The Plant Journal (2007) **51**, 991–1003 doi: 10.1111/j.1365-313X.2007.03197.x

Large-scale identification of microRNAs from a basal eudicot (*Eschscholzia californica*) and conservation in flowering plants

Abdelali Barakat[1,*], Kerr Wall[1], Jim Leebens-Mack[1,†], Yunjiao J. Wang[1], John E. Carlson[2] and Claude W. dePamphilis[1,*]
[1]*Department of Biology, Institute of Molecular Evolutionary Genetics, and The Huck Institutes of the Life Sciences, 403 Life Sciences Building, The Pennsylvania State University, University Park, PA 16802, USA, and*
[2]*The School of Forest Resources, Department of Horticulture, and The Huck Institutes for Life Sciences, Pennsylvania State University, 323 Forest Resources Building, University Park, PA 16802, USA*

Received 2 November 2006; revised 24 April 2007; accepted 9 May 2007.
*For correspondence (fax 814 865 9131; e-mail cwd3@psu.edu or aub14@psu.edu.
†Present address: Department of Plant Biology, University of Georgia Athens, GA 30602-7271, USA.

Summary

MicroRNAs (miRNAs) negatively control gene expression by cleaving or inhibiting the translation of mRNA of target genes, and as such, they play an important role in plant development. Of the 79 plant miRNA families discovered to date, most are from the fully sequenced plant genomes of Arabidopsis, Populus and rice. Here, we identified miRNAs from leaves, roots, stems and flowers at different developmental stages of the basal eudicot species *Eschscholzia californica* (California poppy) using cloning and capillary sequencing, as well as ultrahigh-throughput pyrosequencing using the recently introduced 454 sequencing method. In total, we identified a minimum of 173 unique miRNA sequences belonging to 28 miRNA families and seven *trans*-acting small interfering RNAs (ta-siRNAs) conserved in eudicot and monocot species. miR529 and miR537, which have not yet been reported in eudicot species, were detected in California poppy; loci encoding these miRNAs were also found in Arabidopsis and Populus. miR535, which occurs in the moss *Physcomitrella patens*, was also detected in California poppy, but not in other angiosperms. Several potential miRNA targets were found in cDNA sequences of California poppy. Predicted target genes include transcription factors but also genes implicated in various metabolic processes and in stress defense. Comparative analysis of miRNAs from plants of phylogenetically-critical basal lineages aid the study of the evolutionary gains and losses of miRNAs in plants as well as their conservation, and lead to discoveries about the miRNAs of even well-studied model organisms.

Keywords: ***Eschscholzia californica*, small RNA, microRNA, distribution, high-throughput sequencing.**

Introduction

MicroRNAs (miRNAs) are small non-coding (20–25 nucleotides) RNAs that play a key role in negative control of gene expression, by inhibiting the translation of target genes (Palatnik *et al.*, 2003; Tang *et al.*, 2003) or by cleaving their target mRNAs (Aukerman and Sakai, 2003; Reinhart *et al.*, 2002). As such, miRNAs play an important role in plant development and other physiological processes (for reviews, see Carrington and Ambros, 2003; Jones-Rhoades *et al.*, 2006). For example, in plants the overexpression of different miRNAs results in different developmental defects, such as enlarged boundary size in Arabidopsis meristems (Laufs *et al.*, 2004), early flowering time and lack of specification of floral organ identity (Aukerman and Sakai, 2003), uneven leaf shape and curvature (Palatnik *et al.*, 2003), male sterility (Achard *et al.*, 2004), flower organ fusion (Laufs *et al.*, 2004), seedling arrest and female sterility (Williams *et al.*, 2005), and phosphate accumulation (Aung *et al.*, 2006; Bari *et al.*, 2006; Chiou *et al.*, 2006; Fujii *et al.*, 2005). miRNAs are transcribed as long precursor transcripts, ranging from 70 to 300 nucleotides (nt), that have the capacity to form a fold-back structure. Precursors are processed by the DCL1 endonuclease that generates an miRNA:miRNA* duplex, and then by a helicase that releases mature miRNAs (for a review see Bartel, 2004).

Numerous studies contributed to the identification of miRNAs in Arabidopsis and rice (Adai *et al.*, 2005; Bonnet *et al.*, 2004; Floyd and Bowman, 2004; Jones-Rhoades and Bartel, 2004; Li and Zhang, 2005; Li *et al.*, 2005; Llave *et al.*, 2002; Mette *et al.*, 2002; Palatnik *et al.*, 2003; Park *et al.*, 2002; Reinhart *et al.*, 2002; Sunkar *et al.*, 2005; Wang *et al.*, 2004). miRNAs have also been identified in other plant species, such as the diverse eudicot species *Nicotiana tabacum* (Billoud *et al.*, 2005), *Populus trichocarpa* (Lu *et al.*, 2005a,b; Tuskan *et al.*, 2006), *Medicago truncatula*, *Glycine max*, the grasses *Saccharum officinarum* and *Zea mays* (Dezulian *et al.*, 2005), *Sorghum bicolor* (Bedell *et al.*, 2005) and the moss *Physcomitrella patens* (Arazi *et al.*, 2005). About 71 plant miRNA families have been identified so far, using criteria that include expression, evolutionary conservation, secondary structure and genetic requirements for biogenesis (MIRBASE, release 8; Griffiths-Jones *et al.*, 2006), with half of them from the fully sequenced genomes of Arabidopsis, Populus and rice. Although no miRNAs conserved between animals and plants have been reported so far, miRNAs can be well conserved among distantly related plant species. Indeed, several miRNAs are conserved among angiosperm and gymnosperms species, and some, such as miR165/166, are conserved even among angiosperms, ferns, lycopods and mosses (Arazi *et al.*, 2005; Axtell and Bartel, 2005; Floyd and Bowman, 2004), indicating that the miRNA-mediated gene regulation process is very ancient, and almost certainly existed in the common ancestor of all land plants.

Most miRNAs in plants have been identified by similarity with Arabidopsis and rice homologs. However, not all miRNA genes are conserved among species from different lineages, and some of them seem to be species or clade specific (Axtell and Bartel, 2005). It has been demonstrated that miR158, miR161, miR163, miR173 and miR403 exist in Arabidopsis, but no homolog has been found in rice (Jones-Rhoades and Bartel, 2004; Sunkar and Zhu, 2004). Several miRNA families have been demonstrated to be specific to Populus (Lu *et al.*, 2005a). Furthermore, most miRNAs that have been documented so far are from one family of monocots (grasses) and a few core eudicots from the rosid lineage (Arabidopsis and Populus). To compare the diversity of miRNAs in angiosperm species and analyze their evolution, miRNA sampling across a wide range of different angiosperm lineages is needed.

Unlike animals, the targets of plant miRNAs tend to have a high degree of complementarity with their target miRNA (Jones-Rhoades and Bartel, 2004). This has made the search for plant miRNA target genes a straightforward process. Targets predicted by computational analysis can often be confirmed using different experimental methods, such as 5′ rapid amplification of cDNA ends (5′ RACE) (Jones-Rhoades and Bartel, 2004; Llave *et al.*, 2002), northern hybridization or transformation approaches (Kasschau *et al.*, 2003; Llave *et al.*, 2002). Previous studies (Jones-Rhoades *et al.*, 2006) have shown that plant miRNAs frequently target mRNAs encoding transcription factors involved in different developmental processes, as well as several non-transcription factor genes.

Here we report the identification of miRNAs from the basal eudicot and emerging model species, *Eschscholzia californica* (California poppy) (Becker *et al.*, 2005; Carlson *et al.*, 2006). California poppy belongs to the order Ranunculales, which holds a key position in the angiosperm phylogeny as the sister lineage to the rest of the eudicots (Leebens-Mack *et al.*, 2005). As a result, this plant species is a good choice for comparing the diversity of miRNAs between the core eudicots and monocots, and for helping to interpret the differences among various eudicot species. Furthermore, many resources such as 5713 cDNA unigenes (Albert *et al.*, 2005; Carlson *et al.*, 2006), and 32 000 cDNA contigs derived from pyrosequencing (K. Wall, A. Barakat, J. Leebens-Mack, H. Ma, E.J. Carlson, S. Schuster and C. dePamphilis, unpublished data), were available, making the search for target genes feasible. We used a standard small RNA cloning approach (Lau *et al.*, 2001) and capillary sequencing, as well as high-throughput pyrosequencing (Margulies *et al.*, 2005), to isolate miRNAs. This study enabled us to identify 173 conserved unique miRNA sequences, belonging to 28 families, and seven *trans*-acting siRNA (ta-siRNAs), to compare their distribution in land plants, and to predict their targets.

Results

Cloning of miRNAs from different tissues of California poppy plants

A total of 377 concatemerized clones from the small RNA library from poppy flower buds were sequenced using capillary sequencing. Splitting small RNA (sRNA) concatamers, using adapter sequences that were used for their cloning, enabled the identification of 754 cloned sequences. Sequence sorting based on size showed that 433 sRNAs were greater than 17 nt in length. BLAST against sequences from snoRNA, rRNA and tRNA databases allowed us to remove contaminant rRNA, snoRNA and tRNA sequences. By querying the MIRBASE with the filtered sRNA sequences using PATSCAN (Dsouza *et al.*, 1997), and accepting no more than two substitutions, we identified 19 miRNA sequences (Table S1) belonging to seven previously identified families (miR159, miR167, miR171, miR172, miR319, miR396 and miR398). The isolation frequency of the identified miRNAs varied between one and six, and the most represented are miR167 and miR172 (four and six sequences, respectively).

454 sequences analysis

Pyrosequencing of a half plate of sRNA concatamers from different tissues of California poppy plants generated 166 835 reads. The isolation of each single sRNA sequence by splitting sRNA concatamers enabled the identification of nearly 300 000 cloned sequences. A total of 119 046 reads that were either shorter than 17 nt or longer than 30 nt, which correspond to concatamers that could not be separated or to likely RNA degradation products, were discarded. We also discarded sequences corresponding to other contaminant RNAs by BLASTn (Altschul *et al.*, 1990) against snoRNA, rRNA and tRNA databases. A total of 42 793 sRNAs with sizes between 17 and 30 nt remained for further consideration in the study. The size distribution of sRNA ranging between 17 and 30 nt (Figure 1) shows that 21-nt sRNA is the most abundant. A PATSCAN (Dsouza *et al.*, 1997) search of known miRNAs on sRNA sequences of 17–30 nt in length enabled the identification of 7450 sequence reads that were identical or highly similar (two nucleotide mismatches or less) to known plant miRNAs. After removing redundant and incomplete sequences, we isolated 173 unique sequences belonging to 28 known families, and seven ta-siRNAs (Table S1) homologous with ta-siR2141 and ta-siR2142 (Allen *et al.*, 2005). Based on the annotation criteria of Ambros *et al.* (2003), all miRNAs identified with this process fulfill the expression and the phylogenetic conservation criteria (see below), and should therefore be considered as miRNAs.

An additional 6974 sequence reads ranging between 20 and 24 nt were isolated at least twice – these reads represented 1780 unique sequences. These sequences could correspond either to siRNAs or miRNAs. To determine whether these sequences correspond to miRNAs, we searched cDNA unigenes and 454 pyrosequencing data from California poppy for probable primary transcripts corresponding to these sequences. However, we failed to identify any primary transcripts with a significant free energy value (< –30). This does not necessarily mean that many of these novel, relatively abundant small RNA species are not miRNAs – miRNA precursors are frequently underrepresented in cDNA libraries, probably as a result of rapid processing in the nucleus. We also searched for their homologs in the genome sequences of Arabidopsis, Populus and rice using PATSCAN (0.0.0). Several hits for these miRNA sequences were found, but none passed the MIRCHECK filter.

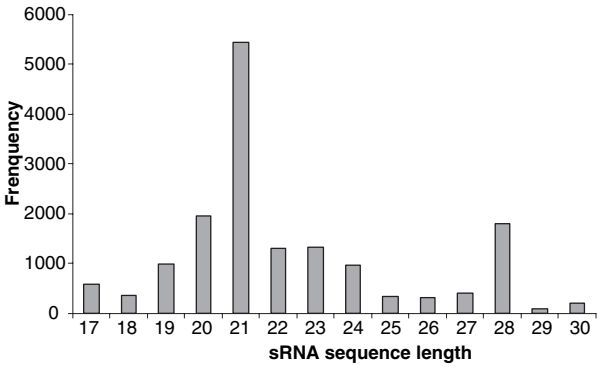

Figure 1. Length distribution of *Eschscholzia californica* (California poppy) small RNA (sRNA) in pyrosequencing data.

Identification of California poppy miRNA and their families

Comparison of identified miRNA families showed that the number of unique sequences (miRNA variants) varies widely between miRNA families (Table S1, Table 1). Some of these variants may be distinct family members, but because an miRNA variant could be encoded by multiple paralogous loci in the genome, allelic variation or errors in the primary sequence data, we will be referring to these unique sequences simply as 'variants' rather than family members. miR172 and miR535, for example, are represented by 27 and 1 variant, respectively. All the miRNA families, except miR162, miR395 and miR535, are represented by more than two variants. Three families (miR166, miR171 and miR172)

Table 1 Number of members per microRNA (miRNA) family identified in Arabidopsis, rice, Populus and the number of miRNA variants detected in *Eschscholzia californica* (California poppy)

Family	Arabidopsis	Rice	Populus	Poppy
miR156	7	12	11	
miR157	4	0	0	5
miR159	2	6	6	10
miR160	3	6	8	5
miR162	2	2	3	1
miR164	3	6	6	2
miR165	2	0	0	5
miR166	7	14	17	25
miR167	4	10	8	9
miR168	2	2	2	4
miR169	15	8	32	6
miR171	1	9	10	23
miR172	6	4	9	27
miR319	3	2	9	6
miR390	2	1	4	3
miR393	2	2	4	2
miR394	2	1	2	5
miR395	6	19	10	1
miR396	2	5	7	10
miR397	2	2	3	7
miR398	3	2	3	8
miR399	6	11	12	2
miR408	1	1	1	2
miR529	1	1	1	2
miR535	0	1	0	1
miR537	1	1	1	1
miR845	2	0	0	1
miR1220	0	0	0	1

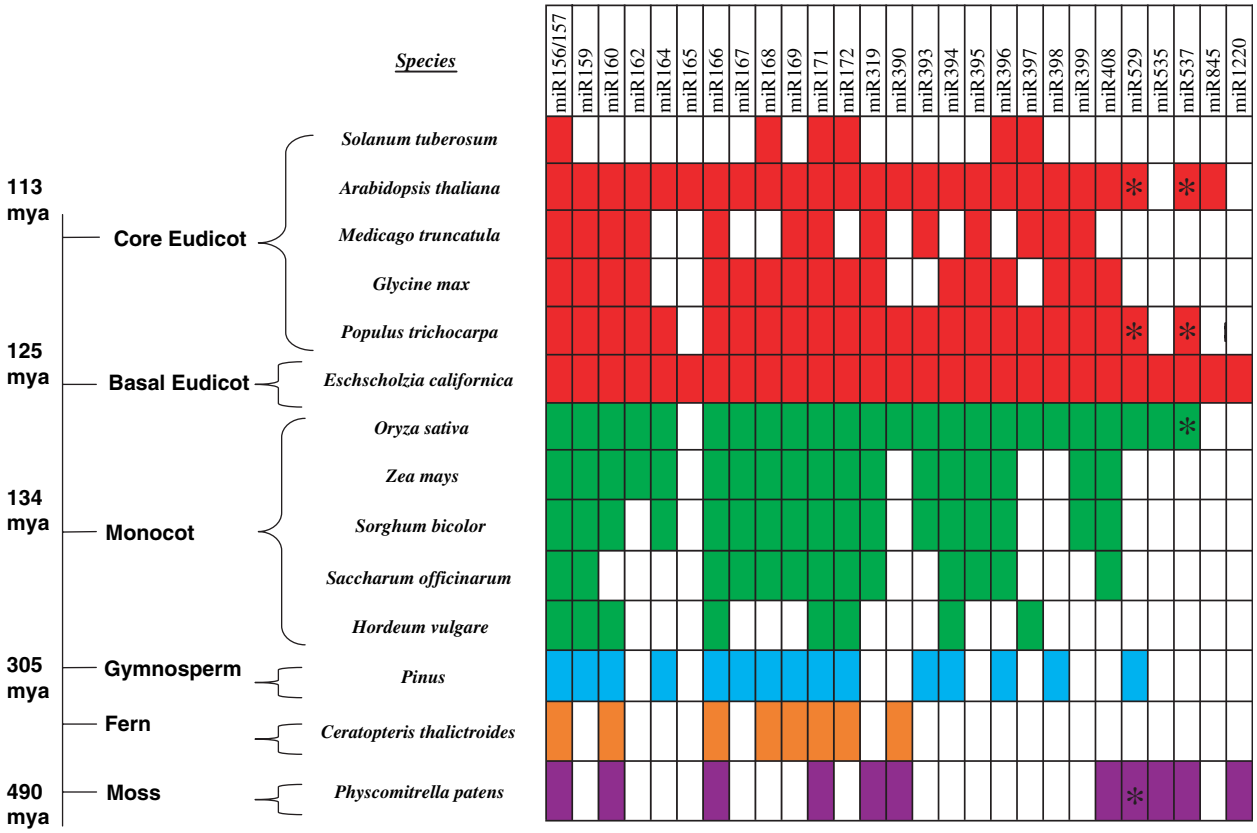

Figure 2. microRNA (miRNA) distribution in plant species. miRNA data are a compilation from MIRBASE, Zhang *et al.* (2005, 2006), Axtell and Bartel (2005), M. J. Axtell and D. P. Bartel (personal communication), and this study. Gymnosperm miRNAs are from different Pinus species: *Pinus taeda*, *P. resinosa* and *P. pinaster*. Divergence dates for land plant clades is from Leebens-Mack *et al.* (2005). *miRNAs discovered in Arabidopsis and Populus following comparative analysis with *Eschscholzia californica* (California poppy).

are represented by more than 20 variants. miR159, miR170 and miR396 are represented by 10–20 variants. Most of the other families (73%) are represented by between two and 10 variants. We found that 24 miRNA families have homologs in both eudicot and the monocot species (Fig. 2). Members of five miRNA families (miR156/157, miR160, miR319 and miR390) were found in monocots, eudicots and bryophytes. Two new miRNAs (named Eca-miR1 and Eca-miR2) were identified; they are conserved in Arabidopsis, but not in Populus and rice.

Three of the miRNAs identified in this study (miR529, miR535 and miR537) present a striking phylogenetic distribution (Fig. 2). miR529 and miR537 were found in California poppy and either Physcomitrella or rice, but no homolog has been reported in core eudicots, which include Arabidopsis and Populus. A search of these miRNAs in Arabidopsis and Populus genomic sequences using PATSCAN, and tolerating two substitutions, did not reveal any similar sequences. By lowering the stringency to three substitutions, we found several potential sequences similar to miR529 and miR537 both in Arabidopsis and in Populus. Significant hits were also found for miR537 in rice, where it has not been found so far. Folding their surrounding genomic sequences (Figure 3) using MFOLD (Zuker, 2003), and analyzing the folded sequences for miRNA features using the MIRCHECK program (Jones-Rhoades and Bartel, 2004) confirmed that these sequences correspond to true miRNAs. miR535 was found in California poppy, rice and Physcomitrella, but not in Arabidopsis and Populus. A search of these miRNAs under the same criteria described above revealed the existence of several potential homologs in Populus and Arabidopsis. However, none of the surrounding genomic sequences for these hits passed the MIRCHECK filter.

California poppy miRNA expression analyses

In order to estimate the relative abundance of the identified miRNA families in California poppy, we examined the relative frequencies of isolation during library sequencing (Figure 4). Analyses of miRNA expression based on the number of sequences detected in the 454 data showed that expression varies widely between families. Overall, the isolation frequencies ranged from 1 (miR395) to 972

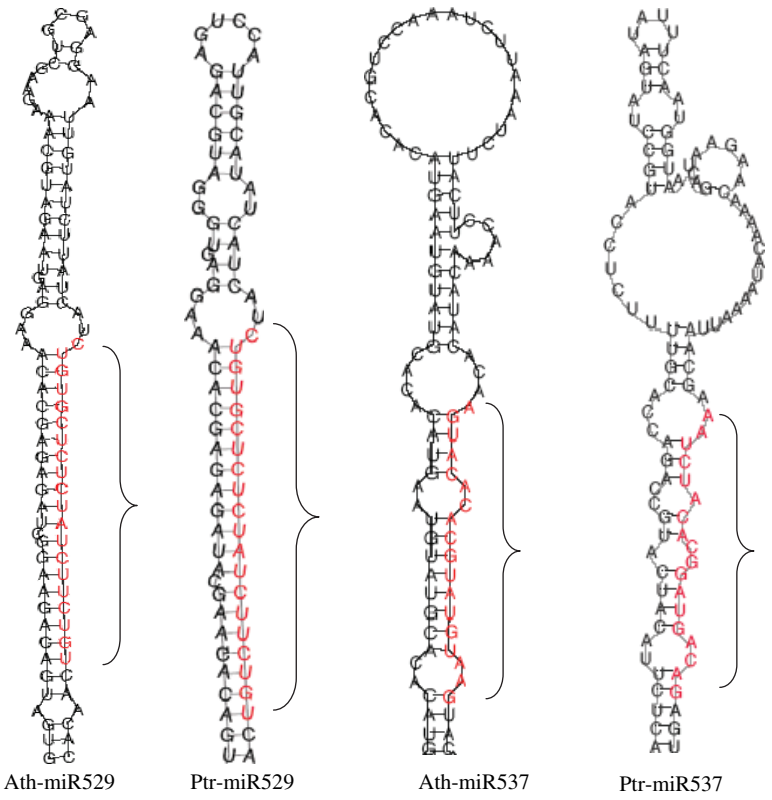

Figure 3. Predicted secondary structure of miR529 and miR537 from Arabidopsis and Populus.

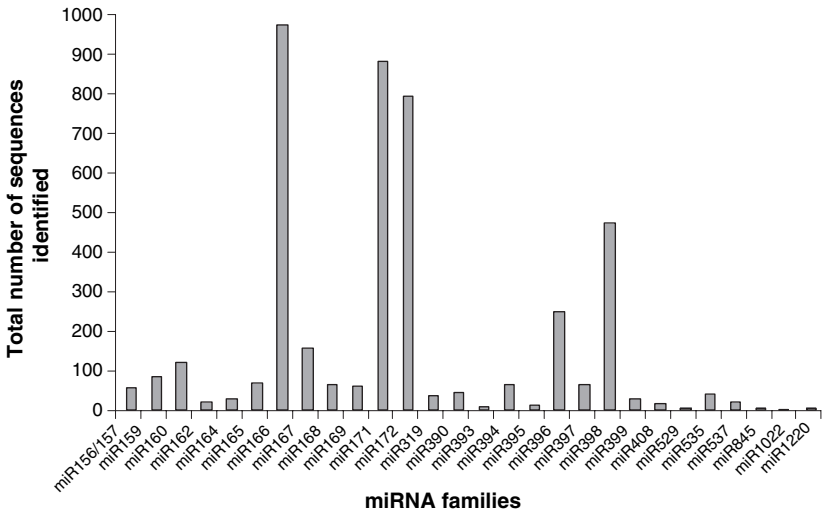

Figure 4. Expression of microRNA (miRNA) families based on the 454 sequences.

(miR166). Some miRNA families (miR160, miR166, miR167, miR171, miR172, miR396 and miR398) are highly expressed, and are represented by more than 100 sequences each. miR156/157, miR159, miR162, miR164, miR165, miR168, miR169, miR319, miR390, miR394, miR395, miR397, miR399, miR408 and miR535 are moderately expressed, and are represented by a number of sequences varying between 10 and 100. The third category, which is lowly expressed (<10 sequences), is represented by miR393 and miR529. The expression also varies widely between variants from the same miRNA family, although there is a general pattern that smaller, less conserved families also tend to be less highly expressed. For example, variants from the phylogenetically ancient miR166 family are represented by a number of sequences varying between two and 725, with five variants being detected 10 or more times.

To confirm the expression of the miRNAs cloned, and assess their potential for playing a role in regulating the

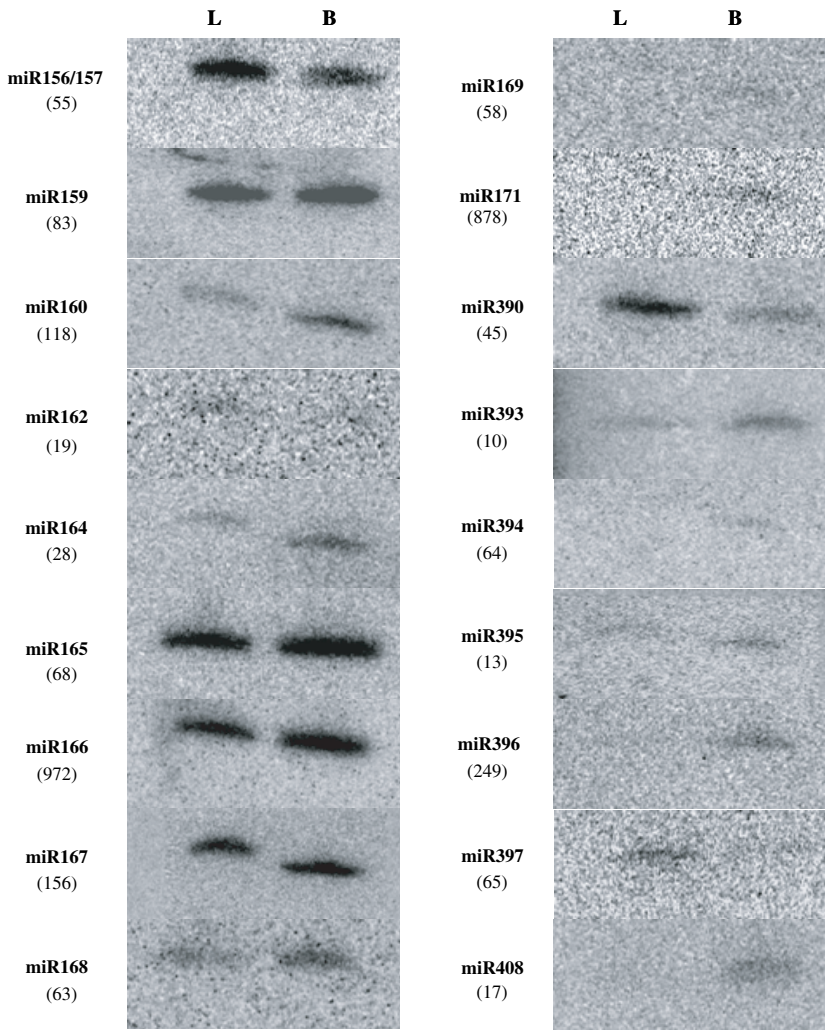

Figure 5. Expression analyses of microRNAs (miRNAs) using Northern blot. Numbers between brackets indicate the number of sequences identified in the 454 library for each miRNA family analyzed. L and B: leaves and buds, respectively.

expression of floral genes, we analyzed the expression of a sample of 25 miRNA sequences (14% of all unique miRNA sequences identified) belonging to 18 families (66% of the families identified) using northern hybridization in flower buds and leaves of California poppy. Expression analysis (Figure 5) showed that 18 (72%) of the 25 miRNAs analyzed present detectable hybridization signal in leaves, flower buds or in both. miRNA families that are represented by a large number of sequences in the 454 library also showed a high level of expression in northern experiments. Most miRNAs analyzed are expressed in leaves and flower buds. However, some miRNAs such as miR169, miR171 and miR408 did not show a detectable hybridization signal in leaves. Some miRNAs, such as miR390 and miR397, are highly expressed in leaves, whereas others (miR160 miR164, miR168, miR393, miR394, miR395, miR396 and miR408) are highly expressed in flower buds, indicating that these miRNAs could function as regulators of floral genes. For seven of the 25 sequences analyzed, we did not detect any expression in both leaves and flower buds. Previous studies (Adai *et al.*, 2005; Jones-Rhoades and Bartel, 2004; Lu *et al.*, 2005a; Sunkar and Zhu, 2004) showed that some cloned or predicted miRNAs that are expressed at a low level could not be detected by northern hybridization, and need to be studied by RT-PCR.

Target prediction

The miRNA sequences were searched against California poppy unigenes and 454 contigs to identify putative targets. Out of the 173 miRNAs analyzed, 59 (34%) and 89 (52%) have at least one potential target in California poppy and Arabidopsis, respectively. 14 of the 59 target sequences found in California poppy could be annotated using the Arabidopsis proteome (Blastx, $< e^{-5}$); this set encodes six genes (Table 2). The number of potential targets in Arabidopsis range from between one to six targets per miRNA family (Table 2). Predicted targets in Arabidopsis include genes from different transcription factor

Table 2 Putative target genes of conserved microRNAs (miRNAs) identified in Arabidopsis and *Eschscholzia californica* (California poppy)

miRNA family	Target gene in Arabidopsis	Target gene in California poppy
miR156/miR157	Squamosa promoter-binding protein-like	X
miR159	Myb family transcription factor	
miR160	Transcription factor B3 family protein	X
miR164	Transcription activator NAC1 (NAC1)	
	Glycosyltransferase family protein 1	
miR165	Homeobox-leucine zipper transcription factor	
	Inositol monophosphatase family protein	
miR166	Homeobox-leucine zipper transcription factor	
miR167	Auxin-responsive factor (ARF8)	
miR168	Argonaute protein (AGO1 YA) family protein	X
miR169	CCAAT-binding transcription factor (CBF-B/NF-YA) family protein	
miR171	Scarecrow transcription factor	
	Sodium-inducible calcium-binding protein (ACP1)	
miR172	bZIP transcription factor family protein	
	Glycosyl transferase family 48 protein	
	Expressed protein; expression supported by MPSS	
	Floral homeotic protein APETALA2 (AP2)	
	F-box family protein (FKF1)	X
	Meprin and TRAF homology domain-containing protein	
miR319	Myb family transcription factor	
miR393	F-box family protein (E3 ubiquitin ligase)	X
miR394	F-box family protein	X
miR396	Cysteine proteinase	
	Nitrilase 1 (NIT1)	
	Endomembrane protein 70	
	Root hair initiation protein root hairless 1 (RHL1)	
	Disease resistance protein (TIR-NBS-LRR class)	
	Ribosomal protein L13 family protein	
miR397	Laccase	
miR399	Ubiquitin-conjugating enzyme family protein	
miR408	Expressed protein	
	Peptide chain release factor	
	Plastocyanin-like domain-containing protein	
	Laccase	
miR529	Zinc finger (C3HC4-type RING finger) family protein	
	Cullin	
	Expressed protein	
	PWWP domain-containing protein	
	U-box domain-containing protein	
ta-siR2141	Transcription factor B3 family protein/auxin-responsive factor	
ta-siR2142	Transcription factor B3 family protein/auxin-responsive factor	

For miR535, the target was predicted from rice.

families such as *MYB*, *SQUAMOSA, NAC*, *AP2*, *zinc finger*, *homeobox* and *B3*), as well as numerous non-transcription factor encoding genes involved in different metabolic processes and stress defense functions (Table 2). Among the targets identified in California poppy, five have not been reported so far. These potential target genes are encoding proteins involved in signaling (*phosphatidylinositol 3- and 4-kinase*), phosphate transport (*phosphate transporter*), glucose synthesis (*aconitate hydratase*), solute transports (*pyrophosphate-energized vacuolar membrane proton pump*) and protein biosynthesis (*elongation factor Tu*). In contrast to rice, where no target has been identified, we identified five genes targeted by miR529 in poppy. miR172, which has been demonstrated to control the expression of only Apetala-like genes (*AP2*), may target in California poppy three other putative genes (*phosphatidylinositol 3-* and *4-kinase*, *phosphate transporter* and *Aconitate hydratase*) (Figure 6).

miR172: Aconitate hydrase

```
miR172          5′AGAAUCUUGAUGAUGUUGCAC3′
                  |  ||||||||||||||||||
Contig29870     3′ACCUAGAACUACUACAACUUG5′
```

miR172: Phosphatidylinositol 3- and 4-kinase family protein

```
miR172          5′AGAAUCUUGAUGAUGCUGUC3′
                  ||||||o|||||||| |  ||
Contig28818     3′UCUUAGGACUACUACUAGAG5′
```

miR172: Phosphate transporter

```
miR172          5′AGAAUCUUGAUGAUGCUGCAC3′
                  ||||  |||||||||  ||o|||
Eca_5_340412    3′UCUUCGAACUACUAAGAUGUG5′
```

Figure 6. Eca-miR172 complementary sequences in three *Eschscholzia californica* (California poppy) putative target genes.

Discussion

Comparison of the 454 sequencing and the ABI sequencing

The sequencing of 377 clones using the ABI machine allowed the identification of 22 known miRNAs belonging to eight families. With the 454 sequencing, we found 7800 conserved miRNAs corresponding to 173 unique sequences belonging to 28 miRNA families. We also identified seven sequences from two ta-siRNA families. All the families identified using di-deoxy sequencing were also found in the pyrosequencing-derived reads. The 28 miRNA families identified by pyrosequencing correspond to 60% of the miRNA families identified so far in Arabidopsis, and 84% of the miRNA families conserved between rice and Arabidopsis (Jones-Rhoades *et al.*, 2006). Two of the miRNA families have not been reported so far in Arabidopsis and Populus. Orthologs of 19 Arabidopsis miRNA families were not isolated from California poppy. They could correspond to lowly expressed miRNAs in the tissues and developmental stages analyzed in this study, miRNAs expressed in developmental stages or physiological conditions not represented in our libraries, or miRNAs that are not conserved between Arabidopsis and poppy. We also identified a large number of other small RNAs that cannot currently be classified as miRNAs, siRNAs or as degradation products of other RNAs. By sequencing a half 454 plate, we also identified a large number of variants from the same families. Some of these may have resulted from sequencing errors generated by the 454 sequencing (Margulies *et al.*, 2005; Moore *et al.*, 2006), but many (99%) of the variants were captured repeatedly, and only rarely did they contain mononucleotide runs that are the expected source of most 454 sequencing errors (Margulies *et al.*, 2005; Moore *et al.*, 2006). For example, the number of variants per family varies between one and 29, with most families having more than 10 variants. The miRNAs identified are represented by a number of sequences varying between one and nearly 1000 sequences in a family. Northern hybridization experiments showed that miRNAs with a high and low number of 454 sequences present high and low hybridization signals, respectively. These observations all suggest that the 454 sequencing is an efficient tool for isolating miRNAs, studying their diversity and analyzing their expression.

Distribution of California poppy miRNAs in land plants

Previous studies identified and confirmed experimentally the presence of 62 miRNA families in Arabidopsis (MIRBASE, release 9). In this work we identified 28 families conserved in monocot, eudicot or moss species, or at least in two of them. All miRNA families identified, except miR535, miR395 and miR162, are represented by a number of variants ranging between two and 29. Three families (miR166, miR171 and miR172) are represented by more than 20 variants. Assuming each variant is encoded by one locus, the number of miRNAs per family in California poppy, which has a genome size of about 500 Mb (Cui *et al.*, 2006), is similar to the ones identified in Arabidopsis, rice, Populus and maize (Jones-Rhoades *et al.*, 2006; Zhang *et al.*, 2005, 2006), which have genome sizes of 125 (The Arabidopsis Genome Initiative, 2000), 389 (The International Rice Genome Sequencing Project, 2005) and 410 $\pm$ 10 Mb (Tuskan *et al.*, 2006), respectively. This indicates that the expansion or the contraction of the genome of the plants compared did not affect the number of miRNA genes per family. For some miRNAs such as miR399, miR395, miR169 and miR164, we identified one or a few variants compared with Arabidopsis, rice and California poppy. The sRNA library sequenced in this project was prepared from multiple tissues of California poppy plants. Thus, we cannot exclude the possibility of less representation in the sRNA library of tissues, or physiological conditions in which these miRNAs are expressed.

The large number of members per miRNA family raises the question of the biological role of these miRNAs. One hypothesis that could explain why plants retain larger miRNA gene families, is that miRNAs families expanded in parallel with the expansion of the gene families they were targeting, allowing plants to control the expression of newly duplicated genes, and thus adapting them to different environments and different physiological conditions. A second hypothesis is that sequences of members from the same family have diverged slightly, and now target diverged members from the same family or different genes with common protein domains.

Some California poppy miRNA families may not exist in core eudicot species

Most (88%) of California poppy identified miRNA families are present in at least one of the core eudicot and monocot species, suggesting their existence in the ancestor of these two large angiosperm clades. Other miRNA families, such as miR529, miR535 and miR537, and miR319, miR390, miR160 and miR156/157, are conserved in monocots, core eudicots and basal eudicots, as well as moss, indicating that pathways of gene regulation involving these miRNAs are very ancient and evolved some time ago in the ancestor of land plants.

Comparative analysis of miRNA distributions led to the discovery of additional conserved miRNAs in Arabidopsis and Populus, and suggests that otherwise conserved miRNAs may sometimes be lost within some angiosperm lineages, whereas other miRNAs may have evolved recently in relatively narrow lineages. miR529 and miR537 were found in California poppy and either rice or Physcomitrella. However, no ortholog has been reported in Populus or Arabidopsis. By lowering the stringency to three substitutions, we found miR529 and miR537 in Arabidopsis and Populus. Moreover, miR537 was also found in rice (this study), and miR529 was found in Physcomitrella (M. J. Axtell and D. P. Bartel, pers. comm.). These miRNAs have not yet been reported in Arabidopsis, Populus or rice, possibly because they are lowly expressed and need deeper sequencing to be discovered. For instance, expression analyses of miR537 in California poppy buds and leaves showed that they are lowly expressed. The previous failure to find miR529 and miR537 in Arabidopsis and Populus is explainable, because these sequences have diverged from their orthologs beyond the divergence limit (two substitutions) of most miRNA orthologs (Jones-Rhoades and Bartel, 2004) used for interspecies comparison. From a functional point of view, the sequence divergence of miR529 and miR537 in Arabidopsis and Populus could reflect the beginning of a loss process, or it could translate into functional divergence and the evolution of new targets. Target analyses showed that miR529 target different genes in California poppy, however no target was found in rice (Liu *et al.*, 2005) and in Arabidopsis (this study). Thus, we could not determine whether the divergence of miR529 in Arabidopsis might be associated with functional divergence. For miR535, no target has been found in California poppy or Arabidopsis, indicating that this miRNA is either not functional, or that it targets a lowly expressed gene that is not represented in Arabidopsis, Populus and California poppy cDNA libraries. Similarly, miR535 was found in California poppy, rice and Physcomitrella, but had not previously been found in Arabidopsis or Populus, both species in which the genomes are completely sequenced. A search of these miRNAs in Arabidopsis, Populus and rice, even with three substitutions, did not reveal the existence of this miRNA in these plant species. Previous studies (Jones-Rhoades and Bartel, 2004; Lu *et al.*, 2005a) have demonstrated the existence of miRNAs specific to Populus and Arabidopsis, but no miRNAs have been reported to be conserved between moss and only one angiosperm species. It is possible that miR537 has been lost repeatedly in different species; in this case, we should find it at least in some angiosperm species. Alternatively, miR537 may have evolved independently in different lineages. Neither hypothesis can explain why this miRNA family has been lost in core eudicots, or why it has evolved independently in moss and in a basal eudicot (California poppy). The target search of miR537 in California poppy did not reveal any target genes that might have given insight into its function and striking distribution.

Comparative analysis of miRNA distributions identified two other miRNA families with complex histories in flowering plants. miR845 was detected in both Arabidopsis and California poppy, but not Populus or rice. This is consistent with its gain early in eudicot history, followed by its loss in at least the lineage that contains Populus, or with independent gains involving ancestors of California poppy and Arabidopsis. Finally, the presence of miRNA1220 in Physcomitrella and California poppy, but not in the sequenced angiosperms, is consistent either with independent gains in these distantly related species, or with miR1220 as an ancient miRNA in plants that has been lost on multiple occasions within the angiosperms. Diagnosis of these alternative possibilities awaits deep sequencing of miRNA complements in additional species.

miRNA expression indicates that most identified California poppy miRNAs are expressed in flowers

Analyses of miRNA expression, based on the number of corresponding sequence reads in the 454 sequences, showed that the expression varied widely between families, ranging from 1 (miR395) to 972 (miR166) tags. Based on their expression, miRNAs could be divided into three categories: highly expressed, moderately expressed and lowly expressed. Similar to earlier results with Massively Parallel Signature Sequencing (Lu *et al.*, 2005a), expression analyses using northern hybridization showed that the number of sequences obtained by 454 sequencing could reflect the actual level of expression of the miRNAs studied. Indeed, miRNAs represented by a high and low number of 454 sequences showed a high and low hybridization signal. However, no correlation was found between 454 data and northern hybridization for moderately expressed miRNAs. This is in agreement with previous work (Lu *et al.*, 2005a) showing that because of cross hybridization between highly similar members from the same family, the hybridization signal may reflect the expression level of the family rather than a single member. Comparison between flower and

leaves showed that most miRNAs analyzed by northern hybridization are expressed in both leaves and flower buds. However, miR169, miR171 and miR408 did not show any expression signal in leaves. The fact that miR169 and miR171 target two genes (CCAAl binding domain and scarecrow transcription factor) involved in controlling floral development might explain their lower expression in leaves. The failure to detect a hybridization signal of miR408 might also be a result of its lower expression in leaves. Indeed, previous work (Sunkar and Zhu, 2004) demonstrates that miR408 is expressed in rice seedlings, but is weakly expressed in all the other plant tissues. However, we saw a high expression of miR408 in poppy flower buds. Other miRNAs, such as miR160, miR164, miR168, miR393, miR394, miR395, miR396 and miR408, which show high expression levels in flower buds, have previously been shown to target floral genes (Jones-Rhoades *et al.*, 2006).

Target prediction

One half (51%) and one third (35%) of the miRNA sequences identified in our analysis also had target genes predicted in Arabidopsis and California poppy, respectively. The number of targets per miRNA identified in Arabidopsis varies between one and six. These numbers are in the range that has been reported in other studies (Jones-Rhoades and Bartel, 2004). The number of targets per gene could not be estimated in California poppy, for which the function of only six targets was identified. Indeed, because of the short size of 454 cDNA reads (100 nt), only 14 of 59 targeted cDNA sequences could be annotated. Most of the predicted target genes are similar to previously identified target genes in Arabidopsis using miRNA from the same species (Jones-Rhoades *et al.*, 2006). A third of the putative target genes encode transcription factors from different families involved in different processes of leaf and flower development. Some of these targets have been previously confirmed in Arabidopsis, such as AP2 (Chen, 2004), CUC (Laufs *et al.*, 2004), ARF (Mallory *et al.*, 2005) and Squamosa-like (Xie *et al.*, 2006). Among the non-transcription factor target genes identified was a superoxide dismutase. This enzyme is involved in resistance to oxidative radicals generated during stress (Jones-Rhoades and Bartel, 2004; Lu *et al.*, 2005a; Sunkar *et al.*, 2006). Several new targets have been predicted for California poppy miRNAs that have not been reported so far in Arabidopsis or in rice (Figure 6). They are involved in various metabolic and cellular processes. For example, miR172 target genes encoding proteins are involved in signaling (phosphatidylinositol 3- and 4-kinase), phosphate transport (phosphate transporter), glucose synthesis (aconitate hydratase), solute transport (pyrophosphate-energized vacuolar membrane proton pump), protein synthesis (elongation factor Tu) and wax synthesis (long-chain-alcohol *O*-fatty-acyltransferase family protein and wax synthase family protein). The failure to discover such targets in Arabidopsis could result from the divergence of some members of these miRNA families and the evolution of new target genes in poppy. For six miRNA families, as well as for half of the identified conserved miRNAs, we did not identify any target genes in California poppy. This could in part result from an incomplete coverage of the California poppy transcriptome by the actual cDNA dataset. It could result from the fact that some of these miRNAs have different modes of target recognition (Jones-Rhoades and Bartel, 2004).

In conclusion, we have shown here the multiple values of large-scale isolation of miRNA using the pyrosequencing in a species that does not yet have a sequenced genome. This study allowed us to: (i) identify 173 miRNAs belonging to 28 families, including seven ta-siRNAs; (ii) discover two divergent miRNAs families not previously reported in Arabidopsis and Populus; (iii) show that miR535 might have been lost systematically from core eudicot species, and to identify other miRNAs that may have also had a complex history of gains and/or losses within angiosperms; (iv) identify genes targeted by some of these miRNAs, including new targets that were not previously reported in model species (rice, Populus and Arabidopsis). Deep sequencing and comparative analysis of miRNA distributions from plants that occupy phylogenetically-critical nodes, such as gymnosperms, basal angiosperms, magnoliids, and other monocots and eudicots, will clarify the ebb and flow of miRNA regulators and their target sequences through evolutionary time.

Experimental procedures

sRNA isolation

The California poppy (*E. californica* cv. 'Aurantiaca Orange') plants used in these experiments were grown in the greenhouse with 16 h of light/8 h of dark and at approximately 23°C at Pennsylvania State University (Carlson *et al.*, 2006). Total RNA was prepared from various tissues (root, leaves, stems, flowering buds and open flowers). The flower buds were between 1.0 and 2.5 mm in diameter, and covered all pre-meiotic stages and post-meiotic stages of flower development, as previously defined by Becker *et al.* (2005). Total RNA was prepared using TRIzol reagent (Invitrogen, http://www.invitrogen.com) according to the manufacturer's recommendations. To assure similar sampling of each tissue, equal quantities of total RNA obtained from each type of tissue of the plants were then mixed. Low molecular weight or sRNAs were purified from total RNA by fractionation on a polyacrylamide gel.

sRNA cloning and sequencing

sRNA cloning was performed as described previously (Lau *et al.*, 2001; Llave *et al.*, 2002). In brief, purified sRNAs were ligated to 5′ and 3′ adapters, reverse transcribed and PCR amplified. We used two approaches to sequence sRNAs from California poppy. The first one consisted of cloning concatamers into TOPO TA vectors (Invitrogen), constructing an sRNA library and sequencing a sample of sRNA clones using capillary sequencing. sRNA inserts were isolated from bacteria clones using the Rolling Circle

Amplification method, as described by Carlson *et al.* (2006), and sequencing was carried out with an ABI3700 sequencing machine. In the second approach, concatamerized cDNA corresponding to sRNA was used to construct a 454 library (Poinar *et al.*, 2006) and was sequenced using the 454 approach (Margulies *et al.*, 2005). The library was constructed as described previously by Poinar *et al.* (2006), and a half plate was sequenced. In summary, 2 μg of concatamerized sRNA was blunt-ended and phosphorylated by enzymatic polishing using T4 DNA polymerase and T4 polynucleotide kinase. The polished DNA fragments were then subjected to adapter ligation, followed by isolation of the single-stranded template DNA (sstDNA). The quality and quantity of the sstDNA library were assessed using an Agilent 2100 Bioanalyzer (Agilent, http://www.agilent.com). Each sstDNA library fragment was captured onto a single DNA capture bead and clonally amplified within individual emulsion droplets. The emulsions were disrupted using isopropanol; beads without an amplified sstDNA fragment were removed, whereas beads with an amplified sstDNA fragment were recovered for sequencing. The recovered sstDNA beads were packed onto a 70 × 75 mm PicoTiterPlate™ and loaded onto the GS 20 Sequencing System (454 Life Sciences, http://www.454.com) as previously described (Margulies *et al.*, 2005).

sRNA analysis and identification of California poppy miRNAs and their targets

Known miRNAs from the MIRBASE (release 9) were used to query pyrosequencing and capillary sequences using the program PATSCAN (Dsouza *et al.*, 1997), with default parameters and two mismatches. This cut-off of two nucleotide mismatches is commonly used for the identification of conserved miRNAs across species (Jones-Rhoades and Bartel, 2004; Lu *et al.*, 2005a,b). The target search was performed using the approach previously described by Schwab *et al.* (2005). All conserved miRNA sequences identified were used to query for potential target sequences using PATSCAN, with default parameters and three mismatches, no insertions, and no deletions against the California poppy cDNA database, which included 5713 traditional expressed sequence tag (EST)-based unigenes (Carlson *et al.*, 2006) and 154 779 454 reads of a non-normalized pre-meiotic floral cDNA library (K. Wall, J. Barakat, A. Leebens-Mack, L. Landherr, S. Schuster, J. Carlson, H. Ma, C. dePamphilis, unpubl. data). Only hits with less than two mismatches in positions 1–9, no mismatches in positions 10 and 11, and less than three mismatches after position 11 (Schwab *et al.*, 2005) in the mature miRNAs were considered good target sequences. The same target search was also performed on the Arabidopsis cDNA dataset downloaded from TIGR (http://www.tigr.org/tdb/e2k1/ath1/), using the same criteria to identify target sequences.

Non-conserved small RNAs were identified in the 454 reads by searching for 17–30-bp sequences flanked at both ends by 12 bp from the adapter sequence used in the cloning approach. Identified sRNAs were then searched for RNA contaminants by querying sRNA isolated on tRNA (http://lowelab.ucsc.edu/GtRNAdb/), rRNA (http://www.psb.ugent.be/rRNA/) or snoRNA (http://www.scri.sari.ac.uk/) databases. Non-conserved sequences of between 20 and 23 nt were clustered to identify variants from the same family. To check for probable conservation of these sequences (which could be either siRNA or miRNA) in other species, we used them to search the genomes of Arabidopsis (Gustafson *et al.*, 2005), rice (*Oryza sativa* ssp. *japonica* cv. Niponbare; http://www.tigr.org/tdb/e2k1/osa1/index.shtml) and Populus (Tuskan *et al.*, 2006) using PATSCAN, tolerating no more than three substitutions. When a probable homolog was identified in any of these three species, the genomic sequence surrounding it was folded using the MFOLD program (Zuker, 2003), and the secondary structures were checked using the program MIRCHECK (Jones-Rhoades and Bartel, 2004).

miRNA expression analysis using Northern blot

Total RNA was prepared from leaves and flowering buds using TRIzol reagent (Invitrogen) according to the manufacturer's recommendations with modifications. Total RNA was fractionated on denaturing (urea 8%) polyacrylamide gel. RNA was then transferred to a nylon membrane using a vacuum transfer system (Bio-Rad, http://www.bio-rad.com). Probes used for hybridization were the reverse complement of mature miRNA sequences. Probes were labeled with $\gamma[^{32}P]$ATP using T4 polynucleotide kinase (New England Biolabs, http://www.neb.com) according to the manufacturer's recommendations. Non-incorporated nucleotides were removed using a Centri-spin-20 column (Princeton Separations, http://www.prinsep.com). Hybridizations were performed at a temperature of 20°C below the melting temperature of the probe in ULTRAhyb-Oligo (Ambion, http://www.ambion.com) buffer as described by the manufacturer. Washing was carried out twice for 30 min at 42°C using 5xSaline-Sodium Phosphate-EDTA (SSPE) and 0.5 SDS. Filters were then exposed and scanned using a phosphoimager (Applied Biosystems, http://www.appliedbiosystems.com).

Acknowledgements

The authors thank our colleagues S. Schuster for access to the 454 sequencer, L. Tomsho for expert technical assistance with the 454 sequencer and A. Omeis for plant growth and care. We also thank M. Axtell and D. Bartel for giving us access to unpublished *P. patens* miRNA data, R. Rajagopalan and D. Bartel for access to Arabidopsis miRNA data, and M. Jones-Rhoades for providing us with the program MIRCHECK. Many thanks to M. Axtell, B. Bliss, H. Ma and K. Müller for their comments on an earlier draft of the manuscript. This work was supported by The Floral Genome Project grant from the NSF Plant Genome Research Program (DBI-0115684), and by funds from the Huck Institutes of Life Sciences of Pennsylvania State University.

Supplementary material

The following supplementary material is available for this article online:

Table S1. miRNAs and *trans*-acting siRNAs identified in California poppy and conserved in other plant species. miRNA names indicate the species name *E. californica*, 'Eca', the miRNA family (miR166 for example) and variant number. All listed sequences were detected via 454 pyrosequencing; the number of times each sequence was sampled is indicated ('454 Counts'). Also shown are the number of times a sequence was sampled using capillary sequencing ('Capillary Counts') and the miRNAs that were subjected to northern analysis ('Expression').

References

Achard, P., Herr, A., Baulcombe, D.C. and Harberd, N.P. (2004) Modulation of floral development by a gibberellin-regulated microRNA. *Development*, **131**, 3357–3365.

Adai, A., Johnson, C., Mlotshwa, S., Archer-Evans, S., Manocha, V., Vance, V. and Sundaresan, V. (2005) Computational prediction

of miRNAs in *Arabidopsis thaliana*. *Genome Res.* **15**, 78–91.

Albert, V.A., Soltis, D.E., Carlson, J.E. *et al.* (2005) Floral gene resources from basal angiosperms for comparative genomics research. *BMC Plant Biol.* **30**, 5–5.

Allen, E., Xie, Z., Gustafson, A.M. and Carrington, J.C. (2005) microRNA-directed phasing during *trans*-acting siRNA biogenesis in plants. *Cell*, **121**, 207–221.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

Ambros, V., Bartel, B., Bartel, D. *et al.* (2003) A uniform system for microRNA annotation. *RNA*, **9**, 277–279.

Arazi, T., Talmor-Neiman, M., Stav, R., Riese, M., Huijser, P. and Baulcombe, D.C. (2005) Cloning and characterization of microRNAs from moss. *Plant J.* **43**, 837–848.

Aukerman, M.J. and Sakai, H. (2003) Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. *Plant Cell*, **15**, 2730–2741.

Aung, K., Lin, S.I., Wu, C.C., Huang, Y.T., Su, C.L. and Chiou, T.J. (2006) pho2, a phosphate overaccumulator, is caused by a nonsense mutation in a microRNA399 target gene. *Plant Physiol.* **141**, 1000–1011.

Axtell, M.J. and Bartel, D.P. (2005) Antiquity of microRNAs and their targets in land plants. *Plant Cell*, **17**, 1658–1673.

Bari, R., Datt Pant, B., Stitt, M. and Scheible, W.R. (2006) PHO2, microRNA399, and PHR1 define a phosphate-signaling pathway in plants. *Plant Physiol.* **141**, 988–999.

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

Becker, A., Gleissberg, S. and Smyth, D. (2005) Floral and vegetative morphogenesis in California Poppy (*Eschoscholzia californica* Cham.). *Int. J. Plant Sci.* **166**, 537–555.

Bedell, J.A., Budiman, M.A., Nunberg, A. *et al.* (2005) Sorghum genome sequencing by methylation filtration. *PLoS Biol.* **3**, e13.

Billoud, B., De Paepe, R., Baulcombe, D. and Boccara, M. (2005) Identification of new small non-coding RNAs from tobacco and Arabidopsis. *Biochimie*, **87**, 905–910.

Bonnet, E., Wuyts, J., Rouze, P. and Van de Peer, Y. (2004) Detection of 91 potential conserved plant microRNAs in *Arabidopsis thaliana* and *Oryza sativa* identifies important target genes. *Proc. Natl. Acad. Sci. USA*, **101**, 11511–11516.

Carlson, J.E., Leebens-Mack, J.H., Wall, P.K. *et al.* (2006) EST database for early flower development in California poppy (*Eschscholzia californica* Cham., Papaveraceae) tags over 6000 genes from a basal eudicot. *Plant Mol. Biol.* **62**, 351–369.

Carrington, J.C. and Ambros, V. (2003) Role of microRNAs in plant and animal development. *Science*, **301**, 336–338.

Chen, X. (2004) A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development. *Science*, **303**, 2022–2025.

Chiou, T.J., Aung, K., Lin, S.I., Wu, C.C., Chiang, S.F. and Su, C.L. (2006) Regulation of phosphate homeostasis by MicroRNA in Arabidopsis. *Plant Cell*, **18**, 412–421.

Cui, L., Wall, P.K., Leebens-Mack, J.H. *et al.* (2006) Widespread genome duplications throughout the history of flowering plants. *Genome Res.* **16**, 738–749.

Dezulian, T., Remmert, M., Palatnik, J.F., Weigel, D. and Huson, D.H. (2005) Identification of plant microRNA homologs. *Bioinformatics*, **22**, 359–360.

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

Floyd, S.K. and Bowman, J.L. (2004) Gene regulation: ancient microRNA target sequences in plants. *Nature*, **428**, 485–486.

Fujii, H., Chiou, T.J., Lin, S.I., Aung, K. and Zhu, J.K. (2005) A miRNA involved in phosphate-starvation response in Arabidopsis. *Curr. Biol.* **15**, 2038–2043.

Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A. and Enright, A.J. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* **34**, D140–D144.

Gustafson, A.M., Allen, E., Givan, S., Smith, D., Carrington, J.C. and Kasschau, K.D. (2005) ASRP: the Arabidopsis small RNA project database. *Nucleic Acids Res.* **33**, D637–D640.

Jones-Rhoades, M.W. and Bartel, D.P. (2004) Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol. Cell*, **14**, 787–799.

Jones-Rhoades, M.W., Bartel, D.P. and Bartel, B. (2006) MicroRNAs and their regulatory roles in plants. *Annu. Rev. Plant Biol.* **57**, 19–53.

Kasschau, K.D., Xie, Z., Allen, E., Llave, C., Chapman, E.J., Krizan, K.A. and Carrington, J.C. (2003) P1/HC-Pro, a viral suppressor of RNA silencing, interferes with Arabidopsis development and miRNA unction. *Dev. Cell*, **4**, 205–217.

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.

Laufs, P., Peaucelle, A., Morin, H. and Traas, J. (2004) MicroRNA regulation of the CUC genes is required for boundary size control in Arabidopsis meristems. *Development*, **131**, 4311–4322.

Leebens-Mack, J., Raubeson, L.A., Cui, L., Kuehl, J.V., Fourcade, M.H., Chumley, T.W., Boore, J.L., Jansen, R.K. and dePamphilis, C.W. (2005) Identifying the basal angiosperm node in chloroplast genome phylogenies: sampling one's way out of the Felsenstein zone. *Mol. Biol. Evol.* **22**, 1948–1963.

Li, X. and Zhang, Y.Z. (2005) Computational detection of microRNAs targeting transcription factor genes in *Arabidopsis thaliana*. *Comput. Biol. Chem.* **29**, 360–367.

Li, Y., Li, W. and Jin, Y.X. (2005) Computational identification of novel family members of microRNA genes in *Arabidopsis thaliana* and *Oryza sativa*. *Acta Biochim Biophys Sin (Shanghai)*, **37**, 75–87.

Liu, B., Li, P., Li, X., Liu, C., Cao, S., Chu, C. and Cao, X. (2005) Loss of function of OsDCL1 affects microRNA accumulation and causes developmental defects in rice. *Plant Physiol.* **139**, 296–305.

Llave, C., Xie, Z., Kasschau, K.D. and Carrington, J.C. (2002) Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. *Science*, **297**, 2053–2056.

Lu, C., Tej, S.S., Luo, S., Haudenschild, C.D., Meyers, B.C. and Green, P.J. (2005a) Elucidation of the small RNA component of the transcriptome. *Science*, **309**, 1567–1569.

Lu, S., Sun, Y.H., Shi, R., Clark, C., Li, L. and Chiang, V.L. (2005b) Novel and mechanical stress-responsive MicroRNAs in *Populus trichocarpa* that are absent from Arabidopsis. *Plant Cell*, **17**, 2186–2203.

Mallory, A.C., Bartel, D.P. and Bartel, B. (2005) MicroRNA-directed regulation of Arabidopsis AUXIN RESPONSE FACTOR17 is essential for proper development and modulates expression of early auxin response genes. *Plant Cell*, **17**, 1360–1375.

Margulies, M., Egholm, M., Altman, W.E. *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**, 376–380.

Mette, M.F., van der Winden, J., Matzke, M. and Matzke, A.J. (2002) Short RNAs can identify new candidate transposable element families in Arabidopsis. *Plant Physiol.* **130**, 6–9.

Moore, M.J., Dhingra, A., Soltis, P.S., Shaw, R., Farmerie, W.G., Folta, K.M. and Soltis, D.E. (2006) Rapid and accurate pyrosequencing of angiosperm plastid genomes. *BMC Plant Biol.* **6**, 17.

Palatnik, J.F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J.C. and Weigel, D. (2003) Control of leaf morphogenesis by microRNAs. *Nature*, **425**, 257–263.

Park, W., Li, J., Song, R., Messing, J. and Chen, X. (2002) CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Curr. Biol.* **12**, 1484–1495.

Poinar, H.N., Schwarz, C., Qi, J. *et al.* (2006) Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science*, **311**, 392–394.

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

Schwab, R., Palatnik, J.F., Riester, M., Schommer, C., Schmid, M. and Weigel, D. (2005) Specific effects of microRNAs on the plant transcriptome. *Dev. Cell*, **8**, 517–527.

Sunkar, R. and Zhu, J.K. (2004) Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis. *Plant Cell*, **16**, 2001–2019.

Sunkar, R., Girke, T., Jain, P.K. and Zhu, J.K. (2005) Cloning and characterization of microRNAs from rice. *Plant Cell*, **17**, 1397–1411.

Sunkar, R., Kapoor, A. and Zhu, J.K. (2006) Posttranscriptional induction of two Cu/Zn superoxide dismutase genes in Arabidopsis is mediated by downregulation of miR398 and important for oxidative stress tolerance. *Plant Cell*, **18**, 2051–2065.

Tang, G., Reinhart, B.J., Bartel, D.P. and Zamore, P.D. (2003) A biochemical framework for RNA silencing in plants. *Genes Dev.* **17**, 49–63.

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

The International Rice Genome Sequencing Project. (2005) The map-based sequence of the rice genome. *Nature*, **436**, 793–800.

Tuskan, G.A., Difazio, S., Jansson, S. *et al.* (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*, **313**, 1596–1604.

Wang, J.F., Zhou, H., Chen, Y.Q., Luo, Q.J. and Qu, L.H. (2004) Identification of 20 microRNAs from *Oryza sativa*. *Nucleic Acids Res.* **32**, 1688–1695.

Williams, L., Grigg, S.P., Xie, M., Christensen, S. and Fletcher, J.C. (2005) Regulation of Arabidopsis shoot apical meristem and lateral organ formation by microRNA miR166 g and its AtHD-ZIP target genes. *Development*, **132**, 3657–3668.

Xie, K., Wu, C. and Xiong, L. (2006) Genomic organization, differential expression, and interaction of SQUAMOSA promoter-binding-like transcription factors and microRNA156 in rice. *Plant Physiol.* **142**, 280–293.

Zhang, B.H., Pan, X.P., Wang, Q.L., Cobb, G.P. and Anderson, T.A. (2005) Identification and characterization of new plant microRNAs using EST analysis. *Cell Res.* **15**, 336–360.

Zhang, B., Pan, X., Cobb, G.P. and Anderson, T.A. (2006) Plant microRNA: a small regulatory molecule with big impact. *Dev. Biol.* **289**, 3–16.

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

