Title: Identification of conserved Aquilegia coerulea microRNAs and their targets.

Article Type: Research Paper

Keywords: Aquilegia; microRNA; evolution; angiosperm

Corresponding Author: Dr. Elena M. Kramer,
Corresponding Author's Institution: Harvard University

First Author: Joshua R Puzey

Order of Authors: Joshua R Puzey; Elena M. Kramer

Manuscript Region of Origin: UNITED STATES

Abstract: Aquilegia is an emerging model organism that is phylogenetically intermediate between the core eudicot and monocot models Arabidopsis and Oryza. In this study, we have used a comparative genomics approach to identify 45 Aquilegia microRNAs that comprise 20 separate plant microRNA families. We have predicted 85 targets of these newly identified Aquilegia microRNAs, including transcription factors and loci involved in metabolism, stress responses, transport, and auxin signaling. microRNA families from 16 plant species and the newly identified microRNAs from Aquilegia were analyzed in a phylogenetic context revealing 40 distantly conserved microRNA families. In addition to these highly conserved plant microRNA families, several families with disjointed phylogenetic distribution were identified. This study provides a phylogenetically important dataset for plant microRNA evolution studies. The current study is the first to identify miRNAs in a lower eudicot in which comprehensive genomic resources are becoming available.
Identification of conserved *Aquilegia coerulea* microRNAs and their targets

Joshua R. Puzey and Elena M. Kramer*

Dept. of Organismic and Evolutionary Biology, Harvard University, Cambridge MA 02138

*Author for Correspondence: 16 Divinity Ave, Cambridge MA 02138; ekramer@oeb.harvard.edu; phone: 617-496-3460, fax: 617-496-5854
ABSTRACT

Aquilegia is an emerging model organism that is phylogenetically intermediate between the core eudicot and monocot models, *Arabidopsis and Oryza*. In this study, we have used a comparative genomics approach to identify 45 *Aquilegia* microRNAs that comprise 20 separate plant microRNA families. We have predicted 85 targets of these newly identified *Aquilegia* microRNAs including transcription factors and loci involved in metabolism, stress responses, transport, and auxin signaling. microRNA families from 16 plant species and the newly identified microRNAs from *Aquilegia* were analyzed in a phylogenetic context revealing 40 distantly conserved microRNA families. In addition to these highly conserved plant microRNA families, several families with disjointed phylogenetic distribution were identified. This study provides a phylogenetically important dataset for plant microRNA evolution studies. The current study is the first to identify miRNAs in a lower eudicot in which comprehensive genomic resources are becoming available.

*Keywords: Aquilegia, microRNA, evolution, angiosperm*
1. Introduction

MicroRNAs (miRNAs) are a set of small (~22 nt) single-stranded non-coding RNAs in plants and animals that play an important role in regulating mRNA targets through cleavage and/or translational repression in a sequence specific manner (Chen, 2004; Jones-Rhoades et al., 2006). In plants, miRNAs have diverse roles including the regulation of leaf development (Palatnik et al., 2003), floral development (Cartolano et al., 2007), phase change (Aukerman and Sakai, 2003; Lauter et al., 2005), male and female reproductive development (Wu et al., 2006), root development (Boualem et al., 2008), and disease and environmental stress response (Shukla et al., 2008; Zhang et al., 2008b; Ding et al., 2009).

Evolutionary studies of plant miRNAs are currently limited by the large phylogenetic distances between plant miRNA datasets (Fig. 1). The four best-annotated plant model species in which significant miRNA datasets have been determined include the core eudicot, Arabidopsis thaliana (A. thaliana), the monocot, Oryza sativa (O. sativa), the lycopod, Selaginella moellendorffii (S. moellendorffii), and the bryophyte Physcomitrella patens (P. patens) (Fig. 1). The eudicot and monocot lineages are estimated to have diverged from one another approximately 140 Myr ago, while lineage containing Aquilegia diverged from other eudicots approximately 100 Myr ago (Fig. 1) (Chaw et al., 2004; Sanderson et al., 2004; Moore et al., 2007). The Aquilegia miRNA dataset therefore helps to break up the large phylogenetic distance that separates major flowering plant lineages, particularly the monocots and dicots.

The goal of this study is to annotate the miRNA profile of Aquilegia. Aquilegia (columbine), a eudicot and a member of the Ranunculales, is an emerging model
organism with a large number of available genetic and genomic tools, including ongoing whole genome sequencing. The phylogenetic position of Aquilegia will provide an important reference point for comparing these core eudicot and monocots models, while allowing us to ask questions about the origin and diversification of major plant miRNA lineages. In addition to having a critical position for deep phylogenetic analyses, Aquilegia has undergone a recent adaptive radiation due to diverse ecological niches (Hodges and Arnold, 1994b; Hodges, 2003; Hodges and Kramer, 2007). This recent adaptive radiation coupled with genomic tools will allow us to elucidate the genetic basis for morphological variation and speciation in the genus.

The evolution and conservation of plant miRNAs has been the subject of significant investigation (Axtell and Bartel, 2005; Zhang et al., 2006a; Axtell and Bowman, 2008). In this study we have identified multiple highly conserved plant miRNAs in Aquilegia as well as miRNA families that have a more disjointed phylogenetic distribution.

2. Materials and Methods

2.1 Aquilegia coerulea database

We searched for miRNAs among the currently available genomic sequences from Aquilegia coerulea 'Goldsmith'. This is a horticultural inbred line derived from several species, but primarily Aquilegia coerulea. This inbred line is referred to as A. coerulea for the remainder of this paper. Currently, whole genome sequence consists of 483,253 reads corresponding to an estimated 0.7x-1x genome coverage and is available at the Trace Archives at NCBI under "Aquilegia coerulea - WGS".
2.2 Criteria for orthologous miRNA annotation

Orthologous miRNAs were identified according to the criteria for conserved plant miRNA annotation established by Meyers et al. (2008). These criteria include conservation of the miRNA precursor hairpin and the mature miRNA sequence. The specific criteria for filtering the stem-loop structure and the mature miRNA sequence conservation are as follows: no more than four mismatches between miRNA/miRNA* were allowed; no bulges in miRNA/miRNA* larger than two bases; and four or fewer mismatched sequences were allowed between the A. coerulea miRNA and the previously identified miRNA (Meyers et al., 2008).

2.3 Mature miRNA dataset

A total of 1023 mature miRNA sequences from Glycine max (G. max) (number of mature sequences = 42), Medicago truncatula (M. truncatula) (34), Gossypium hirsutum (G. hirsutum) (13), A. thaliana (97), Brassica rapa (B. rapa) (15), Brassica napus (B. napus) (41), Populus trichocarpa (P. trichocarpa) (181), Vitis viniferia (V. vinifera) (133), Solanum lycopersicum (S. lycopersicum) (24), Triticum aestivum (T. aestivum) (9), O. sativa (159), Sorghum bicolor (S. bicolor) (78), Zea mays (Z. mays) (97), Pinus taeda (P. taeda) (19), Selaginella moellendorffii (S. moellendorffii) (18), and Physcomitrella patens (P. patens) (64) were obtained from mirBASE (Release 13.0) and the recently published S. bicolor genome (Griffiths-Jones, 2004; Griffiths-Jones et al., 2006; Griffiths-Jones et al., 2008; Paterson et al., 2009). Redundant sequences were removed from this dataset. miRNA families from this dataset that were conserved in two or more plant species were searched for in A. coerulea. The non-redundant sequences were then used as the query in a BLASTn search of the A. coerulea genome (see below).
2.4 BLASTn of A. coerulea genome

We performed a BLASTn search of the A. coerulea whole genome sequence using the following parameters: database was set to Aquilgia coerulea – WGS; BLASTn was chosen; ‘automatically adjust parameters for short input sequences’, which is the default, was deselected to allow us to create our own search profile; the expect threshold was set to 1000; the word size was set to 7; and all other parameters were left at default. The BLASTn was performed and all sequences with four or fewer mismatches across the entire query sequence were extracted and stored in a database for further analysis. Our BLASTn method identified miRNA binding sites in coding genes in addition to miRNAs. True miRNAs were identified by their indicative secondary hairpin structure (see below) while the remaining sequences were filtered out of our database (Fig. 2). These latter sequences could represent target coding genes but the available genome sequence has yet to be annotated.

2.5 miRNA hairpin prediction

All sequences that had four or fewer mismatches with previously identified mature miRNAs were then filtered using their predicted secondary structure (Fig. 2). Mfold, a publicly available online application (http://mfold.bioinfo.rpi.edu), was used to predict the secondary structure of the obtained sequences based on thermodynamic stability (Mathews et al., 1999; Zuker, 2003). The RNA folding application was used and all parameters were left at default. Because the RNA folding application only accepts sequences shorter than 800 nt, for sequences longer than 800 nt we performed two steps: (1) the Nucleic Acid Quikfold application available at Mfold (an application capable of accepting longer sequences) was used to predict secondary structure and (2) CLUSTALw
alignment with mature miRNA sequence was performed. After the *Nucleic Acid Quickfold* and CLUSTALw alignment, the regions furthest from the predicted hairpin and mature sequence site was trimmed off, shortening the sequence to 800 nt. This shortened sequenced was then analyzed in Mfold. The structure with the highest score and lowest free energy was analyzed and the precursor sequence was predicted based on secondary folding structure. The extent of the precursor sequence was predicted by identifying any large loops with little or no nucleotide pairing that followed the end of a region with significant pairing. Secondary structures were then screened for four or fewer mismatches in the miRNA/miRNA* duplex and a folding energy of lower than -15 kcal/mol. Our initial BLAST identified 240 of sequences of which we have only predicted 45 to be miRNAs.

2.6 MFE/AMFE/MFEI

The minimal folding energy (MFE), expressed in kcal/mol, is a method of calculating the thermodynamic stability of the secondary structure of RNA or DNA (Mathews et al., 1999; Zuker, 2003). The lower the MFE of a molecule, the more stable the secondary structure. Because MFE values are strongly correlated with the length of the sequence we normalized the MFE by calculating the adjusted MFE (AMFE) (Zhang et al. 2008) using the following equation: \( \text{AMFE} = \left( \frac{\text{MFE}}{\text{length of RNA sequence}} \right) \times 100 \) (Zhang et al., 2006b). Next, the minimal folding free energy index (MFEI) was calculated for the *A. coerulea* miRNA precursors. The MFEI is an index developed by Zhang et al. (2006) and is used as a criterion to differentiate between miRNAs versus other RNA based on MFE, sequence length, and G+C nucleotide composition (Zhang et al., 2006b; Zhang et al., 2008a). The minimal folding free energy index (MFEI) was calculated by the following equation: \( \text{MFEI} = \left( \frac{\text{AMFE}}{100} \right) / (\text{G\%+C\%}) \) (Zhang et al., 2006b).
2.7 miRNA target prediction

The near-perfect complementarity of plant miRNAs for their targets allows for very accurate prediction of miRNA targets (Rhoades et al., 2002; Jones-Rhoades and Bartel, 2004; Schwab et al., 2005; Schwab et al., 2006). All non-redundant A. coerulea mature miRNA sequences were used as the query in a BLASTn search of the Aquilegia Gene Index database, which consists of 85,039 reads (http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/Blast/index.cgi). The Aquilegia Gene Index (AqGI) was used because opening reading frame (ORF) predictions have been made for these sequences and the A. coerulea whole genome sequence has not yet been assembled. The AqGI is based on a full-length enriched, normalized cDNA library derived from a wide range of tissue collected from Aquilegia formosa X Aquilegia pubescens including shoot meristems, floral apical meristems, flower buds, leaves, and root tissue. The exceptionally low sequence variation between Aquilegia species (Hodges and Arnold, 1994a) allows us to use this database to accurately predict targets for A. coerulea miRNAs.

Targets were identified via BLASTn using the following method: BLASTn expected value was raised to 10,000; sequences were filtered based on four or fewer mismatches with the query; and no gaps were allowed at the binding site. The Aquilegia sequences with four or fewer mismatches were then extracted and a BLASTn search was performed with these sequences against Arabidopsis genome (http://www.arabidopsis.org/Blast/index.jsp). The top hit was selected for each Aquilegia EST and its name, biological process, and molecular function were obtained. We also aligned predicted target ESTs for each miRNA family against one another in order to eliminate redundancies (ESTs that shared greater than 98% sequence identity, usually due to separate annotations of alternative splicing products of the same locus).
For miRNA families not yet identified in *A. coerulea*, potential targets were identified using target BLASTn method described above, but with non-redundant mature miRNA sequences from other species as the query. Although identification of a conserved miRNA binding site in a target is not sufficient evidence for definitively establishing a miRNA family’s presence in the genome, conservation of targets provides supporting evidence for their presence. Targets were only predicted for miRNA families that are conserved across two deeply divergent land plant lineages (e.g. monocots and dicots or dicots and bryophytes). miRNA families that met these criteria were: 162, 164, 390, 393, 394, 397, and 783 (see section 3.5).

3. Results and Discussion

3.1 Identification of *Aquilegia coerulea* miRNAs

A total of 45 miRNAs from 20 miRNA families were identified in *A. coerulea* (Table 1, Supplementary Material (Supp.) Fig. 3A). Family size ranged from one to seven members with seven miRNAs belonging to miR477; six to miR171; five to miR166; three to miR169 and miR482; two to miR156, miR160, miR172, miR395, miR396, and miR398; and one miRNA identified for the remaining families (Supp. Fig. 3A). It is important to note that aqc-miR530 has a three nucleotide bulge on the miRNA* side of the hairpin but was still annotated as a miRNA as it has a MFEI value of 0.88 (see below), indicating that this is likely to be a true miRNA. Similarly, all miRNA annotation criteria as described in Meyers et al. (2008), except for a three nucleotide bulge in the miRNA*, were met for aqc-miR168. aqc-miR168 was included in our miRNA dataset since it has a low folding energy of -40.45 kcal/mol, in spite of the three
nucleotide bulge, indicating it has a relatively stable secondary structure.

The majority of mature miRNAs (66.7%) were 21 nt in length, with 24.4% and 8.9% were 20 nt and 22 nt in length, respectively (Table 1). Nucleotide composition of the mature miRNAs was analyzed. Cytosine is the dominant nucleotide totaling 30.2% of the mature miRNA nucleotide composition; uracil is the next most prevalent comprising 26.7% of the mature sequence, followed by guanine (21.7%), and adenine (21.4%) (Fig. 3).

It has been previously reported that the strong bias of uracil in the first 5’ nucleotide position is due to its important role in the recognition of the miRNA by ARGONAUTE1 (Mi et al., 2008; Montgomery et al., 2008; Takeda et al., 2008; Zhang et al., 2008a). Consistent with this, we found in A. coerulea that uracil (71.1%) was the dominant nucleotide at the first position of the 5’ end of the mature A. coerulea miRNAs (Fig. 3). Conclusions on the importance of position-specific nucleotide preference at sites other than the first nucleotide of the miRNA have varied. Zhang et al. (2008) reported 61% cytosine preference at position 19 in soybean and suggested that this may be important for RISC or Dicer cleavage sites on the miRNA precursor while Mi et al. (2008) reported that, aside from the first position, no other position-specific nucleotide preference could be determined. Similar to Zhang et al (2008), in A. coerulea, we observed a strong preference, 51.1%, 55.6%, and 51.5%, cytosine at positions 18, 19, and 21, respectively (Fig. 3). We agree with Zhang et al. that the most likely explanation for this similarity across otherwise divergent miRNA families is some kind of biochemical constraint related to miRNA processing.

3.2 Precursor Analysis
The length of the predicted Aquilegia miRNA precursors varies from 68 to 180 nt with an average precursor length of 106 ± 32 nt (Table 2). It is important to note that these precursors are predictions based on their secondary folding structure (see Methods). All predicted A. coerulea miRNA hairpins are available in the supplementary information but several representative A. coerulea miRNA hairpin structures are shown in (Fig. 4). The nucleotide composition of the miRNA precursor sequences in order of abundance is, uracil (31.4 ± 3.8%), adenine (25.6 ± 4.4%), guanine (22.9 ± 3.7%), and cytosine (20.1 ± 3.0%) (Table 2). Previous calculations of nucleotide composition of miRNA-precursors have reported similar values (Zhang et al., 2008a).

The MFE for predicted A. coerulea miRNA precursors averaged -46 ± 14.7 (kcal/mol) and ranged from -81.8 kcal/mol to -18.8 kcal/mol (Table 2). The average AMFE of the A. coerulea miRNA precursors is -44.1 ± 7.5 (kcal/mol) (Table 2). The MFEI of A. coerulea miRNA precursors was also analyzed and they scored an average of 1.03 ± 0.18, with the lowest score being 0.60 and the highest score being 1.48 (Table 2). Zhang et al. (2006) showed that a MFEI value greater than or equal to 0.85 is a strong indication of an actual miRNA. Of our identified A. coreulea miRNA precursors, 38 (84.4%) had a MFEI greater or equal to 0.85 and of the remaining seven, two (4.4%) had a MFEI between 0.6 and 0.7, three (6.6%) between 0.71 and 0.8, and two (4.4%) between 0.81 and 0.84. While a MFEI value above 0.85 is highly indicative of an actual miRNA, lower values do not rule out a sequence as a true miRNA (Zhang et al., 2006b; Zhang et al., 2007).

3.3 Target Annotation

We have predicted a total of 85 miRNA targets for A. coerulea. Because the A. coerulea
genome is not yet annotated, these predicted target sequences were compared to the *A. thaliana* database in order to obtain information on potential gene functions. The *A. thaliana* loci with the lowest e-value were selected, corresponding to a total of 71 distinct *A. thaliana* loci. In several cases, multiple *Aquilegia* targets recovered the same top *A. thaliana* hit, resulting in fewer *A. thaliana* loci (71) than predicted *Aquilegia* targets (85). This difference could be due to intervening gene duplications between *A. thaliana* and *Aquilegia* as well as the inexactitude of BLAST searches in assigning orthology. It does not appear that identification of targeted transposable elements is a factor in the multiple redundant BLAST hits as only one such target was characterized. AqGI TC and EST numbers representing redundant sequences (those with greater than 98% similarity) were removed from Table 3 (the numbers for every recovered AqGI TC and EST are present in supplementary table 2). In agreement with the fact that miRNAs are important developmental regulators (Aukerman and Sakai, 2003; Palatnik et al., 2003; Chen, 2004), 28.2% of the predicted targets are inferred to encode transcription factors (Table 3). Of the remaining targets, 26.8% are predicted to be involved in metabolism; 14.1%, stress response; 4.2%, transport; 4.2%, kinases; 2.9%, photo processes; 2.9%, auxin signaling; 1.4%, RNA binding; and 1.4%, protein binding (Table 3).

### 3.4 Related miRNA families

**miR156/529**

Based on our sequence analysis, aqc-miR156 and aqc-miR529 appear to be related (Fig. 5A,B), which has also been suggested for these families in *P. patens* (Axtell et al., 2007). The miR156 family is present in the majority of land plants while miR529 is more narrowly distributed and has only been described in *P. patens*, *S. bicolor*, and *O. sativa*.
(as determined from miRBASE and recently published *S. bicolor* genome) (Fig. 6) (Paterson et al., 2009). Barakat et al. (2007) computationally predicted a miR529 locus in both *A. thaliana* and *P. trichocarpa*, although these sequences are not present in miRBASE. To our knowledge, no other reports have identified miR529 in *A. thaliana* or *P. trichocarpa*, even though small-RNA deep sequencing has been conducted in both these species (Barakat et al., 2007; Fahlgren et al., 2007). *A. thaliana* and *P. trichocarpa* have well annotated genomes and it is unlikely, given the current emphasis on miRNA identification, that miR529’s presence would have been missed in these model systems. Given these facts, we did not include *A. thaliana* or *P. trichocarpa* miR529 in our summary of the phylogenetic distribution of miRNAs (Fig. 6).

One of our identified sequences, 202185620929, could possibly encode both miR156 and miR529 (Fig. 5A, B). Our prediction that it encodes a miR529 family miRNA is based on the CCC repeat at the 3’ end of the mature miRNA, which is conserved in all *P. patens* miR529 members (Axtell et al., 2007) (Fig. 5B). Two other sequences (2185892723 and 2185518132) were placed in the miR156 family since neither of these two sequences have a CCC repeat at the 3’ end (Fig. 5B). In addition, aqc-miR156 sequences have T at the 7th position as opposed to 202185620929, which has a G at the 7th position (Fig. 5B).

Overlapping and specific targets were predicted for each of these miRNAs. aqc-miR156 is predicted to regulate five *SQUAMOSA PROMOTER BINDING PROTEIN-Like* (SBP) genes and one homolog of *Growth Regulating Factor 2* (Table 3). miR156 regulation of the above targets has been shown to play important developmental roles in other plants species, including regulating plastochron length (Wang et al., 2008), organ
size (Wang et al., 2008), cell number (Usami et al., 2009), and phase change (Gandikota et al., 2007). Guo et al. (2008) reported that the miR156 binding site is highly conserved in land plant SBP genes so our Aquilegia predictions are consistent with these observations (Guo et al., 2008). aqc-miR529 shares four predicted SBP targets with aqc-miR156 but, in addition, has five specific predicted targets, including two involved in stress response, one involved in metabolism, and one with an unknown function (Table 3).

\textit{miR159/319}

It has been previously shown that the miR159 and miR319 families are related in sequence and targets but that divergent expression and slight sequence variation allows for specific biological functions for these distinct miRNAs (Palatnik et al., 2007). After our sequence analysis of A. coerulea, it is not possible to determine conclusively if 2183893560 and 2185799162 encode miR159 and/or miR319. However, based solely on their sequences, we predict that 2185799162 encodes aqc-miR319 (Table 1; Fig. 5C). 2183893560 was placed in the miR159 family although its sequence appears to be a hybrid between \textit{A. thaliana} and \textit{O. sativa} miR159 and miR319 (Fig. 5C). 2183893560 has a CT at the 7\textsuperscript{th} and 8\textsuperscript{th} position, which is conserved in the \textit{A. thaliana} and \textit{O. sativa} miR319 family. In contrast it has a TTT repeat at the 5’ end and CTCTA at the 3’ end, which is more similar to the \textit{A. thaliana} and \textit{O. sativa} miR159 family (Fig. 5C). No targets were predicted for aqc-miR159 but a novel locus involved in light sensing is predicted as an aqc-miR319 target.

\textit{3.5 Phylogenetic distribution of microRNAs}

Given that \textit{A. coerulea} is a member of the Ranunculales, a lineage that is roughly
intermediate between the clades that contain *A. thaliana* and *O. sativa*, we thought it pertinent to evaluate the newly identified *A. coerulea* miRNA families in a phylogenetic context (Fig 4). All *A. coerulea* miRNA sequences, along with all miRNA sequences from species used as the query in the BLAST searches, were plotted based on presence or absence and number of miRNAs per family (Fig. 6). This analysis produces a clear pattern. A small group of twenty-one miRNA families appear to be highly conserved across angiosperms (156, 159, 160, 162, 164, 166, 167, 168, 169, 171, 172, 319, 390, 393, 394, 395, 396, 397, 398, 399 and 408). Axtell and Bowman have previously reached the same conclusion in a similar analysis (Axtell and Bowman, 2008). Of these highly conserved angiosperm miRNA families, we were able to identify 15 of the 21 in *A. coerulea* (Fig. 6). In addition, conserved miRNA target sites for five of the six uncharacterized miRNA families were identified, suggesting that they too may be present. The high conservation of these 21 miRNA families and their targets is expected given that these miRNAs have been shown to be involved in critical developmental processes, leading us to expect significant pleiotropy and subsequent selection for the conservation of these developmental modules.

In addition to identifying these highly conserved miRNA families in *A. coerulea*, several miRNA families were predicted that had a more varied phylogenetic distribution (Fig. 6). Axtell and Bowman previously identified 39 miRNA families that are present in two or more distant plant lineages based on miRBASE (version 10.1) (Axtell and Bowman, 2008). Our analysis expands their dataset to 40 conserved miRNA families (Fig. 6) by the addition of miR530, which is conserved in distantly related species: *P. trichocarpa*, *A. coerulea*, and *O. sativa* (Lu et al., 2008). To our knowledge, there are currently no predicted targets of miR530 in *O. sativa* or *P. trichocarpa* (Liu et al., 2005;
Lu et al., 2005; Lu et al., 2008), but we have predicted that aq-miR530 targets a serine carboxypeptidase-like gene. Homologs of this locus were identified in *O. sativa* and *P. trichocarpa* but no miR530-binding site appears to be present in either gene. Moreover, we also included miRNA families that are not as distantly conserved but show representatives across more closely related plant lineages, including two monocot specific miRNA families (miR444 and miR528) (groups previously identified by (Willmann and Poethig, 2007; Sunkar et al., 2008)), a Brassicaceae specific miRNA family (miR824), and three Fabaceae specific miRNA families (mir1507, miR1509, and miR1510).

miR477

miR477 has an interesting phylogenetic distribution as it is present in bryophyes, then absent in the lycopods, gymnosperms, and all surveyed monocots (Axtell et al., 2007), but reported in this paper to be present in *Aquilegia* (Fig. 6). To our knowledge, in addition to its presence in the bryophyte, *P. patens*, miR477 has only been identified in the core eudicots *P. trichocarpa* and *V. vinifera* (Axtell et al., 2007; Lu et al., 2007). The apparent reappearance of miR477 in *Aquilegia*, at the base of the eudicots, prompts several interesting questions: (1) what are the targets of miR477 in *P. patens*, (2) are these targets conserved in *Aquilegia*, *V. vinifera*, or *P. trichocarpa*, and (3) what are the characteristics of the target homologs in the lycopods, gymnosperms, monocots, and core eudicots, i.e., is there a remnant of a miR477 binding site? Targets have been predicted for miR477, but to our knowledge these have only been confirmed in *P. patens*. The only confirmed *P. patens* miR477 target is a basic helix-loop helix transcription factor (Axtell et al., 2007), while several other *P. patens* miR477 target predictions include an abscisic
acid-insensitive-like protein and a kelch motif family protein (Axtell et al., 2007). The poplar miR477 has been predicted to target a GRAS domain containing protein, a NAC protein, a zinc finger protein, and a polygalacturonase protein (Lu et al., 2005). Our predicted targets include a plastidic glucose-6-phosphate dehydrogenase, elongation factor 1B-gamma, and an rRNA processing protein (Table 3). We identified homologs of the predicted A. coerulea miR477 targets in P. trichocarpa, V. vinifera, and P. patens but no miR477 binding site appears to be present in these genes. In answer to question three, the phylogenetic distance among these taxa and complexity of the target gene families, makes one-to-one comparisons of potential targets challenging. It is notable, however, that while members of the NAC and GRAS gene families are known to be regulated by microRNAs in A. thaliana, they are targeted by other families (miR164 and miR170/171, respectively; Rhoades et al. 2002). Seeing that these regulatory interactions appear to be quite highly conserved (Jones-Rhoades et al., 2006), this would seem to suggest that miR477 acquired targeting of GRAS and NAC family members specifically in P. trichocarpa. Similarly, homologs of the predicted Aquilegia targets from rice and A. thaliana did not show conservation of the miR477 binding site (data not shown). Thus, further work is needed to confirm miR477 targets and analyze the evolutionary significance of their distribution. Due to the distance between the last common ancestor of P. patens and A. coerulea as well as the apparent heterogeneity of miR477 targets, we can imagine two evolutionary scenarios. Under the first, miR477 must be significantly less constrained than other ancient miRNAs in terms of target conservation and even presence in the genome. The other alternative is that this family has actually arisen via convergent evolution, although a mechanism for such convergence remains unclear and the sequence similarity among the predicted miR477 representatives is relatively high.
miR529

The phylogenetic distribution of miR529 is also quite interesting. miR529 is present in the lycops, bryophytes, monocots, and the Ranunculids *Escscholzia californica* and *A. coerulea*, but is distinctly absent from all core eudicots (Griffiths-Jones et al., 2006; Griffiths-Jones et al., 2008; Barakat et al., 2007) (Fig. 6; see note above regarding lack miR529 in *A. thaliana* and *P. trichocarpa*). The fact that miR529 appears to be closely related to miR156 raises the possibility that it may have been lost without ill effect, possibility due to some redundancy with miR156. Before this hypothesis is tested, detailed work comparing miR156 and miR529 and their target specificity, similar to Palatnik et al. (2007) where they determined that subtle sequence and expression differences in the related miR159 and miR319 families led to distinct functions, needs to be performed.

miR482

Another miRNA family with a disjunct phylogenetic distribution is miR482 (Fig. 6). mir482 has been identified in the eudicot species *G. max*, *P. trichocarpa*, *V. viniferia* (Barakat et al., 2007; Jaillon et al., 2007; Lu et al., 2008; Zhang et al., 2008a), and now *A. coerulea*. In addition to its presence in the eudicots, miR482 has only been identified in the gymnosperm *P. taeda* (Lu et al., 2007). In *A. coerulea* we identified three members of the family but no target predictions were made. Prediction or confirmation of a target is not critical for the annotation of a miRNA and in this case, the inability to identify a target may be due to incomplete coverage in the EST database or, alternatively, it may be due to the fact that miR482 is evolutionary transient (Axtell and Bowman, 2008). Similar
to the case with miR477, the predicted targets of miR482 are quite diverse. Lu et al. (2005) identified a disease resistance locus in *P. trichocarpa* that is targeted by miR482. We identified a possible *Aquilegia* homolog of this locus but could not identify a miR482 binding site. Approximately 80 miR482 targets have been predicted in *P. taeda*, although they were not analyzed in depth due to their large number (Lu et al., 2007). To our knowledge no targets have been predicted for miR482 in *G. max*.

4. Conclusions

This is the first study to systematically identify and annotate miRNAs in the emerging eudicot model *Aquilegia*. We have identified 45 miRNAs belonging to 20 miRNA families and have determined that, in general, both the miRNA families and predicted targets are highly conserved in *Aquilegia* when compared to other model plant systems. Also, we have mapped, in a phylogenetic context, what is currently known about the miRNA families in 16 other plant species. Due to *Aquilegia’s* critical phylogenetic position at the approximate midpoint between well developed models *A. thaliana* and *O. sativa*, this data set lays the ground work necessary for further evolutionary and developmental studies on the evolution of miRNAs and their role in the angiosperm diversification.

Acknowledgements

We thank members of the Kramer lab, Dr. Scott Hodges and ## anonymous reviewers for comments on the manuscript. This work was supported by grants NSF-BE #0412727 and NSF-IBN #0720240 to EM Kramer.
References


Figure Legends

Fig. 1. Simplified phylogeny and divergence times of the land plants based on
(Sanderson, 2003; Chaw et al., 2004; Sanderson et al., 2004; Moore et al., 2007; Qiu et
al., 2007; Rensing et al., 2008). Major model systems associated with the various land
plant lineages are listed.

Fig. 2 Flow diagram illustrating protocol used to identify conserved *Aquilegia* miRNAs.

Fig. 3. Position-specific nucleotide distribution among *A. coerulea* mature miRNA.
Percentage distribution of nucleotides at position 1-21 of mature *A. coerulea* miRNA.
The average percentage nucleotide composition is depicted as average.

Fig. 4. Three *Aquilegia coerulea* miRNA predicted precursor hairpin structures chosen as
a representative of all *A. coreulea* miRNAs identified. All predicted precursor hairpin
structures are available in the supplementary information.

Fig. 5. miRNA families discussed in greater detail in the text (A) Hairpin structure of
gi20218562092, which appears to encode miR156 and/or miR529. The gray line
delineates the miR156 mature sequence while the black line delineates the miR529
mature sequence. (B) Alignment of miR156/miR529 families of *A. coerulea*
(2185892723, 2185805752, 2185518132, and 2185620929), *O. sativa*, *A. thaliana*, and *P.
apiens*. (C) Alignment of miR159/miR319 families of *A. coerulea* (2183893560 and
2185799162), *O. sativa*, and *A. thaliana*. A. (D) Alignment of miR477 families of *A.
cerulea* (2185477245, 2185552337, 2185755300, 2185768332, 2185874405,
2185853634, and 2185809227), *P. trichocarpa*, *V. vinifera*, and *P. patens* For B-D, osa:
Fig. 6. Phylogenetic distribution of plant miRNAs. All plant miRNAs conserved in two or more plant species [as determined by miRBASE (Release 13.0) and the recently released S. bicolor genome (Paterson et al., 2009)] are plotted. Phylogenetic affinities of each taxon are indicated by the colored bars on the left: Red = core eudicot; Green = eudicot; Dark Blue = monocot; Yellow = gymnosperm; Pink = lycophyte; and Light Blue = bryophyte. Shading of the boxes represent our findings: dark gray = miRNA present, light gray = miRNA target sequence identified, dotted = No target identified. Numbers in boxes represent the number of miRNAs present in a particular miRNA family. † indicates that a miRNA sequence was reported by (Sunkar and Jagadeeswaran, 2008) from the Aquilegia EST database. We searched the EST database for miR394 but failed to identify it, presumably because of stricter annotation criteria recently described in (Meyers et al., 2008).
### Table 1
Conserved *Aquilegia coerulea* microRNAs

<table>
<thead>
<tr>
<th>microRNA family</th>
<th>aqc-miRNA gi</th>
<th>Mature</th>
<th>kcal/m mol</th>
<th>MFE</th>
<th>N M</th>
<th>ML PL</th>
<th>Ar m</th>
<th>A C G U (%)</th>
<th>A (%)</th>
<th>C (%)</th>
<th>G (%)</th>
<th>U (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>aqc-miRNA156a</td>
<td>2.19E+09</td>
<td>GACAGGAGAUAAGAGAGCA</td>
<td>-43.4</td>
<td>0.00</td>
<td>1</td>
<td>20</td>
<td>81</td>
<td>3'</td>
<td>5</td>
<td>37</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>aqc-miRNA160i</td>
<td>2.19E+09</td>
<td>GACAGGAGAUAAGAGAGCA</td>
<td>-44.8</td>
<td>0.00</td>
<td>1</td>
<td>20</td>
<td>83</td>
<td>3'</td>
<td>9</td>
<td>32</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>aqc-miRNA159</td>
<td>2.19E+09</td>
<td>UUGGACUGAGAGACCUU</td>
<td>-18.8</td>
<td>0.00</td>
<td>4</td>
<td>21</td>
<td>71</td>
<td>3'</td>
<td>5</td>
<td>37</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>aqc-miRNA160</td>
<td>2.19E+09</td>
<td>GCCUGGCUCUCCGUGUAGGC</td>
<td>-55.6</td>
<td>0.00</td>
<td>1</td>
<td>21</td>
<td>82</td>
<td>3'</td>
<td>5</td>
<td>37</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>aqc-miRNA156b</td>
<td>2.19E+09</td>
<td>GCGAGCACCCAGUACUUCC</td>
<td>-93.1</td>
<td>0.00</td>
<td>4</td>
<td>21</td>
<td>71</td>
<td>3'</td>
<td>5</td>
<td>37</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>aqc-miRNA156c</td>
<td>2.19E+09</td>
<td>GCGAGCACCCAGUACUUCC</td>
<td>-94.5</td>
<td>0.00</td>
<td>2</td>
<td>21</td>
<td>94</td>
<td>3'</td>
<td>5</td>
<td>37</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>aqc-miRNA156d</td>
<td>2.19E+09</td>
<td>GCGAGCACCCAGUACUUCC</td>
<td>-46.5</td>
<td>0.00</td>
<td>4</td>
<td>21</td>
<td>71</td>
<td>3'</td>
<td>5</td>
<td>37</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>aqc-miRNA156e</td>
<td>2.19E+09</td>
<td>GCGAGCACCCAGUACUUCC</td>
<td>-62.9</td>
<td>0.00</td>
<td>3</td>
<td>21</td>
<td>3'</td>
<td>5</td>
<td>37</td>
<td>6</td>
<td>2</td>
<td>45</td>
</tr>
<tr>
<td>aqc-miRNA156f</td>
<td>2.19E+09</td>
<td>GCGAGCACCCAGUACUUCC</td>
<td>-57.6</td>
<td>0.00</td>
<td>3</td>
<td>21</td>
<td>3'</td>
<td>5</td>
<td>37</td>
<td>6</td>
<td>2</td>
<td>45</td>
</tr>
<tr>
<td>aqc-miRNA156g</td>
<td>2.19E+09</td>
<td>AGAGGAGAGAGAGCACAACCC</td>
<td>-18.8</td>
<td>0.00</td>
<td>4</td>
<td>21</td>
<td>71</td>
<td>3'</td>
<td>5</td>
<td>37</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>aqc-miRNA156h</td>
<td>2.19E+09</td>
<td>AGAGGAGAGAGAGCACAACCC</td>
<td>-18.8</td>
<td>0.00</td>
<td>4</td>
<td>21</td>
<td>71</td>
<td>3'</td>
<td>5</td>
<td>37</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>aqc-miRNA156i</td>
<td>2.19E+09</td>
<td>AGAGGAGAGAGAGCACAACCC</td>
<td>-18.8</td>
<td>0.00</td>
<td>4</td>
<td>21</td>
<td>71</td>
<td>3'</td>
<td>5</td>
<td>37</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>aqc-miRNA156j</td>
<td>2.19E+09</td>
<td>AGAGGAGAGAGAGCACAACCC</td>
<td>-18.8</td>
<td>0.00</td>
<td>4</td>
<td>21</td>
<td>71</td>
<td>3'</td>
<td>5</td>
<td>37</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>aqc-miRNA156k</td>
<td>2.19E+09</td>
<td>AGAGGAGAGAGAGCACAACCC</td>
<td>-18.8</td>
<td>0.00</td>
<td>4</td>
<td>21</td>
<td>71</td>
<td>3'</td>
<td>5</td>
<td>37</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>aqc-miRNA156l</td>
<td>2.19E+09</td>
<td>AGAGGAGAGAGAGCACAACCC</td>
<td>-18.8</td>
<td>0.00</td>
<td>4</td>
<td>21</td>
<td>71</td>
<td>3'</td>
<td>5</td>
<td>37</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>aqc-miRNA156m</td>
<td>2.19E+09</td>
<td>AGAGGAGAGAGAGCACAACCC</td>
<td>-18.8</td>
<td>0.00</td>
<td>4</td>
<td>21</td>
<td>71</td>
<td>3'</td>
<td>5</td>
<td>37</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>aqc-miRNA156n</td>
<td>2.19E+09</td>
<td>AGAGGAGAGAGAGCACAACCC</td>
<td>-18.8</td>
<td>0.00</td>
<td>4</td>
<td>21</td>
<td>71</td>
<td>3'</td>
<td>5</td>
<td>37</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>aqc-miRNA156o</td>
<td>2.19E+09</td>
<td>AGAGGAGAGAGAGCACAACCC</td>
<td>-18.8</td>
<td>0.00</td>
<td>4</td>
<td>21</td>
<td>71</td>
<td>3'</td>
<td>5</td>
<td>37</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>aqc-miRNA156p</td>
<td>2.19E+09</td>
<td>AGAGGAGAGAGAGCACAACCC</td>
<td>-18.8</td>
<td>0.00</td>
<td>4</td>
<td>21</td>
<td>71</td>
<td>3'</td>
<td>5</td>
<td>37</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>aqc-miRNA156q</td>
<td>2.19E+09</td>
<td>AGAGGAGAGAGAGCACAACCC</td>
<td>-18.8</td>
<td>0.00</td>
<td>4</td>
<td>21</td>
<td>71</td>
<td>3'</td>
<td>5</td>
<td>37</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>aqc-miRNA156r</td>
<td>2.19E+09</td>
<td>AGAGGAGAGAGAGCACAACCC</td>
<td>-18.8</td>
<td>0.00</td>
<td>4</td>
<td>21</td>
<td>71</td>
<td>3'</td>
<td>5</td>
<td>37</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>aqc-miRNA156s</td>
<td>2.19E+09</td>
<td>AGAGGAGAGAGAGCACAACCC</td>
<td>-18.8</td>
<td>0.00</td>
<td>4</td>
<td>21</td>
<td>71</td>
<td>3'</td>
<td>5</td>
<td>37</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>aqc-miRNA156t</td>
<td>2.19E+09</td>
<td>AGAGGAGAGAGAGCACAACCC</td>
<td>-18.8</td>
<td>0.00</td>
<td>4</td>
<td>21</td>
<td>71</td>
<td>3'</td>
<td>5</td>
<td>37</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>aqc-miRNA156u</td>
<td>2.19E+09</td>
<td>AGAGGAGAGAGAGCACAACCC</td>
<td>-18.8</td>
<td>0.00</td>
<td>4</td>
<td>21</td>
<td>71</td>
<td>3'</td>
<td>5</td>
<td>37</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>aqc-miRNA156v</td>
<td>2.19E+09</td>
<td>AGAGGAGAGAGAGCACAACCC</td>
<td>-18.8</td>
<td>0.00</td>
<td>4</td>
<td>21</td>
<td>71</td>
<td>3'</td>
<td>5</td>
<td>37</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>aqc-miRNA156w</td>
<td>2.19E+09</td>
<td>AGAGGAGAGAGAGCACAACCC</td>
<td>-18.8</td>
<td>0.00</td>
<td>4</td>
<td>21</td>
<td>71</td>
<td>3'</td>
<td>5</td>
<td>37</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>aqc-miRNA156x</td>
<td>2.19E+09</td>
<td>AGAGGAGAGAGAGCACAACCC</td>
<td>-18.8</td>
<td>0.00</td>
<td>4</td>
<td>21</td>
<td>71</td>
<td>3'</td>
<td>5</td>
<td>37</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>aqc-miRNA156y</td>
<td>2.19E+09</td>
<td>AGAGGAGAGAGAGCACAACCC</td>
<td>-18.8</td>
<td>0.00</td>
<td>4</td>
<td>21</td>
<td>71</td>
<td>3'</td>
<td>5</td>
<td>37</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>aqc-miRNA156z</td>
<td>2.19E+09</td>
<td>AGAGGAGAGAGAGCACAACCC</td>
<td>-18.8</td>
<td>0.00</td>
<td>4</td>
<td>21</td>
<td>71</td>
<td>3'</td>
<td>5</td>
<td>37</td>
<td>6</td>
<td>2</td>
</tr>
</tbody>
</table>

**gi:** gene index number corresponding to *A. coerulea* whole genome sequence database

**MFE:** minimal folding free energy index

**NM:** number of mismatches in miRNA/miRNA* duplex

**ML:** Mature Length

**PL:** Precuror Length

**ARM:** Location of mature miRNA (3' or 5')
<table>
<thead>
<tr>
<th>miRNA precursor characteristics</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Average</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence Length</td>
<td>68</td>
<td>180</td>
<td>106</td>
<td>32.2</td>
</tr>
<tr>
<td>A (%)</td>
<td>16.4</td>
<td>36.3</td>
<td>25.6</td>
<td>4.4</td>
</tr>
<tr>
<td>U (%)</td>
<td>19.7</td>
<td>38.1</td>
<td>31.4</td>
<td>3.8</td>
</tr>
<tr>
<td>G (G)</td>
<td>13.8</td>
<td>31.5</td>
<td>22.9</td>
<td>3.7</td>
</tr>
<tr>
<td>C (%)</td>
<td>13.5</td>
<td>29.9</td>
<td>20.1</td>
<td>3.0</td>
</tr>
<tr>
<td>MFE (-kcal/mol)</td>
<td>18.8</td>
<td>81.8</td>
<td>46</td>
<td>14.7</td>
</tr>
<tr>
<td>AMFE (-kcal/mol)</td>
<td>26.5</td>
<td>58.7</td>
<td>44.1</td>
<td>7.5</td>
</tr>
<tr>
<td>MFEI</td>
<td>0.60</td>
<td>1.48</td>
<td>1.03</td>
<td>0.18</td>
</tr>
<tr>
<td>Target Family</td>
<td>Target Gene</td>
<td>5’END</td>
<td>3’END</td>
<td>ID</td>
</tr>
<tr>
<td>---------------</td>
<td>-------------</td>
<td>-------</td>
<td>-------</td>
<td>----</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT4G35370</td>
<td>AT4G35370</td>
<td>AT4G35370</td>
<td>DT744736</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT5G15690</td>
<td>AT5G15690</td>
<td>AT5G15690</td>
<td>DT744673</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT2G22610</td>
<td>AT2G22610</td>
<td>AT2G22610</td>
<td>DT744736</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT1G72830</td>
<td>AT1G72830</td>
<td>AT1G72830</td>
<td>DT744736</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT4G37970</td>
<td>AT4G37970</td>
<td>AT4G37970</td>
<td>DT744736</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT5G02220</td>
<td>AT5G02220</td>
<td>AT5G02220</td>
<td>DT744736</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT1G72830</td>
<td>AT1G72830</td>
<td>AT1G72830</td>
<td>DT744736</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT4G35370</td>
<td>AT4G35370</td>
<td>AT4G35370</td>
<td>DT744736</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT5G15690</td>
<td>AT5G15690</td>
<td>AT5G15690</td>
<td>DT744673</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT2G22610</td>
<td>AT2G22610</td>
<td>AT2G22610</td>
<td>DT744736</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT1G72830</td>
<td>AT1G72830</td>
<td>AT1G72830</td>
<td>DT744736</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT4G37970</td>
<td>AT4G37970</td>
<td>AT4G37970</td>
<td>DT744736</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT5G02220</td>
<td>AT5G02220</td>
<td>AT5G02220</td>
<td>DT744736</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT1G72830</td>
<td>AT1G72830</td>
<td>AT1G72830</td>
<td>DT744736</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT4G35370</td>
<td>AT4G35370</td>
<td>AT4G35370</td>
<td>DT744736</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT5G15690</td>
<td>AT5G15690</td>
<td>AT5G15690</td>
<td>DT744673</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT2G22610</td>
<td>AT2G22610</td>
<td>AT2G22610</td>
<td>DT744736</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT1G72830</td>
<td>AT1G72830</td>
<td>AT1G72830</td>
<td>DT744736</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT4G37970</td>
<td>AT4G37970</td>
<td>AT4G37970</td>
<td>DT744736</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT5G02220</td>
<td>AT5G02220</td>
<td>AT5G02220</td>
<td>DT744736</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT1G72830</td>
<td>AT1G72830</td>
<td>AT1G72830</td>
<td>DT744736</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT4G35370</td>
<td>AT4G35370</td>
<td>AT4G35370</td>
<td>DT744736</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT5G15690</td>
<td>AT5G15690</td>
<td>AT5G15690</td>
<td>DT744673</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT2G22610</td>
<td>AT2G22610</td>
<td>AT2G22610</td>
<td>DT744736</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT1G72830</td>
<td>AT1G72830</td>
<td>AT1G72830</td>
<td>DT744736</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT4G37970</td>
<td>AT4G37970</td>
<td>AT4G37970</td>
<td>DT744736</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT5G02220</td>
<td>AT5G02220</td>
<td>AT5G02220</td>
<td>DT744736</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT1G72830</td>
<td>AT1G72830</td>
<td>AT1G72830</td>
<td>DT744736</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT4G35370</td>
<td>AT4G35370</td>
<td>AT4G35370</td>
<td>DT744736</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT5G15690</td>
<td>AT5G15690</td>
<td>AT5G15690</td>
<td>DT744673</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT2G22610</td>
<td>AT2G22610</td>
<td>AT2G22610</td>
<td>DT744736</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT1G72830</td>
<td>AT1G72830</td>
<td>AT1G72830</td>
<td>DT744736</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT4G37970</td>
<td>AT4G37970</td>
<td>AT4G37970</td>
<td>DT744736</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT5G02220</td>
<td>AT5G02220</td>
<td>AT5G02220</td>
<td>DT744736</td>
</tr>
<tr>
<td>ATP5A1-like protein</td>
<td>AT1G72830</td>
<td>AT1G72830</td>
<td>AT1G72830</td>
<td>DT744736</td>
</tr>
</tbody>
</table>
nonredundant set of 1023 mature miRNAs

BLASTn

Sequences with 0-4 mismatches vs. BLASTn query

Predicted secondary structure using Mfold

Sequences with 0-4 mismatches in miRNA/miRNA*

Removed sequences with > -15kcal/mol folding energy and predicted precursor based on extent of pairing in secondary structure

Predicted Aquilegia miRNA precursors

Predicted Aquilegia miRNAs
## Supplementary Table 1

<table>
<thead>
<tr>
<th>microRNA family</th>
<th>aqc-miRNA gi</th>
<th>Mature</th>
<th>Genbank Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>aqc-miRNA156a</td>
<td>2185892723</td>
<td>TGACAGAAGATAGAGAGCAC</td>
<td></td>
</tr>
<tr>
<td>aqc-miRNA156b</td>
<td>2185518132</td>
<td>TGACAGAAGATAGAGAGCAC</td>
<td></td>
</tr>
<tr>
<td>aqc-miRNA159</td>
<td>2183893560</td>
<td>TTTTGACTGAAGGGAGCTCTA</td>
<td></td>
</tr>
<tr>
<td>aqc-miRNA160a</td>
<td>2185524891</td>
<td>TGCCCTGGCTCCCTGATGAGCA</td>
<td></td>
</tr>
<tr>
<td>aqc-miRNA160b</td>
<td>2185745223</td>
<td>TGCCCTGGCTCCCTGTATGCA</td>
<td></td>
</tr>
<tr>
<td>aqc-miRNA166a</td>
<td>2185638590</td>
<td>TCGGACCAGGCTTCATTCTCTC</td>
<td></td>
</tr>
<tr>
<td>aqc-miRNA166b</td>
<td>2185550821</td>
<td>TCGGACCAGGCTTCATTCCCCC</td>
<td></td>
</tr>
<tr>
<td>aqc-miRNA166c</td>
<td>2185683719</td>
<td>TCGGACCAGGCTTCATTCTC</td>
<td></td>
</tr>
<tr>
<td>aqc-miRNA166d</td>
<td>2185830185</td>
<td>TCGGACCAGGCTTCATTCTC</td>
<td></td>
</tr>
<tr>
<td>aqc-miRNA166e</td>
<td>2185503206</td>
<td>TCGGACCAGGCTTCATTCCCCC</td>
<td></td>
</tr>
<tr>
<td>aqc-miRNA167</td>
<td>2185494083</td>
<td>TCAAGCTGCCAGCATGATCTA</td>
<td></td>
</tr>
<tr>
<td>aqc-miRNA168</td>
<td>2185845089</td>
<td>TGGCTTAGTGCAGCTCGGGA</td>
<td></td>
</tr>
<tr>
<td>aqc-miRNA169a</td>
<td>2185604956</td>
<td>TAGCCAAAGGATGACTTCCTA</td>
<td></td>
</tr>
<tr>
<td>aqc-miRNA169b</td>
<td>2185687520</td>
<td>TAGCCAAAGGATGACTTGCTG</td>
<td></td>
</tr>
<tr>
<td>aqc-miRNA169c</td>
<td>2185493504</td>
<td>CAGCCAAAGGATGACTTGCCG</td>
<td></td>
</tr>
<tr>
<td>aqc-miRNA171a</td>
<td>2185558457</td>
<td>TGATTGAGCCGTCAGAATATC</td>
<td></td>
</tr>
<tr>
<td>aqc-miRNA171b</td>
<td>2185624710</td>
<td>TGATTGAGCCGTCAGAATATC</td>
<td></td>
</tr>
<tr>
<td>aqc-miRNA171c</td>
<td>2185689689</td>
<td>TAATTGAAACCGACTAAATC</td>
<td></td>
</tr>
<tr>
<td>aqc-miRNA171d</td>
<td>2185735537</td>
<td>TGATTGAGCCGTCAGAATATC</td>
<td></td>
</tr>
<tr>
<td>aqc-miRNA171e</td>
<td>2185815653</td>
<td>TGAATGAAACCGAGCAACATC</td>
<td></td>
</tr>
<tr>
<td>aqc-miRNA171f</td>
<td>2185468845</td>
<td>TAATTGAGCCGTCAGAATATC</td>
<td></td>
</tr>
<tr>
<td>aqc-miRNA172a</td>
<td>2185776282</td>
<td>AGAATCTTGATGATGCTGCAT</td>
<td></td>
</tr>
<tr>
<td>aqc-miRNA172b</td>
<td>2185788415</td>
<td>GGAATCTTGATGATGCTGCAT</td>
<td></td>
</tr>
<tr>
<td>aqc-miRNA319</td>
<td>2185799162</td>
<td>TTGGACTGAAGGGAGCTCCCT</td>
<td></td>
</tr>
<tr>
<td>aqc-miRNA395a</td>
<td>2185894728</td>
<td>CTGAAGGGTTTTGGAGGAACCTC</td>
<td></td>
</tr>
<tr>
<td>aqc-miRNA395b</td>
<td>2185728750</td>
<td>CTGAAGGGTTTTGGAGGAACCTC</td>
<td></td>
</tr>
<tr>
<td>aqc-miRNA396a</td>
<td>2185513426</td>
<td>TTCCACAGCCTTTCTGAACCTG</td>
<td></td>
</tr>
<tr>
<td>aqc-miRNA396b</td>
<td>2185617799</td>
<td>TTCCACAGCCTTTCTGAACCTT</td>
<td></td>
</tr>
</tbody>
</table>
aqc-miRNA398a 2185417724 TGTGTTCCTCAGGTCACCCTT
aqc-miRNA398b 2185579186 TGTGTTCCTCAGGTCGCCCTG
aqc-miRNA399 2185439501 TGCCAAAGGAGAGTGGCCCTA
aqc-miRNA408 2185474487 TGCACTGCCTCTTCTGCAC
aqc-miRNA477a 2185477245 CTCTCCCTCAAGGGCTTCTA
aqc-miRNA477b 2185552337 CTCTCCCTCAAGGGCTTCTA
aqc-miRNA477c 2185755300 CTCTCCCTCAAGTCTTCTCTA
aqc-miRNA477d 2185768332 CTCTCCCTCAAGGGCTTCTA
aqc-miRNA477e 2185874405 CTCTCCCTCAAGGGCTTCTA
aqc-miRNA477f 2185853634 CTCTCCCTCAAGGGCTTCTA
aqc-miRNA477g 2185809227 CTCTCCCTCAAGTCTTCTCTA
aqc-miRNA482a 2185586659 TCTTGCCGACTCCTCCCATACC
aqc-miRNA482b 2185874944 TCTTGCCGACTCCTCCCATACC
aqc-miRNA482c 2185892817 TCTTGCCGACTCCTCCCATACC
aqc-miRNA529 2185620929 AGAAGAGAGAGAGCACAACCC
aqc-miRNA530 2185481266 TGCACTGCCTGCATCTC
aqc-miRNA535 2185616263 TGACAACGAGAGAGCAGCGCG

gi: gene index number corresponding to *A. coerulea* whole genome sequence database
<table>
<thead>
<tr>
<th>microRNA family</th>
<th>Targeted Protein</th>
<th>Target Function</th>
<th>TC#</th>
<th>Top Scoring A.t. locus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Squamosa promoter binding protein-like 2 (SBP)</td>
<td>Transcript factor</td>
<td>TC31380, TC24142</td>
<td>AT5G43270</td>
<td></td>
</tr>
<tr>
<td>Squamosa promoter binding protein-like 6 (SBP)</td>
<td>Transcript factor</td>
<td>TC26823, TC21170, TC1808, DR954092</td>
<td>AT1G69170</td>
<td></td>
</tr>
<tr>
<td>Squamosa promoter binding protein-like 13 (SBP)</td>
<td>Transcript factor</td>
<td>TC32167, DR949672</td>
<td>AT5G50570</td>
<td></td>
</tr>
<tr>
<td>Squamosa promoter binding protein-like 14 (SBP)</td>
<td>Transcript factor</td>
<td>TC27107</td>
<td>AT1G20980</td>
<td></td>
</tr>
<tr>
<td>Squamosa promoter binding protein-like 12 (SBP)</td>
<td>Transcript factor</td>
<td>TC31356</td>
<td>AT3G60030</td>
<td></td>
</tr>
<tr>
<td>Growth Regulating Factor 2 (GRF2)</td>
<td>Transcription factor</td>
<td>TC32167</td>
<td>AT5G50570</td>
<td></td>
</tr>
<tr>
<td>Squamosa promoter binding protein-like 13 (SBP)</td>
<td>Transcription factor</td>
<td>TC32167</td>
<td>AT5G50570</td>
<td></td>
</tr>
<tr>
<td>Squamosa promoter binding protein-like 14 (SBP)</td>
<td>Transcription factor</td>
<td>TC27107</td>
<td>AT1G20980</td>
<td></td>
</tr>
<tr>
<td>Squamosa promoter binding protein-like 12 (SBP)</td>
<td>Transcription factor</td>
<td>TC31356</td>
<td>AT3G60030</td>
<td></td>
</tr>
<tr>
<td>Growth Regulating Factor 2 (GRF2)</td>
<td>Transcription factor</td>
<td>TC32167</td>
<td>AT5G50570</td>
<td></td>
</tr>
<tr>
<td>Squamosa promoter binding protein-like 13 (SBP)</td>
<td>Transcription factor</td>
<td>TC32167</td>
<td>AT5G50570</td>
<td></td>
</tr>
<tr>
<td>Squamosa promoter binding protein-like 14 (SBP)</td>
<td>Transcription factor</td>
<td>TC27107</td>
<td>AT1G20980</td>
<td></td>
</tr>
<tr>
<td>Squamosa promoter binding protein-like 12 (SBP)</td>
<td>Transcription factor</td>
<td>TC31356</td>
<td>AT3G60030</td>
<td></td>
</tr>
<tr>
<td>Growth Regulating Factor 2 (GRF2)</td>
<td>Transcription factor</td>
<td>TC32167</td>
<td>AT5G50570</td>
<td></td>
</tr>
</tbody>
</table>

**None from EST database**

<table>
<thead>
<tr>
<th>microRNA family</th>
<th>Targeted Protein</th>
<th>Target Function</th>
<th>TC#</th>
<th>Top Scoring A.t. locus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cup-Shaped Cotyledon 2 (CUC2)</td>
<td>Transcription factor</td>
<td>DR949423, TC27953</td>
<td>AT5G50570</td>
<td></td>
</tr>
<tr>
<td>NAC Domain containing protein</td>
<td>Metabolism</td>
<td>TC27854</td>
<td>AT3G39610</td>
<td></td>
</tr>
<tr>
<td>Leucine-rich repeat transmembrane protein kinase</td>
<td>Metabolism</td>
<td>DT731408</td>
<td>AT3G28040</td>
<td></td>
</tr>
<tr>
<td>UDP-glucose pyrophosphorylase</td>
<td>Stress</td>
<td>TC21677</td>
<td>AT3G03250</td>
<td></td>
</tr>
<tr>
<td>4-coumarate-CoA ligase family protein</td>
<td>Stress</td>
<td>DT741735, TC24498</td>
<td>AT1G20510</td>
<td></td>
</tr>
<tr>
<td>Class III HD-Zip protein</td>
<td>Transcription factor</td>
<td>TC30761</td>
<td>AT1G30330</td>
<td></td>
</tr>
<tr>
<td>Auxin response factor 6</td>
<td>Transcription factor</td>
<td>TC29952, TC4356</td>
<td>AT3G03250</td>
<td></td>
</tr>
<tr>
<td>Anion Transporter 2</td>
<td>Transport</td>
<td>TC29087</td>
<td>AT4G00370</td>
<td></td>
</tr>
<tr>
<td>CCAAT-binding factor (HAP2)</td>
<td>Transcription factor</td>
<td>DR924236</td>
<td>AT1G72830</td>
<td></td>
</tr>
<tr>
<td>Lipase class 3 family protein</td>
<td>Metabolism</td>
<td>TC31138, DR927955</td>
<td>AT5G18640</td>
<td></td>
</tr>
<tr>
<td>Scarecrow-like protein</td>
<td>Transcription factor</td>
<td>TC21487, TC28359, TC25185</td>
<td>AT2G45160</td>
<td></td>
</tr>
<tr>
<td>Arabidopsis Pumilio 2 (APUM2)-like</td>
<td>RNA binding</td>
<td>TC29397</td>
<td>AT2G29190</td>
<td></td>
</tr>
<tr>
<td>APETALA2</td>
<td>Transcription factor</td>
<td>DT732637, TC28864, TC22964, TC24245</td>
<td>AT4G36920</td>
<td></td>
</tr>
<tr>
<td>Target of EAT 1 (TOE1)</td>
<td>Transcription factor</td>
<td>DT741192, TC28449, TC32032</td>
<td>AT2G28550</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>Unknown</td>
<td>DR954144</td>
<td>AT3G65290</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>Unknown</td>
<td>TC6960</td>
<td>AT3G48070</td>
<td></td>
</tr>
<tr>
<td>Spike1 (SP1)</td>
<td>Metabolism</td>
<td>DT764297</td>
<td>AT4G16340</td>
<td></td>
</tr>
<tr>
<td>SC35-like splicing factor</td>
<td>Metabolism</td>
<td>TC22146</td>
<td>AT3G13570</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>Metabolism</td>
<td>DT749684</td>
<td>AT3G10070</td>
<td></td>
</tr>
<tr>
<td>Ubiquitin-conjugating enzyme</td>
<td>Metabolism</td>
<td>TC25860, TC21155</td>
<td>AT5G05080</td>
<td></td>
</tr>
<tr>
<td>F-box protein</td>
<td>Metabolism</td>
<td>TC28027</td>
<td>AT4G02440</td>
<td></td>
</tr>
<tr>
<td>Selenium binding protein</td>
<td>Stess</td>
<td>TC29593</td>
<td>AT2G24440</td>
<td></td>
</tr>
</tbody>
</table>

**None from EST database**

<table>
<thead>
<tr>
<th>microRNA family</th>
<th>Targeted Protein</th>
<th>Target Function</th>
<th>TC#</th>
<th>Top Scoring A.t. locus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diminuto 1, Enhanced Very-Low-Fluence Responses 1</td>
<td>Photoprocesses</td>
<td>TC25461, DR922754</td>
<td>AT3G19820</td>
<td></td>
</tr>
<tr>
<td>Cation symporter</td>
<td>Transport</td>
<td>TC30528</td>
<td>AT5G03560</td>
<td></td>
</tr>
<tr>
<td>Auxin Signaling F-BOX 2</td>
<td>Auxin</td>
<td>TC29517</td>
<td>AT3G26810</td>
<td></td>
</tr>
<tr>
<td>Transport Inhibitor Response 1 (TIR1)</td>
<td>Auxin</td>
<td>TC20718, TC25434, DR913737</td>
<td>AT3G26810</td>
<td></td>
</tr>
<tr>
<td>F-box protein</td>
<td>Unknown</td>
<td>TC24293, TC26091</td>
<td>AT1G27340</td>
<td></td>
</tr>
<tr>
<td>Kinase</td>
<td>Metabolism</td>
<td>TC32489</td>
<td>AT4G00955</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>Unknown</td>
<td>TC20934</td>
<td>AT3G49590</td>
<td></td>
</tr>
<tr>
<td>ATP sulfurylase</td>
<td>Stress</td>
<td>TC21492</td>
<td>AT3G22890</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>Unknown</td>
<td>TC30052</td>
<td>AT2G02370</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>Unknown</td>
<td>DR926810</td>
<td>AT2G16760</td>
<td></td>
</tr>
<tr>
<td>Calmodulin-binding protein</td>
<td>Stress</td>
<td>TC29396</td>
<td>AT3G16940</td>
<td></td>
</tr>
<tr>
<td>Mildew Resistance Locus</td>
<td>Stress</td>
<td>TC32094</td>
<td>AT1G11000</td>
<td></td>
</tr>
<tr>
<td>Gene Name</td>
<td>Function</td>
<td>Accession</td>
<td>ID</td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------</td>
<td>-------------------</td>
<td>-------------</td>
<td>---------------</td>
<td></td>
</tr>
<tr>
<td>1-aminocyclopropane-1-carboxylate oxidase,</td>
<td>Stress</td>
<td>TC20841</td>
<td>AT1G05010</td>
<td></td>
</tr>
<tr>
<td>Ethylene forming enzyme</td>
<td></td>
<td>TC22590</td>
<td>AT2G18560</td>
<td></td>
</tr>
<tr>
<td>UDP-glucosyl transferase family protein</td>
<td></td>
<td>TC26926</td>
<td>AT5G01190</td>
<td></td>
</tr>
<tr>
<td>Laccase</td>
<td>Metabolism</td>
<td>TC26200</td>
<td>AT2G38080</td>
<td></td>
</tr>
<tr>
<td>Laccase</td>
<td>Metabolism</td>
<td>TC24152</td>
<td>AT2G39805</td>
<td></td>
</tr>
<tr>
<td>Yip1 family protein</td>
<td>Unknown</td>
<td>DT728841</td>
<td>AT5G3740</td>
<td></td>
</tr>
<tr>
<td>None</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UDP-glucosyl transferase family protein</td>
<td></td>
<td>TC22590</td>
<td>AT5G01190</td>
<td></td>
</tr>
<tr>
<td>Laccase</td>
<td>Metabolism</td>
<td>TC26200</td>
<td>AT2G38080</td>
<td></td>
</tr>
<tr>
<td>Yip1 family protein</td>
<td>Unknown</td>
<td>DT728841</td>
<td>AT5G3740</td>
<td></td>
</tr>
<tr>
<td>None</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>beta-fructofuranosidase</td>
<td>Metabolism</td>
<td>TC28894</td>
<td>AT1G65660</td>
<td></td>
</tr>
<tr>
<td>ABA receptor</td>
<td>Metabolism</td>
<td>TC20517</td>
<td>AT2G20770</td>
<td></td>
</tr>
<tr>
<td>Xyloglucan Endotransglycosylase 6</td>
<td>Metabolism</td>
<td>TC30505</td>
<td>AT4G25810</td>
<td></td>
</tr>
<tr>
<td>Plasticid glucose-6-phosphate dehydrogenase</td>
<td>Metabolism</td>
<td>TC30732</td>
<td>AT5G35790</td>
<td></td>
</tr>
<tr>
<td>Elongation Factor 1B-gamma</td>
<td>Metabolism</td>
<td>TC20303</td>
<td>AT1G09640</td>
<td></td>
</tr>
<tr>
<td>rRNA processing protein</td>
<td>Metabolism</td>
<td>TC24529</td>
<td>AT5G61330</td>
<td></td>
</tr>
<tr>
<td>None</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Squamosa promoter binding protein-like 6 (SBP)</td>
<td>Transcription</td>
<td>DR954092,</td>
<td>AT1G69170</td>
<td></td>
</tr>
<tr>
<td>Squamosa promoter binding protein-like 13 (SBP)</td>
<td>Transcription</td>
<td>DR949672,</td>
<td>AT5G0570</td>
<td></td>
</tr>
<tr>
<td>Osephin family protein</td>
<td>Unknown</td>
<td>TC20719</td>
<td>AT5G4130</td>
<td></td>
</tr>
<tr>
<td>Squamosa promoter binding protein-like 14 (SBP)</td>
<td>Transcription</td>
<td>TC21707</td>
<td>AT1G20980</td>
<td></td>
</tr>
<tr>
<td>Squamosa promoter binding protein-like 2 (SBP)</td>
<td>Transcription</td>
<td>TC31380,</td>
<td>AT5G43270</td>
<td></td>
</tr>
<tr>
<td>Threonine endopeptidase</td>
<td>Unknown</td>
<td>TC32968</td>
<td>AT4G01150</td>
<td></td>
</tr>
<tr>
<td>FAD-binding domain-containing protein</td>
<td>Metabolism</td>
<td>DR933604</td>
<td>AT4G20800</td>
<td></td>
</tr>
<tr>
<td>Diacylglycerol kinase gene family</td>
<td>Stress</td>
<td>TC29483</td>
<td>AT5G63770</td>
<td></td>
</tr>
<tr>
<td>Metal Tolerance Protein A2</td>
<td>Stress</td>
<td>DT745966,</td>
<td>AT5G58810</td>
<td></td>
</tr>
<tr>
<td>Serine carboxypeptidase-like 26</td>
<td>Metabolism</td>
<td>TC23269</td>
<td>AT2G35780</td>
<td></td>
</tr>
<tr>
<td>None</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phosphoenolpyruvate carboxylase</td>
<td>Metabolism</td>
<td>DR942629</td>
<td>AT3G14940</td>
<td></td>
</tr>
<tr>
<td>Methionine gamma-lyase</td>
<td>Metabolism</td>
<td>TC32408,</td>
<td>AT1G64660</td>
<td></td>
</tr>
<tr>
<td>Glycoside hydrolase family protein</td>
<td>Metabolism</td>
<td>TC37593</td>
<td>AT3G14940</td>
<td></td>
</tr>
<tr>
<td>Carrier family protein</td>
<td>Transport</td>
<td>TC33228</td>
<td>AT4G27940</td>
<td></td>
</tr>
<tr>
<td>UDP-glucuronic acid decarboxylase</td>
<td>Metabolism</td>
<td>TC27947</td>
<td>AT3G35520</td>
<td></td>
</tr>
</tbody>
</table>
Supplemental Figure 1

A. Size of miRNA families identified in *Aquilegia coerulea*

B. Size distribution of *Aquilegia coerulea* miRNA predicted precursors
microRNA156

Formatted Alignments

| 218589272mature | U G A C A G A A G A | A G A G A G C A C |
| DR949672binding | U G A C A G A A G A | A G A G A G C A C |
| DR954092binding | U G A C A G A A G A | A G A G A G C A C |
| TC21170binding | U G A C A G A A G A | A G A G A G C A C |
| TC21707binding | U G A C A G A A G A | A G A G A G C A C |
| TC21808binding | U G A C A G A A G A | A G A G A G C A C |
| TC24142binding | U G A C A G A A G A | A G A G A G C A C |
| TC26823binding | U G A C A G A A G A | A G A G A G C A C |
| TC21380binding | U G A C A G A A G A | A G A G A G C A C |
| TC21827binding | U G A C A G A A G A | A G A G A G C A C |
| TC32167binding | U G A C A G A A G A | A G A G A G C A C |
| TC25281binding | U G A C A G A A G A | A G A G A G C A C |

Glycine max
Medicago truncatula
Gossypium hirsutum
Arabidopsis thaliana
Brassica rapa
Brassica rapa
Sorghum bicolor
Oryza sativa
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Oryza sativa
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max

Number of miRNAs
microRNA156

dG = -43.40 [initially -43.40] 09Apr06-08-20-30
microRNA156

dG = -44.80 [initially -44.80] 09Apr06-08-46-26
microRNA159

Formatted Alignments

2183893560mature

U U U G G A C U G A A G G G A G C U C U A

U U U G G A C U G A A G G G A G C U C U A

Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sativa
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max

Number of miRNAs

0 1 2 3 4 5 6 7
microRNA159

dG = -18.80 [initially -18.80] 09Mar17-07-23-03

microRNA159
microRNA160

\[ dG = -50.10 \text{ [initially -50.10]} \ 09\text{Mar13-14-54-08} \]
microRNA160

Output of sir_graph (®)
mfold 3.4


microRNA160
microRNA164

- EST1175584 DT741735
- EST1165258 DT731408
- TC27854
- EST1140962 DR949423
- TC27953
- TC21677
- TC24498

20

UGAAGGACAGUAGAGACACAGACG
UGGAGAAGGACAGGUCAGACG
UGGAGAAGGACAGGUCAGACG
UGGAGAAGGACAGGUCAGACG
UGGAGAAGGACAGGUCAGACG
UGGAGAAGGACAGGUCAGACG
UGGAGAAGGACAGGUCAGACG

164

- Physcomitrella patens
- Selaginella moellendorffii
- Pinus taeda
- Zea mays
- Sorghum bicolor
- Oryza sativa
- Triticum aestivum
- Aquilegia coerulea
- Solanum lycopersicum
- Vitis vinifera
- Populus trichocarpa
- Brassica napus
- Brassica rapa
- Arabidopsis thaliana
- Gossypium hirsutum
- Medicago truncatula
- Glycine max

Number of miRNAs
microRNA166

2185503206mature
2185638590mature
gnl|ti|2185550821mature
gnl|ti|2185683719mature
gnl|ti|2185830185mature
TC30761binding

20

U C G G A C C A G G C U U C A U U C C C C
U C G G A C C A G G C U U C A U U C C C C
U C G G A C C A G G C U U C A U U C C C C
U C G G A C C A G G C U U C A U U C C C C
U C G G A C C A G G C U U C A U U C C C C
U C G G A C C A G G C U U C A U U C C C C
C C G G A C C A G G C U U C A U C C C A C G
U C G G A C C A G G C U U C A U U C C C C

Number of miRNAs

Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sativa
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max

Glycine max
Medicago truncatula
Solanum lycopersicum
Aquilegia coerulea
Triticum aestivum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max
microRNA166

$dG = -49.10$ [initially -49.10] 09Mar16-09-34-45
microRNA166

dG = -44.50 [initially -44.50] 09Mar31-10-35-05
microRNA166

\[ dG = -46.60 \text{ [initially -46.60]} \] 09Mar31-10-48-25
$dG = -42.93$ [initially $-45.50$] 09Mar31-15-54-38
microRNA166

Output of sir_graph

mfold 3.4

dG = -57.55 [initially -60.30] 09Mar16-09-20-37

microRNA166
microRNA167

2185494083mature
TC24356binding
TC29952binding

Glycine max
Medicago truncatula
Gossypium hirsutum
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Sorghum bicolor
Solanum lycopersicum
Zea mays
Physcomitrella patens
Selaginella moellendorfii
Pinus taeda

Number of miRNAs

167
microRNA167

\[ dG = -31.50 \text{ [initially -31.50]} \]

09Mar16-06-50-55
microRNA168

2185845089mature
DT747736binding
TC29087binding

<table>
<thead>
<tr>
<th>Phycomitrella patens</th>
<th>Selaginella moellendorffii</th>
<th>Physomitrella patens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pinus taeda</td>
<td>Zea mays</td>
<td>Pinus taeda</td>
</tr>
<tr>
<td>Sorghum bicolor</td>
<td>Oryza sativa</td>
<td>Sorghum bicolor</td>
</tr>
<tr>
<td>Triticum aestivum</td>
<td>Aquilegia coerulea</td>
<td>Triticum aestivum</td>
</tr>
<tr>
<td>Solanum lycopersicum</td>
<td>Vitis vinifera</td>
<td>Solanum lycopersicum</td>
</tr>
<tr>
<td>Populus trichocarpa</td>
<td>Brassica napus</td>
<td>Populus trichocarpa</td>
</tr>
<tr>
<td>Brassica rapa</td>
<td>Arabidopsis thaliana</td>
<td>Brassica rapa</td>
</tr>
<tr>
<td>Gossypium hirsutum</td>
<td>Medicago truncatula</td>
<td>Gossypium hirsutum</td>
</tr>
<tr>
<td>Glycine max</td>
<td></td>
<td>Glycine max</td>
</tr>
</tbody>
</table>

Number of miRNAs

168
microRNA168

\[ \text{dG} = -40.45 \text{ [initially -42.70]} \ 09\text{Apr}01-09-19-45 \]
microRNA169

2185493504mature
| gnl|ti|2185604956mature
| gnl|ti|2185687520mature
| DR924236binding
| DR927955binding
| TC31138binding

- C A G C C A A G G A U G A C U U G C C G G

- U A G C C A A G A U G A C U U G C C U G G

- A A G C C A A G A U G A A U U G C C U A

- A A G U C A A G A U G A C U U C C U A

- A A G U C A A G A U G A C U U C C U A

GWAGCGACUCUCUCUA

169

0 5 10 15 20 25 30 35

Number of miRNAs

Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sativa
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max

microRNA169
dG = -75.30 [initially -75.30] 09Apr02-06-50-01
$dG = -78.15$ [initially -81.10] 09Apr02-07-05-49
microRNA169

$dG = -47.10 \text{ [initially -47.10]}$ 09Mar16-10-45-38

microRNA169
microRNA170/171

2185468845mature
2185568457mature
2185624710mature
2185689689mature
2185735537mature
2185815653mature
TC21487binding
TC25185binding
TC28359binding
TC29397binding

<table>
<thead>
<tr>
<th>Species</th>
<th>Number of miRNAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physcomitrella patens</td>
<td>1</td>
</tr>
<tr>
<td>Selaginella moellendorffii</td>
<td>1</td>
</tr>
<tr>
<td>Pinus taeda</td>
<td>12</td>
</tr>
<tr>
<td>Zea mays</td>
<td>12</td>
</tr>
<tr>
<td>Sorghum bicolor</td>
<td>10</td>
</tr>
<tr>
<td>Oryza sativa</td>
<td>10</td>
</tr>
<tr>
<td>Triticum aestivum</td>
<td>8</td>
</tr>
<tr>
<td>Aquilegia coerulea</td>
<td>8</td>
</tr>
<tr>
<td>Solanum lycopersicium</td>
<td>8</td>
</tr>
<tr>
<td>Vitis vinifera</td>
<td>8</td>
</tr>
<tr>
<td>Populus trichocarpa</td>
<td>12</td>
</tr>
<tr>
<td>Brassica napus</td>
<td>8</td>
</tr>
<tr>
<td>Brassica rapa</td>
<td>8</td>
</tr>
<tr>
<td>Arabidopsis thaliana</td>
<td>8</td>
</tr>
<tr>
<td>Gossypium hirsutum</td>
<td>8</td>
</tr>
<tr>
<td>Medicago truncatula</td>
<td>8</td>
</tr>
<tr>
<td>Glycine max</td>
<td>2</td>
</tr>
</tbody>
</table>

Number of miRNAs
microRNA170/171

Output of sir_graph (®)

mfold 3.4

dG = -46.30 [initially -46.30] 09Mar16-11-29-08

microRNA170/171
microRNA170/171

Output of sir_graph (®)

```
dG = -36.60 [initially -36.60] 09Mar16-11-35-59
```
microRNA170/171

$G = -27.30 \text{ [initially -27.30]} \ 09\text{Mar}16-11-42-19$

microRNA170/171
Output of sir_graph (®)

mfold 3.4


2185735537
microRNA170/171

Output of sir_graph (®)

$dG = -30.50$ [initially -30.50] 09Mar16-11-54-28

microRNA170/171
microRNA170/171

$dG = -41.10 \ [\text{initially} -41.10] \ 09\text{Mar}16-11-20-48$
microRNA172

![Bar chart showing the number of miRNAs in different species.](chart.png)

Species:
- Physcomitrella patens
- Selaginella moellendorffii
- Pinus taeda
- Zea mays
- Sorghum bicolor
- Oryza sativa
- Triticum aestivum
- Aquilegia coerulea
- Solanum lycopersicum
- Vitis vinifera
- Populus trichocarpa
- Brassica napus
- Brassica rapa
- Arabidopsis thaliana
- Gossypium hirsutum
- Medicago truncatula
- Glycine max

Number of miRNAs:
- Physcomitrella patens: 7
- Selaginella moellendorffii: 2
- Pinus taeda: 3
- Zea mays: 5
- Sorghum bicolor: 4
- Oryza sativa: 4
- Triticum aestivum: 4
- Aquilegia coerulea: 2
- Solanum lycopersicum: 4
- Vitis vinifera: 2
- Populus trichocarpa: 9
- Brassica napus: 2
- Brassica rapa: 2
- Arabidopsis thaliana: 5
- Gossypium hirsutum: 2
- Medicago truncatula: 2
- Glycine max: 2

microRNA172
microRNA172

\[ dG = -43.50 \text{ [initially -43.50]} \] 09May26-09-40-08

microRNA172
microRNA319

2185799162mature
DR922754binding
TC25461binding

319

Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sativa
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max
microRNA319

Output of sir_graph (R)
mfold 3.4

\[ dG = -78.60 \text{ [initially -78.60]} \]

09Mar20-05-36-52
MicroRNA390

Formatted Alignments


0 1 2 3 4 5

Number of miRNAs

Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sativa
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max
microRNA393

TC20718binding  U C C A A A G G A U C G C A U U G U C U C
TC25434binding  U C C A A A G G A U C G C A U U G U C U C
DR913737binding U C C A A A G G A U C G C A U U G U C U C
TC29517binding  U C C A A A G G A U C G C A U U G U C U C

<table>
<thead>
<tr>
<th>Plant Species</th>
<th>Number of miRNAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physcomitrella patens</td>
<td>3</td>
</tr>
<tr>
<td>Selaginella moellendorffii</td>
<td>1</td>
</tr>
<tr>
<td>Pinus taeda</td>
<td>1</td>
</tr>
<tr>
<td>Zea mays</td>
<td>1</td>
</tr>
<tr>
<td>Sorghum bicolor</td>
<td>1</td>
</tr>
<tr>
<td>Oryza sativa</td>
<td>2</td>
</tr>
<tr>
<td>Triticum aestivum</td>
<td>2</td>
</tr>
<tr>
<td>Aquilegia coerulea</td>
<td>2</td>
</tr>
<tr>
<td>Solanum lycopersicum</td>
<td>2</td>
</tr>
<tr>
<td>Vitis vinifera</td>
<td>3</td>
</tr>
<tr>
<td>Populus trichocarpa</td>
<td>3</td>
</tr>
<tr>
<td>Brassica napus</td>
<td>1</td>
</tr>
<tr>
<td>Brassica rapa</td>
<td>1</td>
</tr>
<tr>
<td>Arabidopsis thaliana</td>
<td>4</td>
</tr>
<tr>
<td>Gossypium hirsutum</td>
<td>2</td>
</tr>
<tr>
<td>Medicago truncatula</td>
<td>1</td>
</tr>
<tr>
<td>Glycine max</td>
<td>2</td>
</tr>
</tbody>
</table>

Number of miRNAs
microRNA394

Formatted Alignments

TC24293binding
TC26091binding

20

U U G G C A U U C U G U C A A C C U C C
U U G G C A U U C U G U C A A C C U C C
U U G G C A U U C U G U C A A C C U C C

Glycine max
Medicago truncatula
Gossypium hirsutum
Arabidopsis thaliana
Brassica rapa
Brassica napus
Populus trichocarpa
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sativa
Triticum aestivum
Number of miRNAs
microRNA395

Formatted Alignments

<table>
<thead>
<tr>
<th>mature</th>
<th>binding</th>
</tr>
</thead>
<tbody>
<tr>
<td>gnl</td>
<td>ti</td>
</tr>
<tr>
<td>TC20934binding</td>
<td>TC21492binding</td>
</tr>
<tr>
<td>TC30052binding</td>
<td>TC32489binding</td>
</tr>
</tbody>
</table>

20
C U G A A G G G U U U G A G A A C U C

C U G A A G G G U U U G A G A A C U C

G A A A A U U G G A G A A C U C

C U G A A G G G U U U G A G A A C U C

G A A A A U U G G A G A A C U C

395

Number of miRNAs

Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sativa
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max

microRNA395
microRNA395

dG = -43.60 [initially -43.60] 09Apr06-14-21-02

microRNA395
microRNA395

\[ dG = -37.70 \] (initially -37.70) 09Apr06-14:50:03

microRNA395
**Formatted Alignments**

<table>
<thead>
<tr>
<th>miRNA</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>microRNA396</td>
<td></td>
</tr>
<tr>
<td>2185617799mature</td>
<td>U U C A C A G C U U C U G A A C U U</td>
</tr>
<tr>
<td>2185513426mature</td>
<td>U U C A C A G C U U C U G A A C U G</td>
</tr>
<tr>
<td>DR926810binding</td>
<td>U C C A C A G C U U C U U G A A C A U</td>
</tr>
<tr>
<td>TC20841binding</td>
<td>U U C A C A G C U U C U G A A C U C C</td>
</tr>
<tr>
<td>TC22590binding</td>
<td>U U C G A C A G C U U C U U C A U C C C</td>
</tr>
<tr>
<td>TC27254binding</td>
<td>U C G C A A A C U U C U U G A A C U U</td>
</tr>
<tr>
<td>TC29396binding</td>
<td>U U C A C A G C U U C U U G A A C U U</td>
</tr>
<tr>
<td>TC32094binding</td>
<td>U U C A C A G C U U C U U G A A C U U</td>
</tr>
</tbody>
</table>

**Number of miRNAs**

- **Physcomitrella patens**: 3
- **Selaginella moellendorffii**: 2
- **Pinus taeda**: 3
- **Zea mays**: 4
- **Sorghum bicolor**: 3
- **Oryza sativa**: 6
- **Triticum aestivum**: 4
- **Aquilegia coerulea**: 2
- **Solanum lycopersicum**: 3
- **Vitis vinifera**: 4
- **Populus trichocarpa**: 7
- **Brassica napus**: 2
- **Brassica rapa**: 2
- **Arabidopsis thaliana**: 2
- **Gossypium hirsutum**: 2
- **Medicago truncatula**: 2
- **Glycine max**: 3
Output of sir_graph (®)
mfold 3.4

\[ dG = -38.60 \text{ [initially -38.60]} \]

09Mar18-12-06-52

microRNA396
Output of sir_graph (®)
mfold 3.4

dG = -41.23 [initially -47.50] 09Mar18-12-04-12
**Formatted Alignments**

<table>
<thead>
<tr>
<th>Binding</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>TC26200</td>
<td>UUAUGGCA</td>
</tr>
<tr>
<td>TC26926</td>
<td>UCACUGGCA</td>
</tr>
<tr>
<td>DT728841</td>
<td>UCAGUUGCA</td>
</tr>
<tr>
<td>DT743386</td>
<td>GCAGUUGCA</td>
</tr>
<tr>
<td>TC24152</td>
<td>GCAGUUGCA</td>
</tr>
</tbody>
</table>

Number of miRNAs: 397

- Physcomitrella patens
- Selaginella moellendorffii
- Pinus taeda
- Zea mays
- Sorghum bicolor
- Oryza sativa
- Triticum aestivum
- Aquilegia coerulea
- Solanum lycopersicum
- Vitis vinifera
- Populus trichocarpa
- Brassica napus
- Brassica rapa
- Arabidopsis thaliana
- Gossypium hirsutum
- Medicago truncatula
- Glycine max

**Number of miRNAs**

0 1 2 3 4
microRNA398

2185417724mature
2185579186mature

Glycine max
Medicago truncatula
Gossypium hirsutum
Arabidopsis thaliana
Brassica rapa
Brassica napus
Sorghum bicolor
Oryza sativa
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays

Number of miRNAs

0 1 2 3 4
microRNA398

dG = -38.50 [initially -38.50] 09Mar18-21-55-00
Output of sir_graph (®)
mfold 3.4

dG = -42.90 [initially -42.90] 09Mar18-21-59-43
microRNA399

Formatted Alignments

2185439501mature
TC20517binding
TC28894binding

Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sativa
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max

Number of miRNAs
microRNA399

$dG = -49.30$ [initially $-49.30]$ 09Apr10-10-41-59

microRNA399
microRNA408

Formatted Alignments

![Formatted Alignments Image]

Number of miRNAs

<table>
<thead>
<tr>
<th>Species</th>
<th>Number of miRNAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physcomitrella patens</td>
<td>2</td>
</tr>
<tr>
<td>Selaginella moellendorffii</td>
<td>1</td>
</tr>
<tr>
<td>Pinus taeda</td>
<td>1</td>
</tr>
<tr>
<td>Zea mays</td>
<td>1</td>
</tr>
<tr>
<td>Sorghum bicolor</td>
<td>1</td>
</tr>
<tr>
<td>Oryza sativa</td>
<td>1</td>
</tr>
<tr>
<td>Triticum aestivum</td>
<td>1</td>
</tr>
<tr>
<td>Aquilegia coerulea</td>
<td>1</td>
</tr>
<tr>
<td>Solanum lycopersicum</td>
<td>1</td>
</tr>
<tr>
<td>Vitis vinifera</td>
<td>1</td>
</tr>
<tr>
<td>Populus trichocarpa</td>
<td>1</td>
</tr>
<tr>
<td>Brassica napus</td>
<td>1</td>
</tr>
<tr>
<td>Brassica rapa</td>
<td>1</td>
</tr>
<tr>
<td>Arabidopsis thaliana</td>
<td>1</td>
</tr>
<tr>
<td>Gossypium hirsutum</td>
<td>1</td>
</tr>
<tr>
<td>Medicago truncatula</td>
<td>1</td>
</tr>
<tr>
<td>Glycine max</td>
<td>1</td>
</tr>
</tbody>
</table>

microRNA408
microRNA408

\[ dG = -46.50 \text{ [initially -46.50]} \]

09Apr10-10-50-08
microRNA477

Formatted Alignments

2185809227mature C U C U C C C U C A A G U U C U U C U A
2185853634mature C U C U C C U C A A G G C U U C U A
218574405mature C U C U C C C C A A G G G C U U C U A
2185477245mature C U C U C C C C A A G G G C U U C U A
2185552337mature C U C U C C C C A A G G G C U U C U A
2185755300mature C U C U C C C C A A G G G C U U C U A
TC203033binding C U C U C C C C A A G G G C U U C U A
TC24529binding C U C U C U C A A G G G U U C U U U U
TC30732binding C U C U C C C C A A G G G C U U C U A

477

Phycomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sativa
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max

Number of miRNAs
microRNA477

$dG = -40.60 \text{ [initially } -40.60\text{] 09Mar19-10-24-15}$
microRNA477

\[ dG = 36.40 \text{ [initially 36.40]} \]

09Mar19-10-29-38
microRNA477

\[ dG = -38.00 \text{ [initially -38.00]} \]

09Mar19-09-23-29
dG = -27.80 [initially -27.80] 09Mar19-09-30-33
microRNA477

\[ \text{dG} = -40.40 \text{ [initially -40.40]} \]

09Mar19-09-53-09
microRNA477

microRNA477

$\text{dG} = -40.40 \ [\text{initially } -40.40] \ 09\text{Mar19-10-18-53}$
microRNA482

2185586659mature
2185874944mature
2185892817mature

Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sativa
Triticum aestivum
Aquilegia coerula
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max

Number of miRNAs
microRNA482

$dG = -64.20$ [initially -64.20] 09Apr07-09-04-20

microRNA482
dG = -81.80 [initially -81.80] 09Apr07-08-39-28
microRNA482

$dG = -81.50$ (initially -81.50) 09Apr07-09-24-18
microRNA529

2185620929mature

DR933604binding

DR949672binding

DR954092binding

DT745966binding

TC20719binding

TC21170binding

TC21808binding

TC24142binding

TC24319binding

TC26823binding

TC29483binding

TC31356binding

TC31380binding

TC31827binding

TC32167binding

TC32968binding

529

Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sativa
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max

Number of miRNAs
dG = -60.00 [initially -60.00] 09Mar19-15-17-36
microRNA530

2185481266mature
TC23269binding

Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sativa
Triticum aestivum
Aquilegia coerulae
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max

Number of miRNAs

microRNA530
microRNA530

dG = -62.80 [initially -64.40] 09Mar30-12-55-11
microRNA535

Formatted Alignments

```
```

<table>
<thead>
<tr>
<th>Number of miRNAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physcomitrella patens</td>
</tr>
<tr>
<td>Selaginella moellendorffii</td>
</tr>
<tr>
<td>Pinus taeda</td>
</tr>
<tr>
<td>Zea mays</td>
</tr>
<tr>
<td>Sorghum bicolor</td>
</tr>
<tr>
<td>Oryza sativa</td>
</tr>
<tr>
<td>Triticum aestivum</td>
</tr>
<tr>
<td>Aquilegia coerulea</td>
</tr>
<tr>
<td>Solanum lycopersicum</td>
</tr>
<tr>
<td>Vitis vinifera</td>
</tr>
<tr>
<td>Populus trichocarpa</td>
</tr>
<tr>
<td>Brassica napus</td>
</tr>
<tr>
<td>Brassica rapa</td>
</tr>
<tr>
<td>Arabidopsis thaliana</td>
</tr>
<tr>
<td>Gossypium hirsutum</td>
</tr>
<tr>
<td>Medicago truncatula</td>
</tr>
<tr>
<td>Glycine max</td>
</tr>
</tbody>
</table>

535
microRNA535

Output of sir_graph (®)
mfold 3.4

\[ dG = -49.20 \] [initially -49.20] 09Mar19-12-37-36

2185616263

\[ dG = -49.20 \] [initially -49.20] 09Mar19-12-37-36

microRNA535
### microRNA783

**EST1199724_DT765875 binding**

<table>
<thead>
<tr>
<th>Binding</th>
<th>Alignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>TC27593</td>
<td>CUGUUGCGGCUCCGUGCUUUCUCUC</td>
</tr>
<tr>
<td>TC27947</td>
<td>AUGHCGCUCGACUAUUUCU</td>
</tr>
<tr>
<td>TC32408</td>
<td>AUGHCGCUCGACUAUUUCU</td>
</tr>
<tr>
<td>TC33228</td>
<td>AUGHCGCUCGACUAUUUCU</td>
</tr>
</tbody>
</table>

**Formatted Alignments**

<table>
<thead>
<tr>
<th>Alignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUUGCUUGCGACUAUUUCU</td>
</tr>
<tr>
<td>AUUGCUUGCGACUAUUUCU</td>
</tr>
<tr>
<td>AUUGCUUGCGACUAUUUCU</td>
</tr>
<tr>
<td>AUUGCUUGCGACUAUUUCU</td>
</tr>
</tbody>
</table>

### 783

<table>
<thead>
<tr>
<th>Species</th>
<th>Number of miRNAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physcomitrella patens</td>
<td></td>
</tr>
<tr>
<td>Selaginella moellendorffii</td>
<td></td>
</tr>
<tr>
<td>Pinus taeda</td>
<td>1</td>
</tr>
<tr>
<td>Zea mays</td>
<td></td>
</tr>
<tr>
<td>Sorghum bicolor</td>
<td></td>
</tr>
<tr>
<td>Oryza sativa</td>
<td></td>
</tr>
<tr>
<td>Triticum aestivum</td>
<td></td>
</tr>
<tr>
<td>Aquilegia coerulea</td>
<td></td>
</tr>
<tr>
<td>Solanum lycopersicum</td>
<td></td>
</tr>
<tr>
<td>Vitis vinifera</td>
<td></td>
</tr>
<tr>
<td>Populus trichocarpa</td>
<td></td>
</tr>
<tr>
<td>Brassica napus</td>
<td></td>
</tr>
<tr>
<td>Brassica rapa</td>
<td></td>
</tr>
<tr>
<td>Arabidopsis thaliana</td>
<td>1</td>
</tr>
<tr>
<td>Gossypium hirsutum</td>
<td></td>
</tr>
<tr>
<td>Medicago truncatula</td>
<td></td>
</tr>
<tr>
<td>Glycine max</td>
<td></td>
</tr>
</tbody>
</table>

**Number of miRNAs**