 1 (DCL1 [27,28]).

**miRNA precursors**
Animal miRNA hairpin precursors have a rigidly defined secondary structure, with little variation in either the loop sizes or the positions of miRNA–miRNA* duplexes with respect to the secondary structure [29]. Plant miRNA precursors differ in that they display a huge variation in their lengths and secondary structures. It seems likely that these variations reflect different molecular mechanisms for the recognition and processing of miRNA precursors.

**Target regulation**
Often, complementarity to nucleotides two to seven of an animal miRNA is sufficient for measurable repression of an mRNA target. Such short target-sites most often occur in 3′ untranslated regions (3′ UTRs) in a combinatorial fashion and cause repression of translation independent of Argonaute-mediated target cleavage. The conservation or avoidance of these short miRNA-complementary sites appears to be a major force driving the evolution of animal 3′ UTR sequences [30–32]. By contrast, plant miRNA targets are typified by having a single complementary site with near-perfect base pairing; these sites enabled Argonaute-catalyzed cleavage of the miRNA targets [33–37].

**Conservation of miRNAs**
Many animal miRNAs are conserved even between greatly diverged species [38], especially when considering the crucial ‘seed’ regions, which are often solely responsible for targeting specificity [39]. Several plant miRNAs are also universal among land plants [40–42]. However, the evidence for any miRNAs conserved between animals and plants is slim.

The apparent restriction of miRNAs to animals and land plants, coupled with the clear differences between miRNAs...
in the two lineages, can be explained by the independent evolution of miRNAs in the two lineages from an ancestral siRNA machinery. However, several studies complicate this simple hypothesis. In one recent study [43], it was suggested that miR854 and miR855 are present in both animals and land plants; this implies common ancestry, which in turn implies that the miRNA system of gene regulation was in place at the time of the last common ancestor of today’s plants and animals. Both miR854 and miR855 arise from long terminal repeat (LTR) retrotransposons, and two independent large-scale small RNA sequencing efforts in *Arabidopsis thaliana* failed to recover either mature miRNA [9,44]. However, broad regions surrounding the *A. thaliana* miR854 and miR855 loci produced numerous 24-nt small RNAs corresponding to both genomic strands – a pattern typical of siRNAs, but not miRNAs (M.J. Axtell, unpublished). Similarly, neither mature miR854 nor miR855 have yet been sequenced from any other species, animal or plant. Thus, the conserved small RNA expression observed from these loci [43] probably reflects the ubiquitous deployment of the ancient siRNA system against repetitive elements and does not by itself provide convincing evidence against the ‘two-origin’ hypothesis for animal and plant miRNAs.

Recent data demonstrate that single-celled organisms also produce miRNAs: both *C. reinhardtii* and *D. discoideum* produce small RNAs from imperfectly base-paired, single-stranded precursors [45–47]. Although the nature of the *D. discoideum* miRNA targets remains obscure, the *C. reinhardtii* miRNAs can direct the cleavage of target miRNAs with near-perfect complementarity, reminiscent of the activity of land plant miRNAs [46,47]. However, neither *D. discoideum* nor *C. reinhardtii* miRNA has any detectable sequence similarity with known animal or plant miRNAs. These observations leave open the possibility of two independent origins for miRNAs – one in the last common ancestor of animals, fungi and slime molds, and the other before the divergence of the land plant lineage and the chlorophytic algae (Figure 1). Given that miRNAs are not apparent among any fungi, this particular two-origin scenario requires the secondary loss of miRNAs from the fungal lineage. However, more than two independent derivations of miRNA-mediated gene regulation can also be rationalized. A third, less parsimonious scenario is that miRNA-mediated gene regulation arose just once in eukaryotic evolution, and that this was followed by extensive molecular diversification in animals, land plants and single-celled organisms. Increased sampling of small RNAs from currently unexplored eukaryotic lineages might resolve the deepest origins of miRNAs.

**Conserved and non-conserved miRNAs in the green lineage**

A small set of miRNAs has been detected in several major lineages of land plants (Figure 2). These miRNAs have been identified through sequencing of small RNAs, computational predictions, hybridizations, and the discovery of cleaved targets [37,40–42,48–53]. In total, the current miRBase release (version 10.1) lists 39 miRNA families present in two or more phylogenetically distant plant species; many of these annotations are supported by robust

---

**Figure 2.** Conserved land plant miRNA families (rows) present in different land plant species (columns). All families annotated in more than one plant species in miRBase (version 10.1) are listed and scored with respect to the amount of experimental support; one point each was tallied for each of the following: (a) miRNA sequenced from a small RNA library; (b) evidence of miRNA–miRNA* accumulation, diagnostic of small RNA excision from a stem-loop precursor; (c) detection on an RNA blot; (d) experimental validation of one or more predicted targets; and (e) documented phenotype due to perturbation of miRNA–target relationship. Abbreviations: ath, *Arabidopsis thaliana*; bna, *Brassica napus*; mtr, *Medicago truncatula*; osa, *Oryza sativa*; pta, *Pinus taeda*; ptc, *Populus trichocarpa*; smo, *Selaginella moellendorffii*; tae, *Triticum aestivum*; vli, *Vitis vinifera*; zma, *Zea mays*. Some sparsely sampled species were omitted for brevity, as were species in which miRNAs have been analyzed but not registered in miRBase.
experimental data in multiple species, whereas some have required more experimental data to support their annotated conservation patterns (Figure 2). In particular, 21 miRNA families (156, 159, 160, 162, 164, 166–169, 171, 172, 319, 390, 393–399 and 408) seem to be universally expressed among diverse angiosperms. A subset of these miRNA families is more ancient, because it is also present in gymnosperms, lycopsids and bryophytes. For the purposes of this review, we use the term ‘conserved’ for miRNAs that are present throughout at least one major ancient clade of land plants, for example angiosperms. One common feature of conserved miRNAs is that they are often represented by multiple loci in sequenced genomes; many of them were spawned through large-scale genome duplication events, giving some indication of their antiquity [54]. Strikingly, 13 of 21 conserved angiosperm miRNAs target transcription factor mRNAs with demonstrated or predicted roles in developmental control. In cases where the evolutionary histories of both the miRNA and its target genes have been analyzed, the target gene predates the evolution of the miRNA. For example, class III HD-Zip (homeodomain-leucine zipper) genes are present in all land plants and Chara, a Charaphycean alga in a clade that is the sister of the land plant clade. However, miR166 has only been detected in land plants, with miR165 being derived from miR166, perhaps within the evolution of flowering plants [42,55]. Although some miRNAs evolved before the initial diversification of land plants, none of the miRNAs found to date in Chlamydomonas are shared with land plants.

Advances in techniques of DNA sequencing have enabled more-thorough explorations of the complement of miRNA genes (miRome) of individual species, and have revealed a surprising number of ‘non-conserved’ miRNAs (at least 48 in Arabidopsis), outnumbering the ‘conserved’ miRNAs [9,44]. ‘Non-conserved’ miRNAs are defined as those with a limited phylogenetic distribution and are characterized by primarily being single-copy genes. The significant non-overlap between the non-conserved miRNAs in these two studies suggests that additional non-conserved miRNAs are likely to be identified in Arabidopsis. A similar diversity of seemingly lineage-specific miRNAs has been observed in the bryophyte Physcomitrella patens and the lycophyte Selaginella moellendorffii [41,51,53], indicating that non-conserved miRNAs might be a universal feature within diverse land plants. Many (16 out of 48 [44]) non-conserved Arabidopsis miRNA genes exhibit significant sequence similarity outside the miRNA binding sites within their putative target genes. These features suggest that non-conserved miRNAs are evolutionarily recent, and imply that miRNA genes have high birth and death rates [9,44]. In this respect, they might not differ from protein coding genes, because the estimated rate of duplication of 0.01 per million years for protein coding genes also implies that they have a high death rate [56]. Only those miRNAs stably integrated into indispensable genetic networks make the evolutionary transition from being non-conserved to being conserved. Based on the number of conserved miRNAs, this transition occurs only infrequently. That many conserved miRNAs regulate developmentally important transcription factors, while non-conserved miRNAs have a diverse array of targets is likely a reflection of the infrequent transition to an evolutionarily conserved function.

The birth of plant miRNAs
One plausible hypothesis for the evolutionary origin of plant miRNAs is based on the observed sequence similarity between evolutionarily recent miRNA genes and their target genes where the sequence similarity extends beyond the sequence of the miRNA and its cognate binding site [57]. In this scenario – the inverted duplication hypothesis – miRNA genes arise from inverted duplications of target genes, or fragments of target genes (Figure 3), and transcription of the inverted duplication produces hairpin transcripts that can act as substrates for DCL activities. Consistent with the inverted duplication hypothesis, a significant fraction of non-conserved miRNAs in Arabidopsis exhibit sequence similarity with their putative precursor genes. This similarity extends into the hairpin arms beyond the sequence of the miRNA itself [9,44,57]. Initially, since the transcribed hairpin RNAs would likely exhibit near perfect self-complementarity, they may be processed by DCL enzymes other than DCL1 since DCL1 has limited activity on such substrates. Subsequent acquisition of mutations due to genetic drift would render the hairpin with imperfect complementarity and canalize its processing to DCL1. Simultaneously, this would erase the extended sequence similarity such that little identity outside the miRNA and its complementary target sequence persists. In support of this hypothesis, the processing of some non-conserved miRNAs, such as miR161, is heterogeneous, and some other non-conserved miRNAs appear to be processed by DCL4 [9,44,57].

The retention or loss of the newly evolved miRNA gene would then depend upon its selective advantage. Considering the paucity of phylogenetically conserved miRNAs compared with the large number of non-conserved miRNAs, the most common fate of newly evolved miRNAs seems to be decay to non-functionality through the accumulation of mutations, either driven by selection or through drift. Indeed, some of the non-conserved miRNAs appear to have drifted to the point at which they do not interact with any target transcripts [44]. However, if miRNA-mediated negative regulation of the gene from which the miRNA was derived provides a selective advantage, the miRNA gene might be retained with specificity for the gene of origin. The new miRNA would need to acquire regulatory sequences from the genomic location of the inverted duplication, but this is not an uncommon occurrence, based on the gene expression patterns of random insertions of organelle DNA into the nuclear genome [58]. If selection is maintained, the target pairing of miRNA–gene of origin can co-evolve as a regulatory unit. Subsequent gene duplications of either the miRNA or its target can lead to sub-functionalization — through changes in their respective regulatory sequences – such that expression patterns of duplicated miRNA loci differ both qualitatively and quantitatively [54].

Evolution of miRNA targeting after birth
In theory, the subsequent loss of miRNA regulation could occur through mutations in the miRNA binding site in the
target gene. Such losses can be detected through phyloge-netic analysis of target genes and their close relatives. Instances in which miRNA targeted genes exhibit a paraphyletic relationship suggest a loss of miRNA targeting or, alternatively, a gain of miRNA targeting in closely related genes. The latter scenario is less likely unless recombinational events or gene conversion events are involved. One case of paraphyletic distribution of predicted miRNA regulation is miR395 and its ATP sulfurylase-encoding target genes. It has been predicted that one such gene, *Glycine max*, is no longer regulated by miRNA, owing to changes in its miRNA binding site [59]. However, this has not been confirmed experimentally. Two other cases exhibiting a paraphyletic distribution of target genes are the miRNA–target gene pairs miR319(JAW)–TCP [TEOSINTE BRANCHED, CYCLOIDEA, PCF (=PROLIFERATING CELL FACTOR)] and miR164–CUC (CUP-SHAPED COTYLEDON). In these cases, the target genes do not form a monophyletic clade when considering the gene families in *Arabidopsis*, implying losses and/or gains of miRNA regulation in some target gene family members.

In contrast to the loss of miRNA regulation, which can occur through single nucleotide mutations in miRNA binding sites, gain of miRNA regulation is less likely to be due to the requirement for near complete complementarity between miRNA and target gene pairs. However, one
Arabidopsis miRNA, miR395, targets transcripts generated from two unrelated genes encoding an ATP sulfurylase and a sulfate transporter, which are involved in sulfate uptake and assimilation [60,61]. Although both target sites are within coding regions, they are not in the same reading frame, suggesting convergent evolution of miRNA binding sites rather than acquisition by exon shuffling. miR395 evolved before the divergence of the moss and flowering plant lineages [53], and therefore tracing its evolutionary history and that of its two targets might clarify the origin of the dual specificity of miR395. The evolutionary history of miR319(JAW)–TCP might also be complex, with miR319 evolving before the divergence between the moss and flowering plant lineages, but neither the Physcomitrella nor Selaginella TCP genes (the targets of miR319 in angiosperms) have a recognizable miR319 binding site.

In most cases, conserved plant miRNAs regulate homologous targets at identical target sites in every species in which they exist. The lack of sequence drift is probably due to the rarity of the simultaneous, compensatory mutations that would be required to maintain a given regulatory interaction. However, there are a few instances in which such compensatory changes might have occurred. miR390-dependent trans-acting siRNAs (ta-siRNAs) target auxin response factor genes in both mosses and angiosperms but have completely different ta-siRNA and target sequences [41,62]. Likewise, angiosperm miR168 and moss miR904 both target AGO1 genes despite vastly different sequences [41]. Both of these examples might be explained by sequence drift of ancestral small RNA-target interactions; alternatively, they might represent examples of convergent evolution of small RNAs towards recurrent molecular functions.

The proposed mechanism for the evolution of plant miRNAs and their target sites differs from that proposed for animal miRNAs and their target sites [57,63]. In animals, most miRNA binding sites exhibit perfect complementarity for only 6–8 bases towards the 5′ end of the miRNA, in contrast to near perfect complementarity across the entire miRNA in plants. In addition, animal miRNA binding sites occur typically in the 3′ UTRs of the target genes, whereas in plants they are most often within the coding region. Given that there are many fewer constraints in terms of both sequence length and sequence identity for animal miRNA binding sites, convergent evolution of 3′ UTR miRNA binding sites is thought to be common. Most animal miRNA targets do not appear to be evolutionarily related to the miRNAs that control them, as they are in plants [63]. The ease with which animal miRNA binding sites might evolve has two major consequences. First, a large percentage of animal genes might be regulated by miRNA-mediated control, with up to one-third of all human genes proposed to be miRNA regulated. Second, the sequences of 3′ UTRs of animal genes experience evolutionary constraint such that they do not come under miRNA-mediated regulation. By contrast, owing to the difficulty of convergent evolution of miRNA binding sites in plants, the number of targets of an miRNA are limited, and targets are usually evolutionarily related to the miRNA itself.

Prospects

miRNA-mediated gene regulation appears to have arisen more than once during eukaryotic diversification from pre-existing small RNA-based immune systems. Most ancient plant miRNA-target relationships have already been discovered and are overwhelmingly biased towards regulation of developmentally potent transcription factor genes. By contrast, the discovery of further, less conserved miRNAs with more diverse regulatory targets seems likely to continue. Understanding how these two types of miRNAs interact and function in multiple species will be a major focus of ongoing research.

Acknowledgements

We apologize to colleagues whose work was not discussed, owing to space constraints. Research in M.J.A.’s laboratory is supported by the US National Science Foundation (MCB-0718051), and research in J.L.B.’s laboratory is supported by the Australian Research Council (DP0771232, FP0561326) and the US National Science Foundation (IOB-0515435).

References

2 Sijen, T. et al. (2007) Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. Science 315, 244–247
Review

29 Han, J. et al. (2006) Molecular basis for the recognition of primary microRNAs by the Drosha–DGCR8 complex. Cell 125, 887–901
31 Stark, A. et al. (2005) Animal microRNAs confer robustness to gene expression and have a significant impact on 3’UTR evolution. Cell 123, 1133–1146
35 Llave, C. et al. (2002) Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. Science 297, 2053–2056
47 Zhao, T. et al. (2007) A complex system of small RNAs in the unicellular green alga Chlamydomonas reinhardtii. Genes Dev. 21, 1190–1203
49 Barakat, A. et al. (2007) Large-scale identification of microRNAs from a basal eudicot (Eschscholzia californica) and conservation in flowering plants. Plant J. 51, 991–1003