Prediction of Plant MicroRNA Targets

Matthew W. Rhoades,[1,2] **Brenda J. Reinhart,**[1]
Lee P. Lim,[1,2] **Christopher B. Burge,**[2]
Bonnie Bartel,[3,4] **and David P. Bartel**[1,2,4]
[1]Whitehead Institute for Biomedical Research
9 Cambridge Center
Cambridge, Massachusetts 02142
[2]Department of Biology
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139
[3]Department of Biochemistry and Cell Biology
Rice University
6100 Main Street
Houston, Texas 77005

Summary

We predict regulatory targets for 14 *Arabidopsis* mi-
croRNAs (miRNAs) by identifying mRNAs with near
complementarity. Complementary sites within pre-
dicted targets are conserved in rice. Of the 49 pre-
dicted targets, 34 are members of transcription factor
gene families involved in developmental patterning or
cell differentiation. The near-perfect complementarity
between plant miRNAs and their targets suggests that
many plant miRNAs act similarly to small interfering
RNAs and direct mRNA cleavage. The targeting of de-
velopmental transcription factors suggests that many
plant miRNAs function during cellular differentiation
to clear key regulatory transcripts from daughter cell
lineages.

Introduction

Nearly 200 genes for tiny, noncoding RNAs termed mi-
croRNAs (miRNAs) have been identified in animals and
plants (Lagos-Quintana et al., 2001, 2002; Lau et al.,
2001; Lee and Ambros, 2001; Llave et al., 2002; Moure-
latos et al., 2002; Reinhart et al., 2002). Two miRNAs,
lin-4 and *let-7* RNAs, have been studied in detail; both
control developmental timing in *C. elegans* through a
mechanism that involves imperfect base pairing to the
3′ UTRs of target mRNAs (Lee et al., 1993; Wightman
et al., 1993; Ha et al., 1996; Moss et al., 1997; Reinhart
et al., 2000; Slack et al., 2000). The remaining miRNAs
have unknown functions. Nonetheless, their sequences
are typically conserved among different species, and
many have intriguing expression patterns in different
tissues or stages of development, indicating that these
other miRNAs have important functions and might also
modulate gene expression. This idea is supported by
the observation that Dicer and Argonaute proteins,
which are known to be crucial for normal plant and
animal development, are needed for proper miRNA ac-
cumulation (Robinson-Beers et al., 1992; Ray et al.,
1996a, 1996b; Jacobsen et al., 1999; Grishok et al., 2001;

[4]Correspondence: dbartel@wi.mit.edu (D.P.B.), bartel@rice.edu (B.B.)

Hutvágner et al., 2001; Ketting et al., 2001; Knight and
Bass, 2001; Reinhart et al., 2002).

The major challenge in determining miRNA functions
is to identify their regulatory targets. By analogy to *lin-4*
and *let-7* RNAs, it is reasonable to suppose that miRNAs
generally recognize their regulatory targets through
base pairing. However, the small size of the mature
miRNAs (20–24 nt) and the imperfect nature of
miRNA:mRNA base pairing have hampered the general
prediction of mRNA targets for animal miRNAs. Thus
far, prediction of animal miRNA targets has been
achieved only after experimental evidence narrowed the
number of candidate mRNAs to a small set, either by
placing the mRNAs within the same regulatory pathway
as the miRNA or by identifying regulatory elements
within mRNA 3′ UTRs (Lee et al., 1993; Wightman et al.,
1993; Moss et al., 1997; Reinhart et al., 2000; Slack et
al., 2000; Lai, 2002). An indication that target prediction
for certain plant miRNAs might be more straightforward
came with the recent identification of miR171, a plant
miRNA with perfect antisense complementarity to the
mRNAs of three SCARECROW-like transcription factors
(Llave et al., 2002; Reinhart et al., 2002).

Here we report that near complementarity to mRNAs,
particularly transcription factor mRNAs, is a general
trend for plant miRNAs. We have been able to identify
potential regulatory targets for 14 of the 16 miRNAs
studied by searching for mRNAs capable of base pairing
with three or fewer mismatches to one of the miRNAs.
The fact that many of these potential targets are mem-
bers of gene families with roles in plant development
supports the idea that the function of miRNAs in mediat-
ing development is conserved across kingdoms. Partic-
ularly compelling targets include the *PHABULOSA* and
PHAVULOTA mRNAs, for which the identification of
miRNA complementary sites may explain the ectopic
expression previously described for mutations in these
genes (McConnell et al., 2001). Similar analysis of animal
miRNAs did not predict animal regulatory targets, sug-
gesting mechanistic differences between plant and ani-
mal miRNA function.

Results and Discussion

Plant MicroRNAs Have Significant
Complementarity to Messenger RNAs

To identify potential regulatory targets, we searched for
Arabidopsis mRNAs that were complementary, with four
or fewer mismatches, to at least one of 16 recently identi-
fied *Arabidopsis* miRNAs (Reinhart et al., 2002). Gaps
were not allowed, and G:U and other noncanonical pairs
were treated as mismatches. To evaluate the signifi-
cance of these hits to annotated mRNAs, parallel analy-
ses were performed using cohorts of randomly per-
muted sequences that had identical sizes and base
compositions as the set of authentic miRNAs. There
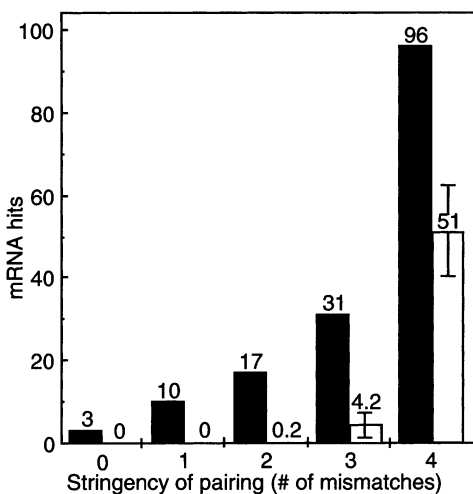were substantially more antisense hits to the authentic

Figure 1. Antisense Hits between *Arabidopsis* miRNAs and Annotated mRNAs

Annotated *Arabidopsis* mRNAs were searched for sites complementary to 16 *Arabidopsis* miRNAs with 0–4 mismatches (solid bars). Identical searches with cohorts of 16 randomized RNAs were also performed (open bars, mean values from ten cohorts; error bars, one standard deviation). Note that two hits by similar miRNAs to the same complementary site within an mRNA were counted as separate hits (Table 1).

miRNAs than to the randomized sequences (Figure 1). This difference was especially striking at higher stringency; when summing the hits with two or fewer mismatches, the number of hits to the authentic miRNA set outnumbered those to the randomized cohorts by a ratio of 30:0.2 (Figure 1). Considering the low probability of so many antisense hits occurring by chance, we suggest that these complementary sites reflect a functional relationship between the miRNAs and the identified mRNAs—that these protein-coding genes are regulatory targets of the miRNAs to which they can potentially base pair.

At lower stringencies, there were also significantly more hits with the authentic set of miRNAs than with the randomized cohorts. Most of the 31 hits with three mismatches are viable miRNA target candidates, although a few are likely to be mRNAs with fortuitous complementarity, as judged by the observation that on average the randomized cohorts hit 4.2 mRNAs when three mismatches were permitted (Figure 1). Some hits with four mismatches might also be genuine targets. However, they are not included in the present analysis because of the greater likelihood that their complementarity is fortuitous or occurs because they are targets of unidentified miRNAs related to our query set of 16 miRNAs.

Potential regulatory targets with three or fewer mismatches were found for 14 of the 16 miRNAs (Table 1). Targets for the other two miRNAs might be identified through slight changes in the search algorithm. For example, miR163, one of the two miRNAs without predicted targets in Table 1, has extensive complementarity to members of the AtPP-like gene family (At1g66690, At1g66700, At1g66720, At3g44860, At3g44870), which have unknown functions (Cui et al., 1999). All 24 nucleo-

tides of this miRNA paired to complementary sites within these mRNAs when a single-nucleotide gap was permitted near the 3′ terminus of the miRNA. Nonetheless, when searching for miRNA targets, permitting gaps did not substantially increase the number of targets predicted for the other miRNAs (data not shown). Perhaps a bulge is accommodated near the miRNA terminus more readily for miR163 because this miRNA is 24 nt in length, which is 3 nt longer than the other miRNAs queried.

In all cases where an miRNA was complementary to more than one mRNA, most of the potential targets were members of the same gene family (Table 1). The fraction of the gene family members with miRNA complementary sites varied considerably. Of the 16 Squamosa-promoter Binding Protein (SBP)-like genes in *Arabidopsis* (Riechmann et al., 2000), ten have miR156 complementary sites. In contrast, the *MYB* and *NAC* families each have over 100 members in *Arabidopsis* (Riechmann et al., 2000), of which five in each case have sites complementary to miR159 or miR164, respectively. As more miRNAs are identified, it will be interesting to learn whether remaining members of these gene families have complementary sites to other miRNAs. In support of this possibility, unrelated miRNAs can be complementary to different members of the same gene family, as illustrated by miR160 and miR167, which apparently target different members of the Auxin Response Factor family (Ulmasov et al., 1999).

When considering the significance of multiple hits to the same gene family, it is important to address the possibility that these hits are merely the consequence of complementarity to a nucleotide sequence that encodes a critical protein motif. Indeed, for miR161, miR165, miR170, and miR171, the miRNA complementary sites were within the context of a domain strongly conserved among family members, as shown for the miR165 complementary sites (Figure 2A). Therefore, we cannot rule out the possibility that only a subset of the hits for these miRNAs are authentic targets. This possibility is less likely in the cases of miR156, miR157, miR159, miR160, miR164, and miR169. The complementary sites for these miRNAs fell outside the conserved domains that define the families and instead fell within sequence contexts that were only weakly conserved among the family members, as shown for the miR156 sites within *SBP*-like mRNAs (Figure 2B). Indeed, there are examples where the conservation of the miRNA complementary sites among family members must be independent of conserved protein function. In the case of the *MYB* genes with miR159 complementary sites, four genes translate the complementary site in the same reading frame, while the fifth gene translates the site in a different reading frame. In four other cases (miR156/157 to At1g53160, miR156 to At2g33810, and miR169 to At1g17590 and At1g54160), the miRNA complementary sites are not in the coding regions at all but rather in the 3′ UTRs, as illustrated for miR156 and its complementary sites (Figure 2B).

MicroRNA Complementary Sites Are Conserved among Flowering Plants

Many complementary sites observed in *Arabidopsis* are conserved in rice (*Oryza sativa*). Analysis of rice homo-

Table 1. Potential Regulatory Targets of *Arabidopsis* miRNAs

MicroRNA	Target Protein Family	Target Gene Names (Number of Mismatches)
miR156	Squamosa-promoter Binding Protein (SBP)-like proteins	At3g57920 (1), At2g42200/SPL9 (1), At5g50570 (1), At5g50670 (1), At1g53160/SPL4 (2), At2g33810/SPL3 (2), At1g27370/SPL10 (2), At5g43270/SPL2 (2), At1g69170/SPL6 (2), At1g27360/SPL11 (2)
miR157	Squamosa-promoter Binding Protein (SBP)-like proteins	At1g27370/SPL10 (1), At3g57920 (1), At2g42200/SPL9 (1), At5g43270/SPL2 (1), At1g27360/SPL11 (1), At1g69170/SPL6 (2), At5g50570 (2), At5g50670 (2), At1g53160/SPL4 (3)
	Putative RNA helicase	At5g08620 (3)
	Unknown proteins	At3g47170 (3), At1g22000 (3)
miR158	Unknown protein	At1g64100 (3)
miR159	MYB proteins	At2g32460/AtMYB101 (2), At3g60460 (3), At2g26950/AtMYB104 (3), At5g06100/AtMYB33 (3), At3g11440/AtMYB65 (3)
	Unknown protein	At1g29010 (3)
miR160	Auxin Response Factors	At1g77850/ARF17 (1), At2g28350/ARF10 (2), At4g30080/ARF16 (3)
miR161	PPR repeat proteins	At1g63150 (3), At1g63400 (3), At1g06580 (3), At1g64580 (3), At5g16640 (3), At1g62670 (3), At1g62720 (3), At5g41170 (3), At1g63080 (3)
miR164	NAC domain proteins	At5g61430 (2), At5g07680 (2), At1g56010/NAC1 (2), At3g15170 (3), At5g53950/CUC2 (3)
miR165	HD-Zip transcription factors	At5g60690/REV (3), At3g34710/PHB (3), At4g32880/ATHB-8 (3), At1g30490/PHV (3)
miR166	HD-Zip transcription factor	At1g52150/ATHB-15 (3)
miR167	Auxin Response Factor	At5g37020/ARF8 (3)
miR168	ARGONAUTE	At1g48410/AGO (3)
miR169	CCAAT Binding Factor (CBF)-HAP2-like proteins	At1g17590 (3), At1g54160 (3)
miR170	GRAS domain proteins (SCARECROW-like)	At2g45160 (2), At3g60630 (2), At4g00150/SCL6 (2)
miR171	GRAS domain proteins (SCARECROW-like)	At2g45160 (0), At3g60630 (0), At4g00150/SCL6 (0)

For each gene, the number of mismatches between the miRNA and the mRNA is indicated in parentheses. The sequences of three pairs of miRNAs (miR156/miR157, miR165/miR166, and miR170/miR171) are closely related and therefore are sometimes complementary to the same sites within the target mRNAs. Sites complementary to miR158 had an additional mismatch added to compensate for the fact that miR158 is at least 1 nt shorter than the other miRNAs.

logs focused on the seven miRNAs perfectly conserved in *Oryza* (Reinhart et al., 2002) for which complementary sites had been identified in *Arabidopsis* (Table 1). When using a three-mismatch cutoff, six of the seven conserved miRNAs (miR156, miR160, miR164, miR167, miR169, and miR171) have at least one potential target gene in *Oryza* homologous to a corresponding *Arabidopsis* target. As a control, an analogous study was performed using *Arabidopsis* hits to the cohorts of randomized miRNAs; no miRNA complementary sites were found in rice homologs of these *Arabidopsis* hits, even when four mismatches were allowed.

The location of the miRNA complementary sites within the mRNAs was conserved between *Arabidopsis* and rice. Importantly, when there were differences between *Arabidopsis* and rice complementary sites within homologous genes, these differences were distributed evenly across the three codon positions (Table 2). Homologous regions under selection only at the protein level tend to exhibit a higher frequency of differences at codon position 3. Thus, the even distribution of mismatches across the codon positions indicates selection occurring at the nucleic acid level, in addition to any selection at the protein level, as would be expected if these segments act in miRNA recognition.

Most Predicted MicroRNA Targets Are Members of Transcription Factor Families Involved in Development

Perhaps the most intriguing evidence that these genes are regulatory targets of the miRNAs is the identity of the genes themselves. Sixty-one miRNA/mRNA pairings are reported, which, due to overlap between similar miRNAs, represent 49 unique genes (Table 1). Of these 49 predicted targets, 34 are known or putative transcription factors (Table 1), even though transcription factors are thought to represent only 6% of protein-coding genes in *Arabidopsis* (Riechmann et al., 2000). Many of these genes specify shoot and floral meristem development or, for those with unknown functions, are in families that have members involved in meristem development. For example, the predicted targets of miR164 include *CUP-SHAPED COTYLEDON2* (*CUC2*), which is required for shoot apical meristem formation (Aida et al., 1997), and miR165 predicted targets include *PHABULOSA* (*PHB*) and *PHAVOLUTA* (*PHV*), which encode HD-Zip transcription factors that regulate axillary meristem initiation and leaf development (McConnell et al., 2001). A miR159 predicted target, AtMYB33, can bind to the promoter of the floral meristem identity gene *LEAFY* (Gocal et al., 2001). Homologs of the SBPs, which are thought to regulate the *Antirrhinum* floral meristem identity gene *SQUAMOSA* (Klein et al., 1996), may in turn be regulated by miR156 and miR157.

Genetic evidence supports the regulatory roles of miR165 complementary sites within *PHB* and *PHV* (Figure 2A). Multiple gain-of-function alleles have been isolated for both genes, and each of these mutations disrupts the miR165 complementary site, usually as a single-nucleotide substitution (McConnell et al., 2001). In the mutant examined, *phb* mRNA expression extends more broadly than in wild-type (McConnell et al., 2001),
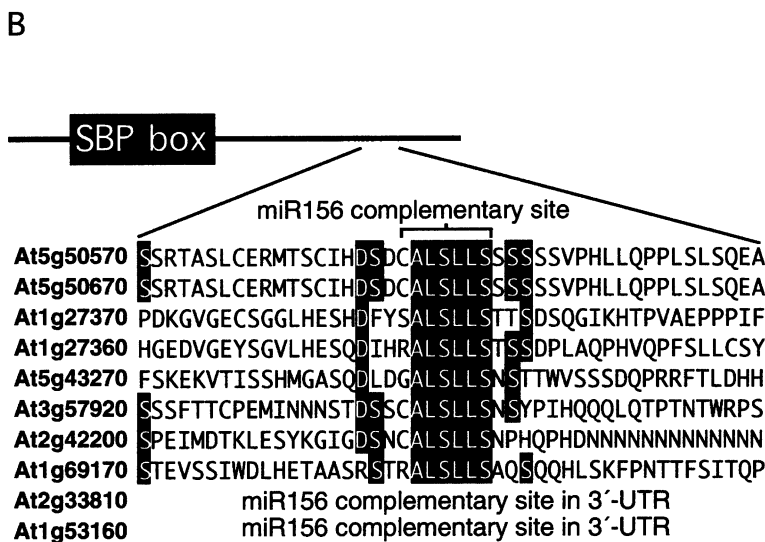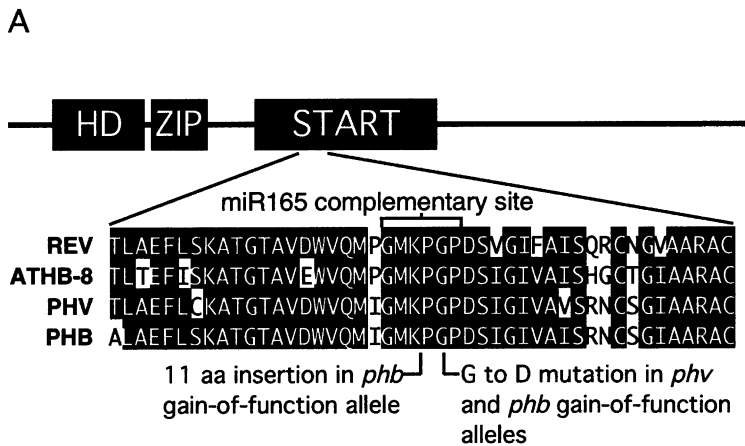
A

B

suggesting that complementarity to miR165 is required for confining *PHB* mRNA accumulation to the proper cell types.

A connection between miRNAs and meristem development is consistent with the phenotypes of the *Arabidopsis carpel factory* (*caf*) mutant. Dicer and CAF are homologous RNaseIII domain proteins required for the accumulation of mature miRNAs in animals and plants, respectively (Hutvágner and Zamore, 2002; Reinhart et al., 2002). Mutant alleles of *CAF*, which is also known as *SHORT INTEGUMENT1* (*SIN1*), delay the meristem switch from vegetative to floral development and cause overproliferation of the floral meristem (Ray et al., 1996a; Jacobsen et al., 1999). Other genes required for miRNA accumulation in animals are homologs of the *Arabidopsis* gene *ARGONAUTE* (*AGO1*), which is required for axillary shoot meristem formation and leaf development in *Arabidopsis* (Bohmert et al., 1998). While *AGO1* has not yet been reported to influence miRNA accumulation in plants, it is a predicted target of miR168 (Table 1),

suggesting a negative-feedback mechanism for controlling expression of the *AGO1* gene.

Other predicted targets of miRNAs do not have direct roles in meristem identity but rather could have roles in cell division or differentiation. For example, miR160 and miR167 are predicted to target auxin response factors, DNA binding proteins that are thought to control transcription in response to the phytohormone auxin (Ulmasov et al., 1999). Transcriptional regulation is important for many of the diverse developmental responses to auxin signals, which include cell elongation, division, and differentiation in both roots and shoots (Rogg and Bartel, 2001; Liscum and Reed, 2002). The predicted targets of miR170 and miR171 are three SCARECROW-like proteins, a family of transcription factors whose members have been implicated in radial patterning in roots, signaling by the phytohormone gibberellin, and light signaling (Di Laurenzio et al., 1996; Peng et al., 1997; Silverstone et al., 1998; Bolle et al., 2000; Helariutta et al., 2000). Overall, the high percentage of predicted

Table 2. MicroRNA Complementary Sites in Potential mRNA Targets Conserved Between *Arabidopsis* and *Oryza*

Target Gene	RNA Sequence of Complementary Site	Peptide Sequence
miR156	UGU GCU CAC UCU CUU CUG UCA	
At5g50570 (1)	UGU GCU CuC UCU CUU CUG UCA	CALSLLS
At5g50670 (1)	UGU GCU CuC UCU CUU CUG UCA	CALSLLS
At3g57920 (1)	UGU GCU CuC UCU CUU CUG UCA	CALSLLS
At2g42200 (1)	UGU GCU CuC UCU CUU CUG UCA	CALSLLS
At1g27370 (2)	aGU GCU CuC UCU CUU CUG UCA	SALSLLS
At1g27360 (2)	cGU GCU CuC UCU CUU CUG UCA	RALSLLS
At5g43270 (2)	gGU GCU CuC UCU CUU CUG UCA	GALSLLS
At1g69170 (2)	cGU GCU CuC UCU CUU CUG UCA	RALSLLS
At2g33810 (2)	UuU GCU uAC UCU CUU CUG UCA	3′ UTR
At1g53160 (2)	UcU GCU CuC UCU CUU CUG UCA	3′ UTR
Os 20095 (1)	UGU GCU CuC UCU CUU CUG UCA	CALSLLS
Os 06618 (1)	UGU GCU CuC UCU CUU CUG UCA	CALSLLS
Os 02878 (1)	UGU GCU CuC UCU CUU CUG UCA	CALSLLS
Os 25470 (2)	gGU GCU CuC UCU CUU CUG UCA	GALSLLS
miR160	U GGC AUA CAG GGA GCC AGG CA	
At1g77850 (1)	U GGC AUg CAG GGA GCC AGG CA	AGMQGARQ
At2g28350 (2)	a GGa AUA CAG GGA GCC AGG CA	AGIQGARQ
At4g30080 (3)	g GGu uUA CAG GGA GCC AGG CA	VGLQGARH
OsTC73519 (1)	a GGC AUA CAG GGA GCC AGG CA	AGIQGARH
OsTC70631 (1)	a GGC AUA CAG GGA GCC AGG CA	AGIQGARH
Os 17478 (1)	a GGC AUA CAG GGA GCC AGG CA	AGIQGARH
Os 02679 (1)	a GGC AUA CAG GGA GCC AGG CA	AGIQGARH
miR164	UG CAC GUG CCC UGC UUC UCC A	
At1g56010 (2)	aG CAC GUa CCC UGC UUC UCC A	EHVPCFSN
At5g07680 (2)	Uu uAC GUG CCC UGC UUC UCC A	VYVPCFSN
At5g61430 (2)	Uc uAC GUG CCC UGC UUC UCC A	VYVPCFSN
At3g15170 (3)	aG CAC GUG uCC UGu UUC UCC A	EHVSCFSN
At5g53950 (3)	aG CAC GUG uCC UGu UUC UCC A	EHVSCFST
Os 00116 (2)	cG CAC GUG aCC UGC UUC UCC A	AHVTCFSN
miR167	U AGA UCA UGC UGG CAG CUU CA	
At5g37020 (3)	U AGA UCA gGC UGG CAG CUU gu	LRSGWQLV
OsTC79868 (3)	U AGA UCA gGC UGG CAG CUU gu	DRSGWQLV
miR169	UCG GCA AGU CAU CCU UGG CUG	
At1g17590 (3)	aaG GgA AGU CAU CCU UGG CUG	3′ UTR
At1g54160 (3)	aCG GgA AGU CAU CCU UGG CUa	3′ UTR
Os 04048 (3)	UaG GCA AcU CAU uCU UGG CUG	3′ UTR
Os 09843 (3)	UaG GCA AuU CAU CCU UGG CUu	3′ UTR
miR171	G AUA UUG GCG CGG CUC AAU CA	
At2g45160 (0)	G AUA UUG GCG CGG CUC AAU CA	GILARLNH
At3g60630 (0)	G AUA UUG GCG CGG CUC AAU CA	GILARLNH
At4g00150 (0)	G AUA UUG GCG CGG CUC AAU CA	GILARLNQ
OsTC76755 (0)	G AUA UUG GCG CGG CUC AAU CA	EILARLNQ
OsTC81772 (0)	G AUA UUG GCG CGG CUC AAU CA	EILARLNH
Os 00711 (0)	G AUA UUG GCG CGG CUC AAU CA	EILARLNQ
Os 12185 (0)	G AUA UUG GCG CGG CUC AAU CA	EILARLNQ
OsTC75254 (1)	G AUA UUG GCG CGG CUC AAU uA	EILARLNY

For each gene, the nucleotide sequence of the miRNA complementary site is broken into codons corresponding to the reading frame of the mRNA. The reverse complement is shown for each miRNA, and for each complementary site, mismatches are shown in lower case and tallied in parentheses. The peptide sequence of the miRNA complementary site is shown. *Oryza* genes are labeled either by their tentative consensus (TC) numbers from the TIGR rice gene index (version 9.0) or by the genomic contig of the mRNA predicted by GenomeScan.

miRNA targets that act as developmental regulators suggest that miRNAs are involved in a wide range of cell division and cell fate decisions throughout the plant.

Mechanistic and Functional Models for Regulation by MicroRNAs in Plants

The success in identifying potential miRNA targets in *Arabidopsis* prompted us to examine whether our simple computational approach could also identify miRNA targets in *C. elegans* and *D. melanogaster*. In both organisms, the miRNAs had few mRNA hits with complementary sites—essentially the same number of hits as seen for randomized cohorts (data not shown). While the possibility that a few animal miRNAs do recognize their targets with near-perfect complementarity cannot be excluded, the general phenomenon of near-perfect
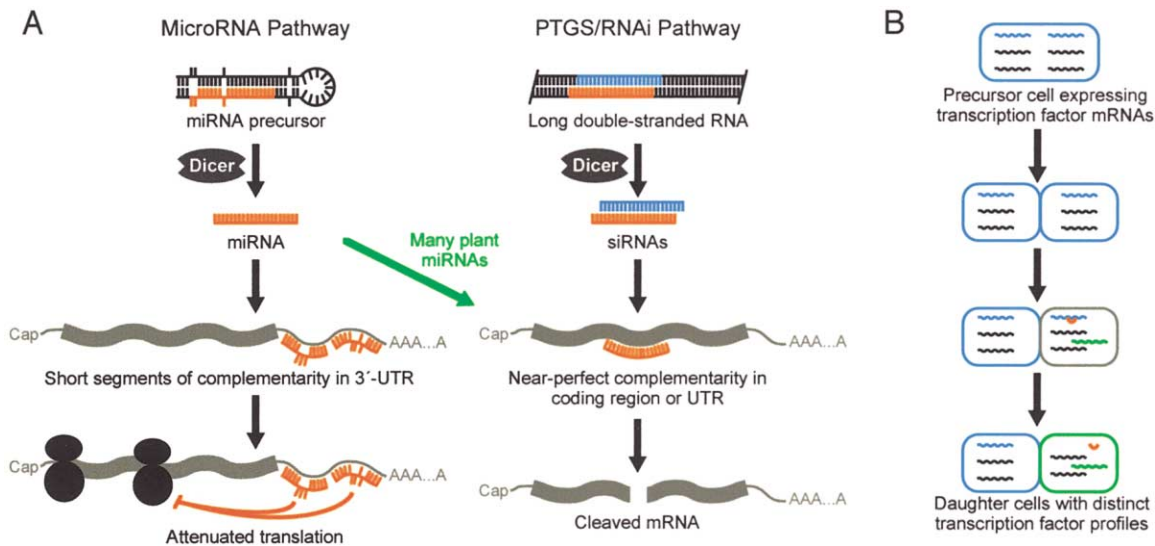
Figure 3. Models for the Biogenesis, Action, and Roles of miRNAs in Plants

(A) Although plant miRNAs are apparently generated through the classical miRNA pathway (Reinhart et al., 2002), we propose that many act as classical siRNAs, pairing with near-perfect complementarity to their mRNA targets to specify mRNA cleavage.

(B) Plant miRNAs might target transcription factor mRNAs for cleavage following cell divisions that require rapid implementation of new transcription factor programs. Following cell division, the daughter cells inherit transcription factor mRNAs from the precursor cell. At the onset of differentiation, one daughter might express not only new transcription factor mRNAs (green) but also miRNAs (red) complementary to certain transcription factor mRNAs (blue) inherited from the precursor cell. The miRNAs could direct the cleavage of the inherited transcription factor mRNA (as well as any newly transcribed messages from this gene), preventing the inappropriate expression of the transcription factor protein, thus enabling the rapid differentiation of the daughter cell.

complementarity appears to be specific to plants. Two other key differences emerge when comparing the predicted target sites of plant miRNAs with those of the *C. elegans lin-4* and *let-7* miRNAs. First, the plant complementary sites are primarily, though not exclusively, within the ORFs, whereas the only proposed *lin-4* and *let-7* sites are within 3′ UTRs (Lee et al., 1993; Wightman et al., 1993; Moss et al., 1997; Reinhart et al., 2000; Slack et al., 2000). Second, multiple sites within the same target mRNA are not detected in plants, whereas there are typically multiple *lin-4* and *let-7* sites within each mRNA target (Lee et al., 1993; Wightman et al., 1993; Ha et al., 1996; Reinhart et al., 2000; Slack et al., 2000).

These differences observed between plant and animal miRNA target recognition have intriguing mechanistic implications for plant miRNA function (Figure 3A). Namely, plant miRNA target recognition appears to resemble that of small interfering RNAs (siRNAs) much more than that of animal miRNAs. During RNA interference (RNAi), long double-stranded RNA is processed by Dicer into ~22 nt siRNAs, which serve as guide RNAs to target homologous mRNA sequences for cleavage (Bernstein et al., 2001; Hutvágner and Zamore, 2002). Importantly, targeting either the ORF or the UTRs is effective (McManus et al., 2002), provided that the siRNA has near-perfect complementarity to the targeted mRNA (Elbashir et al., 2001). Plants also have siRNAs. Indeed, these tiny RNAs were first observed in plants and are associated with a process related to RNAi, known as posttranscriptional gene silencing (PTGS), which leads to the destruction of mRNA from plant viruses and transgenes (Hamilton and Baulcombe, 1999; Matzke et al., 2001). Plant miRNAs resemble animal miRNAs in their

biogenesis, in that they are derived from endogenous, evolutionarily conserved genes and are processed from stem-loop precursors by a Dicer homolog, with accumulation of mature miRNA from only one arm of the precursor stem-loop (Reinhart et al., 2002). However, plant miRNAs resemble siRNAs in their target recognition, suggesting that they might also resemble siRNAs in their mechanism of action (Figure 3A). We propose that many plant miRNAs hybridize to mRNAs with near-perfect complementarity and target the mRNAs for cleavage. A function in mediating RNA cleavage might allow the plant miRNAs to target any region of the mRNA, whereas the animal miRNAs that mediate translational attenuation might be relegated to 3′ UTRs in order to avoid the mRNA-clearing activity of ribosomes. The efficiency and finality of mRNA cleavage might require only a single complementary site in each message, whereas the regulatory mechanism of *lin-4* and *let-7* miRNAs, which leaves the mRNA intact, might generally require multiple target sites.

In presenting this hypothesis, we leave open the possibility that some plant miRNAs might not specify cleavage of their regulatory targets, and some might specify cleavage of some targets but employ other mechanisms to regulate other targets. Targets with many mismatches, analogous to the targets of lin-4 and let-7 miRNAs, would not have been detected in our analysis. Furthermore, some mismatches for the predicted targets are near the center of the complementary sites (Table 2, data not shown) and might be expected to abrogate siRNA-mediated mRNA cleavage (Elbashir et al., 2001). However, it is difficult to know whether these mismatches are incompatible with mRNA cleavage be-

cause the types and locations of mismatches permissive for siRNA-mediated cleavage are still being determined in animals and have not yet been explored in plants. In those cases where the miRNAs might not be mediating mRNA cleavage, they might attenuate translation (Olsen and Ambros, 1999), act as guide RNAs for mRNA modifications (Kiss, 2002), or target DNA for epigenetic modifications, such as methylation (Matzke et al., 2001). Although DNA targeting cannot be excluded as an additional miRNA function for some miRNAs, two observations argue strongly for a role in targeting mRNAs in addition to any possible role in targeting DNA. First, plant miRNAs are complementary to the sense rather than antisense strands of mRNAs (data not shown). Second, the complementary sites for miR165 and miR166 span a splice junction within each of the HD-Zip mRNAs.

The observation that many plant miRNAs potentially target the mRNAs of transcription factors involved in development suggests that some miRNAs might function to clear key regulatory transcripts from certain daughter cell lineages (Figure 3B). Through the action of miRNAs, these inherited mRNAs could be eliminated without relying on constitutively unstable messages, and any remaining transcription from these genes could be neutralized. Now that potential miRNA binding sites in some of these developmentally important transcription factor mRNAs have been identified, it should be possible to test this speculative model by disrupting the miRNA complementarity site in the mRNA without changing the protein sequence of the transcription factor.

The miRNAs analyzed here are likely to be only a small fraction of the miRNAs in *Arabidopsis* (Llave et al., 2002; Reinhart et al., 2002). Nonetheless, the discovery that so many of these plant miRNAs appear to have readily identifiable regulatory targets will greatly facilitate experimental investigation of the functions of these tiny noncoding RNAs and the many other miRNAs remaining to be found in plants. With the ability to computationally identify candidate targets, the presumed roles of miRNAs in development can be more readily explored, and roles of miRNAs in other processes can be more readily uncovered.

Experimental Procedures

Identification of miRNA Complementary Sites in Annotated mRNAs

The set of annotated *Arabidopsis* mRNA sequences was extracted from the genomic GenBank files, January 2002 release (*Arabidopsis* Genome Initiative, 2000). This set was searched for complementary sites to any of 16 miRNAs (GenBank accession numbers AJ493620–AJ493656) using PatScan (Dsouza et al., 1997). When the miRNA was cloned as both a 20 and 21 nt RNA, the 21 nt RNA was used (Reinhart et al., 2002). Thus, the miR158 sequence was 20 nt, the miR163 sequence was 24 nt, and the remaining 14 miRNA sequences were 21 nt. One mismatch was added to all miR158 complementary sites to compensate for their smaller size and the correspondingly greater chance of fortuitous complementarity. Complementary sites were also identified for 10 cohorts of 16 randomly permuted sequences that had identical sizes and base compositions to the authentic miRNAs. One mismatch was added to the sites complementary to the randomly permuted versions of miR158. Analogous searches for animal miRNA complementary sites queried annotated mRNAs in the *D. melanogaster* genome (GenBank October 2000 release) and annotated coding regions in the *C. elegans* genome (GenBank April 1999 release).

Identification of Homologous miRNA Complementary Sites in *Oryza* mRNAs

For each *Arabidopsis* target mRNA, the mRNAs of up to ten homologous *Oryza* proteins were predicted from the unannotated *Oryza* contigs (Yu et al., 2002) by GenomeScan, a program that identifies genes within genomic sequence using homology to input protein sequences combined with an ab initio gene-finding algorithm (Yeh et al., 2001). Complementary sites in this data set were identified by PatScan searches, and homology to the *Arabidopsis* targets was confirmed by alignment of the inferred protein sequences (ClustalX). One additional target homolog (TC79868) was found by searching the TIGR Rice Gene Index (9.0). For the control study, the identical GenomeScan/PatScan procedure was applied to the 44 *Arabidopsis* mRNAs with sites complementary (allowing up to three mismatches) to the 160 sequences in the 10 cohorts of randomized miRNAs.

References

Aida, M., Ishida, T., Fukaki, H., Fujisawa, H., and Tasaka, M. (1997). Genes involved in organ separation in *Arabidopsis*: an analysis of the *cup-shaped cotyledon* mutant. Plant Cell *9*, 841–857.

Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature *408*, 796–815.

Bernstein, E., Denli, A.M., and Hannon, G.J. (2001). The rest is silence. RNA *7*, 1509–1521.

Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M., and Benning, C. (1998). AGO1 defines a novel locus of *Arabidopsis* controlling leaf development. EMBO J. *17*, 170–180.

Bolle, C., Koncz, C., and Chua, N.H. (2000). PAT1, a new member of the GRAS family, is involved in phytochrome A signal transduction. Genes Dev. *14*, 1269–1278.

Cui, Y., Brugière, N., Jackman, L., Bi, Y.-M., and Rothstein, S. (1999). Structural and transcriptional comparative analysis of the *S* locus regions in two self-incompatible *Brassica napus* lines. Plant Cell *11*, 2217–2231.

Di Laurenzio, L., Wysocka-Diller, J., Malamy, J.E., Pysh, L., Helariutta, Y., Freshour, G., Hahn, M.G., Feldmann, K.A., and Benfey, P.N. (1996). The SCARECROW gene regulates an asymmetric cell division that is essential for generating the radial organization of the *Arabidopsis* root. Cell *86*, 423–433.

Dsouza, M., Larsen, N., and Overbeek, R. (1997). Searching for patterns in genomic data. Trends Genet. *13*, 497–498.

Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W., and Tuschl, T. (2001). Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. EMBO J. *20*, 6877–6888.

Gocal, G.F., Sheldon, C.C., Gubler, F., Moritz, T., Bagnall, D.J., MacMillan, C.P., Li, S.F., Parish, R.W., Dennis, E.S., Weigel, D., and King, R.W. (2001). GAMYB-like genes, flowering, and gibberellin signalling in *Arabidopsis*. Plant Physiol. *127*, 1682–1693.

Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G., and Mello, C.C. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. Cell *106*, 23–34.

Ha, I., Wightman, B., and Ruvkun, G. (1996). A bulged *lin-4/lin-14* RNA duplex is sufficient for *Caenorhabditis elegans lin-14* temporal gradient formation. Genes Dev. *10*, 3041–3050.

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

Helariutta, Y., Fukaki, H., Wysocka-Diller, J., Nakajima, K., Jung, J., Sena, G., Hauser, M.T., and Benfey, P.N. (2000). The SHORT-ROOT gene controls radial patterning of the *Arabidopsis* root through radial signaling. Cell *101*, 555–567.

Hutvágner, G., and Zamore, P.D. (2002). RNAi: nature abhors a double-strand. Curr. Opin. Genet. Dev. *12*, 225–232.

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

Jacobsen, S.E., Running, M.P., and Meyerowitz, E.M. (1999). Disruption of an RNA helicase/RNAseIII gene in *Arabidopsis* causes unregulated cell division in floral meristems. Development *126*, 5231–5243.

Ketting, R.F., Fischer, S.E.J., Bernstein, E., Sijen, T., Hannon, G.J., and Plasterk, R.H.A. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. Genes Dev. *15*, 2654–2659.

Kiss, T. (2002). Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. Cell *109*, 145–148.

Klein, J., Saedler, H., and Huijser, P. (1996). A new family of DNA binding proteins includes putative transcriptional regulators of the *Antirrhinum majus* floral meristem identity gene *SQUAMOSA*. Mol. Gen. Genet. *250*, 7–16.

Knight, S.W., and Bass, B.L. (2001). A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. Science *293*, 2269–2271.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. Science *294*, 853–858.

Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., and Tuschl, T. (2002). Identification of tissue-specific microRNAs from mouse. Curr. Biol. *12*, 735–739.

Lai, E.C. (2002). MicroRNAs are complementary to 3′UTR motifs that mediate negative post-transcriptional regulation. Nat. Genet. *30*, 363–364.

Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. Science *294*, 858–862.

Lee, R.C., and Ambros, V. (2001). An extensive class of small RNAs in *Caenorhabditis elegans*. Science *294*, 862–864.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. Cell *75*, 843–854.

Liscum, E., and Reed, J.W. (2002). Genetics of Aux/IAA and ARF action in plant growth and development. Plant Mol. Biol. *49*, 387–400.

Llave, C., Kasschau, K.D., Rector, M.A., and Carrington, J.C. (2002). Endogenous and silencing-associated small RNAs in plants. Plant Cell *14*, 1605–1619.

Matzke, M.A., Matzke, A.J., Pruss, G.J., and Vance, V.B. (2001). RNA-based silencing strategies in plants. Curr. Opin. Genet. Dev. *11*, 221–227.

McConnell, J.R., Emery, J., Eshed, Y., Bao, N., Bowman, J., and Barton, M.K. (2001). Role of PHABULOSA and PHAVOLUTA in determining radial patterning in shoots. Nature *411*, 709–713.

McManus, M.T., Petersen, C.P., Haines, B.B., Chen, J., and Sharp, P.A. (2002). Gene silencing using micro-RNA designed hairpins. RNA *8*, 842–850.

Moss, E.G., Lee, R.C., and Ambros, V. (1997). The cold shock domain protein LIN-28 controls developmental timing in *C. elegans* and is regulated by the *lin-4* RNA. Cell *88*, 637–646.

Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M., and Dreyfuss, G. (2002). miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes Dev. *16*, 720–728.

Olsen, P.H., and Ambros, V. (1999). The *lin-4* regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. Dev. Biol. *216*, 671–680.

Peng, J., Carol, P., Richards, D.E., King, K.E., Cowling, R.J., Murphy, G.P., and Harberd, N.P. (1997). The *Arabidopsis* GAI gene defines a signaling pathway that negatively regulates gibberellin responses. Genes Dev. *11*, 3194–3205.

Ray, A., Lang, J.D., Golden, T., and Ray, S. (1996a). *SHORT INTEGUMENT (SIN1)*, a gene required for ovule development in *Arabidopsis*, also controls flowering time. Development *122*, 2631–2638.

Ray, S., Golden, T., and Ray, A. (1996b). Maternal effects of the *short integument* mutation on embryo development. Dev. Biol. *180*, 365–369.

Reinhart, B.J., Slack, F.J., Basson, M., Bettinger, J.C., Pasquinelli, A.E., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. (2000). The 21 nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. Nature *403*, 901–906.

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., and Bartel, D.P. (2002). MicroRNAs in plants. Genes Dev. *16*, 1616–1626.

Riechmann, J.L., Heard, J., Martin, G., Reuber, L., Jiang, C.-Z., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O.J., Samaha, R.R., et al. (2000). *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. Science *290*, 2105–2110.

Robinson-Beers, K., Pruitt, R.E., and Gasser, C.S. (1992). Ovule development in wild-type *Arabidopsis* and two female-sterile mutants. Plant Cell *4*, 1237–1249.

Rogg, L.E., and Bartel, B. (2001). Auxin signaling: derepression through regulated proteolysis. Dev. Cell *1*, 595–604.

Silverstone, A.L., Ciampaglio, C.N., and Sun, T. (1998). The *Arabidopsis* RGA gene encodes a transcriptional regulator repressing the gibberellin signal transduction pathway. Plant Cell *10*, 155–169.

Slack, F.J., Basson, M., Liu, Z., Ambros, V., Horvitz, H.R., and Ruvkun, G. (2000). The *lin-41* RBCC gene acts in the *C. elegans* heterochronic pathway between the *let-7* regulatory RNA and the LIN-29 transcription factor. Mol. Cell *5*, 659–669.

Ulmasov, T., Hagen, G., and Guilfoyle, T.J. (1999). Dimerization and DNA binding of auxin response factors. Plant J. *19*, 309–319.

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. Cell *75*, 855–862.

Yeh, R.F., Lim, L.P., and Burge, C.B. (2001). Computational inference of homologous gene structures in the human genome. Genome Res. *11*, 803–816.

Yu, J., Hu, S., Wang, J., Wong, G.K., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science *296*, 79–92.