DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

Identification of Conserved Aquilegia Coerulea MicroRNAs and Their Targets

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters.

Citation	Puzey, Joshua Robert, and Elena M. Kramer. 2009. Identification of conserved Aquilegia coerulea microRNAs and their targets. Gene 448(1): 46-56.
Published Version	doi:10.1016/j.gene.2009.08.005
Accessed	February 18, 2015 5:21:09 PM EST
Citable Link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:8519268
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP

(Article begins on next page)

Elsevier Editorial System(tm) for Gene
Manuscript Draft

Manuscript Number: GENE-D-09-00303R1

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.

* Revised Manuscript (unmarked)

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 miR*NA 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 m*iRNA 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/mold. 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: AMFE=[(MFE/length of RNA sequence) x 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: MFEI= [(AMFE) x 100]/(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 (eg. moncots 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 obervations (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).

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 *A. thaliana* and *O. sativa* miR159 and miR319 (Fig. 5C). 2183893560 has a *CT* at the 7th and 8th position, which is conserved in the *A. thaliana* and *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 *A. thaliana* and *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.

3.5 Phylogenetic distribution of microRNAs

Given that *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 aqc-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

(Fig. 5D).

miR529

The phylogenetic distribution of miR529 is also quite interesting. miR529 is present in the lycopods, 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

Aukerman, M.J., 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.

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

Axtell, M.J., Bowman, J.L., 2008. Evolution of plant microRNAs and their targets. Trends Plant Sci. 13, 343-349.

Axtell, M.J., Snyder, J.A., Bartel, D.P., 2007. Common functions for diverse small RNAs of land plants. Plant Cell 19, 1750-1769.

Barakat, A., Wall, P.K., Diloreto, S., Depamphilis, C.W., Carlson, J.E., 2007. Conservation and divergence of microRNAs in Populus. BMC Genomics 8, 481.

Barakat, A., Wall, K., Leebens-Mack, J., Wang, Y.J., Carlson, J.E.Depamphilis, C.W., 2007. Large-scale identification of microRNAs from a basal eudicot (Eschscholzia californica) and conservation in flowering plants. Plant J. 51, 991-1003.

Boualem, A., Laporte, P., Jovanovic, M., Laffont, C., Plet, J., Combier, J.P., Niebel, A., Crespi, M., Frugier, F., 2008. MicroRNA166 controls root and nodule development in Medicago truncatula. Plant J. 54, 876-887.

Cartolano, M., Castillo, R., Efremova, N., Kuckenberg, M., Zethof, J., Gerats, T., Schwarz-Sommer, Z., Vandenbussche, M., 2007. A conserved microRNA module exerts homeotic control over Petunia hybrida and Antirrhinum majus floral organ identity. Nat. Genet. 39, 901-905.

Chaw, S.M., Chang, C.C., Chen, H.L., Li, W.H., 2004. Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. J. Mol. Evol. 58, 424-441.

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

Ding, D., Zhang, L., Wang, H., Liu, Z., Zhang, Z., Zheng, Y., 2009. Differential expression of miRNAs in response to salt stress in maize roots. Ann. Bot. (Lond.) 103, 29-38.

Fahlgren, N., Howell, M.D., Kasschau, K.D., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., Law, T.F., Grant, S.R., Dangl, J.L., et al., 2007. High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. PLoS ONE 2, e219.

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

Griffiths-Jones, S., 2004. The microRNA Registry. Nucleic Acids Res. 32, D109-111.

Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A.Enright, A.J., 2006. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. 34, D140-144.

Griffiths-Jones, S., Saini, H.K., van Dongen, S., Enright, A.J., 2008. miRBase: tools for microRNA genomics. Nucleic Acids Res. 36, D154-158.

Guo, A.Y., Zhu, Q.H., Gu, X., Ge, S., Yang, J., Luo, J., 2008. Genome-wide identification and evolutionary analysis of the plant specific SBP-box transcription factor family. Gene 418, 1-8.

Hodges, S.A., Arnold, M.L., 1994a. Columbines - a Geographically Widespread Species Flock. Proc. Natl. Acad. Sci. 91, 5129-5132.

Hodges, S.A., Arnold, M.L., 1994b. Floral and ecological isolation between Aquilegia formosa and Aquilegia pubescens. Proc. Natl. Acad. Sci. U S A 91, 2493-2496.

Hodges, S.A., Fulton M., Yang JY, Whittal JB., 2003. Verne Grant and evolutionary studies of *Aquilegia*. New Phytologist 113-120.

Hodges, S.A., Kramer, E.M., 2007. Columbines. Curr. Biol. 17, R992-994.

Jaillon, O., Aury, J.M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., et al., 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449, 463-467.

Jones-Rhoades, M.W., 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., Bartel, B., 2006. MicroRNAS and their regulatory roles in plants. Annu. Rev. Plant. Biol. 57, 19-53.

Lauter, N., Kampani, A., Carlson, S., Goebel, M., Moose, S.P., 2005. microRNA172 down-regulates glossy15 to promote vegetative phase change in maize. Proc. Natl. Acad. Sci. U S A 102, 9412-9417.

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

Lu, S., Sun, Y.H., Amerson, H., Chiang, V.L., 2007. MicroRNAs in loblolly pine (Pinus taeda L.) and their association with fusiform rust gall development. Plant J. 51, 1077-1098.

Lu, S., Sun, Y.H., Chiang, V.L., 2008. Stress-responsive microRNAs in Populus. Plant J. 55, 131-151.

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

Mathews, D.H., Sabina, J., Zuker, M., Turner, D.H., 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J. Mol. Biol. 288, 911-940.

Meyers, B.C., Axtell, M.J., Bartel, B., Bartel, D.P., Baulcombe, D., Bowman, J.L., Cao, X., Carrington, J.C., Chen, X., Green, P.J., et al., 2008. Criteria for Annotation of Plant MicroRNAs. Plant Cell

Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C., et al., 2008. Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. Cell 133, 116-127.

Montgomery, T.A., Howell, M.D., Cuperus, J.T., Li, D., Hansen, J.E., Alexander, A.L., Chapman, E.J., Fahlgren, N., Allen, E., Carrington, J.C., 2008. Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. Cell 133, 128-141.

Moore, M.J., Bell, C.D., Soltis, P.S., Soltis, D.E., 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proc. Natl. Acad. Sci. 104, 19363-19368.

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

Palatnik, J.F., Wollmann, H., Schommer, C., Schwab, R., Boisbouvier, J., Rodriguez, R., Warthmann, N., Allen, E., Dezulian, T., Huson, D., et al., 2007. Sequence and expression differences underlie functional specialization of arabidopsis microRNAs miR159 and miR319. Dev. Cell 13, 115-125.

Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., et al., 2009. The Sorghum bicolor genome and the diversification of grasses. Nature 457, 551-556.

Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B., Bartel, D.P., 2002. Prediction of plant microRNA targets. Cell 110, 513-520.

Sanderson, M.J., Thorne, J.L., Wikstrom, N., Bremer, K., 2004. Molecular evidence on plant divergence times. American Journal of Botany 91, 1656-1665.

Schwab, R., Ossowski, S., Riester, M., Warthmann, N., Weigel, D., 2006. Highly specific gene silencing by artificial microRNAs in Arabidopsis. Plant Cell 18, 1121-1133.

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

Shukla, L.I., Chinnusamy, V., Sunkar, R., 2008. The role of microRNAs and other endogenous small RNAs in plant stress responses. Biochim. Biophys. Acta 1779, 743-748.

Sunkar, R., Zhou, X., Zheng, Y., Zhang, W., Zhu, J.K., 2008. Identification of novel and candidate miRNAs in rice by high throughput sequencing. BMC Plant Biol. 8, 25.

Takeda, A., Iwasaki, S., Watanabe, T., Utsumi, M., Watanabe, Y., 2008. The mechanism selecting the guide strand from small RNA duplexes is different among argonaute proteins. Plant Cell Physiol. 49, 493-500.

Usami, T., Horiguchi, G., Yano, S., Tsukaya, H., 2009. The more and smaller cells mutants of Arabidopsis thaliana identify novel roles for SQUAMOSA PROMOTER BINDING PROTEIN-LIKE genes in the control of heteroblasty. Development 136, 955-964.

Wang, J.W., Schwab, R., Czech, B., Mica, E., Weigel, D., 2008. Dual effects of miR156-targeted SPL genes and CYP78A5/KLUH on plastochron length and organ size in Arabidopsis thaliana. Plant Cell 20, 1231-1243.

Willmann, M.R., Poethig, R.S., 2007. Conservation and evolution of miRNA regulatory programs in plant development. Curr. Opin. Plant Biol. 10, 503-511.

Wu, M.F., Tian, Q., Reed, J.W., 2006. Arabidopsis microRNA167 controls patterns of ARF6 and ARF8 expression, and regulates both female and male reproduction. Development 133, 4211-4218.

Zhang, B., Pan, X., Cannon, C.H., Cobb, G.P., Anderson, T.A., 2006a. Conservation and divergence of plant microRNA genes. Plant J. 46, 243-259.

Zhang, B., Pan, X., Stellwag, E.J., 2008a. Identification of soybean microRNAs and their targets. Planta 229, 161-182.

Zhang, B., Wang, Q., Wang, K., Pan, X., Liu, F., Guo, T., Cobb, G.P., Anderson, T.A., 2007. Identification of cotton microRNAs and their targets. Gene 397, 26-37.

Zhang, B.H., Pan, X.P., Cox, S.B., Cobb, G.P., Anderson, T.A., 2006b. Evidence that miRNAs are different from other RNAs. Cell Mol. Life Sci. 63, 246-254.

Zhang, J.F., Yuan, L.J., Shao, Y., Du, W., Yan, D.W., Lu, Y.T., 2008b. The disturbance of small RNA pathways enhanced abscisic acid response and multiple stress responses in Arabidopsis. Plant Cell Environ. 31, 562-574.

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

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. patens*. (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. coerulea* (2185477245, 2185552337, 2185755300, 2185768332, 2185874405, 2185853634, and 2185809227), *P. trichocarpa, V. vinifera,* and *P. patens* For B-D, osa:

Oryza sativa, ath: *Arabidopsis thaliana*, ptc: *Populus trichocarpa*, ppt: *Physcomitrella patens*, vvi: *Vitis vinifera*.

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

microRNA family	aqc-miRNA gi	Mature	kcal/mol	MFEI	NM	ML	PL	Arm	A	C	G	U	A (%)	C (%)	G (%)	U (%)
aqc-miRNA156a	2.19E+09	GACAGAAGAUAGAGAGCA	-43.4	0.00	1	20	81	5'	9	3	6	2	45	15	30	10
aqc-miRNA156b	2.19E+09	GACAGAAGAUAGAGAGCA	-44.8	0.00	1	20	83	5'	9	3	6	2	45	15	30	10
aqc-miRNA159	2.18E+09	UUGGACUGAAGGGAGCUCU	-18.8	0.00	4	21	71	3'	5	3	7	6	24	14	33	29
aqc-miRNA160a	2.19E+09	GCCUGGCUCCCUGGAUGCC	-45.6	0.00	1	21	83	5'	2	8	6	5	10	38	29	24
aqc-miRNA160b	2.19E+09	GCCUGGCUCCCUGUAUGCC	-48.1	0.00	1	21	82	5'	2	8	5	6	10	38	24	29
aqc-miRNA166a	2.19E+09	CGGACCAGGCUUCAUUCCU	-49.1	0.00	4	21	##	3'	3	8	4	6	14	38	19	29
aqc-miRNA166b	2.19E+09	CGGACCAGGCUUCAUUCCC	-44.5	0.00	2	21	94	3'	3	9	4	5	14	43	19	24
aqc-miRNA166c	2.19E+09	UCGGACCAGGCUUCAUUCCU	-46.6	0.00	4	20	##	3'	3	7	4	6	15	35	20	30
aqc-miRNA166d	2.19E+09	CGGACCAGGCUUCAUUCCU	-42.9	0.00	3	21	##	3'	3	8	4	6	14	38	19	29
aqc-miRNA166e	2.19E+09	CGGACCAGGCUUCAUUCCC	-57.6	0.00	3	21	##	3'	3	9	4	5	14	43	19	24
aqc-miRNA167	2.19E+09	CAAGCUGCCAGCAUGAUCU	-31.5	0.00	4	21	68	5'	6	6	4	5	29	29	19	24
aqc-miRNA168	2.19E+09	GGCUUAGUGCAGCUCGGGG	-40.5	0.00	4	21	##	5'	3	4	9	5	14	19	43	24
aqc-miRNA169a	2.19E+09	AGCCAAGGAUGACUUGCCU	-75.3	0.00	3	21	##	5'	6	5	5	5	29	24	24	24
aqc-miRNA169b	2.19E+09	AGCCAAGGAUGACUUGCCU	-78.2	0.00	2	21	##	5'	5	5	6	5	24	24	29	24
aqc-miRNA169c	2.19E+09	AGCCAAGGAUGACUUGCCG	-47.1	0.00	4	21	##	5'	5	6	7	3	24	29	33	14
aqc-miRNA171a	2.19E+09	GAUUGAGCCGUGCCAAUAU	-46.3	0.00	2	21	##	3'	5	5	5	6	24	24	24	29
aqc-miRNA171b	2.19E+09	GAUUGAGCCGUGCCAAUAU	-36.6	0.00	1	21	77	3'	5	5	5	6	24	24	24	29
aqc-miRNA171c	2.19E+09	AAUUGAACCGCACUAAUAU	-27.3	0.00	3	21	80	3'	8	5	2	6	38	24	10	29
aqc-miRNA171d	2.19E+09	GAUUGAGCCGUGCCAAUAU	-36	0.00	2	21	##	3'	5	5	5	6	24	24	24	29
aqc-miRNA171e	2.19E+09	GAAUGAACCGAGCCAACAU	-30.5	0.00	2	21	##	3'	8	6	4	3	38	29	19	14
aqc-miRNA171f	2.19E+09	AAUUGAGCCGUGCCAAUAU	-41.1	0.00	3	21	88	3'	6	5	4	6	29	24	19	29
aqc-miRNA172a	2.19E+09	GAAUCUUGAUGAUGCUGCA	-33.6	0.00	4	21	##	3'	6	3	5	7	29	14	24	33
aqc-miRNA172b	2.19E+09	GAAUCUUGAUGAUGCUGCA	-43.7	0.00	1	21	##	3'	5	3	6	7	24	14	29	33
aqc-miRNA319	2.19E+09	UGGACUGAAGGGAGCUCCC	-78.6	0.00	1	21	##	3'	4	5	7	5	19	24	33	24
aqc-miRNA395a	2.19E+09	UGAAGGGUUUGGAGGAACU	-43.6	0.00	2	21	##	3'	5	3	8	5	24	14	38	24
aqc-miRNA395b	2.19E+09	UGAAGGGUUUGGAGGAACU	-37.7	0.00	2	21	80	3'	5	3	8	5	24	14	38	24
aqc-miRNA396a	2.19E+09	UCCACAGCUUUCUUGAACU	-41.2	0.00	1	21	##	5'	4	6	3	8	19	29	14	38
aqc-miRNA396b	2.19E+09	UCCACAGCUUUCUUGAACU	-38.6	0.00	1	21	89	5'	4	6	2	9	19	29	10	43
aqc-miRNA398a	2.19E+09	GUGUUCUCAGGUCACCCCU	-42.9	0.00	1	21	92	3'	2	7	4	8	10	33	19	38
aqc-miRNA398b	2.19E+09	GUGUUCUCAGGUCGCCCCU	-38.5	0.00	1	21	93	3'	1	7	6	7	5	33	29	33
aqc-miRNA399	2.19E+09	GCCAAAGGAGAGUUGCCCU	-49.3	0.00	3	21	98	3'	6	5	6	4	29	24	29	19
aqc-miRNA408	2.19E+09	GCACUGCCUCUUCCCUGCA	-46.5	0.00	3	21	87	3'	2	#	3	6	10	48	14	29
aqc-miRNA477a	2.19E+09	CUCUCCCUCAAGGGCUUCU	-38	0.00	2	20	73	5'	3	8	3	6	15	40	15	30
aqc-miRNA477b	2.19E+09	CUCUCCCUCAAGGGCUUCU	-27.8	0.00	2	20	72	5'	3	8	3	6	15	40	15	30
aqc-miRNA477c	2.19E+09	CUCUCCCUCAAGUUCUUCU	-40.4	0.00	2	20	76	5'	3	8	1	8	15	40	5	40
aqc-miRNA477d	2.19E+09	CUCUCCCUCAAGGGCUUCU	-24.2	0.00	2	20	68	5'	3	8	3	6	15	40	15	30
aqc-miRNA477e	2.19E+09	CUCUCCCUCAAGGGCUUCU	-36.4	0.00	2	20	84	5'	3	8	3	6	15	40	15	30
aqc-miRNA477f	2.19E+09	CUCUCCUUCAAAGGCUUCU	-40.6	0.00	2	20	80	5'	4	7	2	7	20	35	10	35
aqc-miRNA477g	2.19E+09	CUCUCCCUCAAGUUCUUCU	-40.4	0.00	2	20	76	5'	3	8	1	8	15	40	5	40
aqc-miRNA482a	2.19E+09	CUUGCCGACUCCUCCCAUA	-81.5	0.00	3	22	##	3'	3	#	2	6	14	50	9	27
aqc-miRNA482b	2.19E+09	CUUGCCGACUCCUCCCAUA	-64.2	0.00	3	22	##	3'	3	#	2	6	14	50	9	27
aqc-miRNA482c	2.19E+09	CUUGCCGACUCCUCCCAUA	-81.8	0.00	3	22	##	3'	3	#	2	6	14	50	9	27
aqc-miRNA529	2.19E+09	GAAGAGAGAGAGCACAAC	-60	0.00	1	21	##	5'	#	5	6	0	48	24	29	0
aqc-miRNA530	2.19E+09	UGCAUUUGCACCUGCAUCU	-62.8	0.00	2	20	##	5'	3	7	3	7	15	35	15	35
aqc-miRNA535	2.19E+09	ACAACGAGAGAGAGCACG	-49.2	0.00	0	22	87	5'	8	5	8	1	36	23	36	5

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

MFEI: minimal folding free energy index

NM: number of mismatches in miRNA/miRNA* duplex

ML: Mature Length

PL: Precursor Length

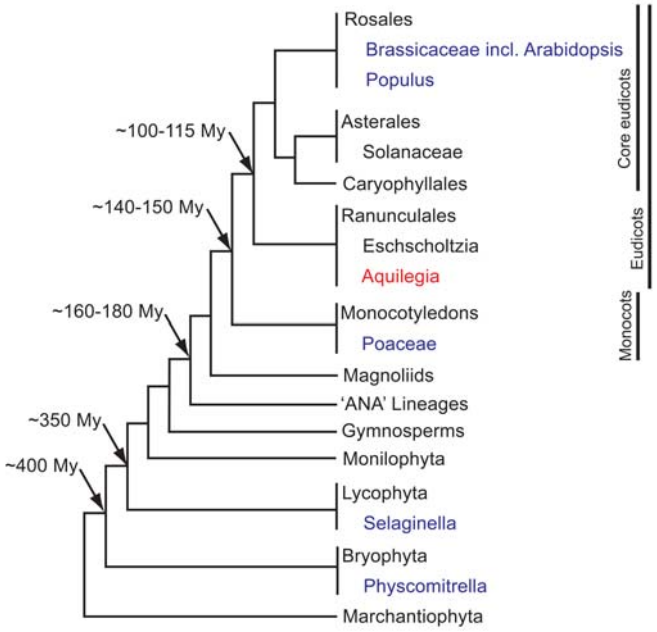
ARM: Location of mature miRNA (3' or 5')

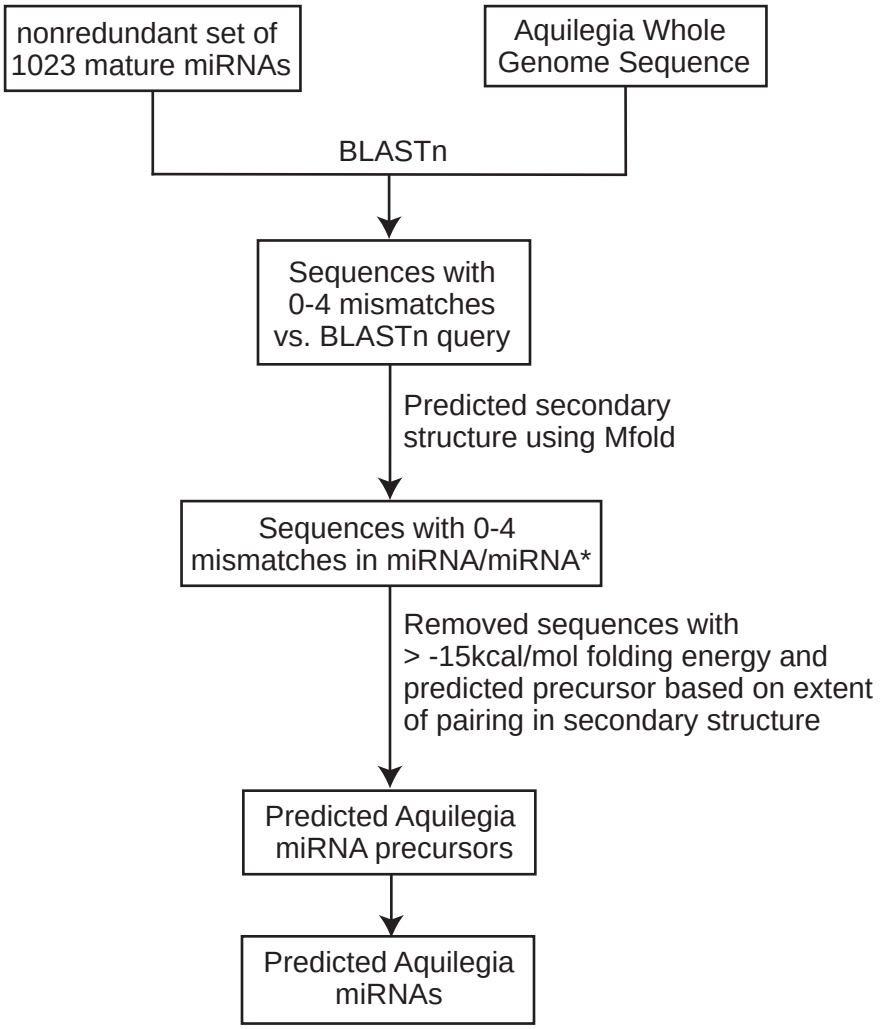
Table 2

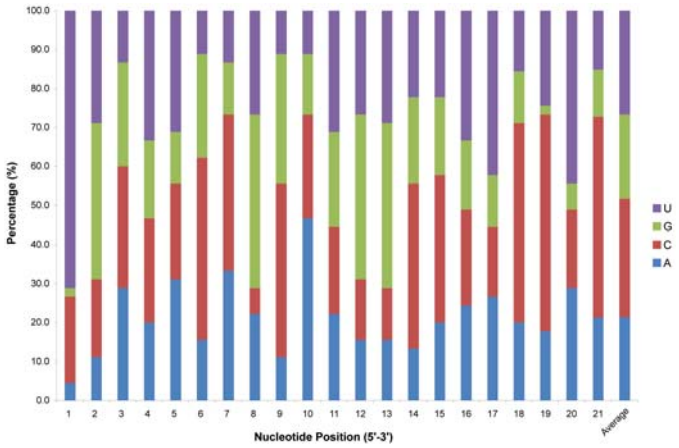
miRNA precursor characteristics	Minimum	Maximum	Average	Std Dev
Sequence Length	68	180	106	32.2
A (%)	16.4	36.3	25.6	4.4
U (%)	19.7	38.1	31.4	3.8
G (5)	13.8	31.5	22.9	3.7
C (%)	13.5	29.9	20.1	3.0
MFE (-kcal/mol)	18.8	81.8	46	14.7
AMFE (-kcal/mol)	26.5	58.7	44.1	7.5
MFEI	0.60	1.48	1.03	0.18

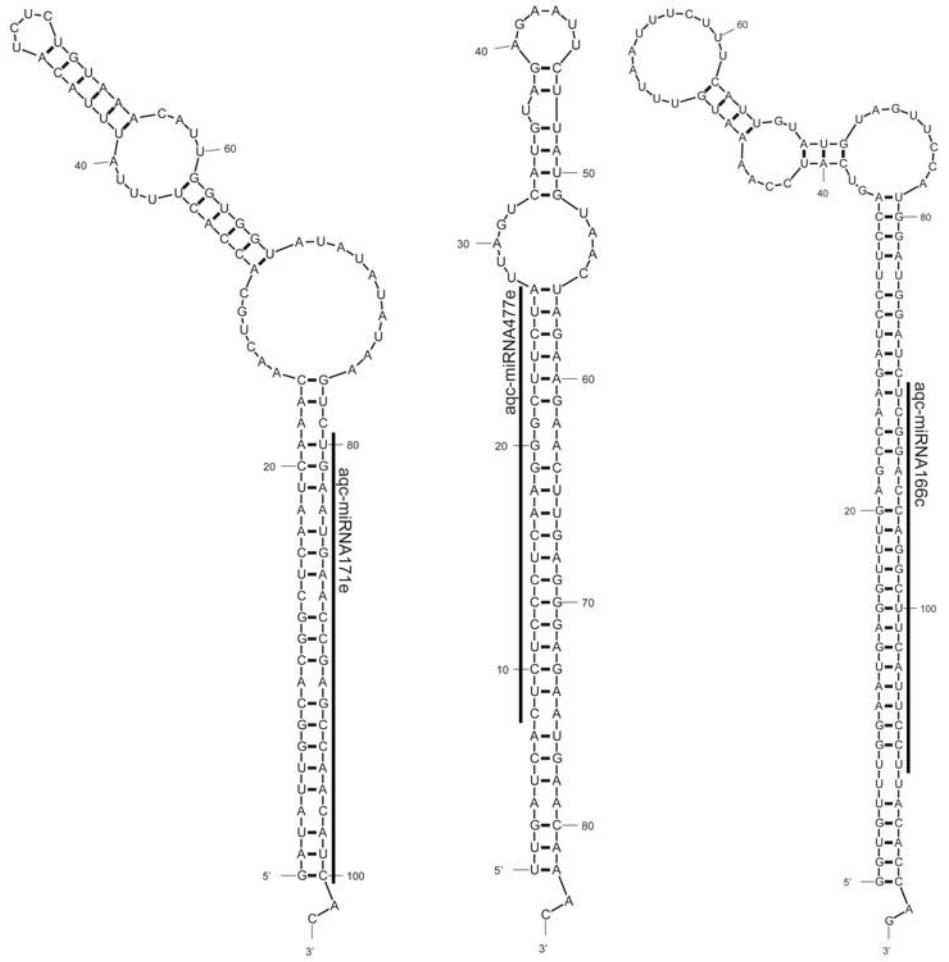
Table 3
Potential targets of *Aquilegia coerulea* microRNAs

microRNA family	Targeted Protein	Target Function	TC#	Top Scoring A.t. locus
156	Squamosa promoter binding protein-like 2 (SBP)	Transcription factor	TC24142	AT5G43270
	Squamosa promoter binding protein-like 6 (SBP)	Transcription factor	TC26823, TC21170, TC21808, DR954092	AT1G69170
	Squamosa promoter binding protein-like 13 (SBP)	Transcription factor	TC32167, DR949672	AT5G50570
	Squamosa promoter binding protein-like 14 (SBP)	Transcription factor	TC21707	AT1G20980
	Squamosa promoter binding protein-like 12 (SBP)	Transcription factor	TC31356	AT3G60030
	Growth Regulating Factor 2 (GRF2)	Transcription factor	TC25281	AT4G37740
	Unknown	Unknown	TC31827	AT1G40133
159	None			
160	Transposable element		DT734946	AT5G28526
	Psuedo-response regulator 2 (APRR2)	Transcription factor	TC22019	AT4G18020
162	None			
164	Cup-Shaped Cotyledon 2 (CUC2)	Transcription factor	TC27953	AT5G53950
	NAC Domain containing protein	Transcription factor	TC27854	AT5G39610
	Leucine-rich repeat transmembrane protein kinase	Metabolism	DT731408	AT3G28040
	UDP-glucose pyrophosphorylase	Stress	TC21677	AT3G03250
	4-coumarate-CoA ligase family protein	Stress	DT741735, TC24498	AT1G20510
166	Class III HD-Zip protein	Transcription factor	TC30761	AT1G52150
167	Auxin response factor 6	Transcription factor	TC29952, TC24356	AT1G30330
168	Unknown	Metabolism	DT747736	AT3G48870
	Anion Transporter 2	Transport	TC29087	AT4G00370
169	CCAAT-binding factor (HAP2)	Transcription factor	DR924236	AT1G72830
	Ligase class 3 family protein	Metabolsim	TC31138	AT5G18640
170/171	Scarecrow-like protein	Transcription factor	TC21487, TC28359, TC25185	AT2G45160
	Arabidopsis Pumilio 2 (APUM2)-like	RNA binding	TC29397	AT2G29190
172	APETALA2	Transcription factor	TC24245	AT4G36920
	Target of EAT 1 (TOE1)	Transcription factor	TC28449, TC32032	AT2G28550
	Unknown	Unknown	DR954144	AT5G65290
	Unknown	protein binding, zinc ion binding	TC26960	AT3G48070
	Spike1 (SPK1)	Metabolism	DT764297	AT4G16340
	SC35-like splicing factor	Metabolism	TC22146	AT3G13570
	Unknown	Transcription factor	DT749684	AT3G10070
	Ubitiquitin-conjugating enzyme	Metabolism	TC21155	AT5G05080
	F-box protein	Photoprocesses	TC28027	AT4G02440
	Selenium binding protein	Stess	TC29593	AT2G24440
319	Diminuto 1, Enhanced Very-Low-Fluence Responses 1	Photoprocesses	TC25461	AT3G19820
390	Cation symporter	Transport	TC30528	AT5G03560
393	Auxins Signalizng F-BOX 2	Auxin	TC29517	AT3G26810
	Transport Inhibitor Response 1 (TIR1)	Auxin	TC20718, TC25434	AT3G62980
394	F-box protein	Unknown	TC24293	AT1G27340
395	Kinase	Metabolism	TC32489	AT4G00955
	Unknown	Unknown	TC20934	AT3G49590
	ATP sulfurylase	Stress	TC21492	AT3G22890
	Unknown	Unknown	TC30052	AT2G02370
396	Unknown	Unknown	DR926810	AT2G16760
	Calmodulin-binding protein	Stress	TC29396	AT3G16940
	Mildew Resistance Locus	Stress	TC32094	AT1G11000
	1-aminocyclopropane-1-carboxylate oxidase, Ethylene forming enzyme	Stress	TC20841	AT1G05010
	UDP-glucosyl transferase family protein	Stress	TC22590	AT2G18560
397	Laccase	Metabolism	TC26926	AT5G01190
	Laccase	Metabolism	TC26200	AT2G38080
	Yip1 family protein	Unknown	TC24152	AT2G39805
	Unknown	Unknown	DT728841	AT5G53740
398	None			
399	beta-fructofuranosidase	Metabolism	TC28894	AT1G56560
	ABA receptor	Metabolism	TC20517	AT2G20770
408	Xyloglucan Endotransglycosylase 6	Metabolism	TC30505	AT4G25810
477	Plastidic glucose-6-phosphate dehydrogenase	Metabolism	TC30732	AT5G35790
	Elongation Factor 1B-gamma	Metabolism	TC20303	AT1G09640
	rRNA processing protein	Metabolism	TC24529	AT5G61330
482	None			
529	Squamosa promoter binding protein-like 6 (SBP)	Transcription factor	DR954092, TC26823, TC21808, TC21170	AT1G69170
	Squamosa promoter binding protein-like 13 (SBP)	Transcription factor	DR949672, TC32167	AT5G50570
	Osephin family protein	Unknown	TC20719	AT3G54130
	Squamosa promoter binding protein-like 14 (SBP)	Transcription factor	TC21707	AT1G20980
	Squamosa promoter binding protein-like 2 (SBP)	Transcription factor	TC24142	AT5G43270
	Threonine endopeptidase	Unknown	TC32968	AT4G01150
	FAD-binding domain-containing protein	Metabolism	DR933604	AT4G20800
	Diacylglycerol kinase gene family	Stress	TC29483	AT5G63770
	Metal Tolerance Protein A2	Stress	TC24319	AT3G58810
530	Serine carboxypeptidase-like 26	Metabolism	TC23269	AT2G35780
535	None			
783	Phosphoenolpyruvate carboxylase	Metabolism	DR942629	AT3G14940
	Methionine gamma-lyase	Metabolism	TC32408, TC27593	AT1G64660
	Glycoside hydrolase family protein	Metabolism	DT65875	AT5G43710
	Carrier family protein	Transport	TC33228	AT4G27940
	UDP-glucuronic acid decarboxylase	Metabolism	TC27947	AT3G53520

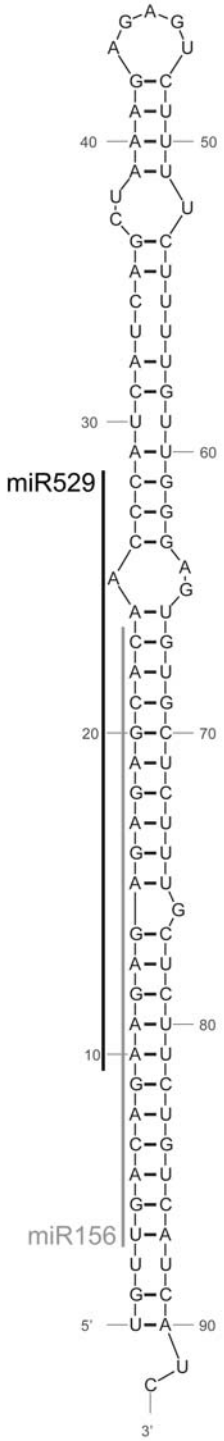

Rosales
Brassicaceae incl. Arabidopsis
Populus
Asterales
Solanaceae
Caryophyllales
Ranunculales
Eschscholtzia
Aquilegia
Monocotyledons
Poaceae
Magnoliids
'ANA' Lineages
Gymnosperms
Monilophyta
Lycophyta
Selaginella
Bryophyta
Physcomitrella
Marchantiophyta
~100-115 My
~140-150 My
~160-180 My
~350 My
~400 My
Core eudicots
Eudicots
Monocots


nonredundant set of 1023 mature miRNAs
Aquilegia Whole Genome Sequence
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

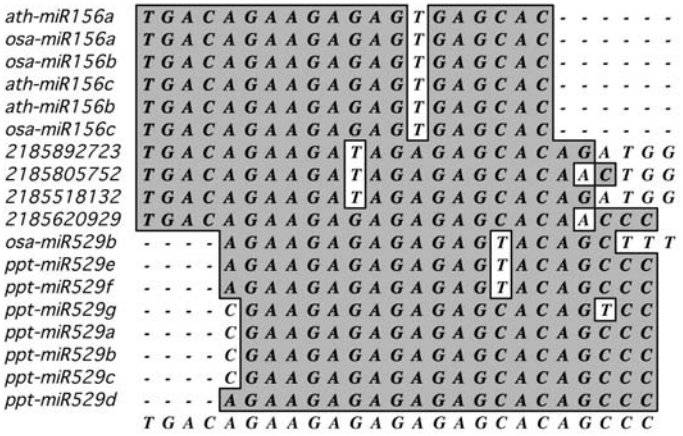

100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
Percentage (%)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
Average
Nucleotide Position (5'-3')
U
G
C
A


aqc-miRNA171e
aqc-miRNA477e
aqc-miRNA166c
5'
3'

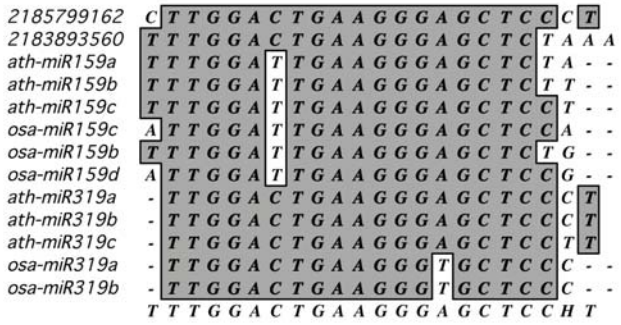
A.


miR529
miR156
5'
3'
10
20
30
40
50
60
70
80
90

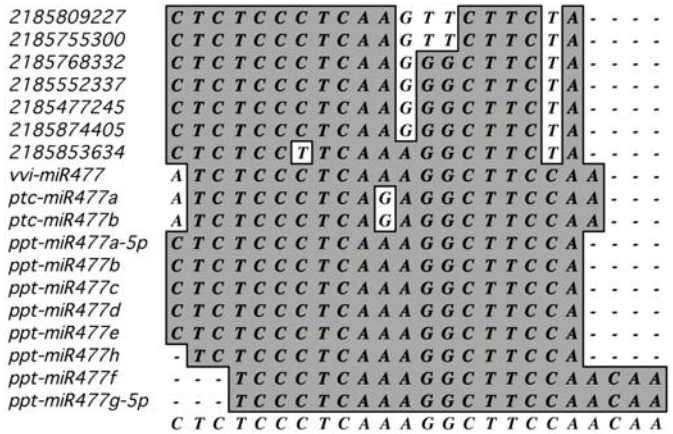
B.

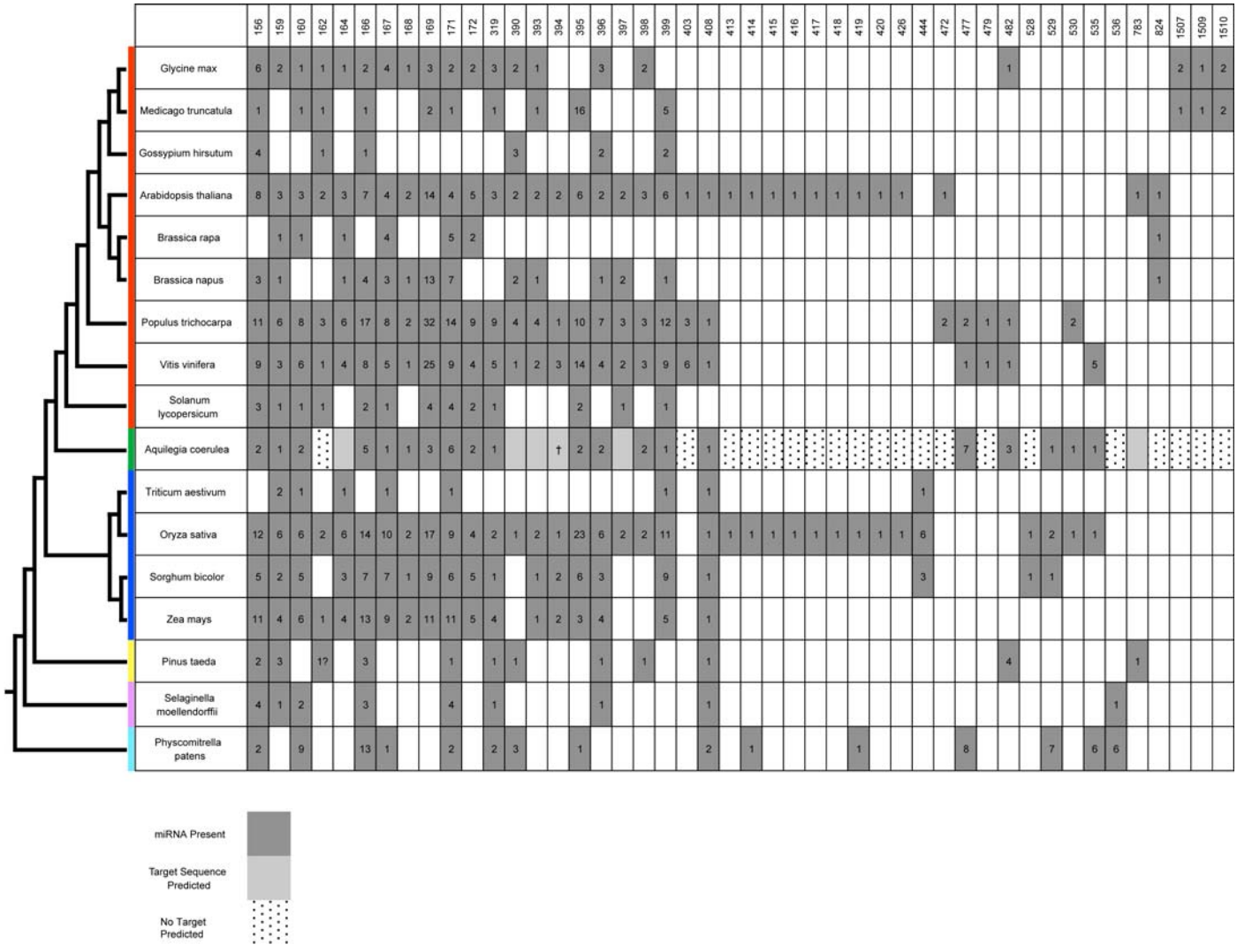

ath-miR156a
osa-miR156a
osa-miR156b
ath-miR156c
ath-miR156b
osa-miR156c
2185892723
2185805752
2185518132
2185620929
osa-miR529b
ppt-miR529e
ppt-miR529f
ppt-miR529g
ppt-miR529a
ppt-miR529b
ppt-miR529c
ppt-miR529d

C.


2185799162
2183893560
ath-miR159a
ath-miR159b
ath-miR159c
osa-miR159c
osa-miR159b
osa-miR159d
ath-miR319a
ath-miR319b
ath-miR319c
osa-miR319a
osa-miR319b

D.


2185809227
2185755300
2185768332
2185552337
2185477245
2185874405
2185853634
vvi-miR477
ptc-miR477a
ptc-miR477b
ppt-miR477a-5p
ppt-miR477b
ppt-miR477c
ppt-miR477d
ppt-miR477e
ppt-miR477h
ppt-miR477f
ppt-miR477g-5p


156 159 160 162 164 166 167 168 169 171 172 319 390 393 394 395 396 397 398 399 403 408 413 414 415 416 417 418 419 420 426 444 472 477 479 482 528 529 530 535 536 783 824 1507 1509 1510
Glycine max
Medicago truncatula
Gossypium hirsutum
Arabidopsis thaliana
Brassica rapa
Brassica napus
Populus trichocarpa
Vitis vinifera
Solanum lycopersicum
Aquilegia coerulea
Triticum aestivum
Oryza sativa
Sorghum bicolor
Zea mays
Pinus taeda
Selaginella moellendorffii
Physcomitrella patens
miRNA Present
Target Sequence Predicted
No Target Predicted

Supplementary Table 1

microRNA family	aqc-miRNA gi	Mature	Genbank Accession Number
aqc-miRNA156a	2185892723	TGACAGAAGATAGAGAGCAC	
aqc-miRNA156b	2185518132	TGACAGAAGATAGAGAGCAC	
aqc-miRNA159	2183893560	TTTGGACTGAAGGGAGCTCTA	
aqc-miRNA160a	2185524891	TGCCTGGCTCCCTGGATGCCA	
aqc-miRNA160b	2185745223	TGCCTGGCTCCCTGTATGCCA	
aqc-miRNA166a	2185638590	TCGGACCAGGCTTCATTCCTC	
aqc-miRNA166b	2185550821	TCGGACCAGGCTTCATTCCCC	
aqc-miRNA166c	2185683719	TCGGACCAGGCTTCATTCCT	
aqc-miRNA166d	2185830185	TCGGACCAGGCTTCATTCCTC	
aqc-miRNA166e	2185503206	TCGGACCAGGCTTCATTCCCC	
aqc-miRNA167	2185494083	TCAAGCTGCCAGCATGATCTA	
aqc-miRNA168	2185845089	TGGCTTAGTGCAGCTCGGGGA	
aqc-miRNA169a	2185604956	TAGCCAAGGATGACTTGCCTA	
aqc-miRNA169b	2185687520	TAGCCAAGGATGACTTGCCTG	
aqc-miRNA169c	2185493504	CAGCCAAGGATGACTTGCCGG	
aqc-miRNA171a	2185568457	TGATTGAGCCGTGCCAATATC	
aqc-miRNA171b	2185624710	TGATTGAGCCGTGCCAATATC	
aqc-miRNA171c	2185689689	TAATTGAACCGCACTAATATC	
aqc-miRNA171d	2185735537	TGATTGAGCCGTGCCAATATC	
aqc-miRNA171e	2185815653	TGAATGAACCGAGCCAACATC	
aqc-miRNA171f	2185468845	TAATTGAGCCGTGCCAATATC	
aqc-miRNA172a	2185776282	AGAATCTTGATGATGCTGCAT	
aqc-miRNA172b	2185788415	GGAATCTTGATGATGCTGCAT	
aqc-miRNA319	2185799162	TTGGACTGAAGGGAGCTCCCT	
aqc-miRNA395a	2185894728	CTGAAGGGTTTGGAGGAACTC	
aqc-miRNA395b	2185728750	CTGAAGGGTTTGGAGGAACTC	
aqc-miRNA396a	2185513426	TTCCACAGCTTTCTTGAACTG	
aqc-miRNA396b	2185617799	TTCCACAGCTTTCTTGAACTT	

aqc-miRNA398a	2185417724	TGTGTTCTCAGGTCACCCCTT
aqc-miRNA398b	2185579186	TGTGTTCTCAGGTCGCCCCTG
aqc-miRNA399	2185439501	TGCCAAAGGAGAGTTGCCCTA
aqc-miRNA408	2185474487	TGCACTGCCTCTTCCCTGCAC
aqc-miRNA477a	2185477245	CTCTCCCTCAAGGGCTTCTA
aqc-miRNA477b	2185552337	CTCTCCCTCAAGGGCTTCTA
aqc-miRNA477c	2185755300	CTCTCCCTCAAGTTCTTCTA
aqc-miRNA477d	2185768332	CTCTCCCTCAAGGGCTTCTA
aqc-miRNA477e	2185874405	CTCTCCCTCAAGGGCTTCTA
aqc-miRNA477f	2185853634	CTCTCCTTCAAAGGCTTCTA
aqc-miRNA477g	2185809227	CTCTCCCTCAAGTTCTTCTA
aqc-miRNA482a	2185586659	TCTTGCCGACTCCTCCCATACC
aqc-miRNA482b	2185874944	TCTTGCCGACTCCTCCCATACC
aqc-miRNA482c	2185892817	TCTTGCCGACTCCTCCCATACC
aqc-miRNA529	2185620929	AGAAGAGAGAGAGCACAACCC
aqc-miRNA530	2185481266	TGCATTTGCACCTGCATCTC
aqc-miRNA535	2185616263	TGACAACGAGAGAGAGCACGCG

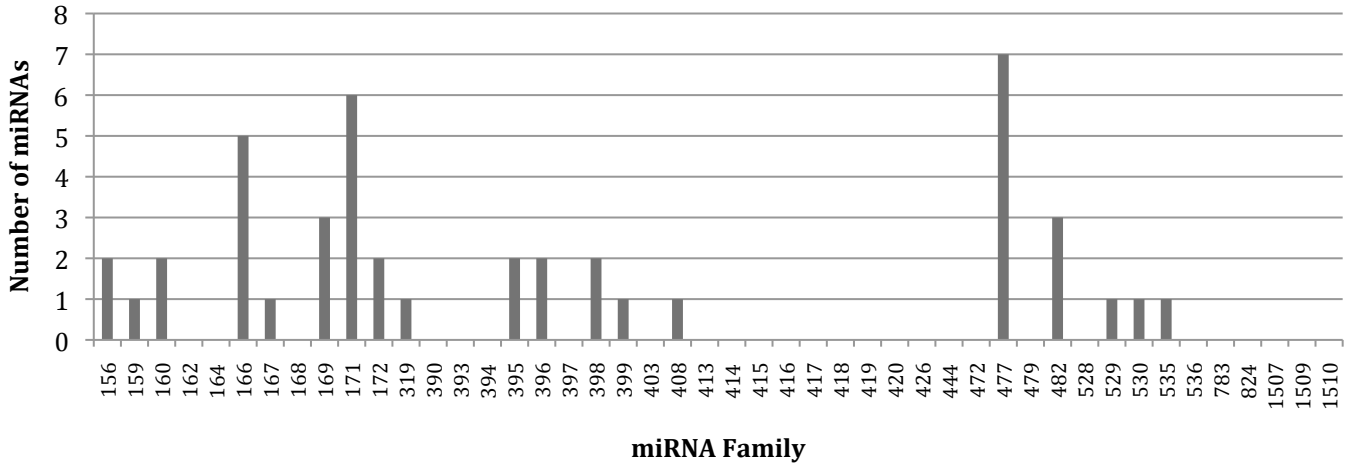
gi: gene index number corresponding to *A. coerulea* whole genome sequence databas

Supplementary Table 2
Potential targets of *Aquilegia coerulea* microRNAs

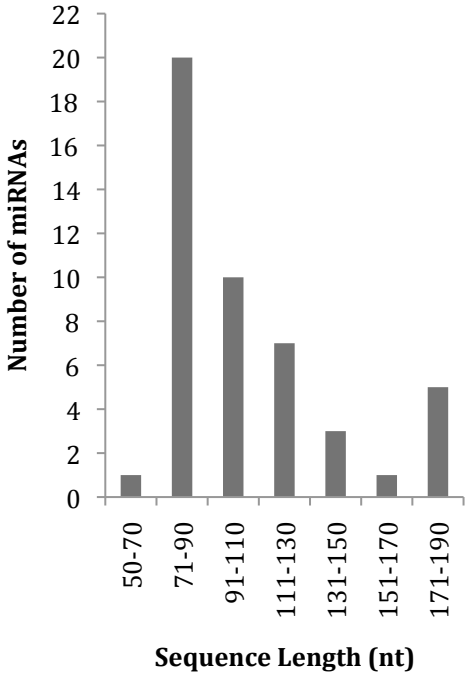
microRNA family	Targeted Protein	Target Function	TC#	Top Scoring A.t. locus
156	Squamosa promoter binding protein-like 2 (SBP)	Transcription factor	TC31380, TC24142	AT5G43270
	Squamosa promoter binding protein-like 6 (SBP)	Transcription factor	TC26823, TC21170, TC21808, DR954092	AT1G69170
	Squamosa promoter binding protein-like 13 (SBP)	Transcription factor	TC32167, DR949672	AT5G50570
	Squamosa promoter binding protein-like 14 (SBP)	Transcription factor	TC21707	AT1G20980
	Squamosa promoter binding protein-like 12 (SBP)	Transcription factor	TC31356	AT3G60030
	Growth Regulating Factor 2 (GRF2)	Transcription factor	TC25281	AT4G37740
	Unknown	Unknown	TC31827	AT1G40133
159	None			
160	Transposable element		DT734946	AT5G28526
	Psuedo-response regulator 2 (APRR2)	Transcription factor	TC22019	AT4G18020
162	None from EST database		None from EST database	
164	Cup-Shaped Cotyledon 2 (CUC2)	Transcription factor	DR949423, TC27953	AT5G53950
	NAC Domain containing protein	Transcription factor	TC27854	AT5G39610
	Leucine-rich repeat transmembrane protein kinase	Metabolism	DT731408	AT3G28040
	UDP-glucose pyrophosphorylase	Stress	TC21677	AT3G03250
	4-coumarate-CoA ligase family protein	Stress	DT741735, TC24498	AT1G20510
166	Class III HD-Zip protein	Transcription factor	TC30761	AT1G52150
167	Auxin response factor 6	Transcription factor	TC29952, TC24356	AT1G30330
168	Unknown	Metabolism	DT747736, DT74773	AT3G48870
	Anion Transporter 2	Transport	TC29087	AT4G00370
169	CCAAT-binding factor (HAP2)	Transcription factor	DR924236	AT1G72830
	Lipase class 3 family protein	Metabolsim	TC31138, DR927955	AT5G18640
170/171	Scarecrow-like protein	Transcription factor	TC21487, TC28359, TC25185	AT2G45160
	Arabidopsis Pumilio 2 (APUM2)-like	RNA binding	TC29397	AT2G29190
172	APETALA2	Transcription factor	DT732637, TC28864, TC22964, TC24245	AT4G36920
	Target of EAT 1 (TOE1)	Transcription factor	DT741192, TC28449, TC32032	AT2G28550
	Unknown	Unknown	DR954144	AT5G65290
	Unknown	protein binding, zinc ion binding	TC26960	AT3G48070
	Spike1 (SPK1)	Metabolism	DT764297	AT4G16340
	SC35-like splicing factor	Metabolism	TC22146	AT3G13570
	Unknown	Transcription factor	DT749684	AT3G10070
	Ubitiquitin-conjugating enzyme	Metabolism	TC25860, TC21155	AT5G05080
	F-box protein	Photoprocesses	TC28027	AT4G02440
	Selenium binding protein	Stess	TC29593	AT2G24440
319	Diminuto 1, Enhanced Very-Low-Fluence Responses 1	Photoprocesses	TC25461, DR922754	AT3G19820
390	Cation symporter	Transport	TC30528	AT5G03560
393	Auxing Signalizng F-BOX 2	Auxin	TC29517	AT3G26810
	Transport Inhibitor Response 1 (TIR1)	Auxin	TC20718, TC25434, DR913737	AT3G62980
394	F-box protein	Unknown	TC24293, TC26091	AT1G27340
395	Kinase	Metabolism	TC32489	AT4G00955
	Unknown	Unknown	TC20934	AT3G49590
	ATP sulfurylase	Stress	TC21492	AT3G22890
	Unknown	Unknown	TC30052	AT2G02370
396	Unknown	Unknown	DR926810	AT2G16760
	Calmodulin-binding protein	Stress	TC29396	AT3G16940
	Mildew Resistance Locus	Stress	TC32094	AT1G11000

	1-aminocyclopropane-1-carboxylate oxidase, Ethylene forming enzyme	Stress	TC20841	AT1G05010
	UDP-glucosyl transferase family protein	Stress	TC22590	AT2G18560
397	Laccase	Metabolism	TC26926	AT5G01190
	Laccase	Metabolism	TC26200	AT2G38080
	Yip1 family protein	Unknown	TC24152	AT2G39805
	Unknown	Unknown	DT728841	AT5G53740
398	None			
399	beta-fructofuranosidase	Metabolism	TC28894	AT1G56560
	ABA receptor	Metabolism	TC20517	AT2G20770
408	Xyloglucan Endotransglycosylase 6	Metabolism	TC30505	AT4G25810
477	Plastidic glucose-6-phosphate dehydrogenase	Metabolism	TC30732	AT5G35790
	Elongation Factor 1B-gamma	Metabolism	TC20303	AT1G09640
	rRNA processing protein	Metabolism	TC24529	AT5G61330
482	None			
529	Squamosa promoter binding protein-like 6 (SBP)	Transcription factor	DR954092, TC26823, TC21808, TC21170	AT1G69170
	Squamosa promoter binding protein-like 13 (SBP)	Transcription factor	DR949672, TC32167	AT5G50570
	Osephin family protein	Unknown	TC20719	AT3G54130
	Squamosa promoter binding protein-like 14 (SBP)	Transcription factor	TC21707	AT1G20980
	Squamosa promoter binding protein-like 2 (SBP)	Transcription factor	TC31380, TC24142	AT5G43270
	Threonine endopeptidase	Unknown	TC32968	AT4G01150
	FAD-binding domain-containing protein	Metabolism	DR933604	AT4G20800
	Diacylglycerol kinase gene family	Stress	TC29483	AT5G63770
	Metal Tolerance Protein A2	Stress	DT745966, TC24319	AT3G58810
530	Serine carboxypeptidase-like 26	Metabolism	TC23269	AT2G35780
535	None			
783	Phosphoenolpyruvate carboxylase	Metabolism	DR942629	AT3G14940
	Methionine gamma-lyase	Metabolism	TC32408, TC27593	AT1G64660
	Glycoside hydrolase family protein	Metabolism	DT65875	AT5G43710
	Carrier family protein	Transport	TC33228	AT4G27940
	UDP-glucuronic acid decarboxylase	Metabolism	TC27947	AT3G53520

A.



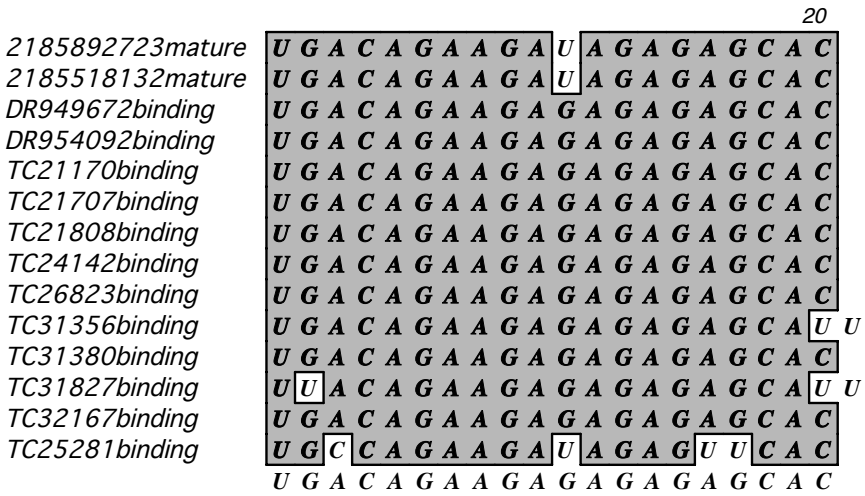
B.

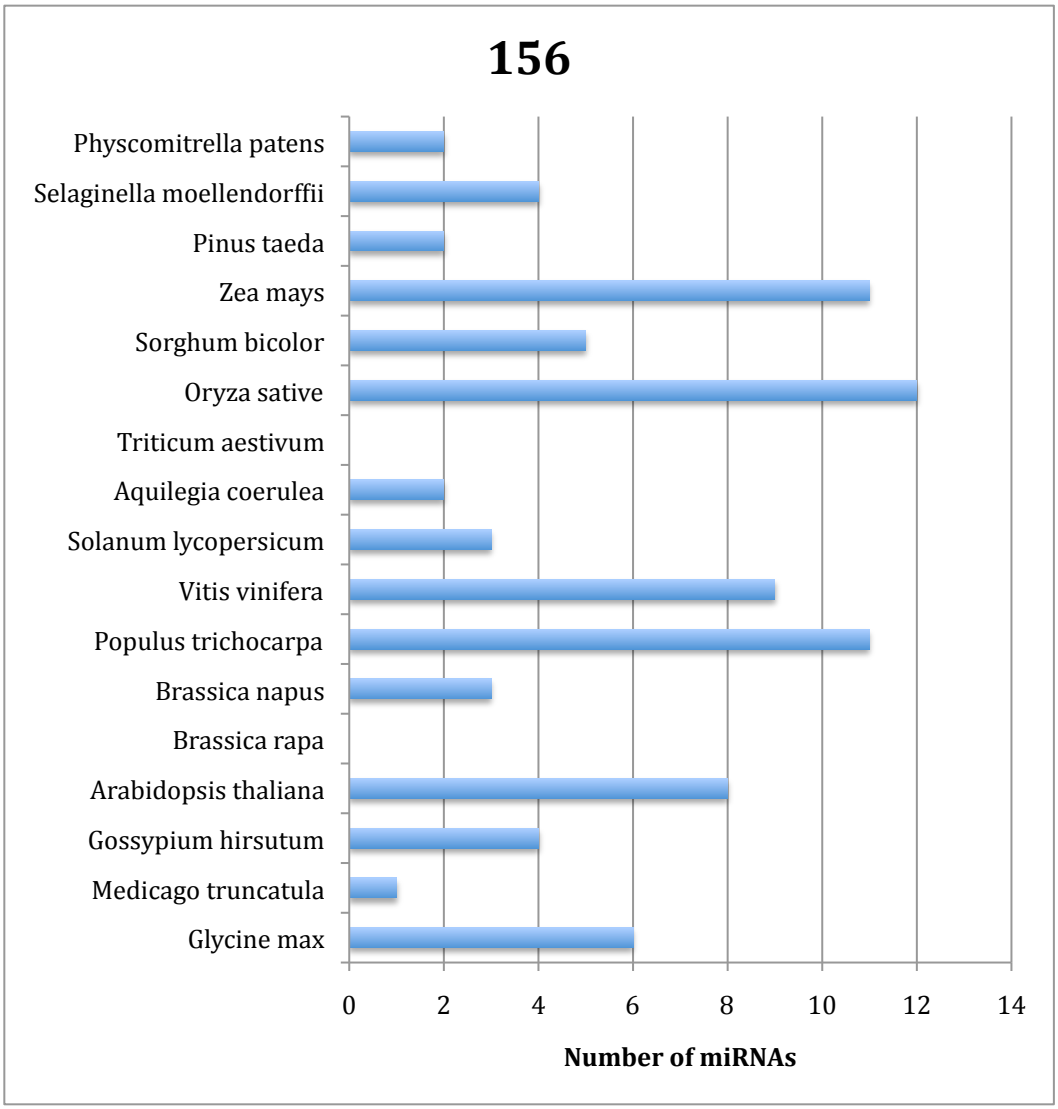


Supplemental Figure 1

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

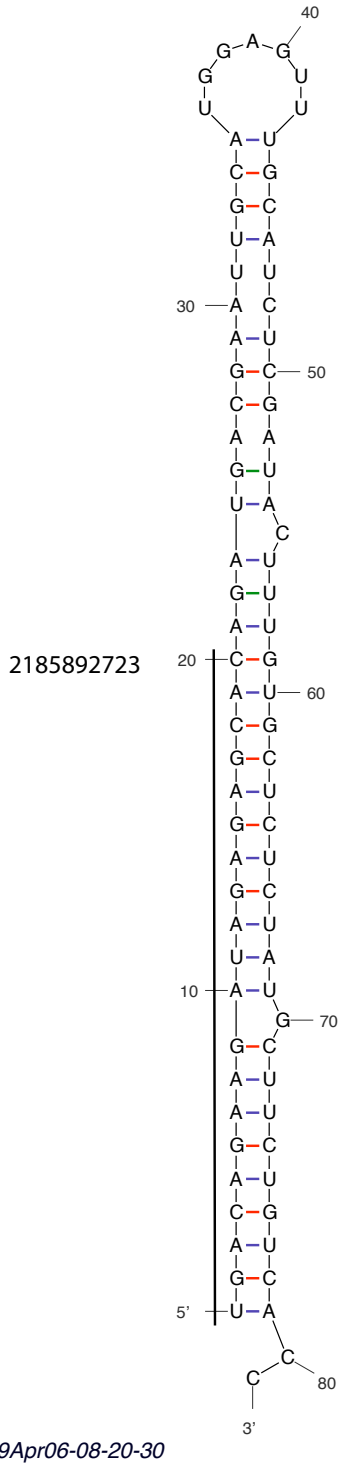
B. Size distribution of *Aquilegia coerulea miRNA* predicted precursors


20
2185892723mature U G A C A G A A G A U A G A G A G C A C
2185518132mature U G A C A G A A G A U A G A G A G C A C
DR949672binding U G A C A G A A G A G A G A G A G C A C
DR954092binding U G A C A G A A G A G A G A G A G C A C
TC21170binding U G A C A G A A G A G A G A G A G C A C
TC21707binding U G A C A G A A G A G A G A G A G C A C
TC21808binding U G A C A G A A G A G A G A G A G C A C
TC24142binding U G A C A G A A G A G A G A G A G C A C
TC26823binding U G A C A G A A G A G A G A G A G C A C
TC31356binding U G A C A G A A G A G A G A G A G C A U U
TC31380binding U G A C A G A A G A G A G A G A G C A C
TC31827binding U U A C A G A A G A G A G A G A G C A U U
TC32167binding U G A C A G A A G A G A G A G A G C A C
TC25281binding U G C C A G A A G A U A G A G U U C A C
U G A C A G A A G A G A G A G A G C A C


156
Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sative
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max
0
2
4
6
8
10
12
14
Number of miRNAs

Output of sir_graph (©)
mfold_util_ng 4.1

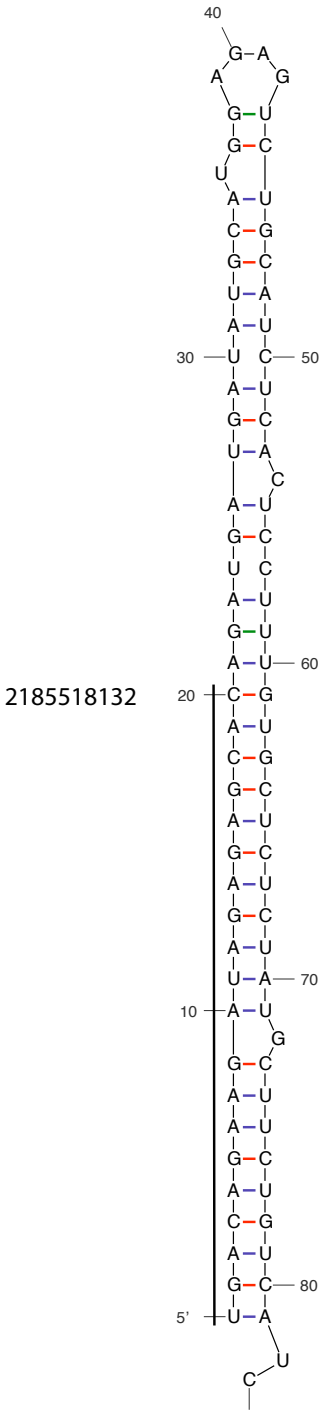
Created Mon Apr 6 08:20:43 2009



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

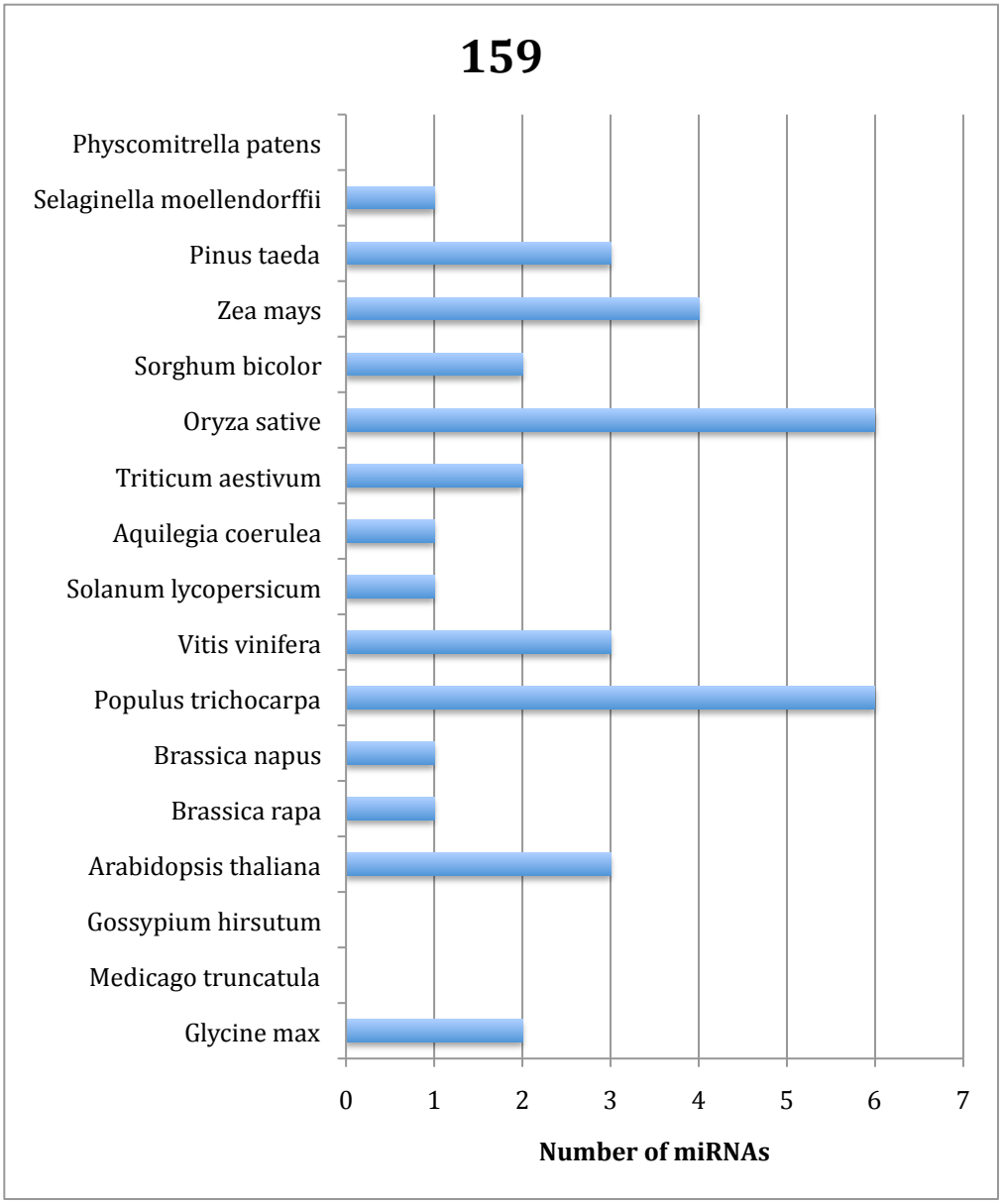
Output of sir_graph (©)
mfold_util_ng 4.1

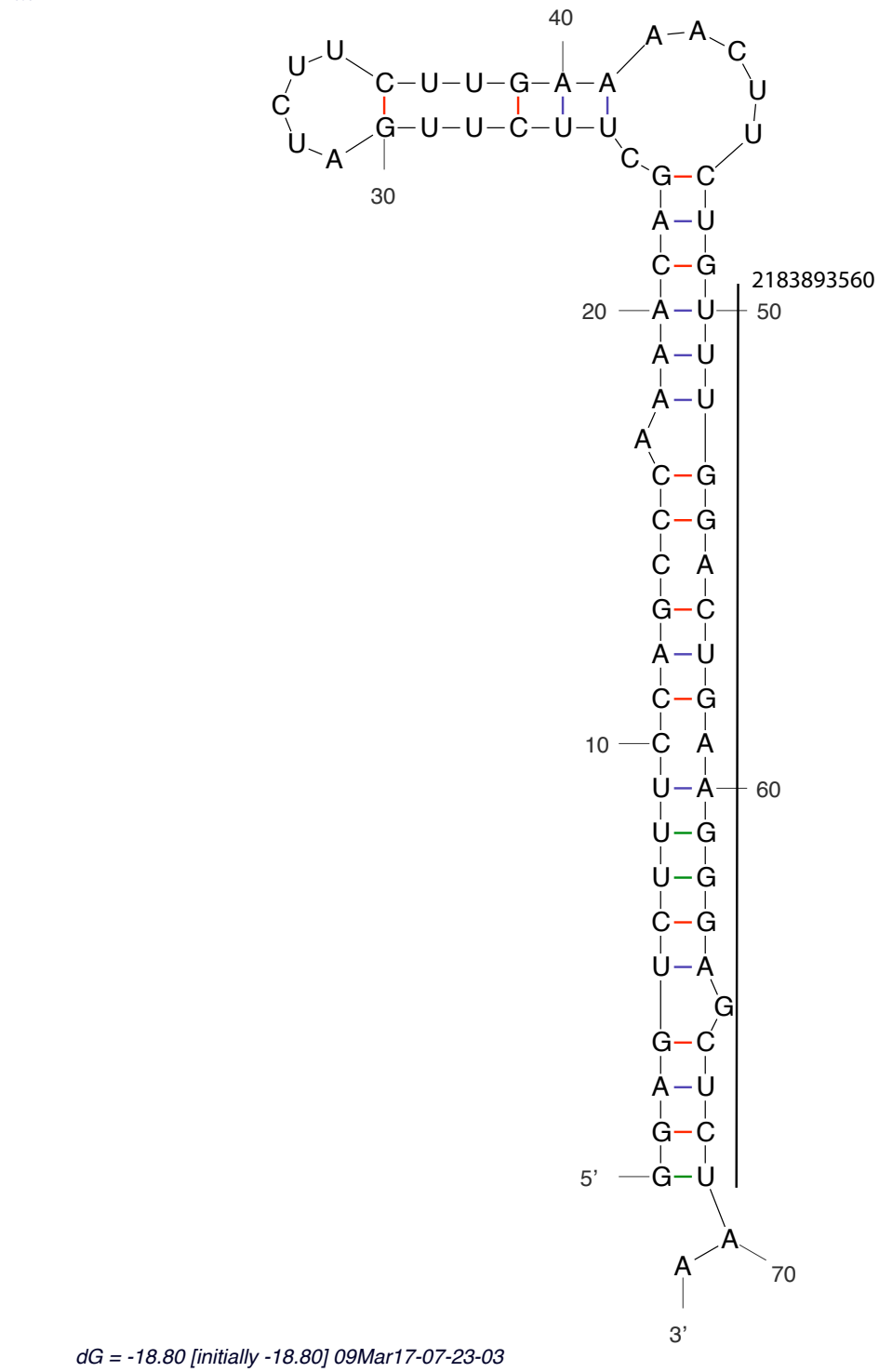
Created Mon Apr 6 08:46:33 2009



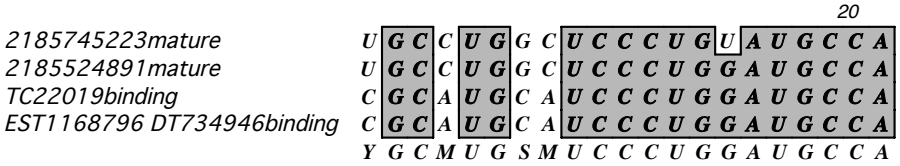
dG = -44.80 [initially -44.80] 09Apr06-08-46-26

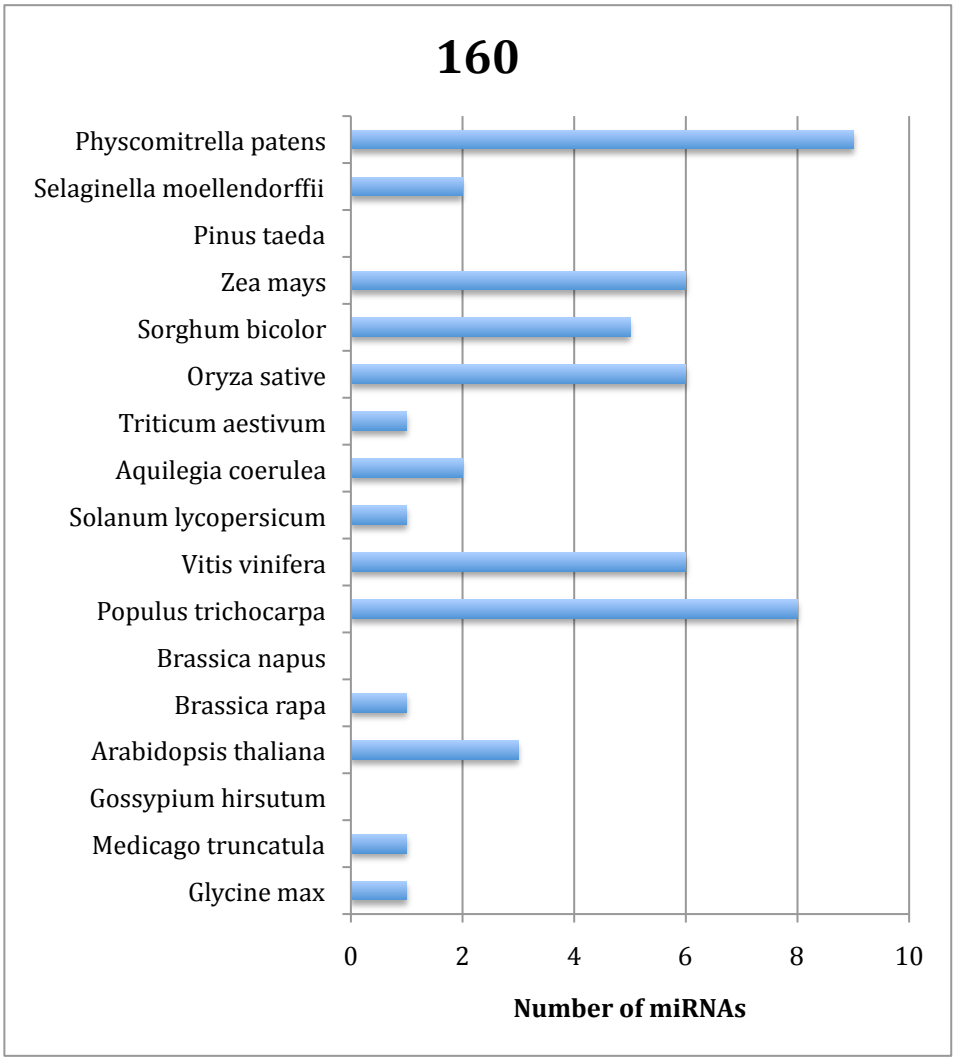
20
2183893560mature
U U U G G A C U G A A G G G A G C U C U A
U U U G G A C U G A A G G G A G C U C U A

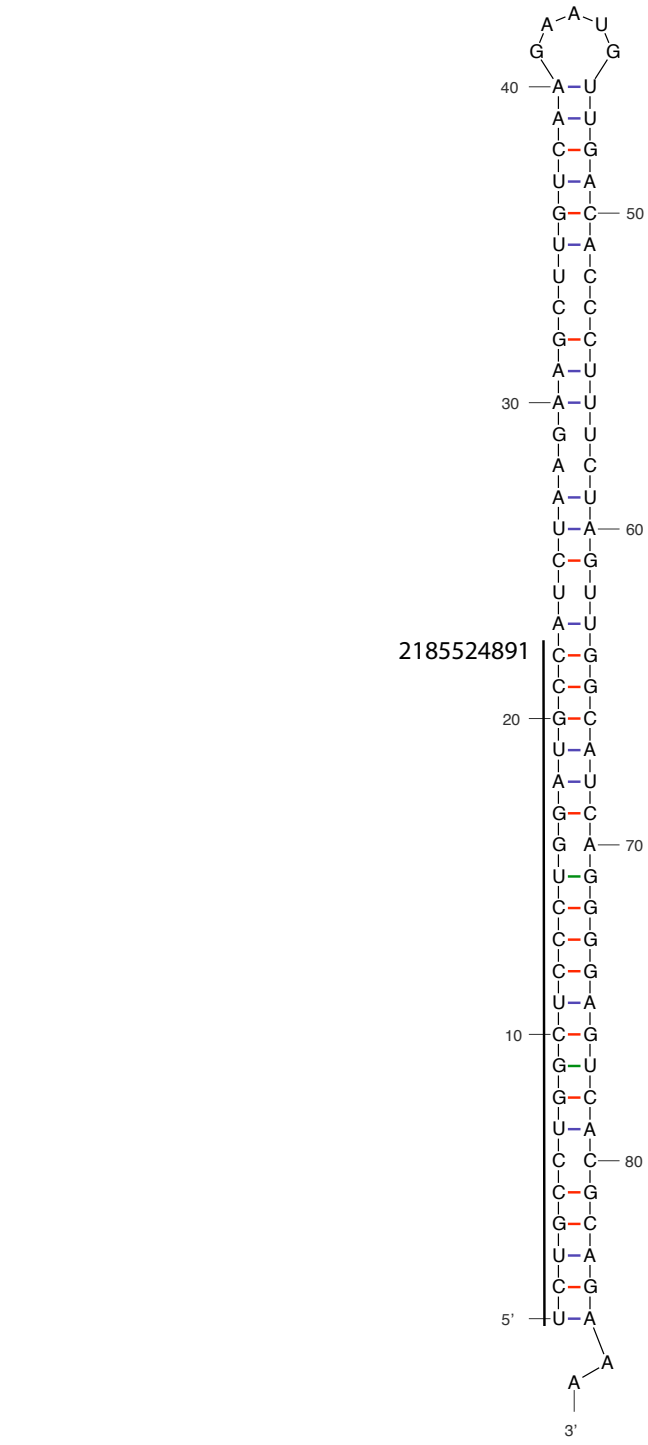

159
Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sative
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max
0
1
2
3
4
5
6
7
Number of miRNAs



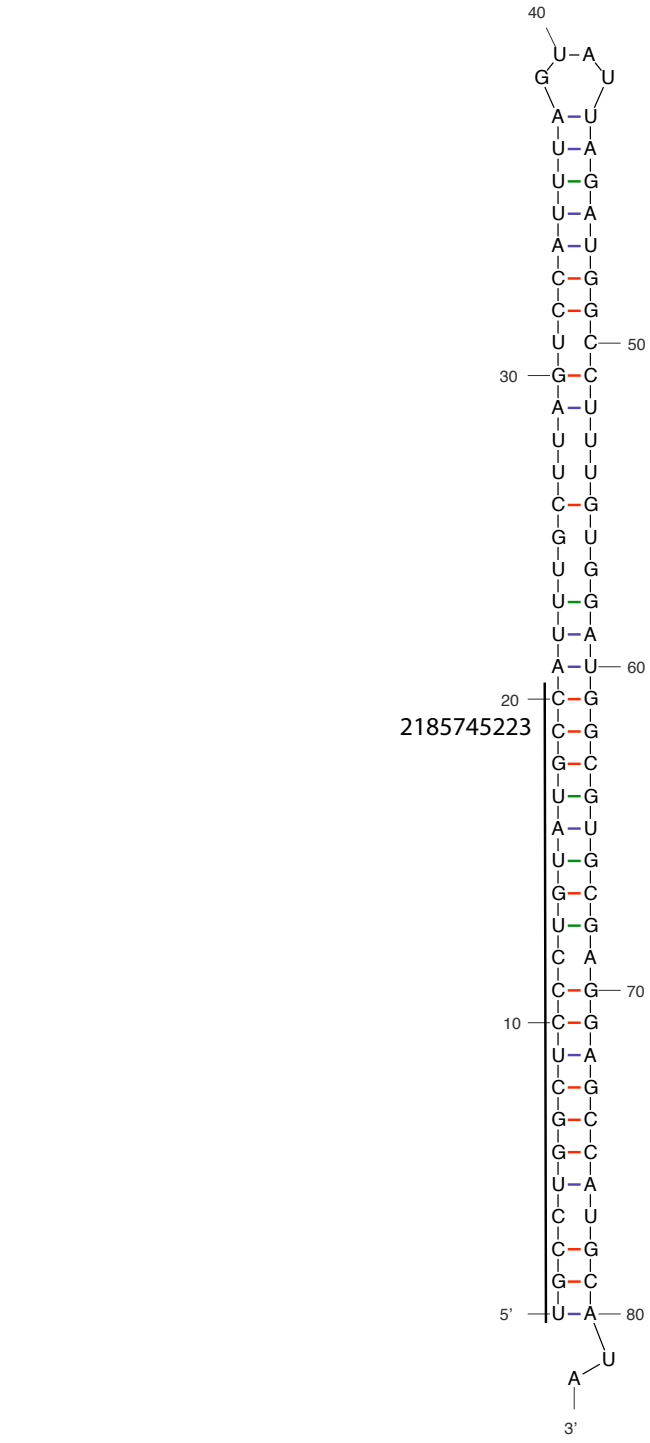
dG = -18.80 [initially -18.80] 09Mar17-07-23-03

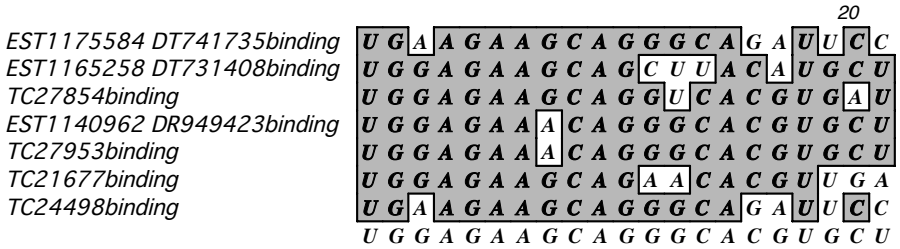

20
2185745223mature
2185524891mature
TC22019binding
EST1168796 DT734946binding
U G C C U G G C U C C C U G U A U G C C A
U G C C U G G C U C C C U G G A U G C C A
C G C A U G C A U C C C U G G A U G C C A
C G C A U G C A U C C C U G G A U G C C A
Y G C M U G S M U C C C U G G A U G C C A

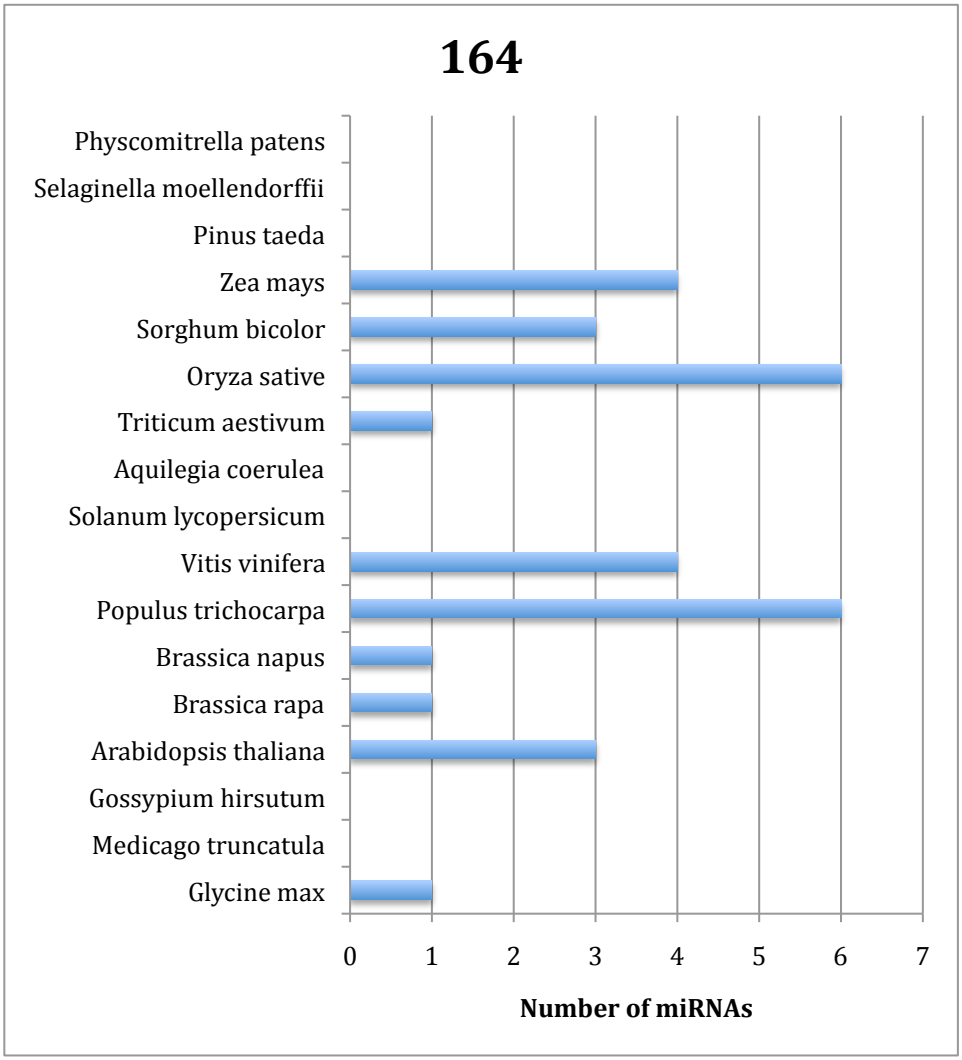

160
Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sative
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max
0
2
4
6
8
10
Number of miRNAs

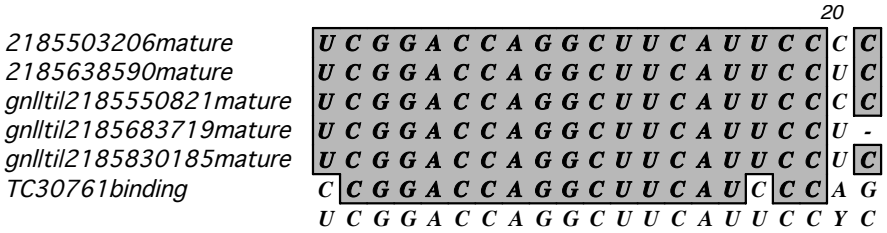


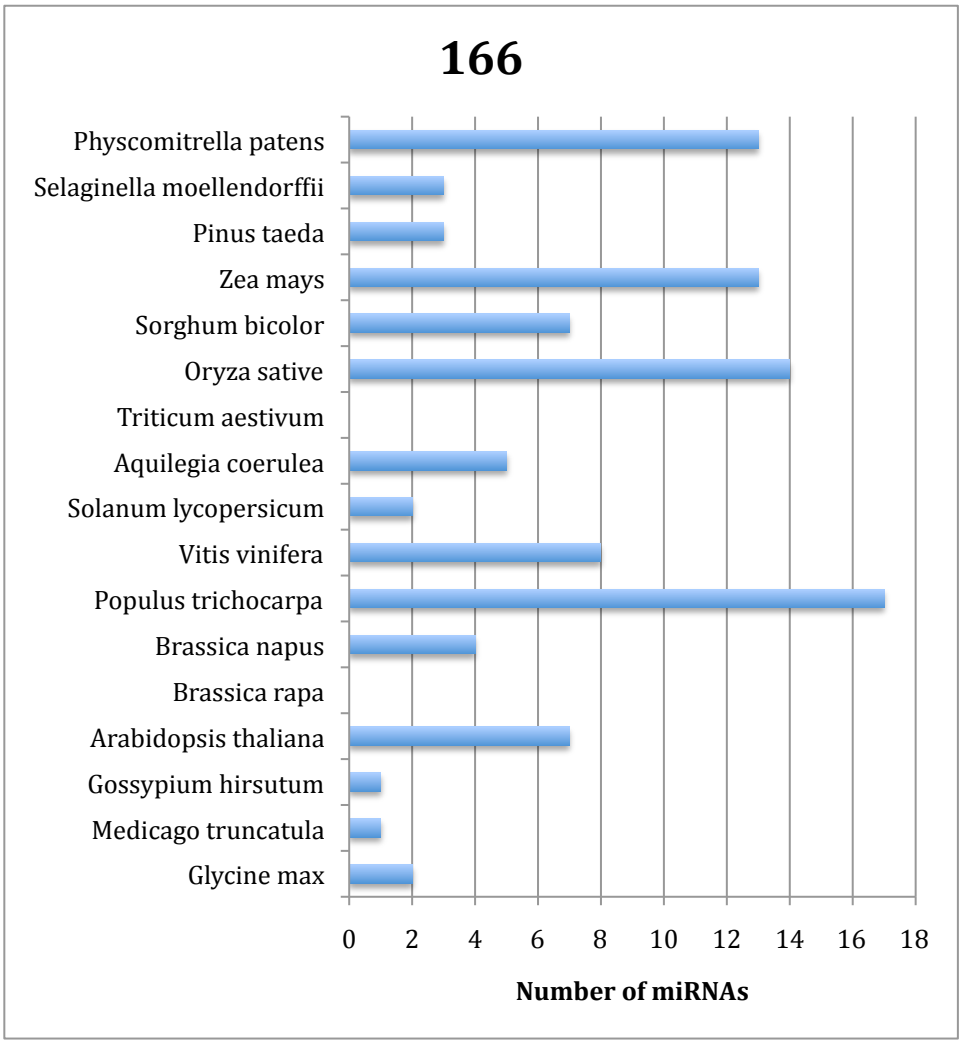
dG = -50.10 [initially -50.10] 09Mar13-14-54-08

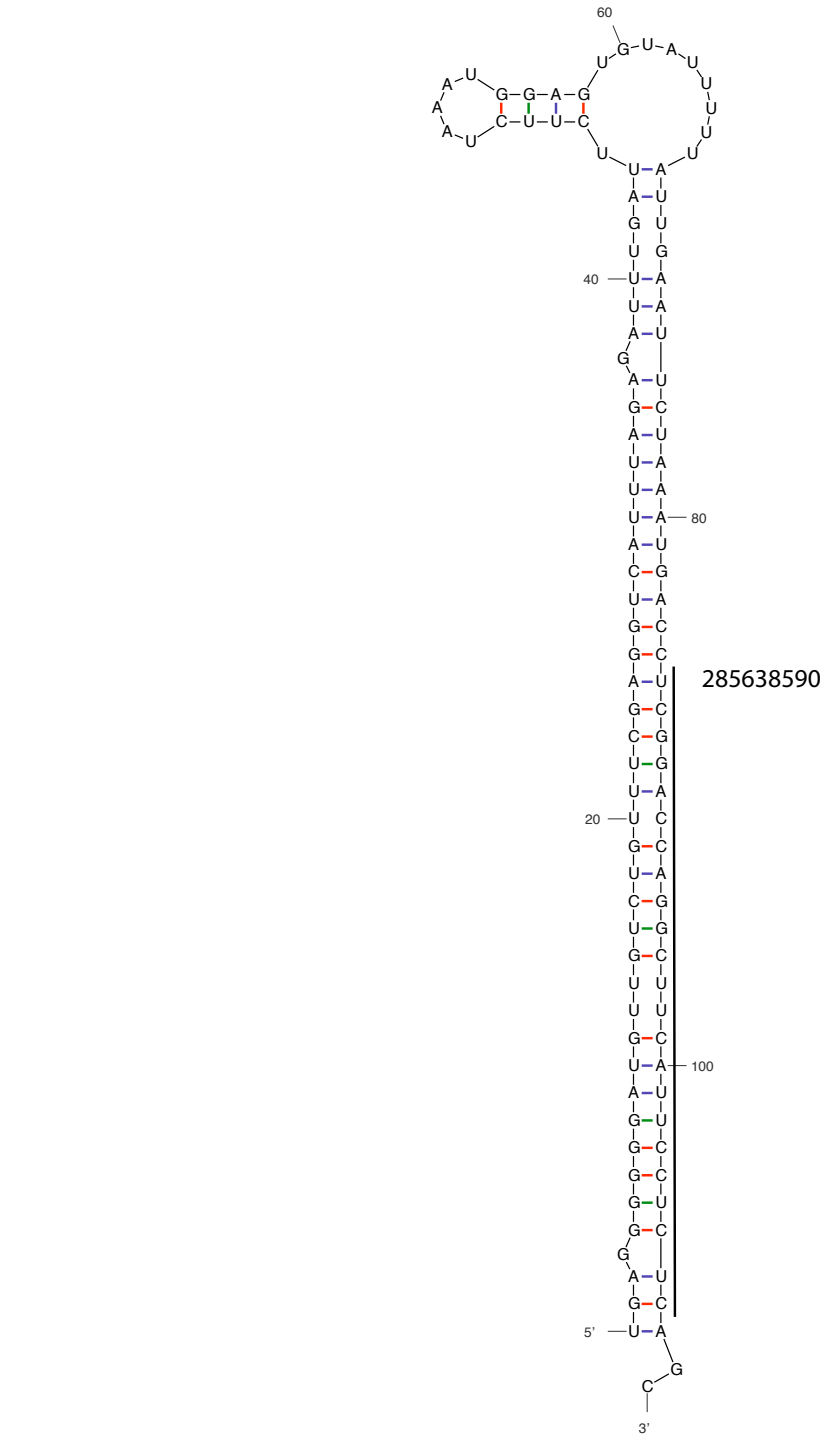

Output of sir_graph (®)
mfold 3.4
2185745223
dG = -48.10 [initially -48.10] 09Mar13-15-10-05


20
EST1175584 DT741735binding
EST1165258 DT731408binding
TC27854binding
EST1140962 DR949423binding
TC27953binding
TC21677binding
TC24498binding
U G A A G A A G C A G G G C A G A U U C C
U G G A G A A G C A G C U U A C A U G C U
U G G A G A A G C A G G U C A C G U G A U
U G G A G A A A C A G G G C A C G U G C U
U G G A G A A A C A G G G C A C G U G C U
U G G A G A A G C A G A A C A C G U U G A
U G A A G A A G C A G G G C A G A U U C C
U G G A G A A G C A G G G C A C G U G C U


164
Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sative
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max
0
1
2
3
4
5
6
7
Number of miRNAs


20
2185503206mature U C G G A C C A G G C U U C A U U C C C C
2185638590mature U C G G A C C A G G C U U C A U U C C U C
gnl|til2185550821mature U C G G A C C A G G C U U C A U U C C C
gnl|til2185683719mature U C G G A C C A G G C U U C A U U C C U -
gnl|til2185830185mature U C G G A C C A G G C U U C A U U C C U C
TC30761binding C C G G A C C A G G C U U C A U C C C A G
U C G G A C C A G G C U U C A U U C C Y C

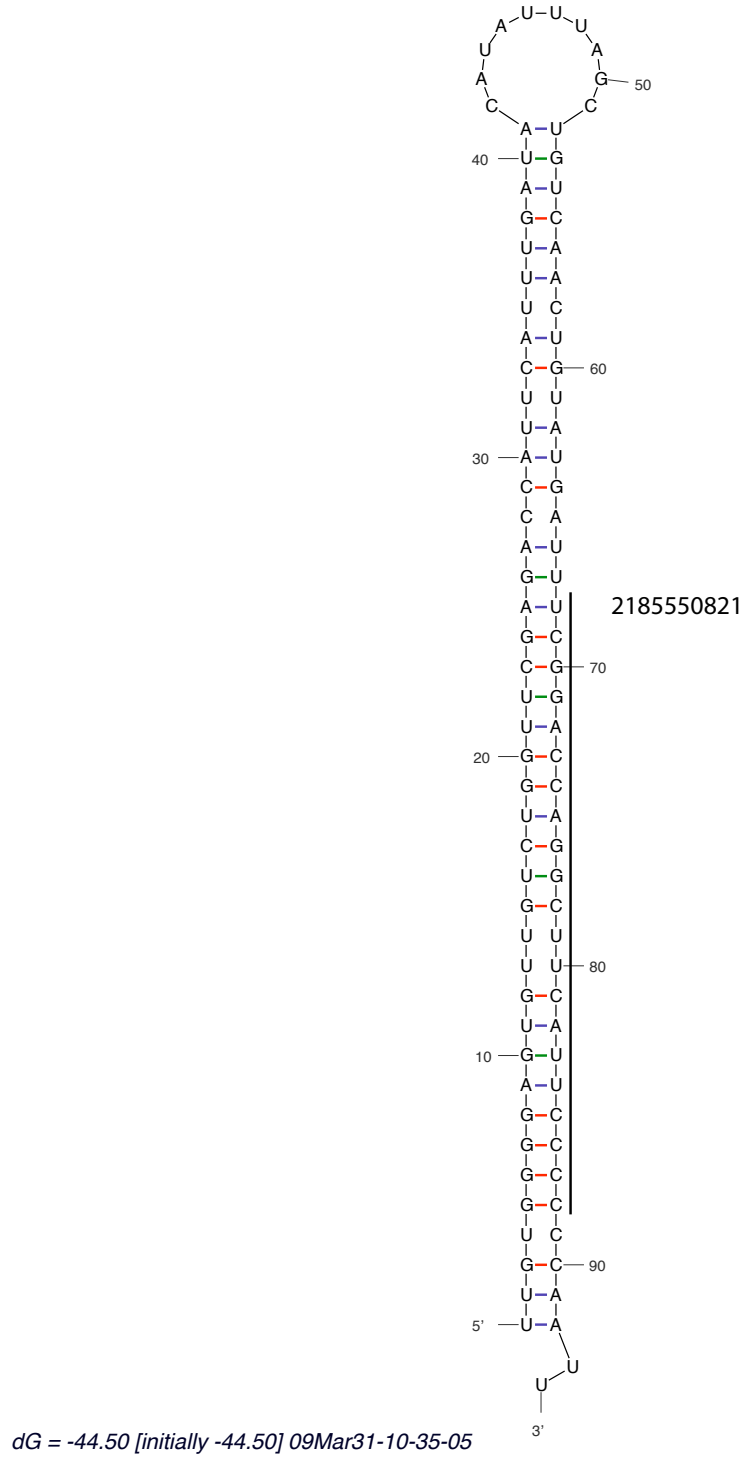

166
Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sative
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max
0 2 4 6 8 10 12 14 16 18
Number of miRNAs



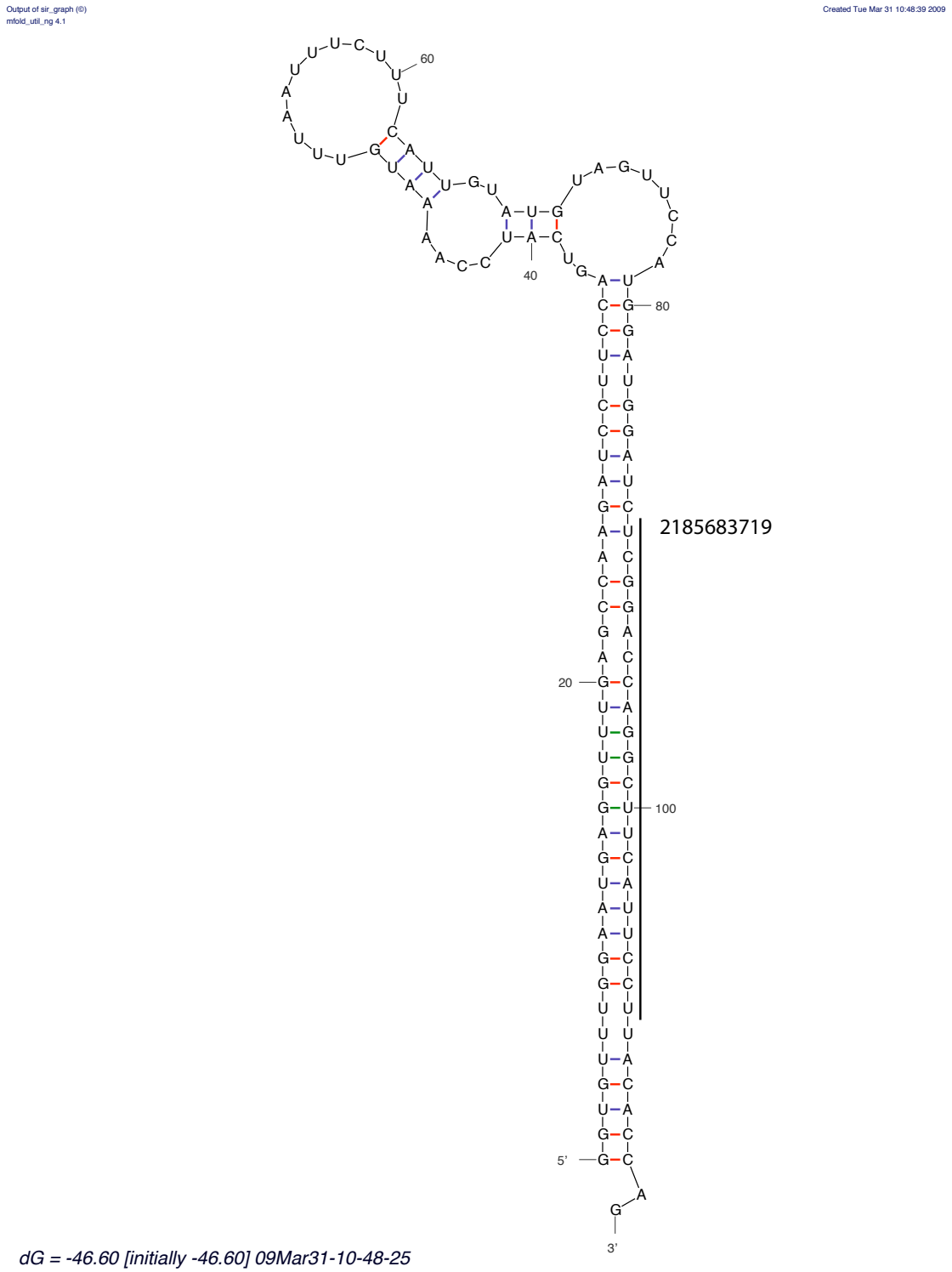
dG = -49.10 [initially -49.10] 09Mar16-09-34-45

Output of sir_graph (©)
mfold_util_ng 4.1

Created Tue Mar 31 10:35:09 2009



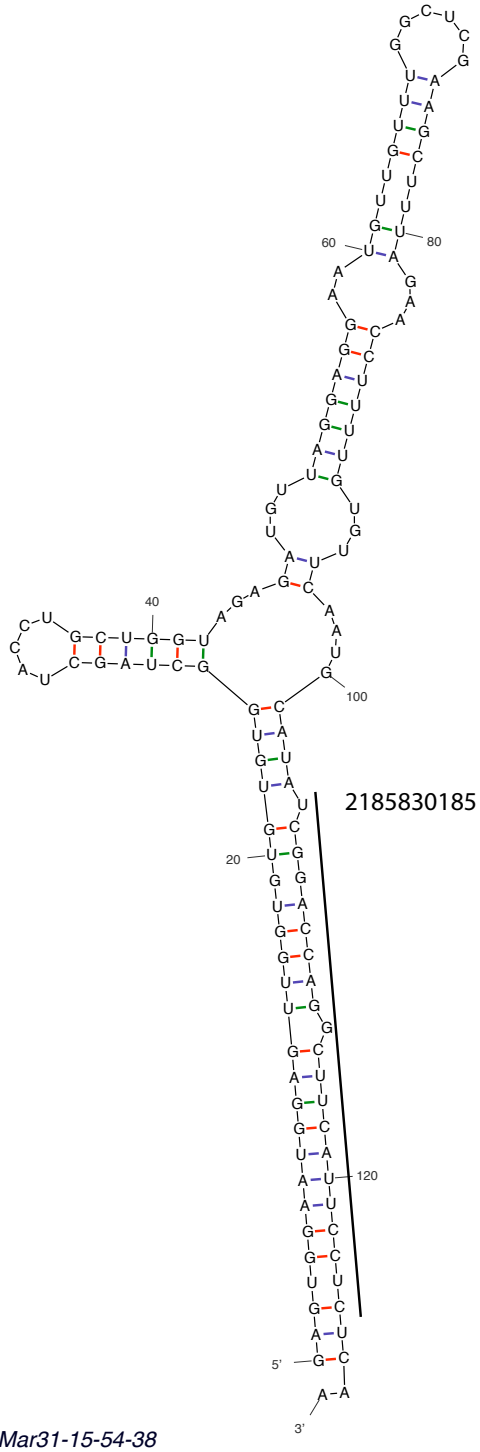
dG = -44.50 [initially -44.50] 09Mar31-10-35-05



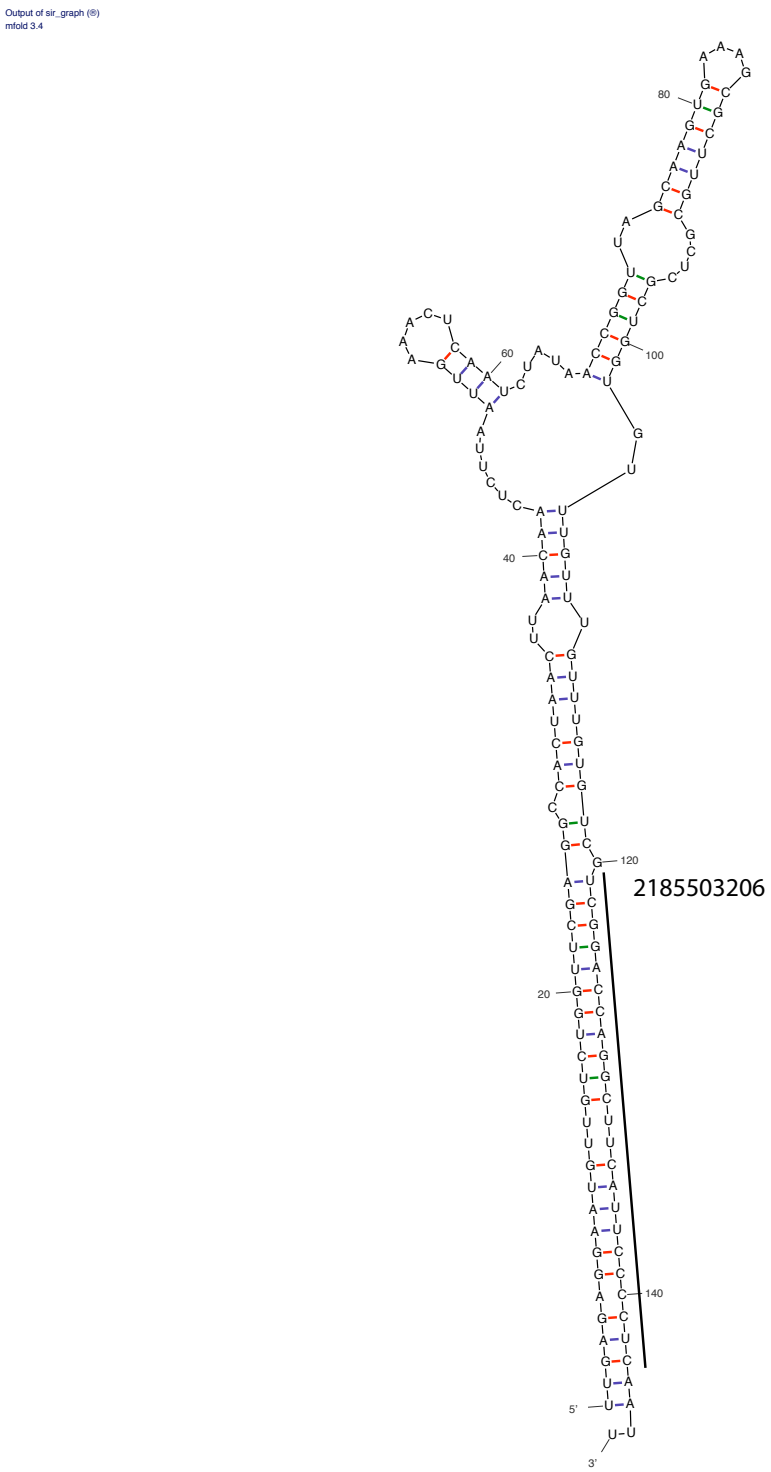
dG = -46.60 [initially -46.60] 09Mar31-10-48-25

Output of sir_graph (©)
mfold_util_ng 4.1

Created Tue Mar 31 15:54:48 2009

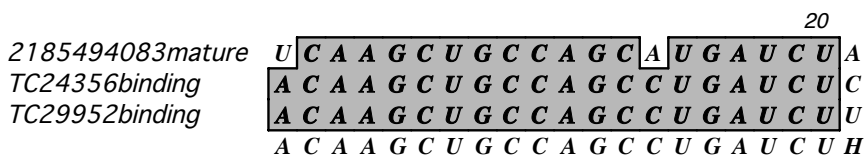


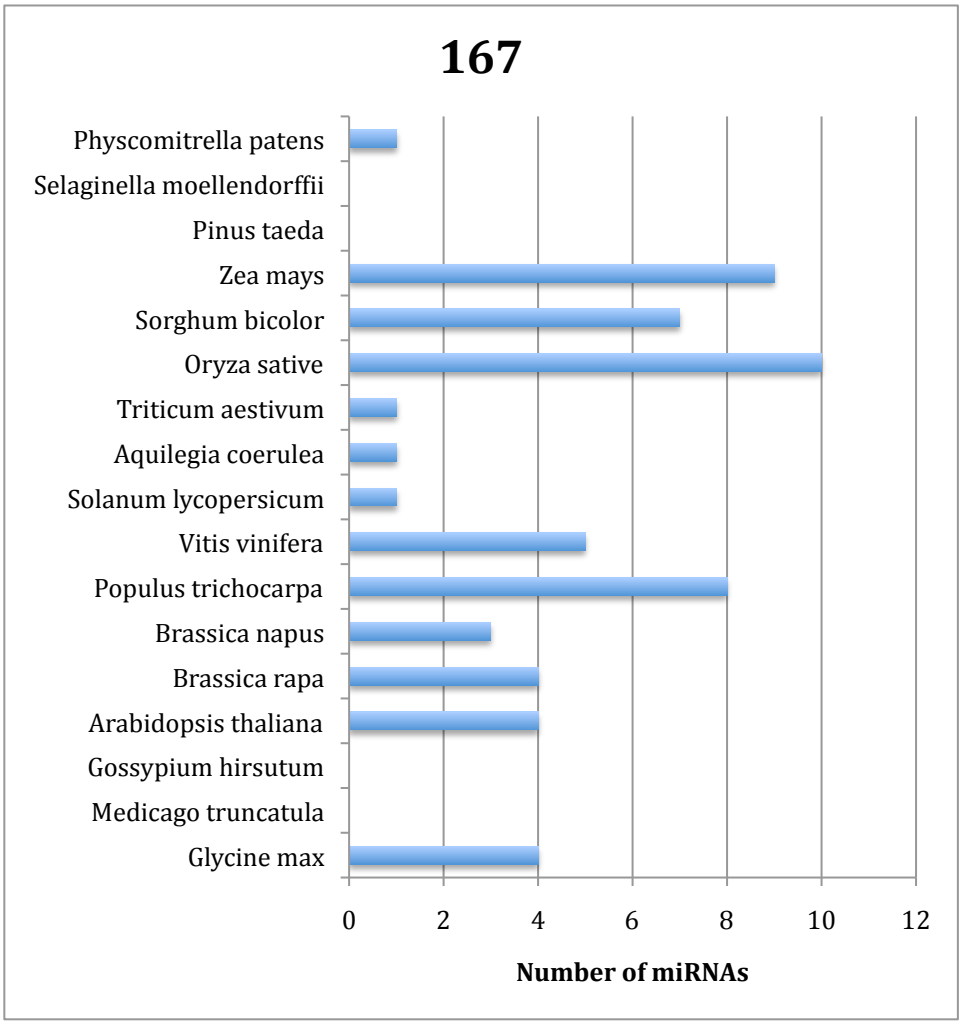
dG = -42.93 [initially -45.50] 09Mar31-15-54-38



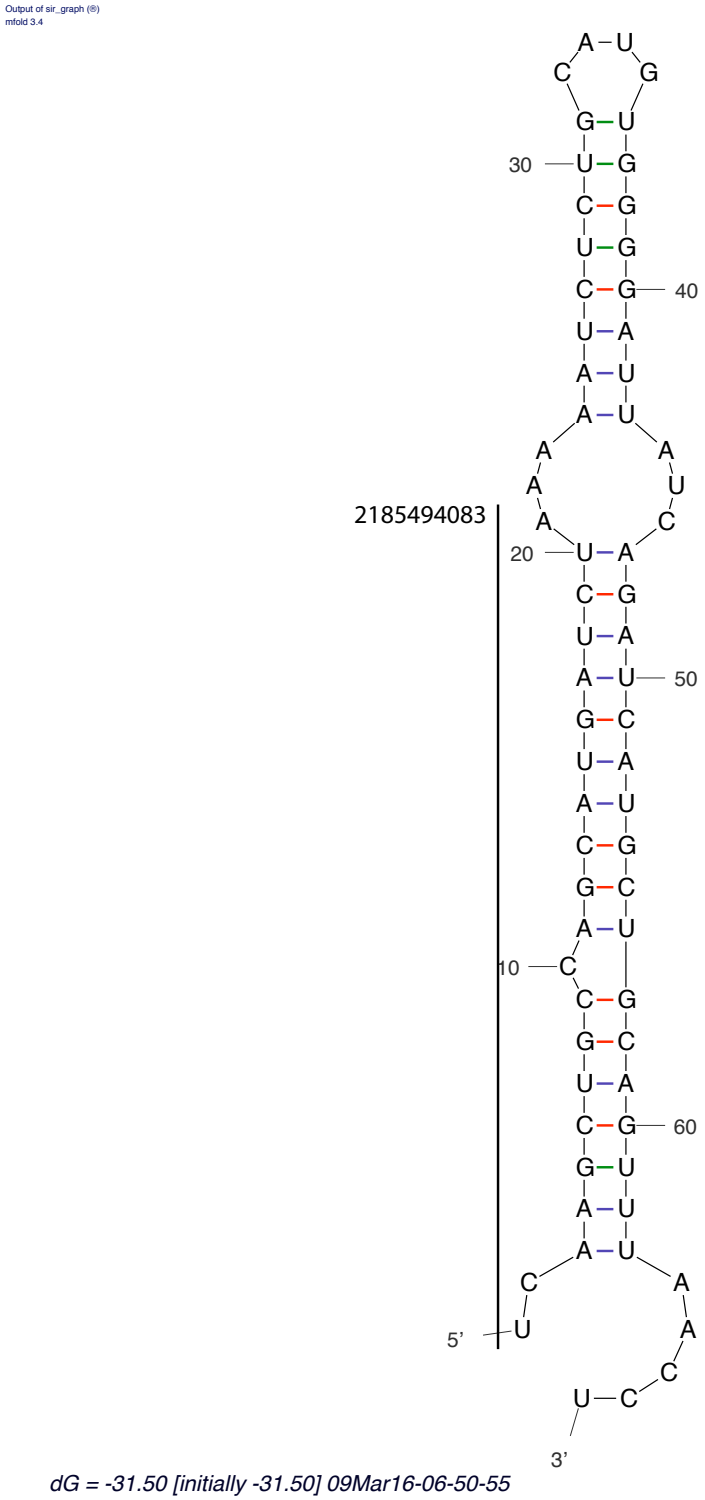
dG = -57.55 [initially -60.30] 09Mar16-09-20-37

microRNA167

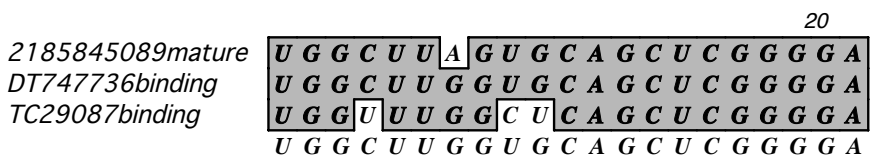


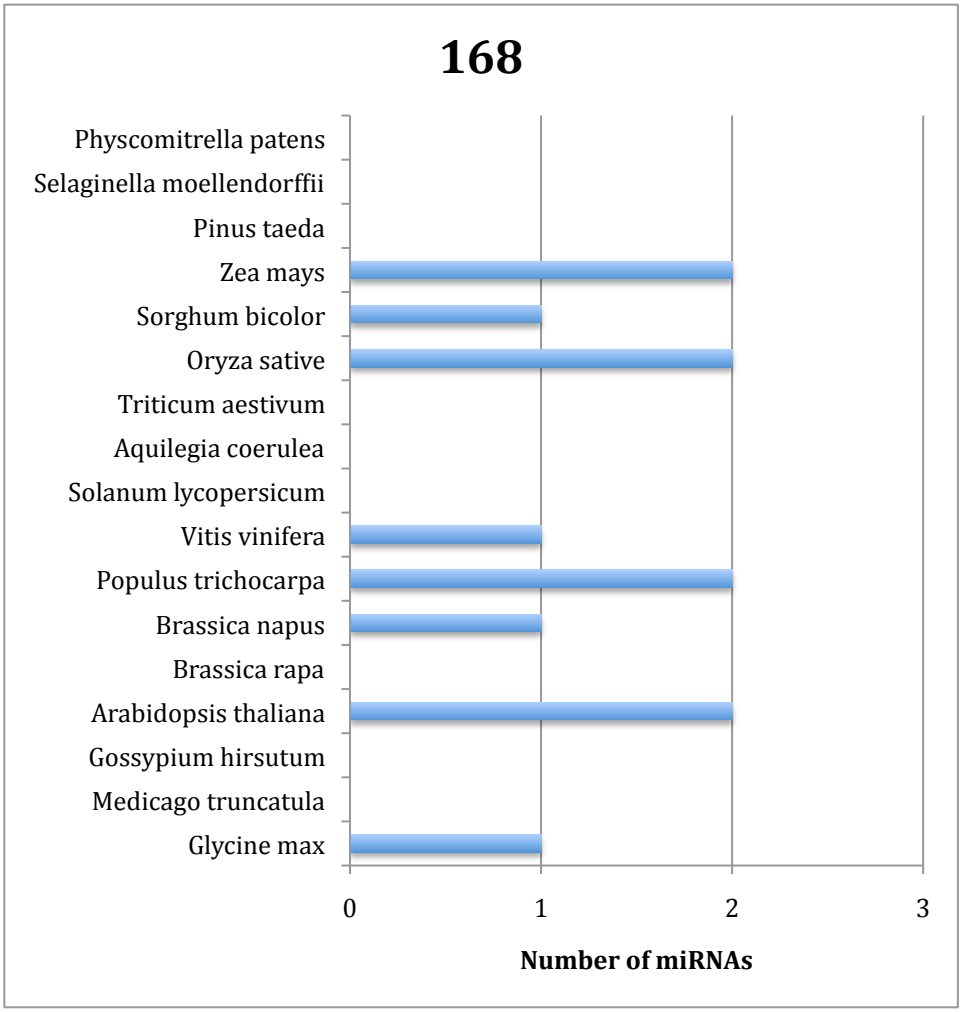


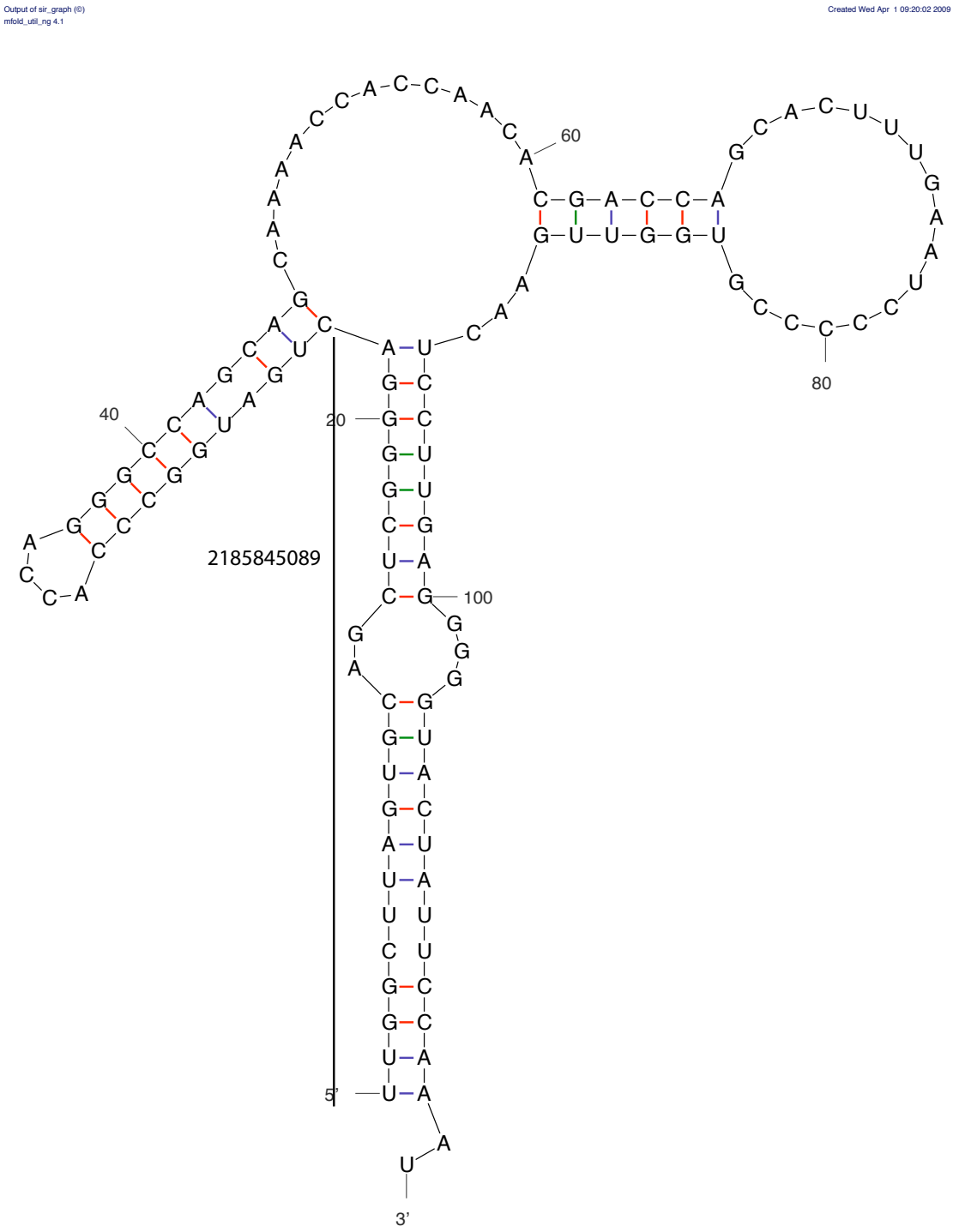
microRNA167



dG = -31.50 [initially -31.50] 09Mar16-06-50-55

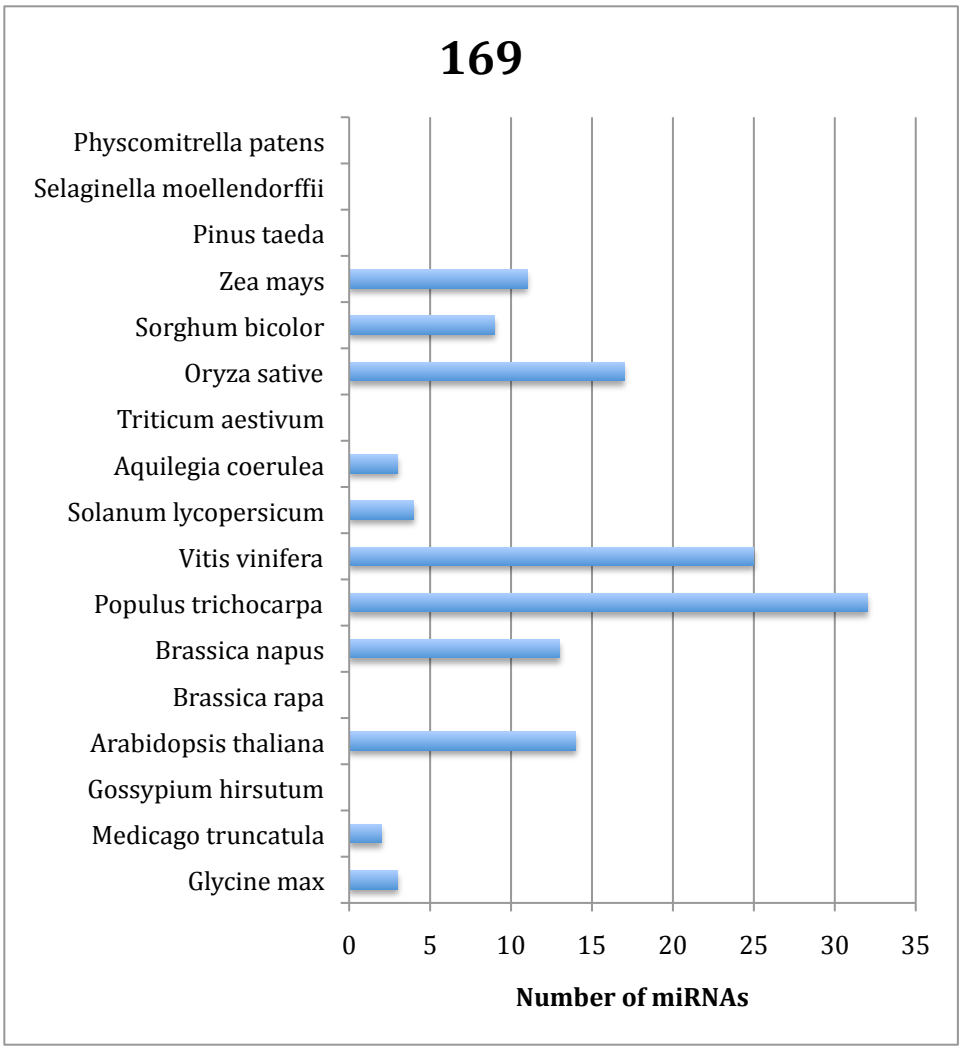

20
2185845089mature
DT747736binding
TC29087binding
U G G C U U A G U G C A G C U C G G G G A
U G G C U U G G U G C A G C U C G G G G A
U G G U U U G G C U C A G C U C G G G G A
U G G C U U G G U G C A G C U C G G G G A


168
Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sative
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max
0
1
2
3
Number of miRNAs



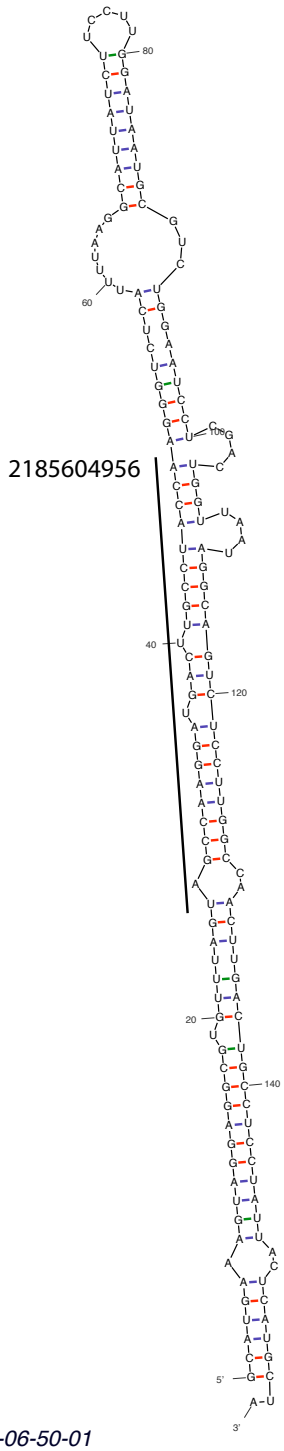
dG = -40.45 [initially -42.70] 09Apr01-09-19-45

```
                                                                  20
2185493504mature          - C A G C C A A G G A U G A C U U G C C G G G
gnlltil2185604956mature   G U A G C C A A G G A U G A C U U G C C U - -
gnlltil2185687520mature   - U A G C C A A G G A U G A C U U G C C U G G
DR924236binding           - A A G C C A A G G A U G A A U U G C C U A
DR927955binding           - A A G U C A A G G A U G A C U U U C C U A
TC31138binding            - A A G U C A A G G A U G A C U U U C C U A
                          G W A G C C A A G G A U G A C U U G C C U A
```

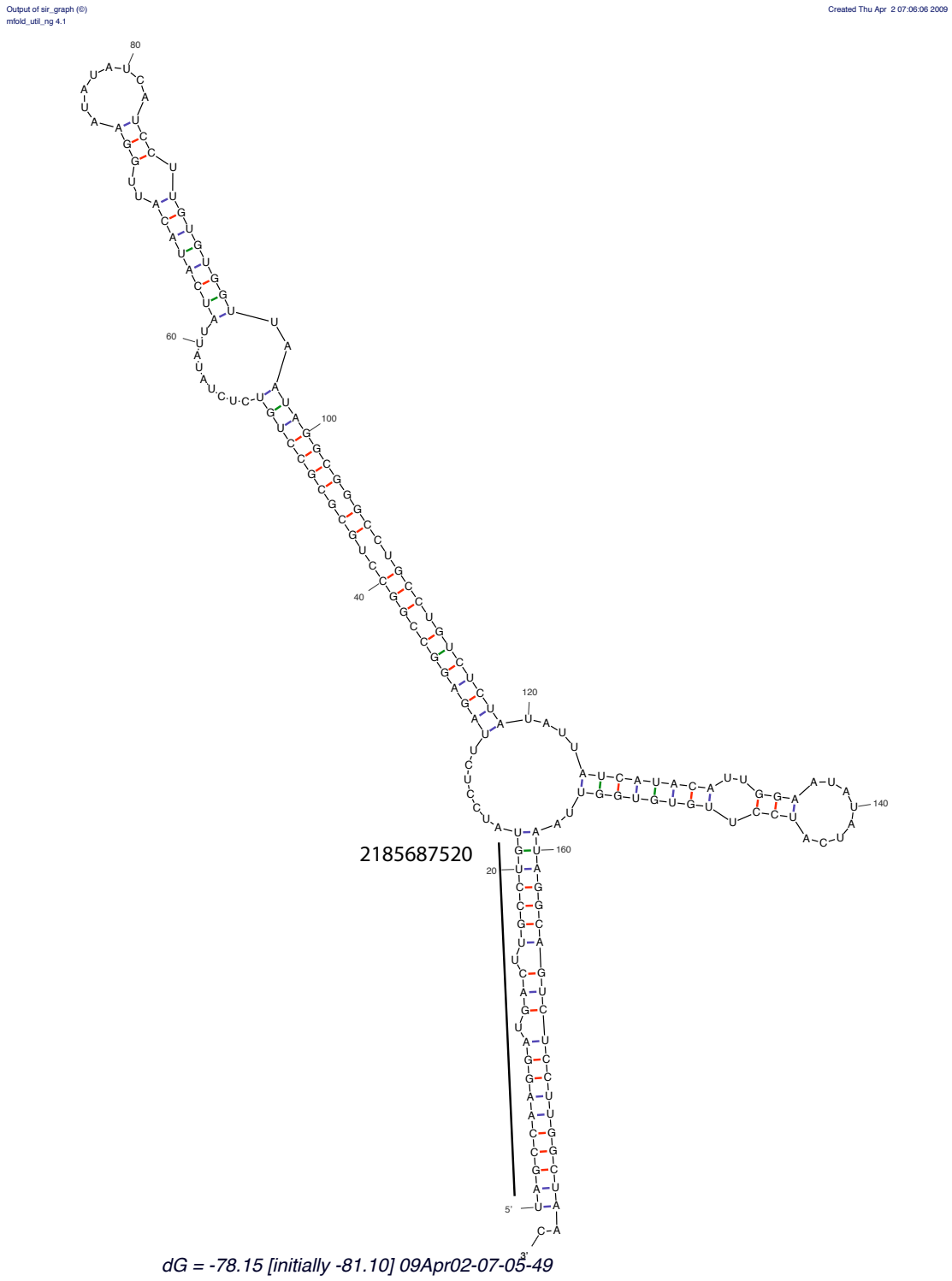


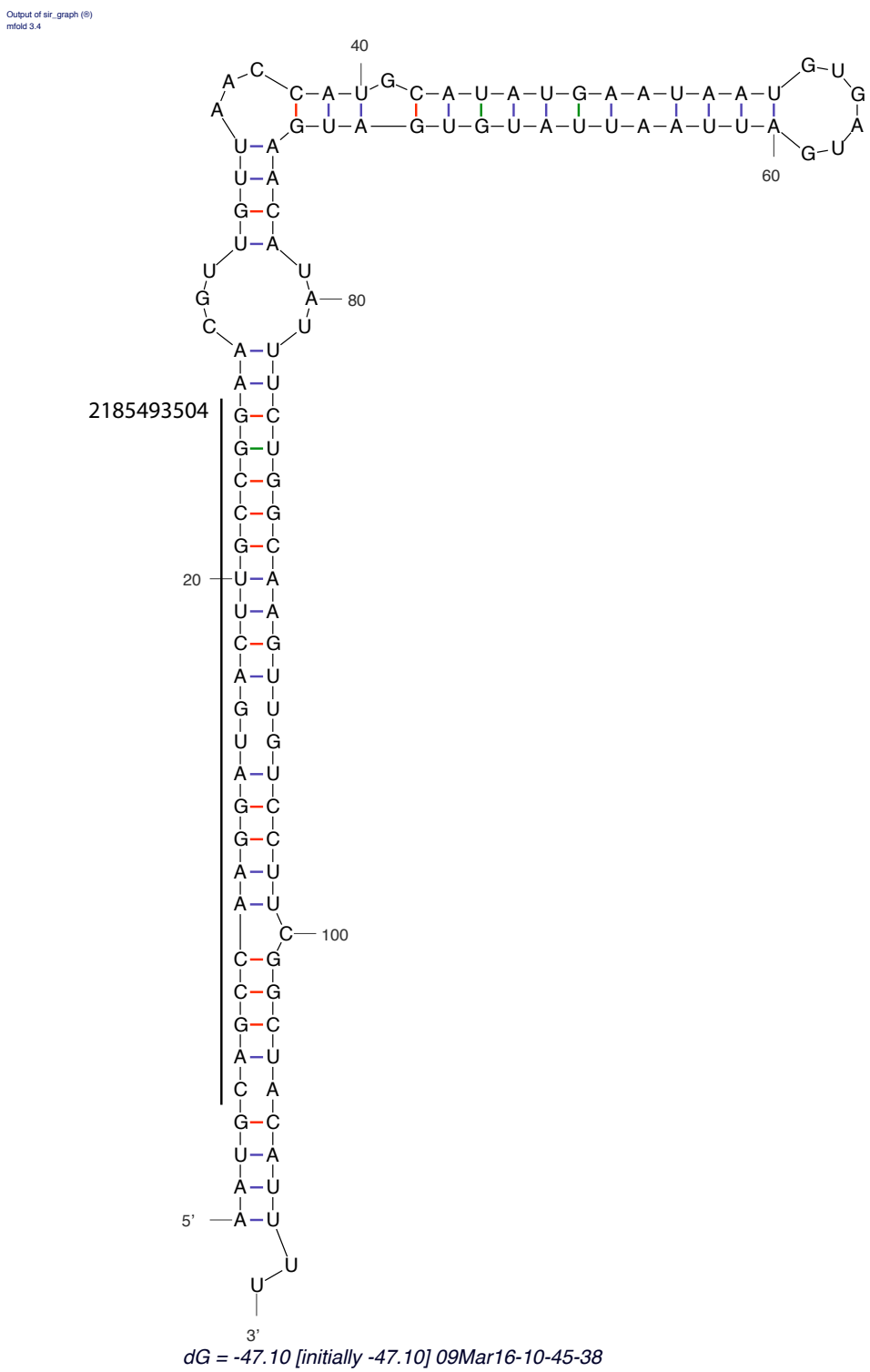
Output of sir_graph (©)
mfold_util_ng 4.1

Created Thu Apr 2 06:50:07 2009

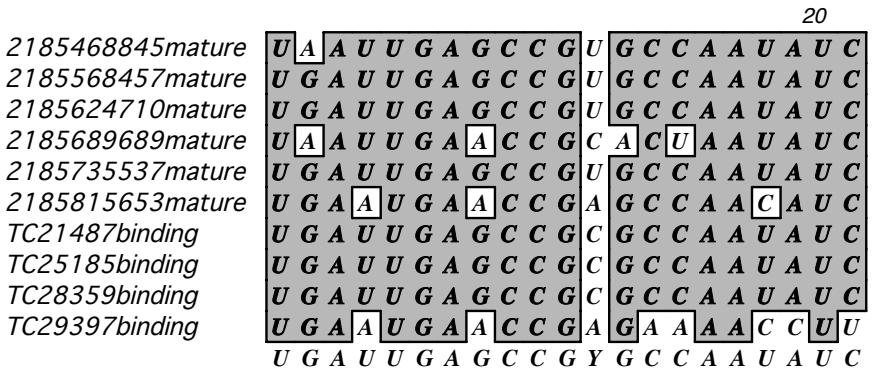


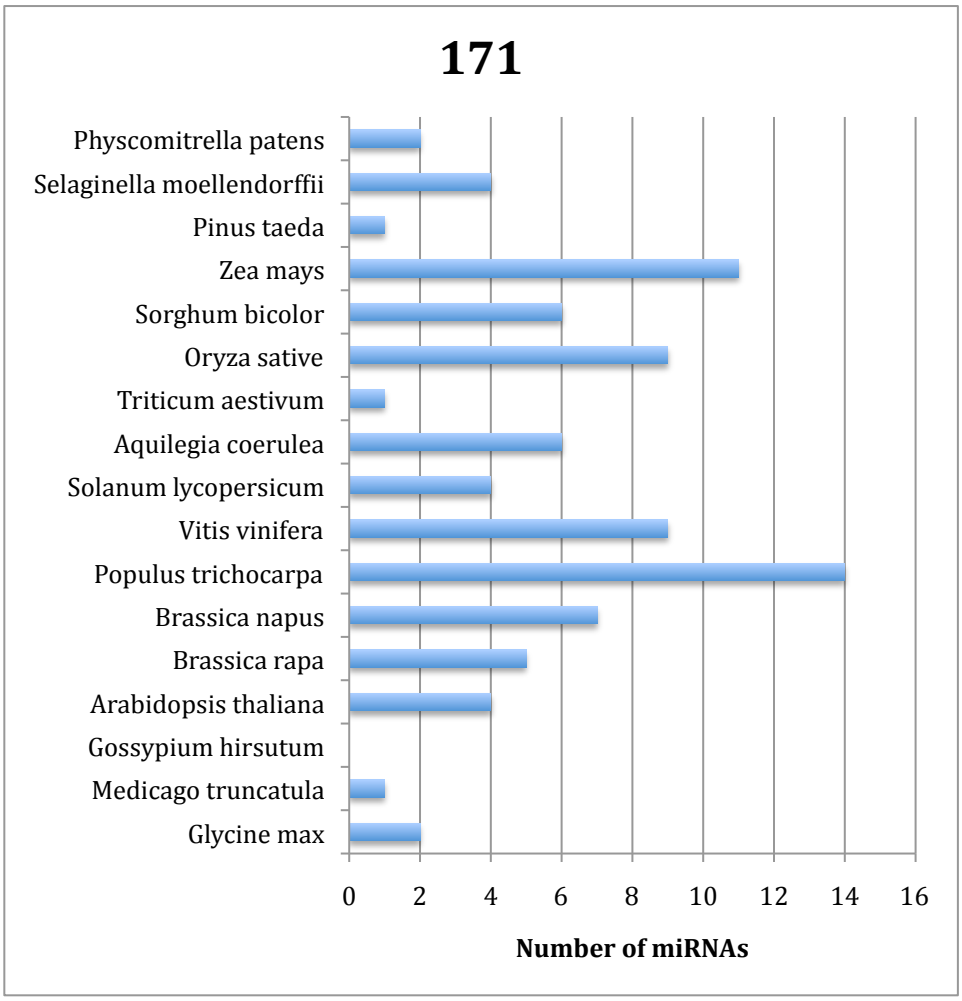
dG = -75.30 [initially -75.30] 09Apr02-06-50-01

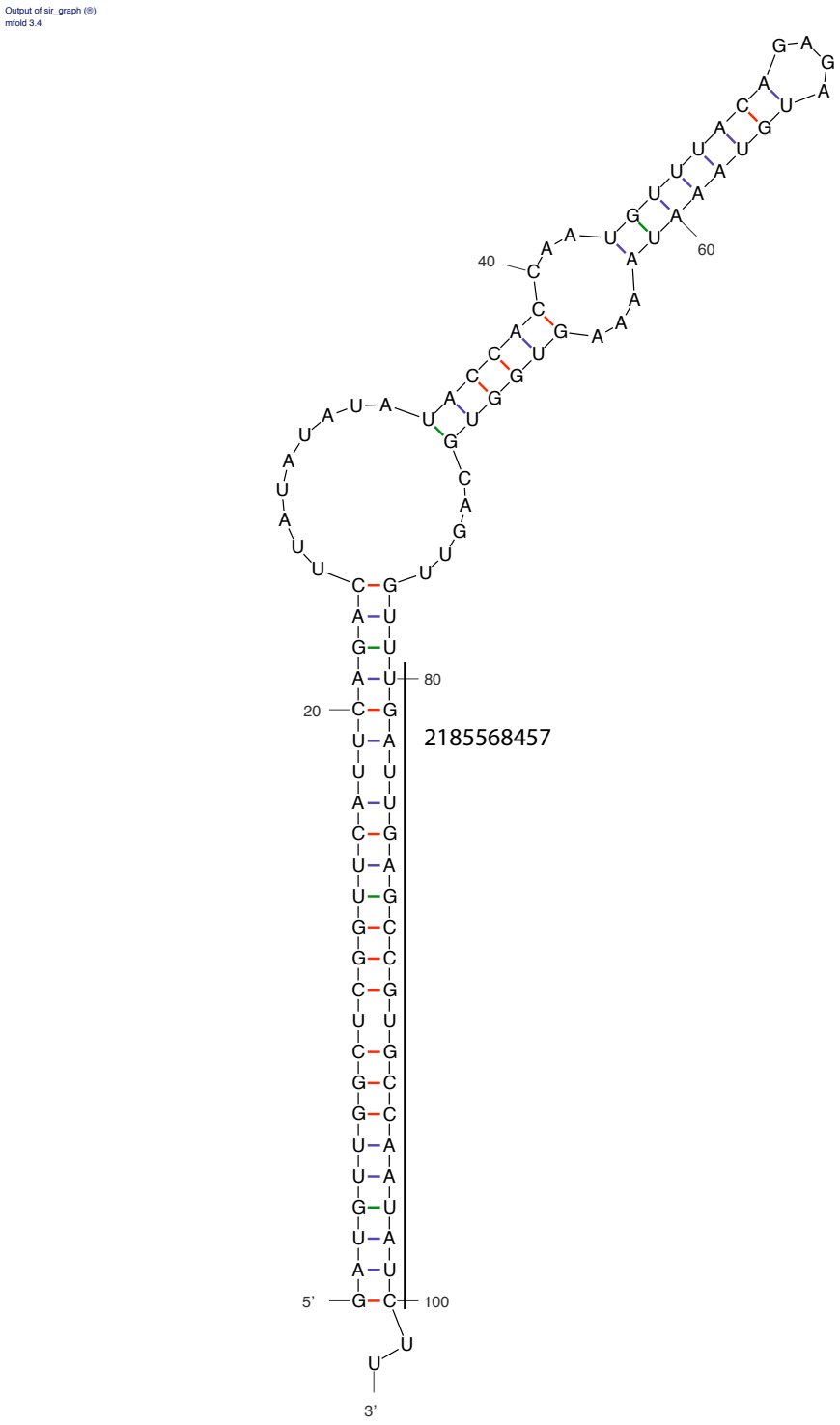

Output of sir_graph (©)
mfold_util_ng 4.1
Created Thu Apr 2 07:06:06 2009
2185687520
dG = -78.15 [initially -81.10] 09Apr02-07-05-49



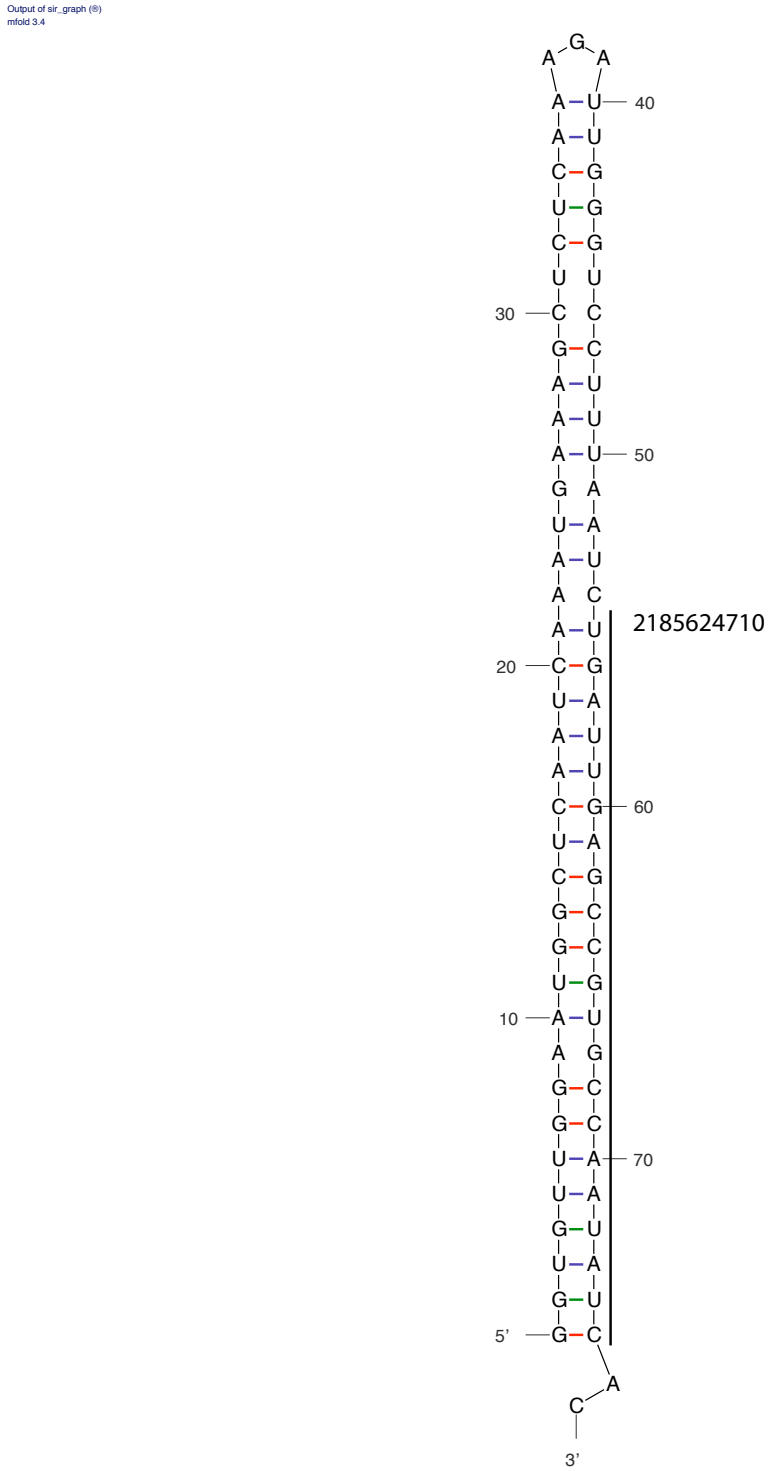
dG = -47.10 [initially -47.10] 09Mar16-10-45-38


20
2185468845mature
2185568457mature
2185624710mature
2185689689mature
2185735537mature
2185815653mature
TC21487binding
TC25185binding
TC28359binding
TC29397binding
U G A U U G A G C C G Y G C C A A U A U C


171
Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sative
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max
0
2
4
6
8
10
12
14
16
Number of miRNAs

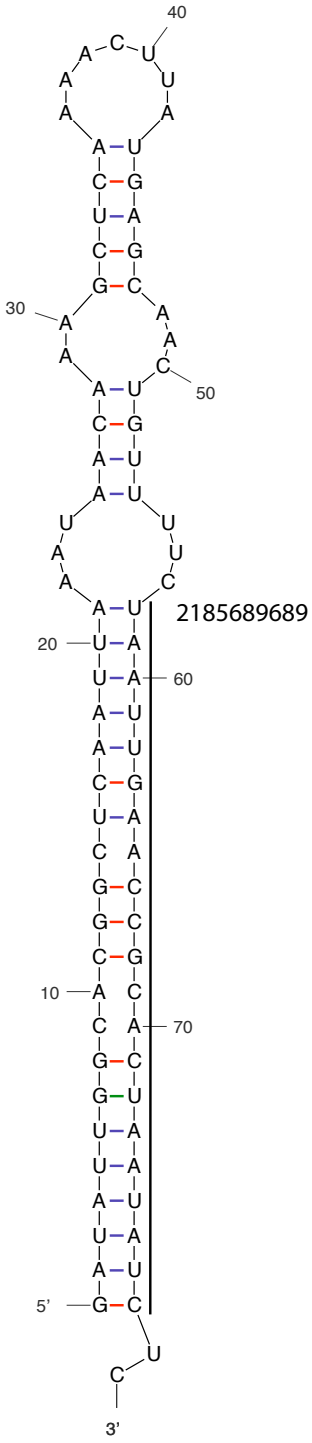


dG = -46.30 [initially -46.30] 09Mar16-11-29-08

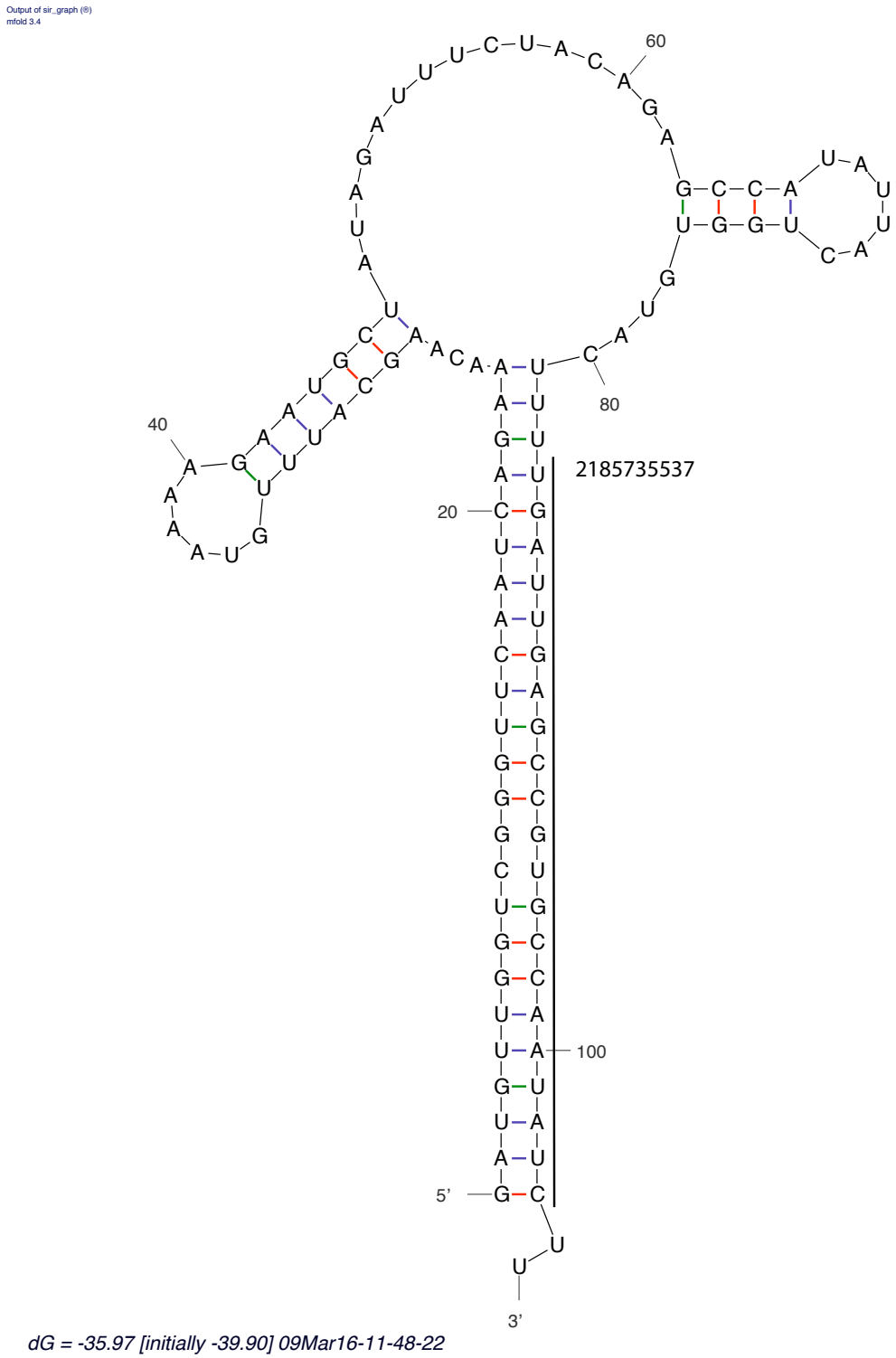


dG = -36.60 [initially -36.60] 09Mar16-11-35-59

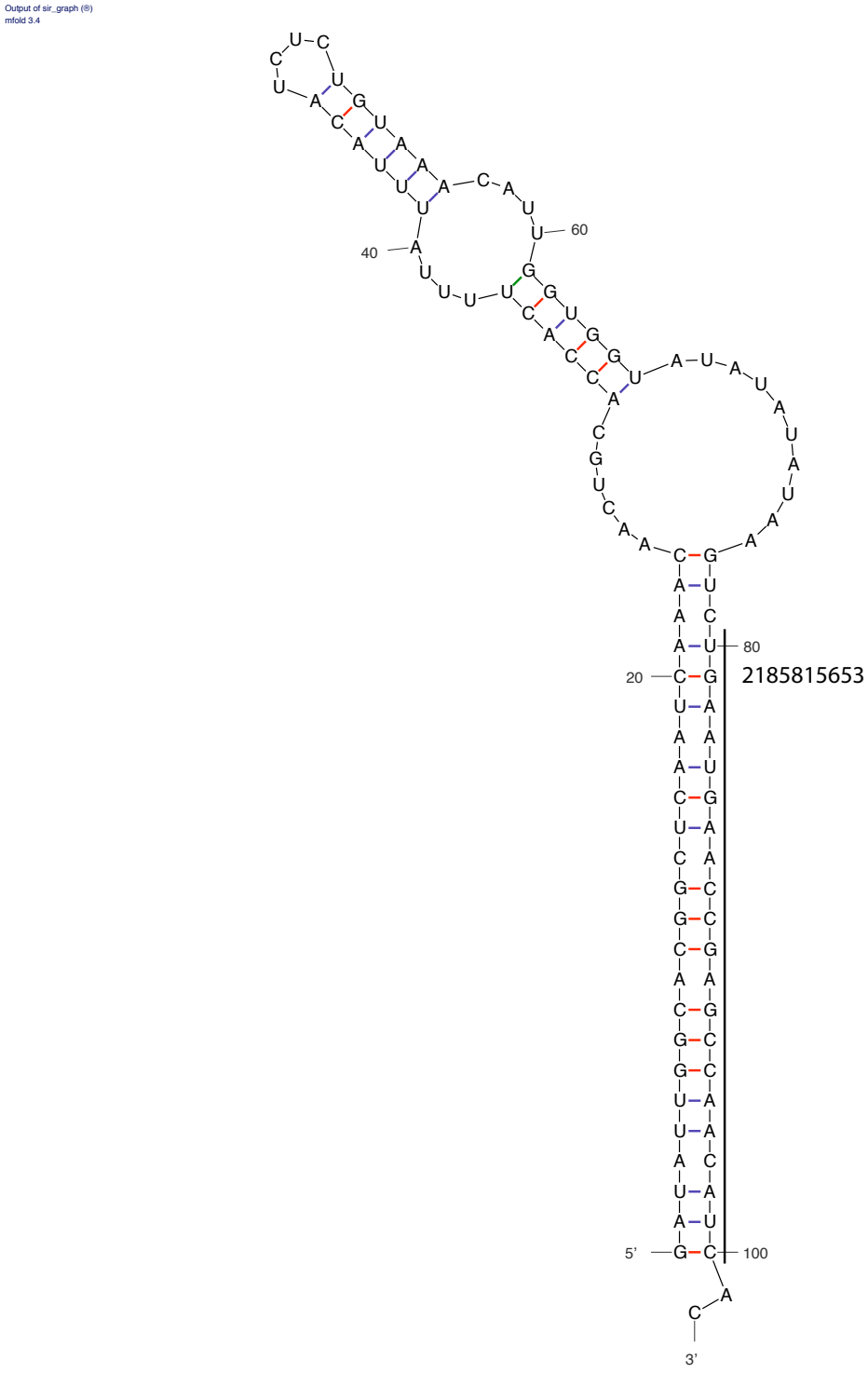
Output of sir_graph (®)
mfold 3.4



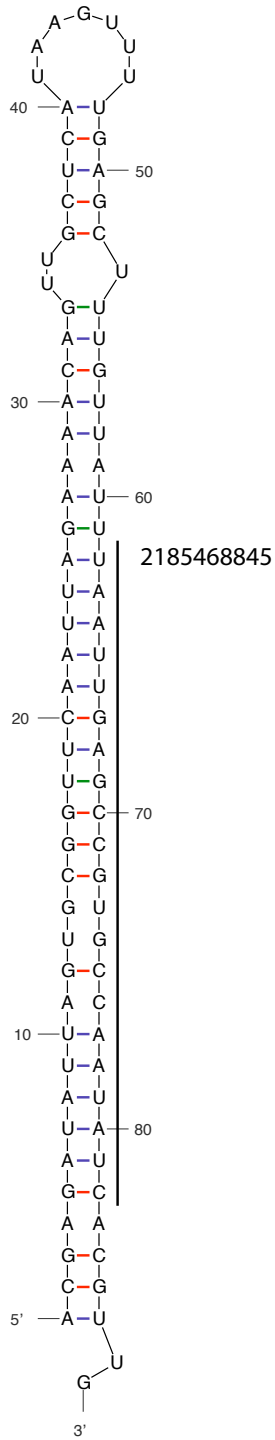
dG = -27.30 [initially -27.30] 09Mar16-11-42-19



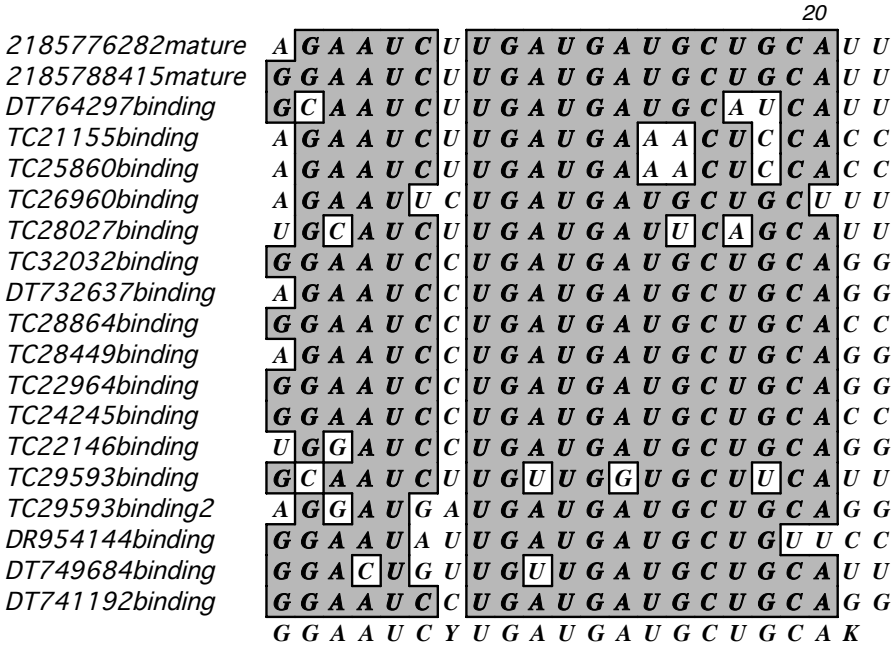
dG = -35.97 [initially -39.90] 09Mar16-11-48-22

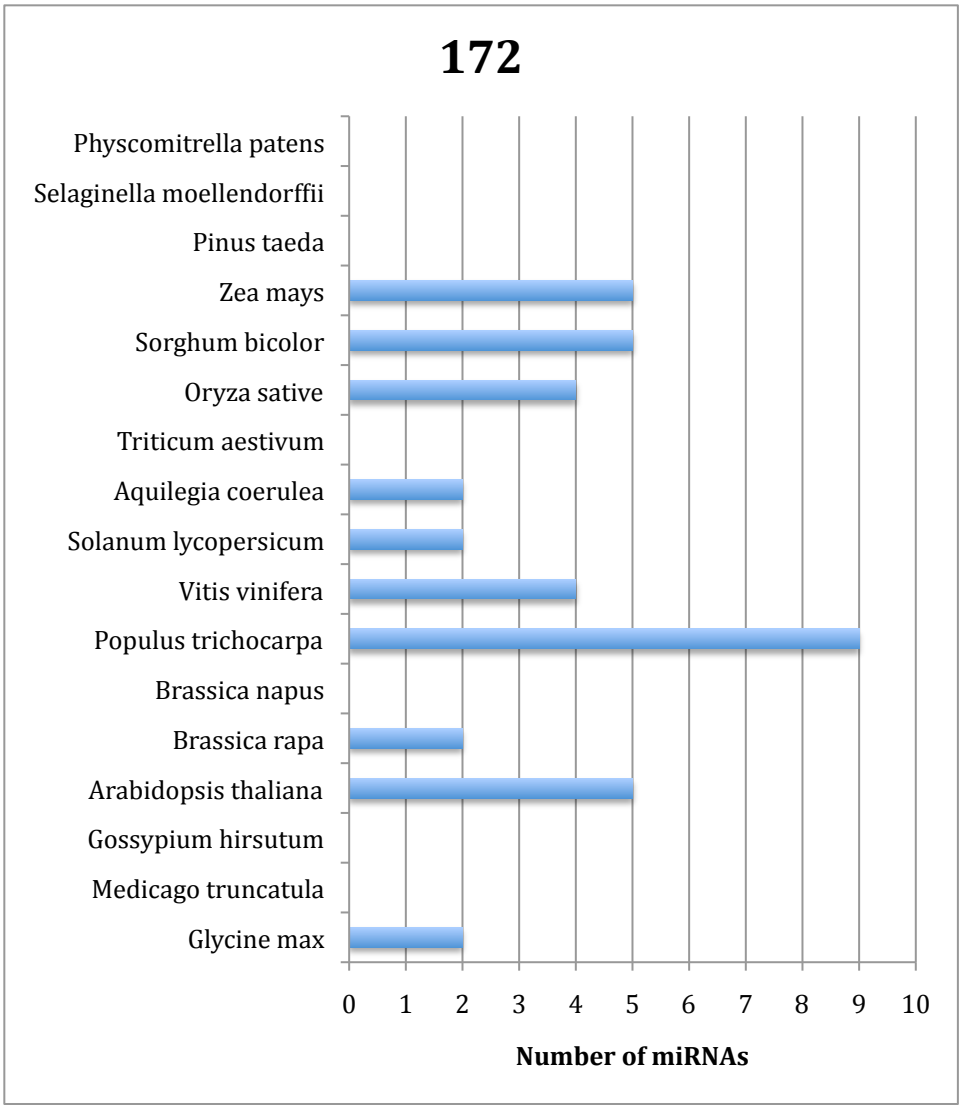

Output of sir_graph (®)
mfold 3.4
2185815653
dG = -30.50 [initially -30.50] 09Mar16-11-54-28

Output of sir_graph (®)
mfold 3.4



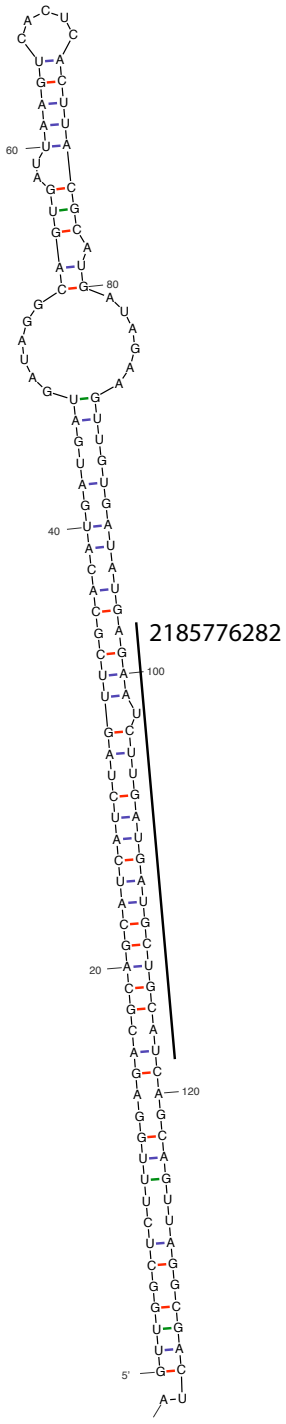
dG = -41.10 [initially -41.10] 09Mar16-11-20-48


20
2185776282mature
2185788415mature
DT764297binding
TC21155binding
TC25860binding
TC26960binding
TC28027binding
TC32032binding
DT732637binding
TC28864binding
TC28449binding
TC22964binding
TC24245binding
TC22146binding
TC29593binding
TC29593binding2
DR954144binding
DT749684binding
DT741192binding
G G A A U C Y U G A U G A U G C U G C A K

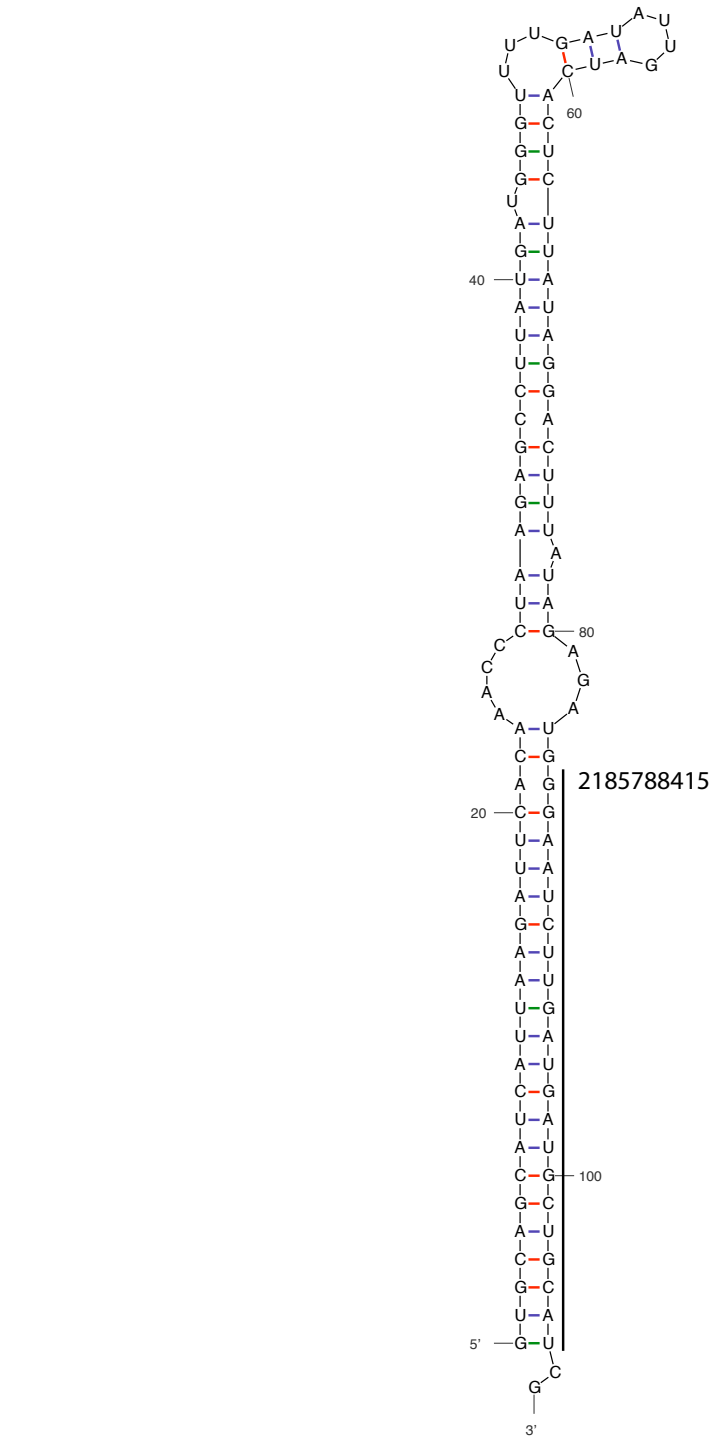

172
Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sative
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max
0 1 2 3 4 5 6 7 8 9 10
Number of miRNAs

Output of sir_graph (©)
mfold_util 4.4

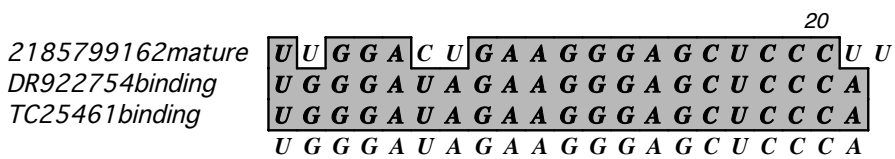
Created Tue May 26 09:40:16 2009

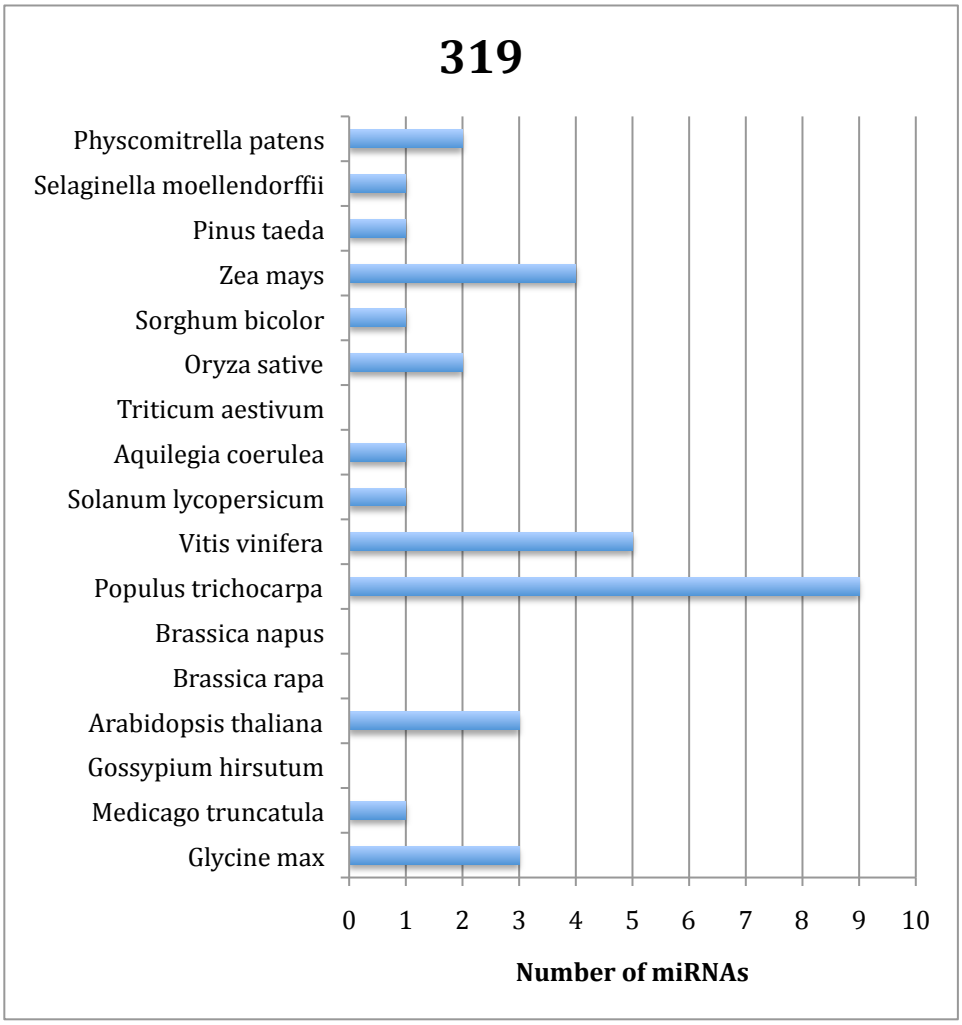


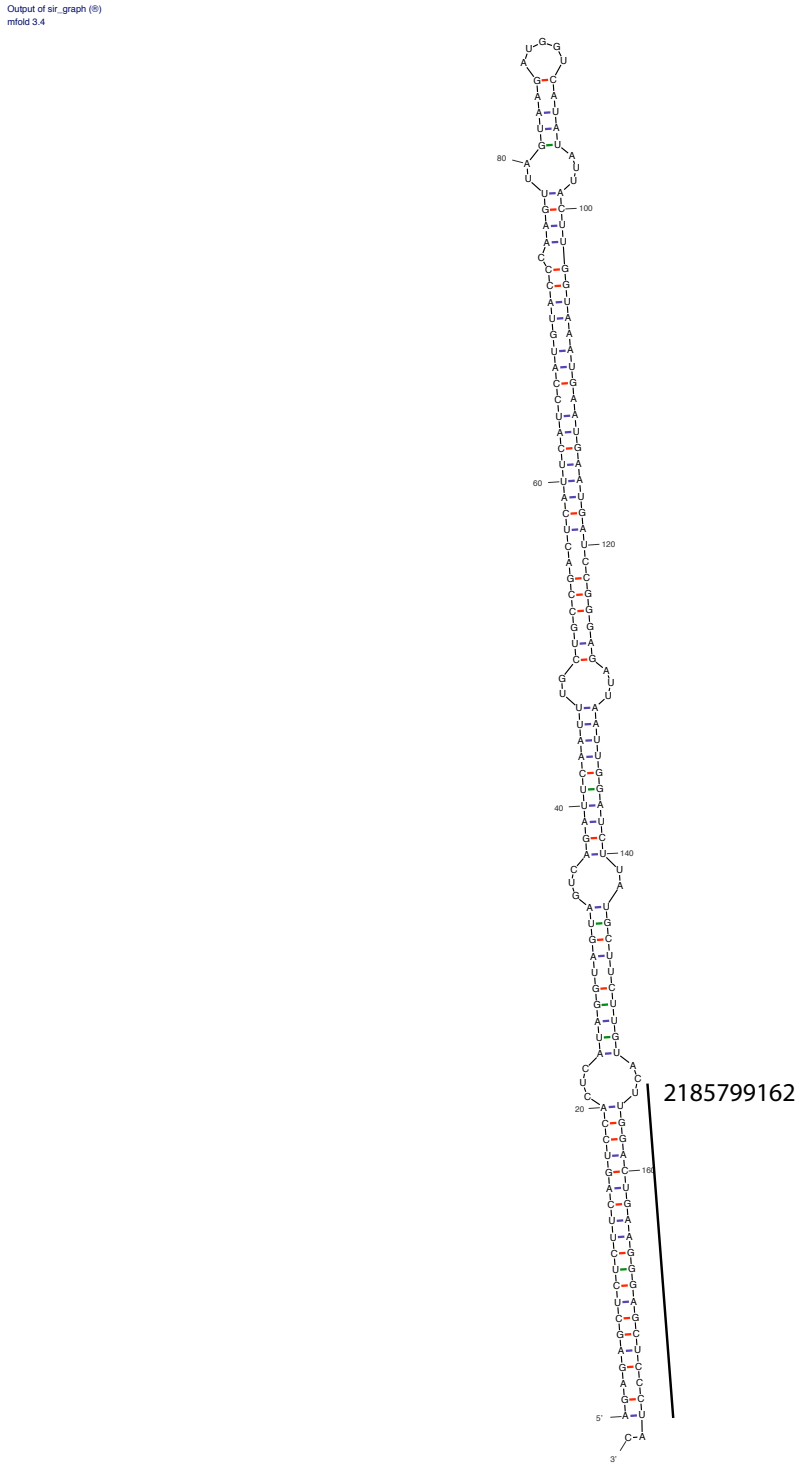
dG = -43.50 [initially -43.50] 09May26-09-40-08



dG = -43.70 [initially -43.70] 09Mar15-21-30-04


20
2185799162mature
DR922754binding
TC25461binding
U U G G A C U G A A G G G A G C U C C C U U
U G G G A U A G A A G G G A G C U C C C A
U G G G A U A G A A G G G A G C U C C C A
U G G G A U A G A A G G G A G C U C C C A


319
Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sative
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max
0 1 2 3 4 5 6 7 8 9 10
Number of miRNAs

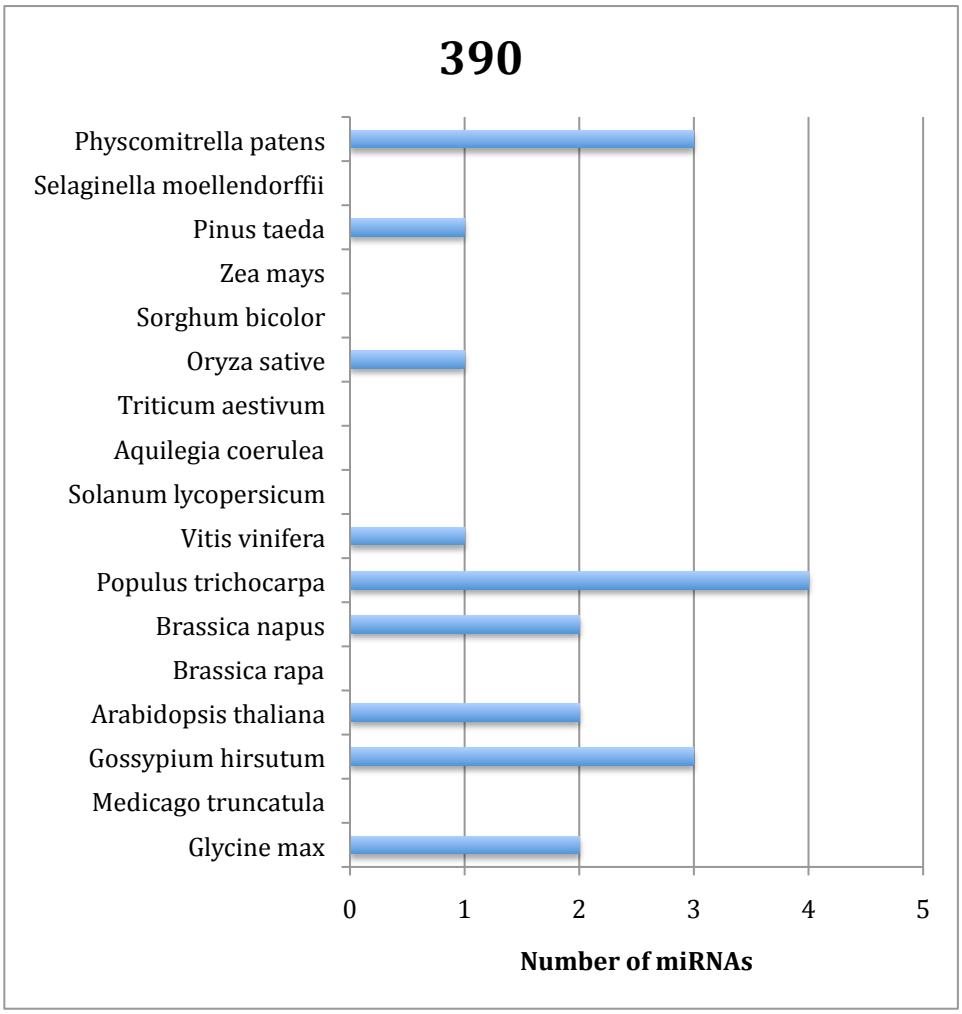


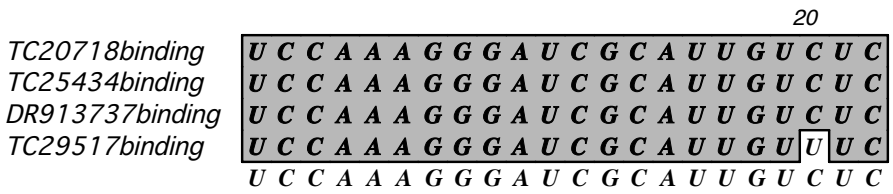
dG = -78.60 [initially -78.60] 09Mar20-05-36-52

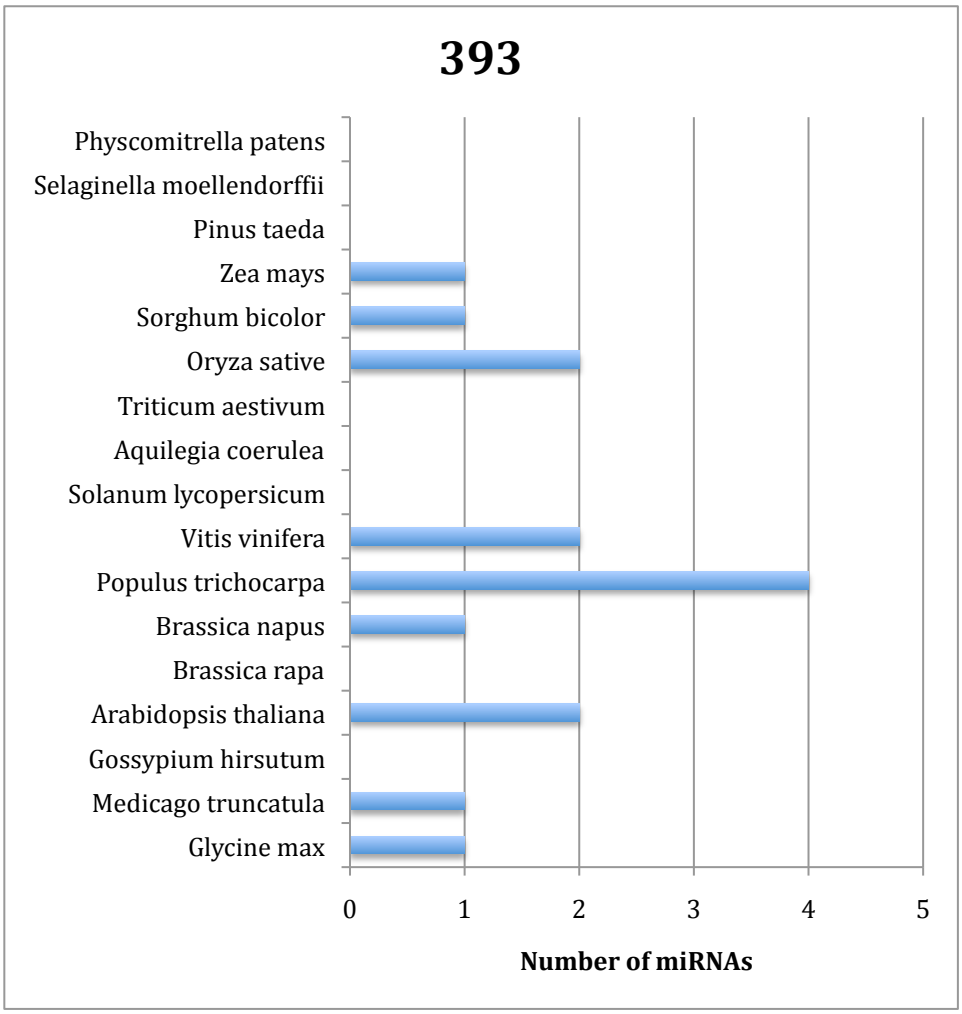
20

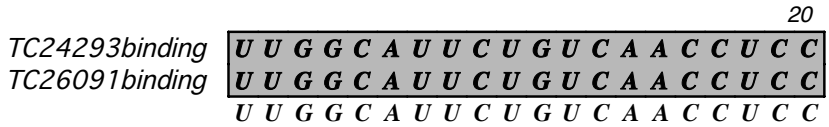
TC30528binding A A G C U C A G G A G G G A G U G C G A U

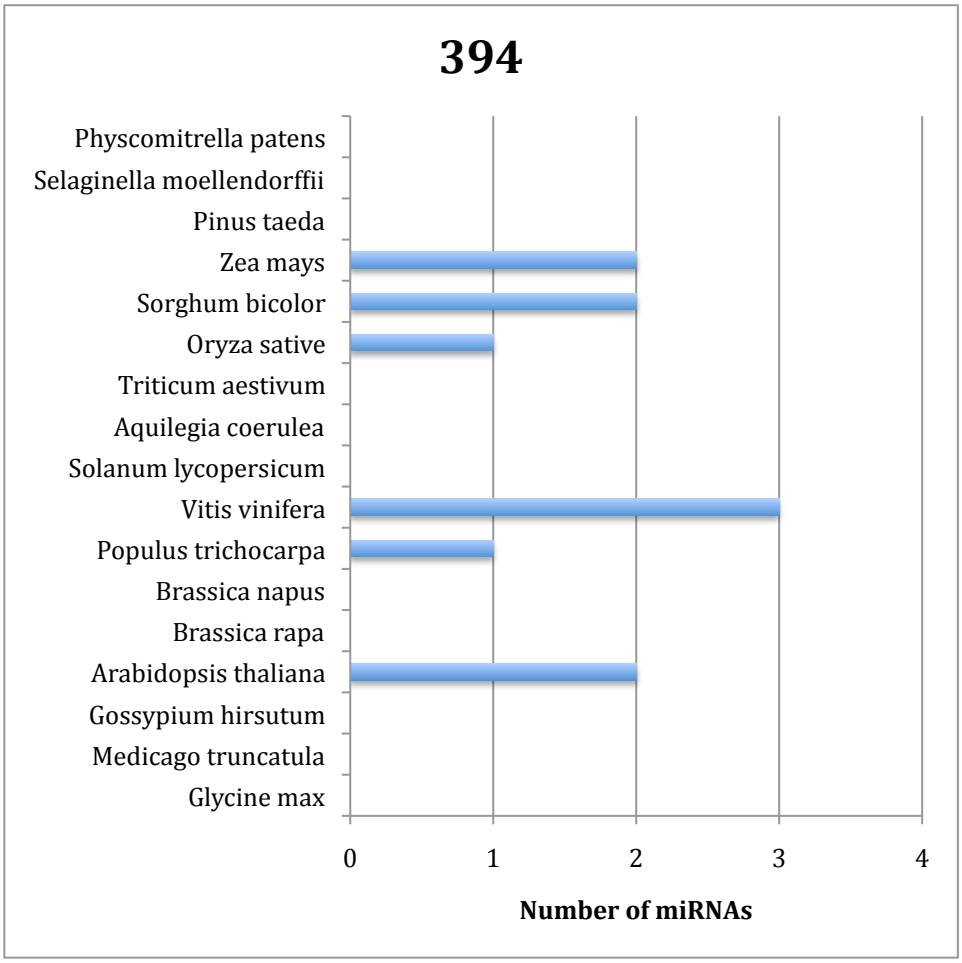
A A G C U C A G G A G G G A G U G C G A U

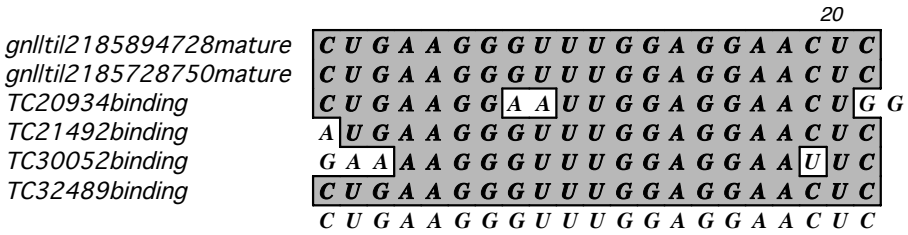


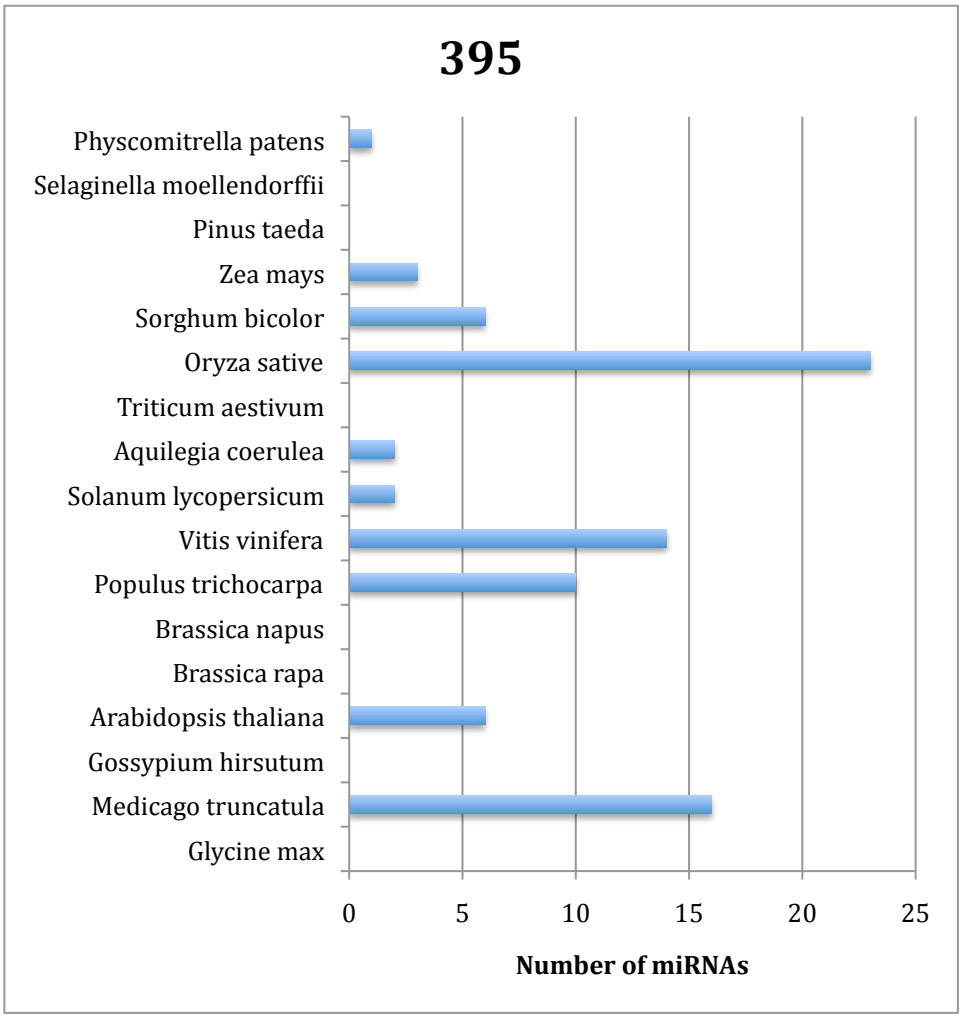

20
TC20718binding
TC25434binding
DR913737binding
TC29517binding
U C C A A A G G G A U C G C A U U G U C U C
U C C A A A G G G A U C G C A U U G U C U C
U C C A A A G G G A U C G C A U U G U C U C
U C C A A A G G G A U C G C A U U G U U C
U C C A A A G G G A U C G C A U U G U C U C

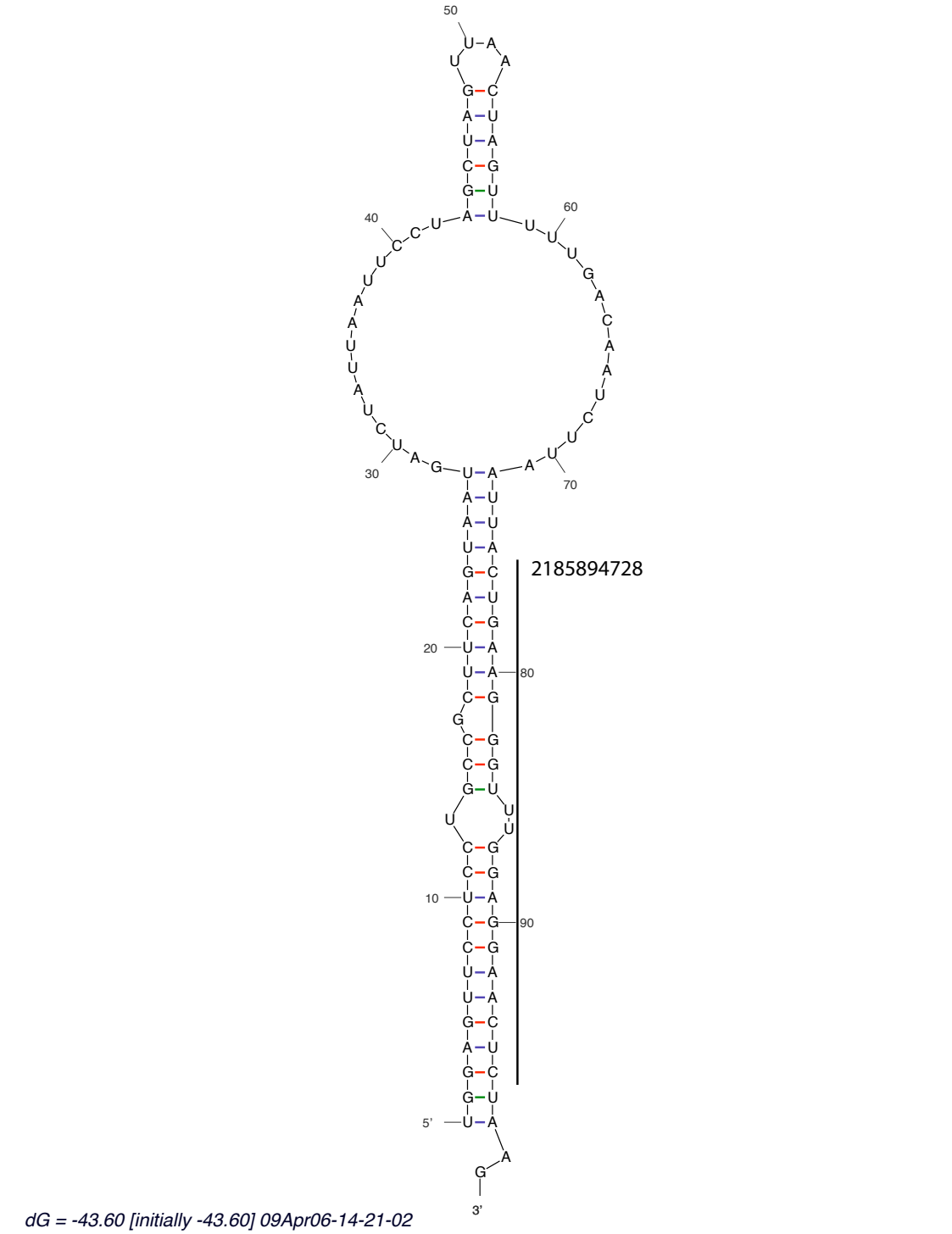

393
Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sative
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max
0
1
2
3
4
5
Number of miRNAs


20
TC24293binding
TC26091binding
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


394
Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sative
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max
0
1
2
3
4
Number of miRNAs

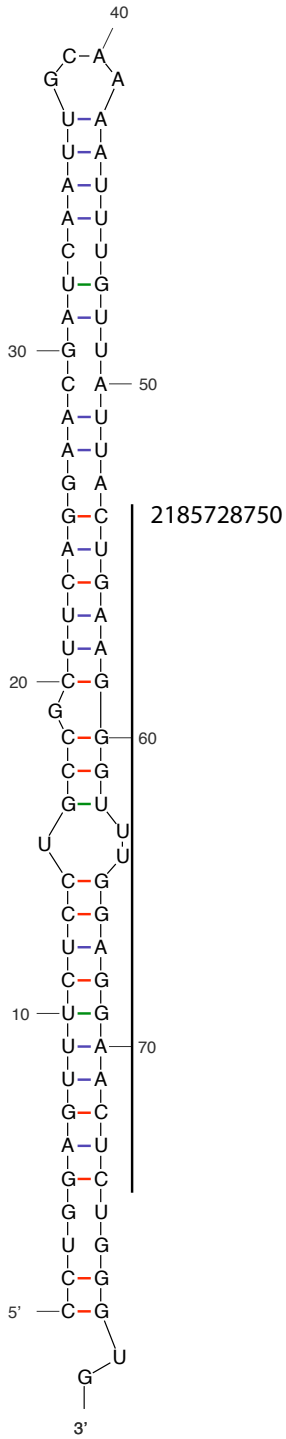

20
gnlltil2185894728mature
gnlltil2185728750mature
TC20934binding
TC21492binding
TC30052binding
TC32489binding
C U G A A G G G U U U G G A G G A A C U C
C U G A A G G G U U U G G A G G A A C U C
C U G A A G G A A U U G G A G G A A C U G G
A U G A A G G G U U U G G A G G A A C U C
G A A A A G G G U U U G G A G G A A U U C
C U G A A G G G U U U G G A G G A A C U C
C U G A A G G G U U U G G A G G A A C U C


395
Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sative
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max
0
5
10
15
20
25
Number of miRNAs

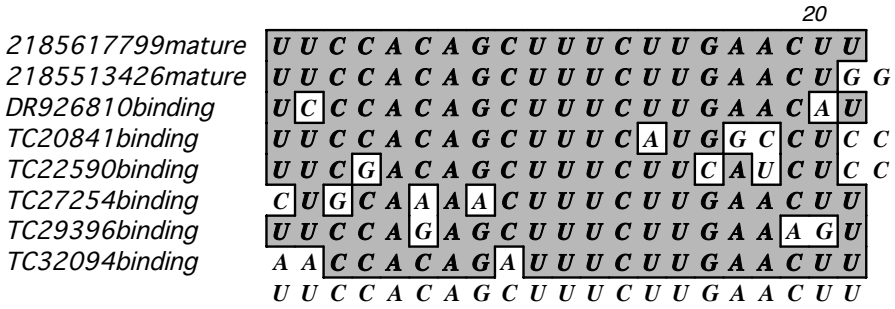

Output of sir_graph (©)
mfold_util_ng 4.1
Created Mon Apr 6 14:21:10 2009
2185894728
dG = -43.60 [initially -43.60] 09Apr06-14-21-02

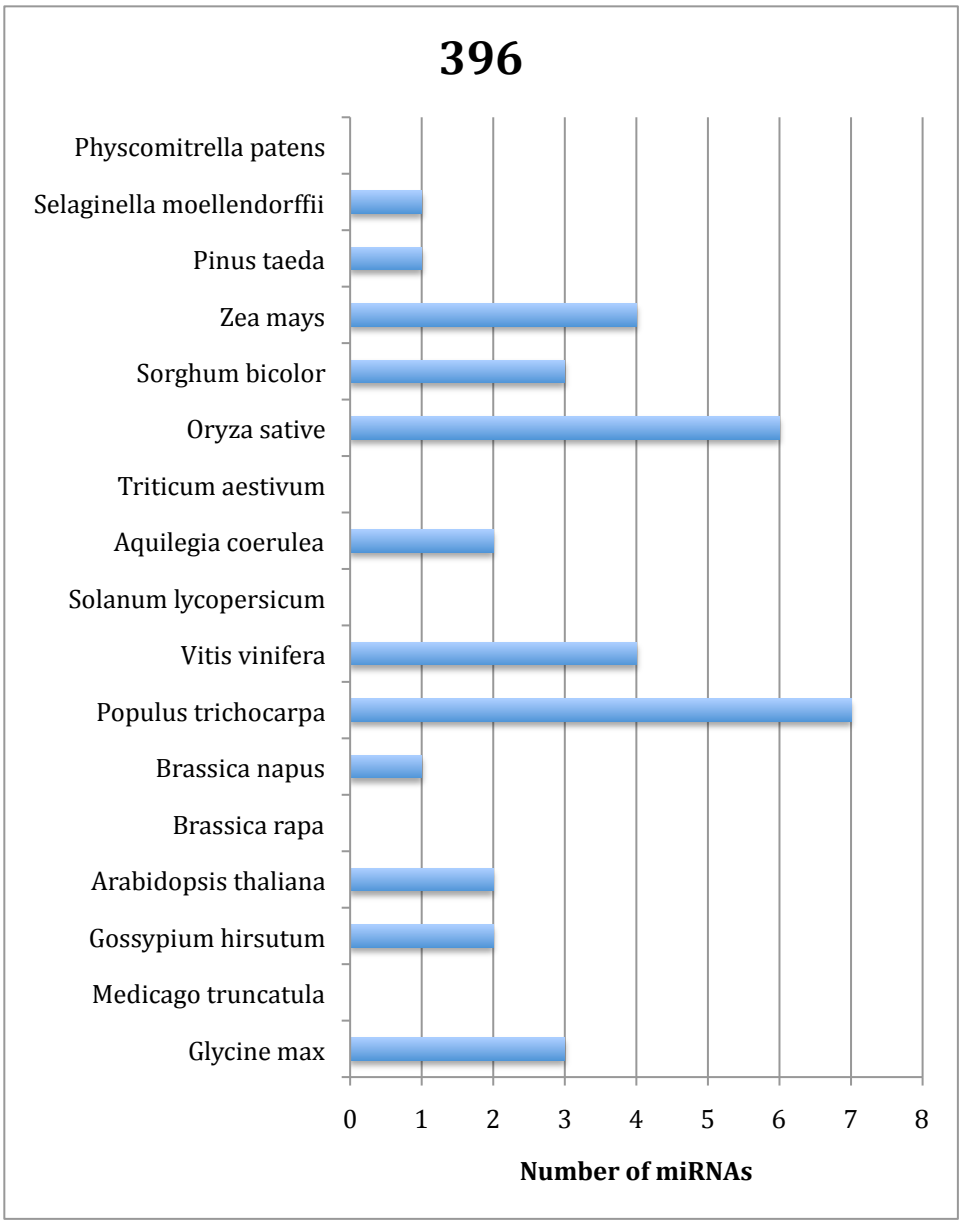
Output of sir_graph (©)
mfold_util_ng 4.1

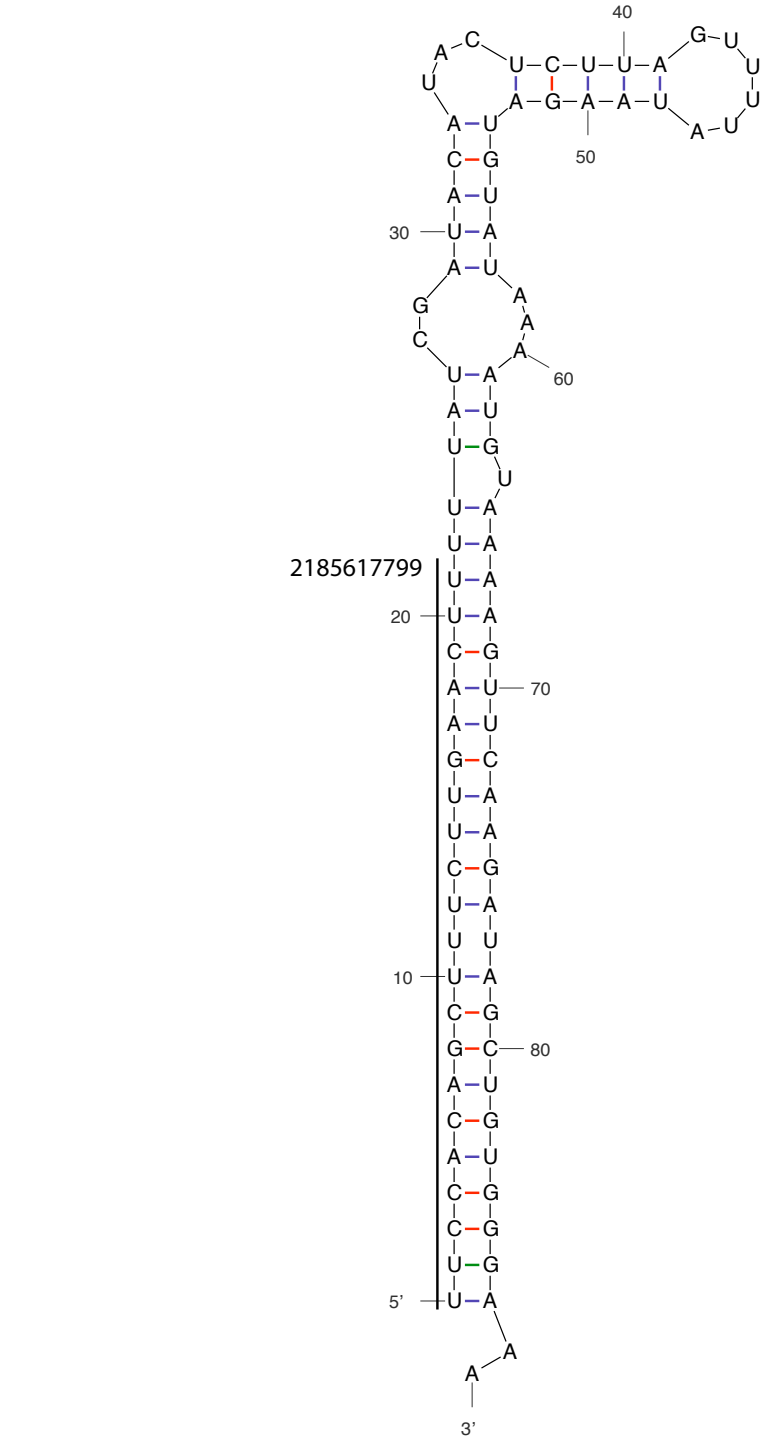
Created Mon Apr 6 14:50:17 2009



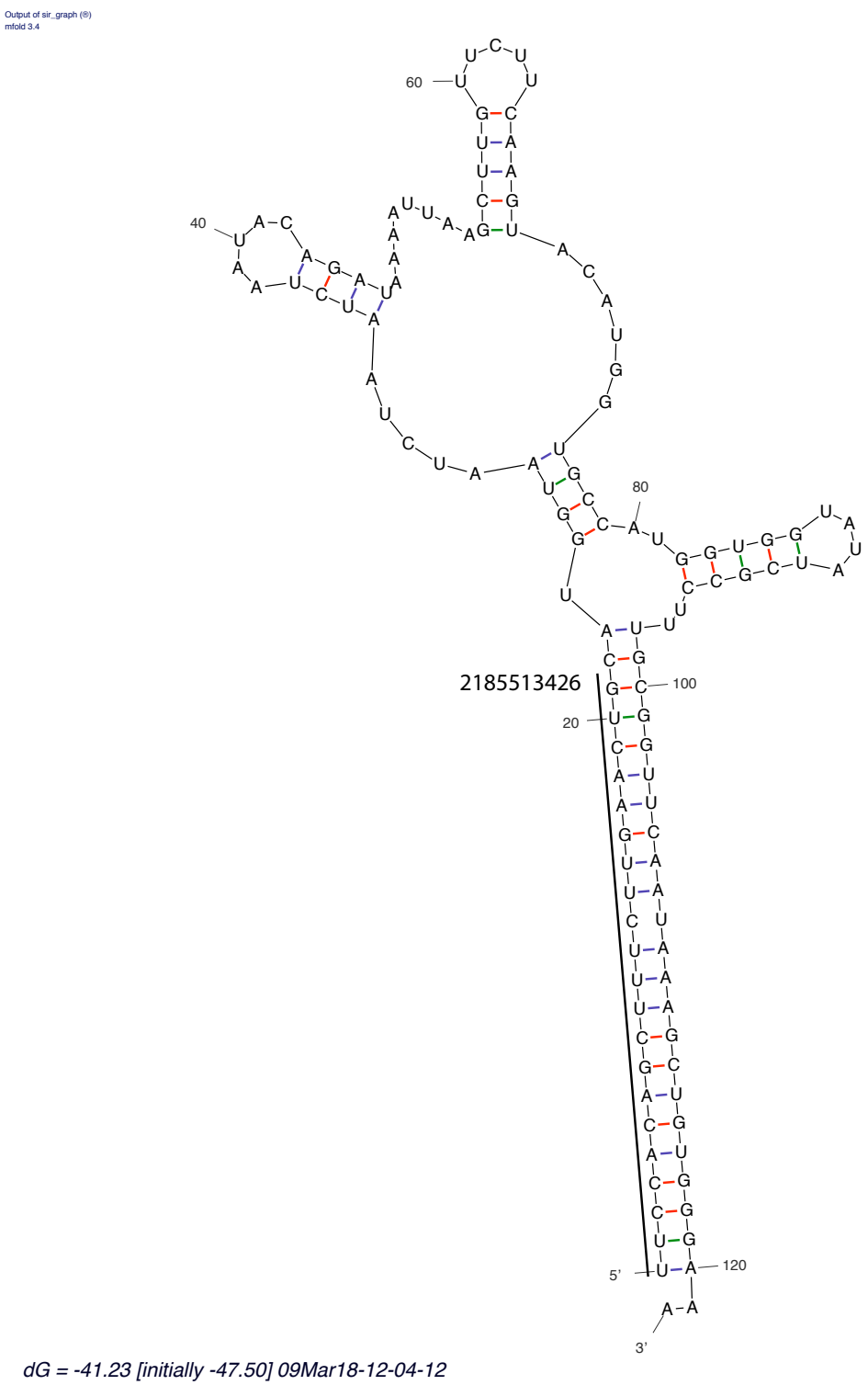
dG = -37.70 [initially -37.70] 09Apr06-14-50-03


20
2185617799mature U U C C A C A G C U U U C U U G A A C U U
2185513426mature U U C C A C A G C U U U C U U G A A C U G G
DR926810binding U C C C A C A G C U U U C U U G A A C A U
TC20841binding U U C C A C A G C U U U C A U G G C C U C C
TC22590binding U U C G A C A G C U U U C U U C A U C U C C
TC27254binding C U G C A A A A C U U U C U U G A A C U U
TC29396binding U U C C A G A G C U U U C U U G A A A G U
TC32094binding A A C C A C A G A U U U C U U G A A C U U
U U C C A C A G C U U U C U U G A A C U U

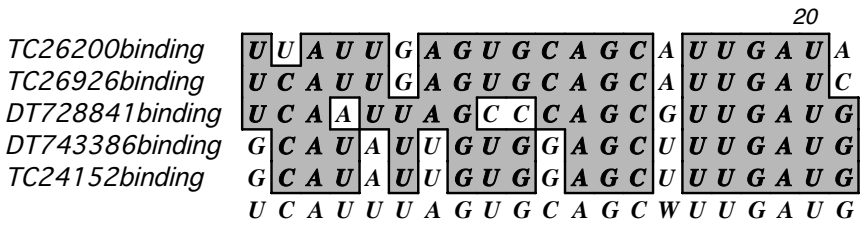

396
Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sative
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max
0
1
2
3
4
5
6
7
8
Number of miRNAs

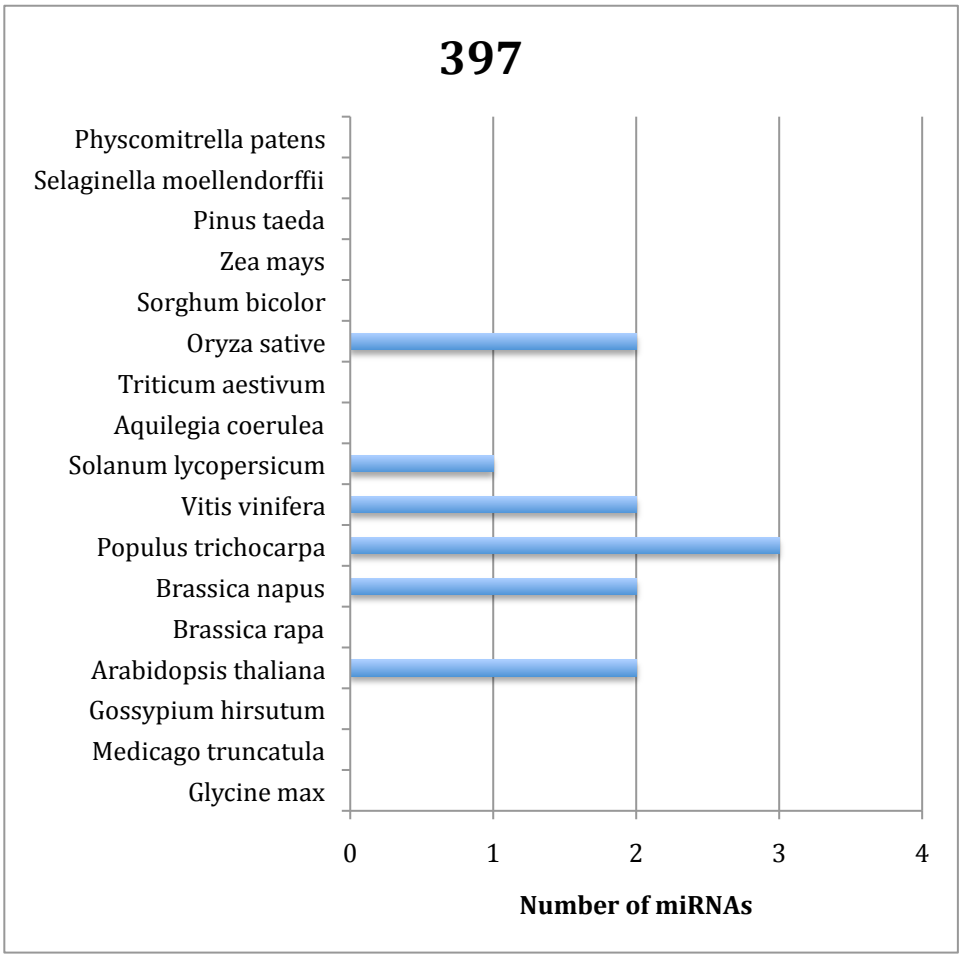


dG = -38.60 [initially -38.60] 09Mar18-12-06-52

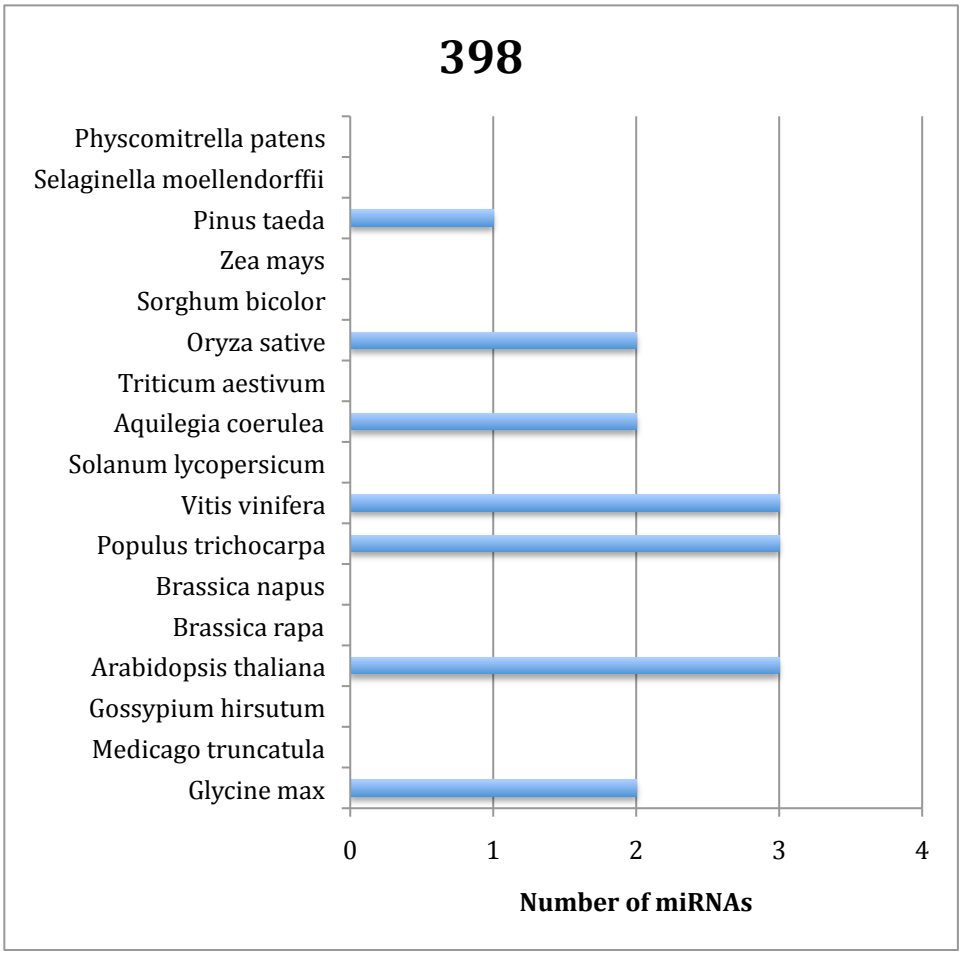


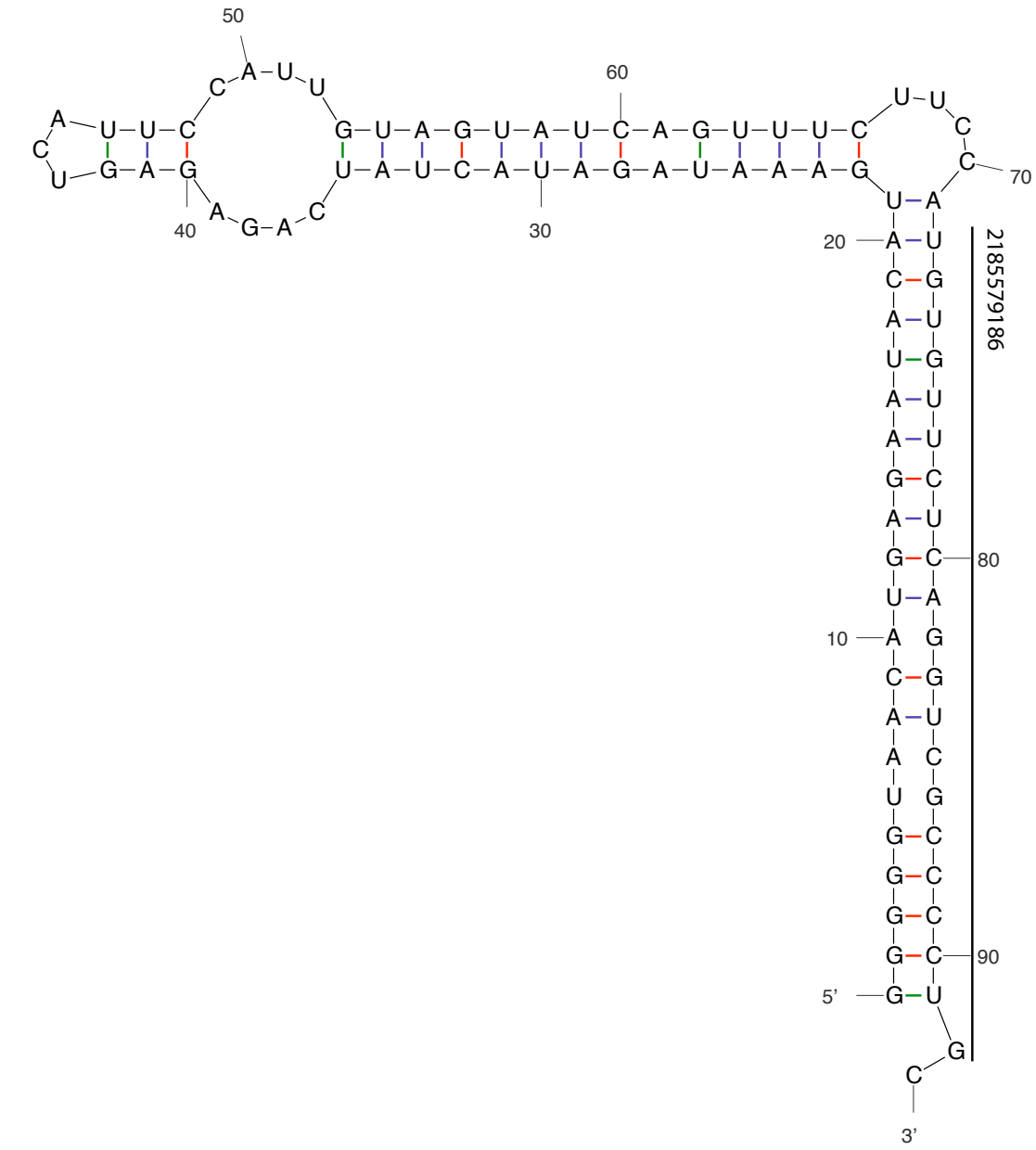
dG = -41.23 [initially -47.50] 09Mar18-12-04-12


20
TC26200binding
TC26926binding
DT728841binding
DT743386binding
TC24152binding
U U A U U G A G U G C A G C A U U G A U A
U C A U U G A G U G C A G C A U U G A U C
U C A A U U A G C C C A G C G U U G A U G
G C A U A U U G U G G A G C U U U G A U G
G C A U A U U G U G G A G C U U U G A U G
U C A U U U A G U G C A G C W U U G A U G

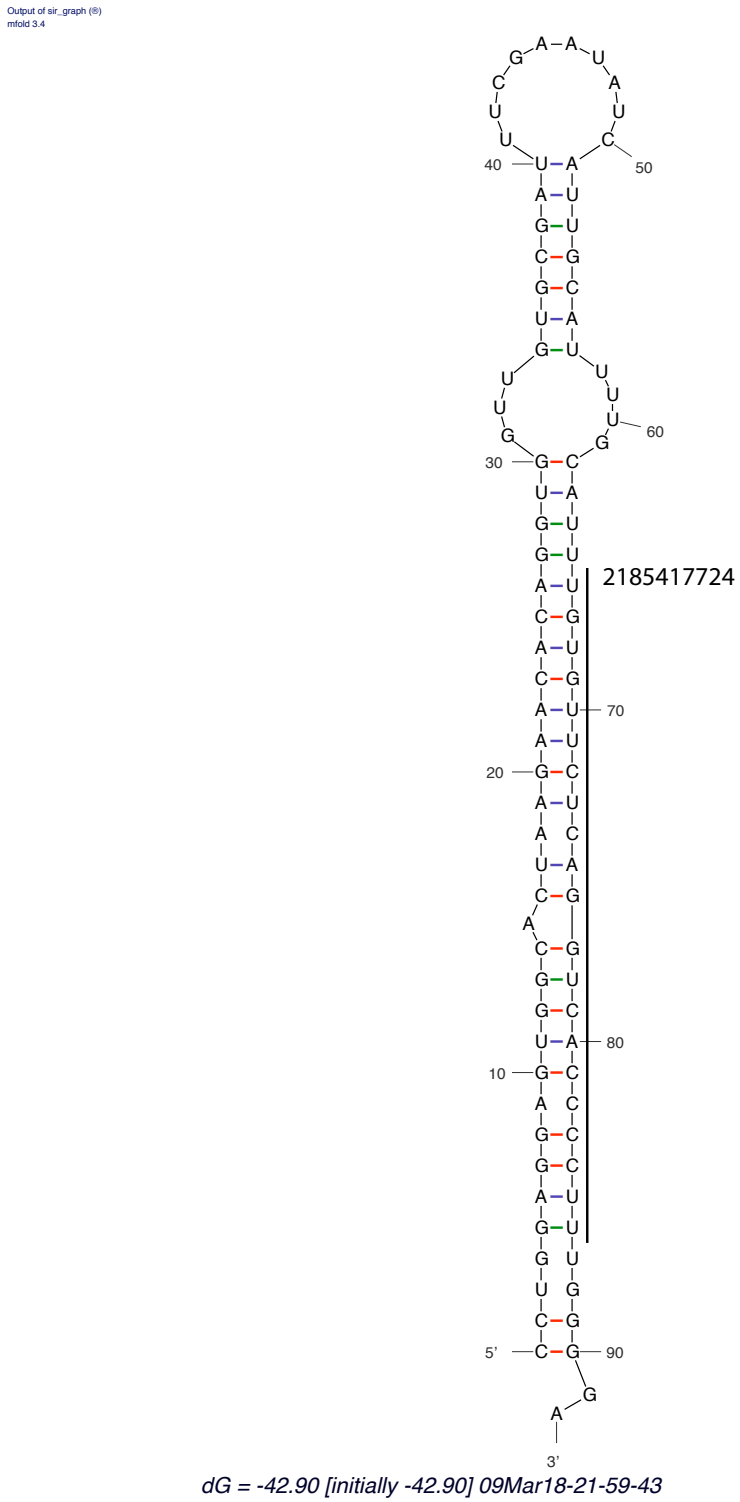

397
Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sative
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max
0
1
2
3
4
Number of miRNAs

```
                                                                                20
2185417724mature  U G U G U U C U C A G G U C A C C C C U U U
2185579186mature  U G U G U U C U C A G G U C G C C C C U G G
                  U G U G U U C U C A G G U C R C C C C U K
```

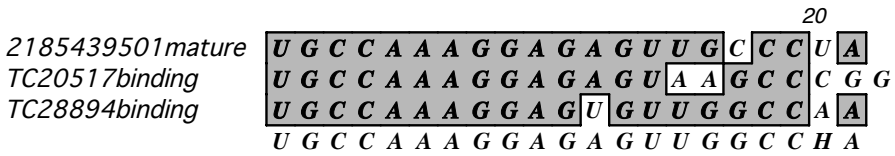


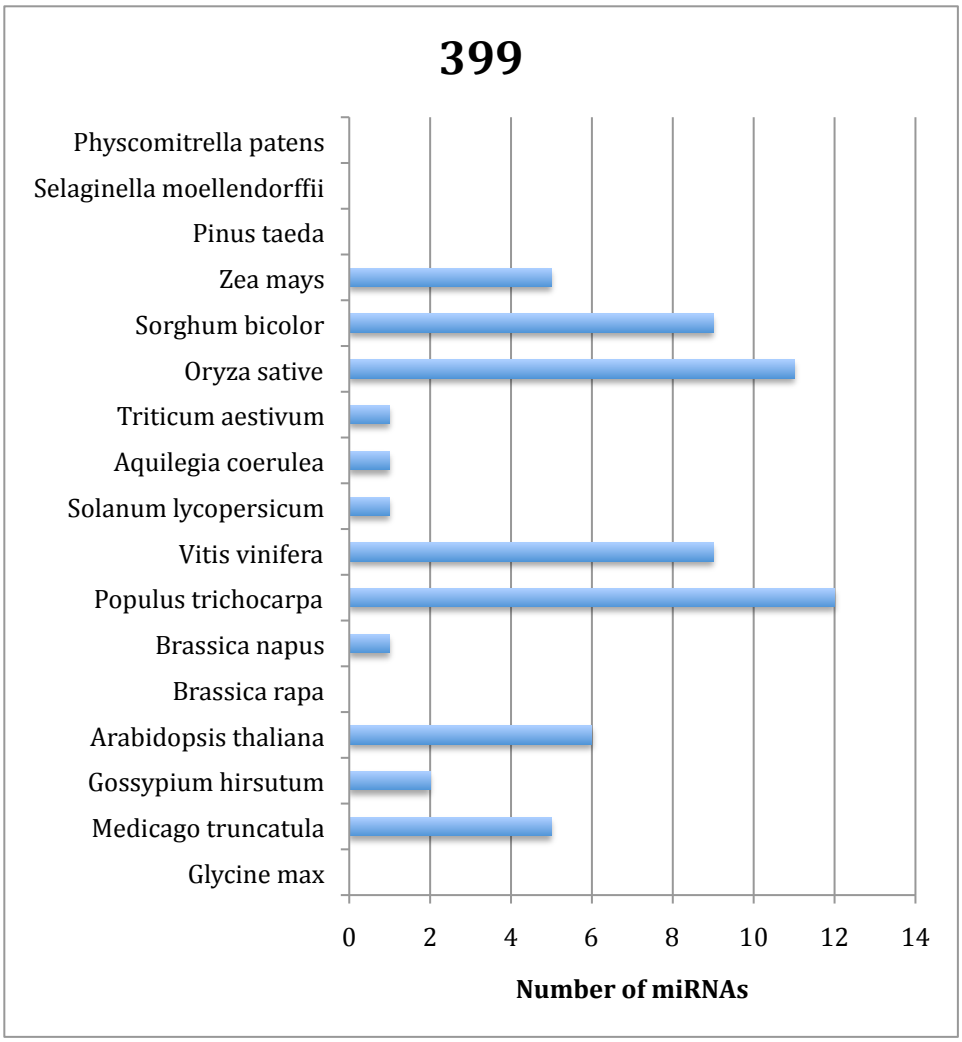


dG = -38.50 [initially -38.50] 09Mar18-21-55-00



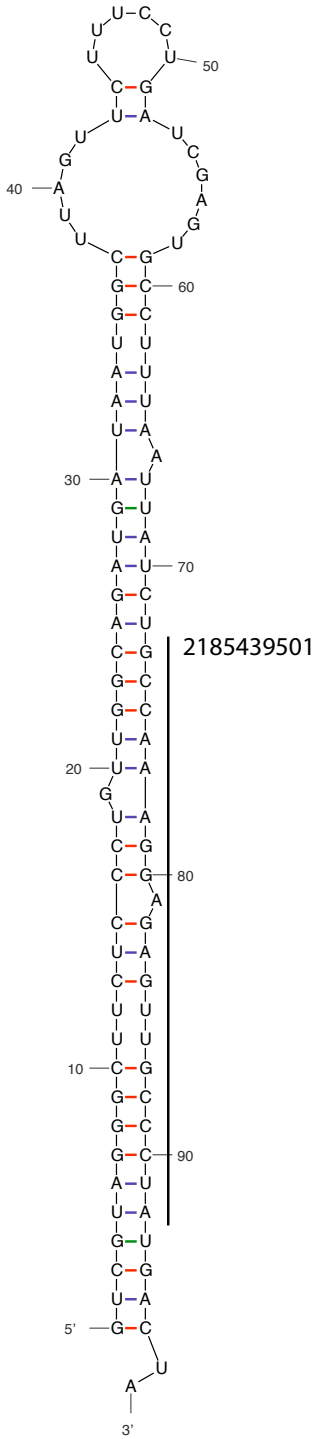
dG = -42.90 [initially -42.90] 09Mar18-21-59-43


20
2185439501mature
TC20517binding
TC28894binding
U G C C A A A G G A G A G U U G C C C U A
U G C C A A A G G A G A G U A A G C C C G G
U G C C A A A G G A G U G U U G G C C A A
U G C C A A A G G A G A G U U G G C C H A

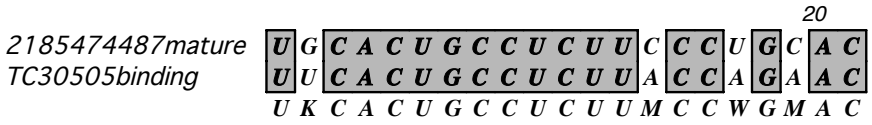

399
Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sative
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max
0
2
4
6
8
10
12
14
Number of miRNAs

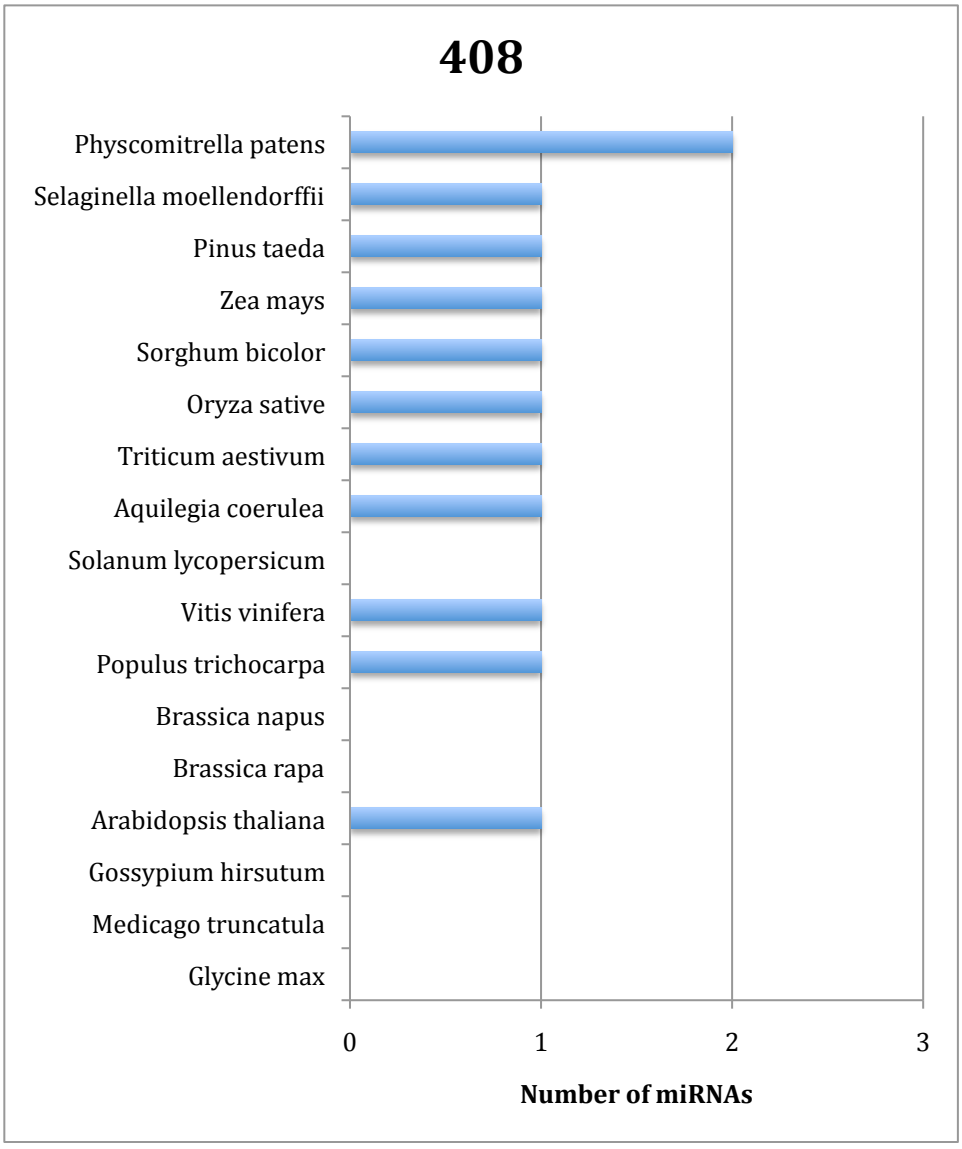
Output of sir_graph (©)
mfold_util 4.4

Created Fri Apr 10 10:42:17 2009



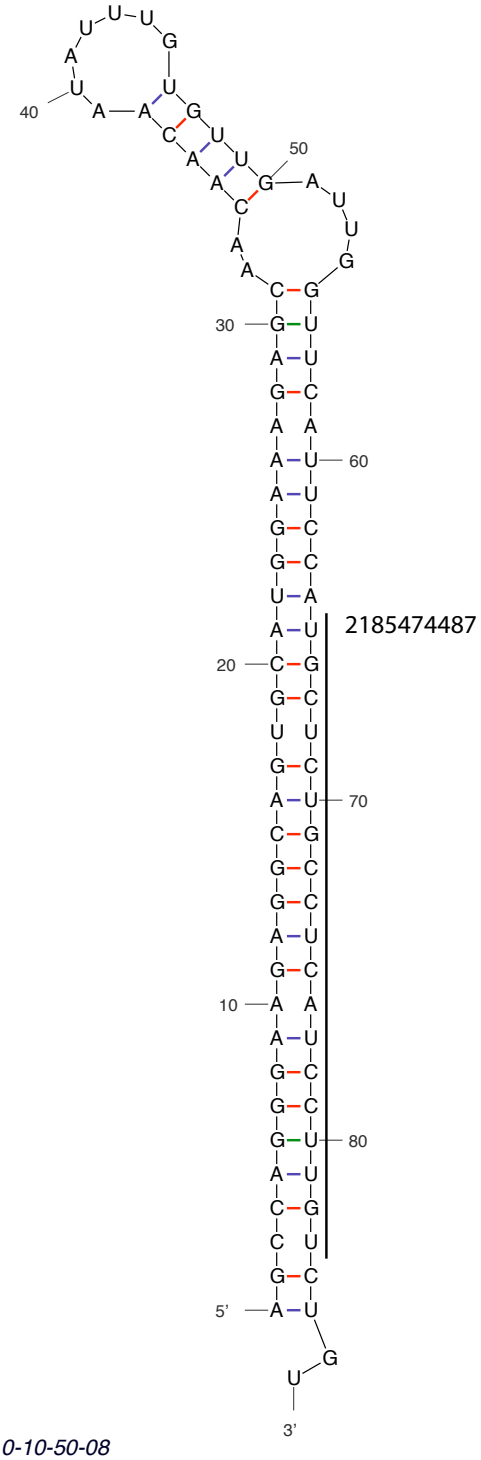
dG = -49.30 [initially -49.30] 09Apr10-10-41-59


20
2185474487mature
TC30505binding
U G C A C U G C C U C U U C C C U G C A C
U U C A C U G C C U C U U A C C A G A A C
U K C A C U G C C U C U U M C C W G M A C

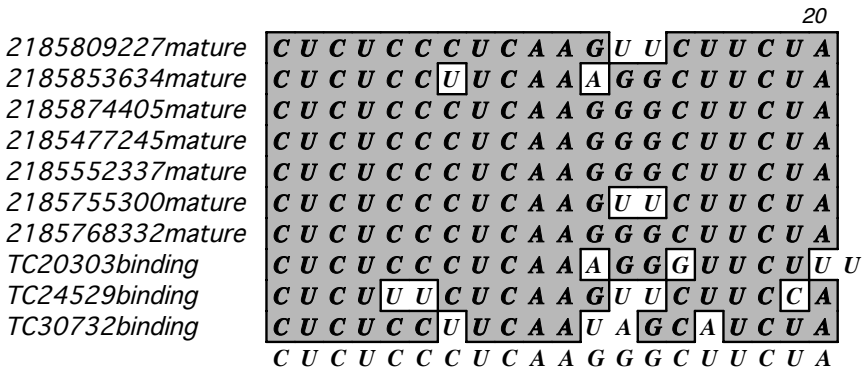

408
Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sative
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max
0
1
2
3
Number of miRNAs

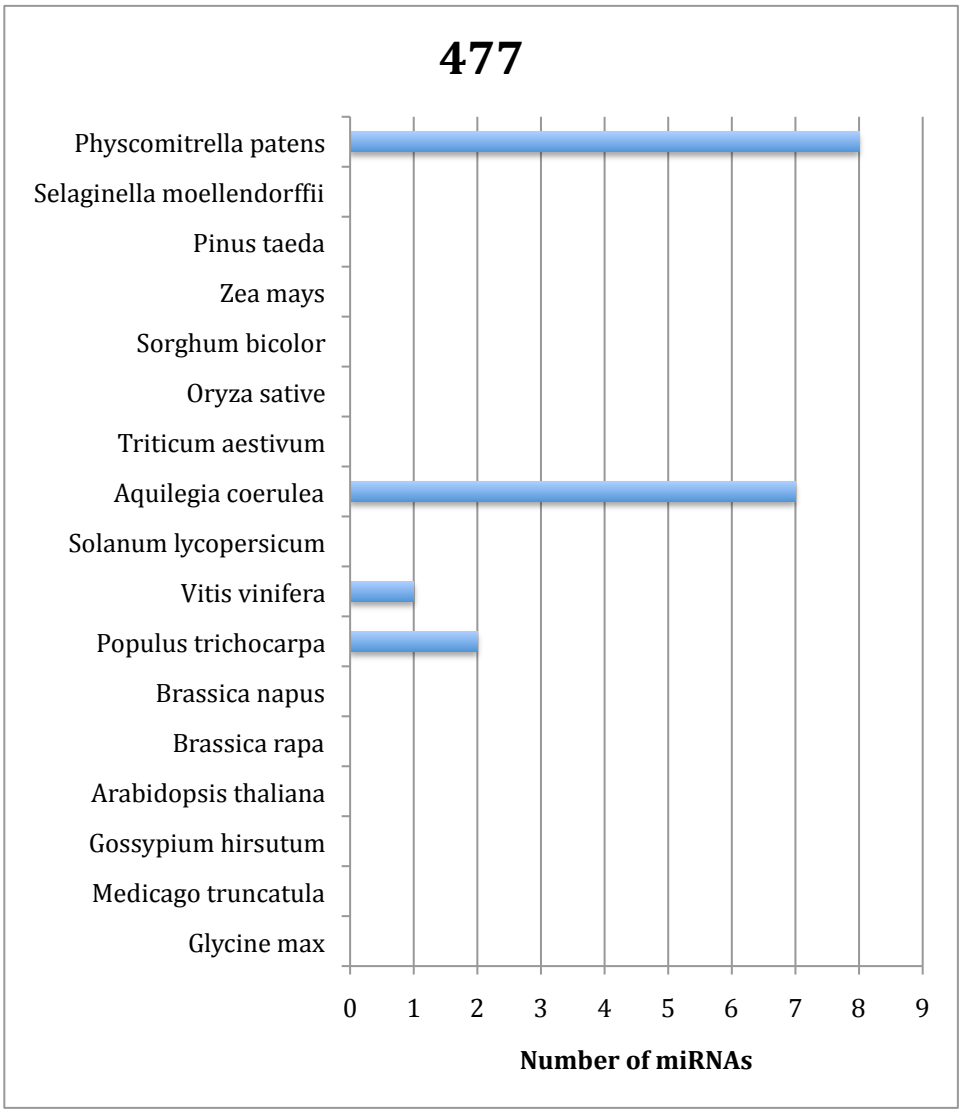
Output of sir_graph (©)
mfold_util 4.4

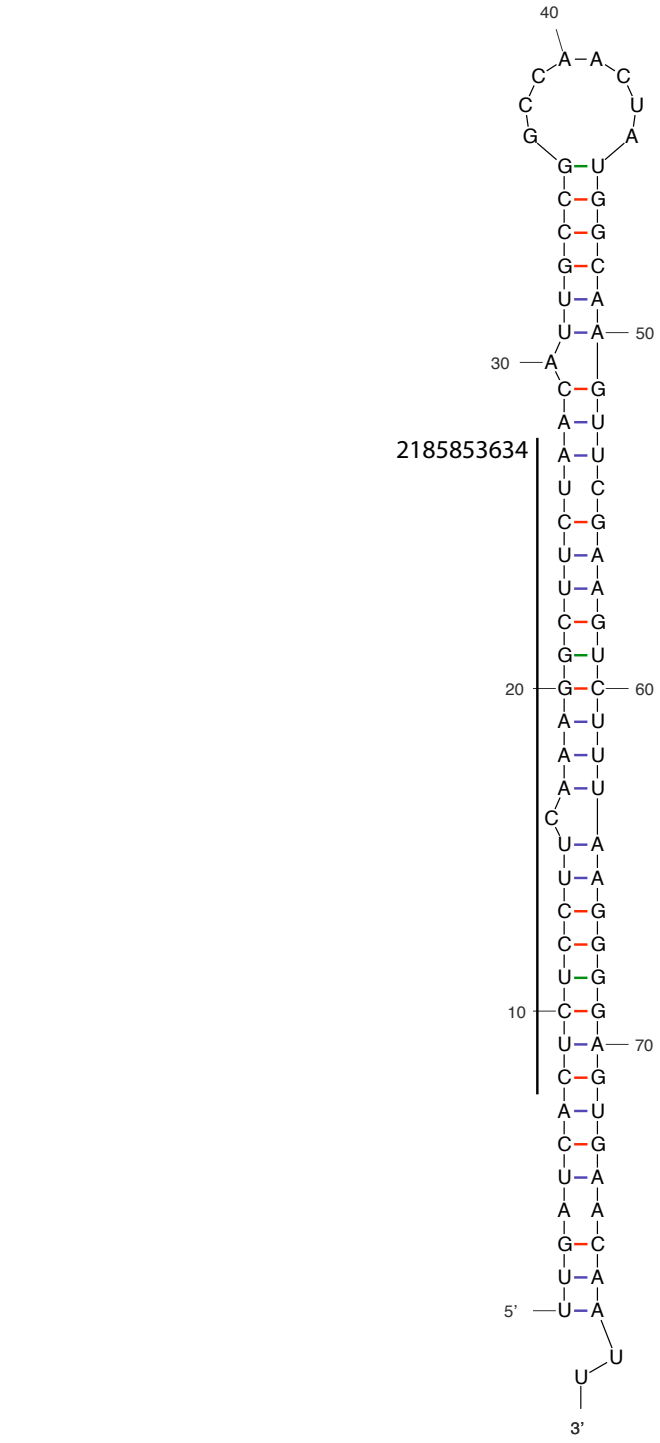
Created Fri Apr 10 10:50:27 2009



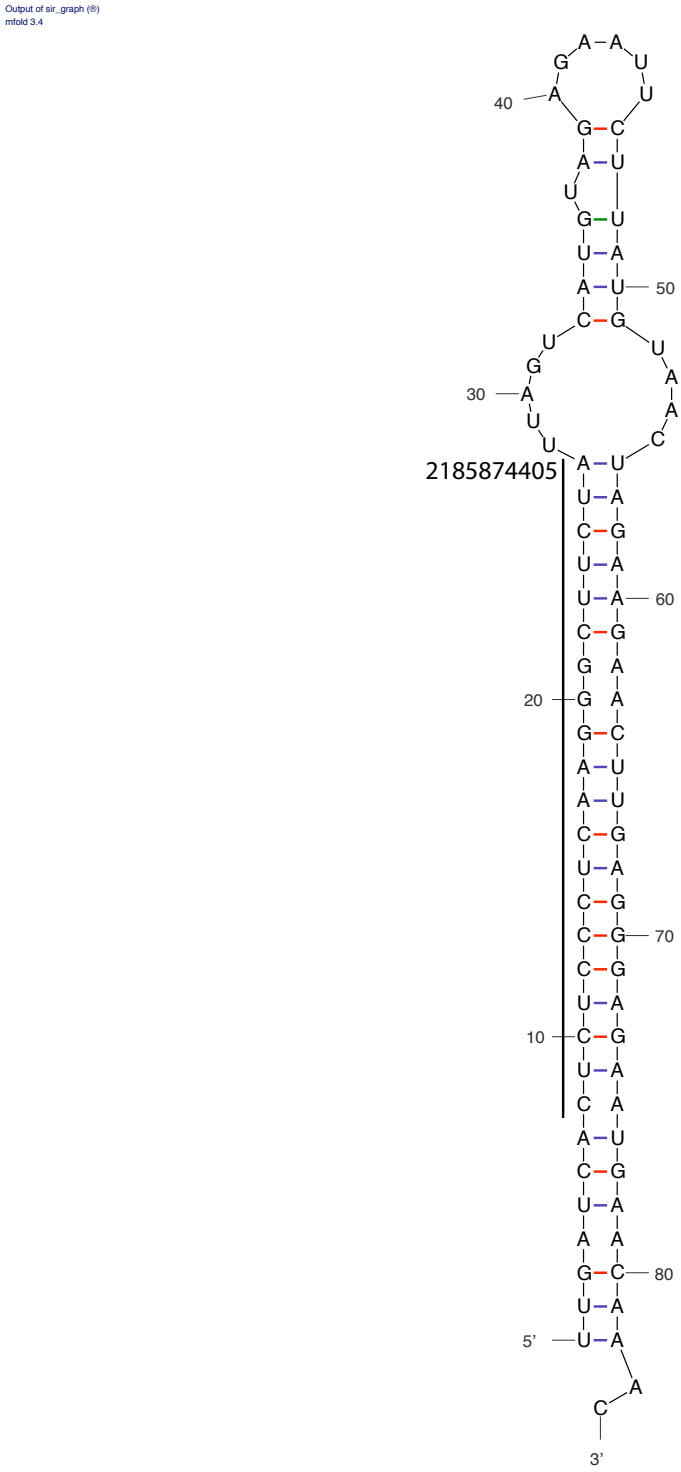
dG = -46.50 [initially -46.50] 09Apr10-10-50-08


20
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 U C A A A G G C U U C U A
2185874405mature C U C U C C C U C A A G G G C U U C U A
2185477245mature C U C U C C C U C A A G G G C U U C U A
2185552337mature C U C U C C C U C A A G G G C U U C U A
2185755300mature C U C U C C C U C A A G U U C U U C U A
2185768332mature C U C U C C C U C A A G G G C U U C U A
TC20303binding C U C U C C C U C A A A G G G U U C U U U
TC24529binding C U C U U U C U C A A G U U C U U C C A
TC30732binding C U C U C C U U C A A U A G C A U C U A
C U C U C C C U C A A G G G C U U C U A

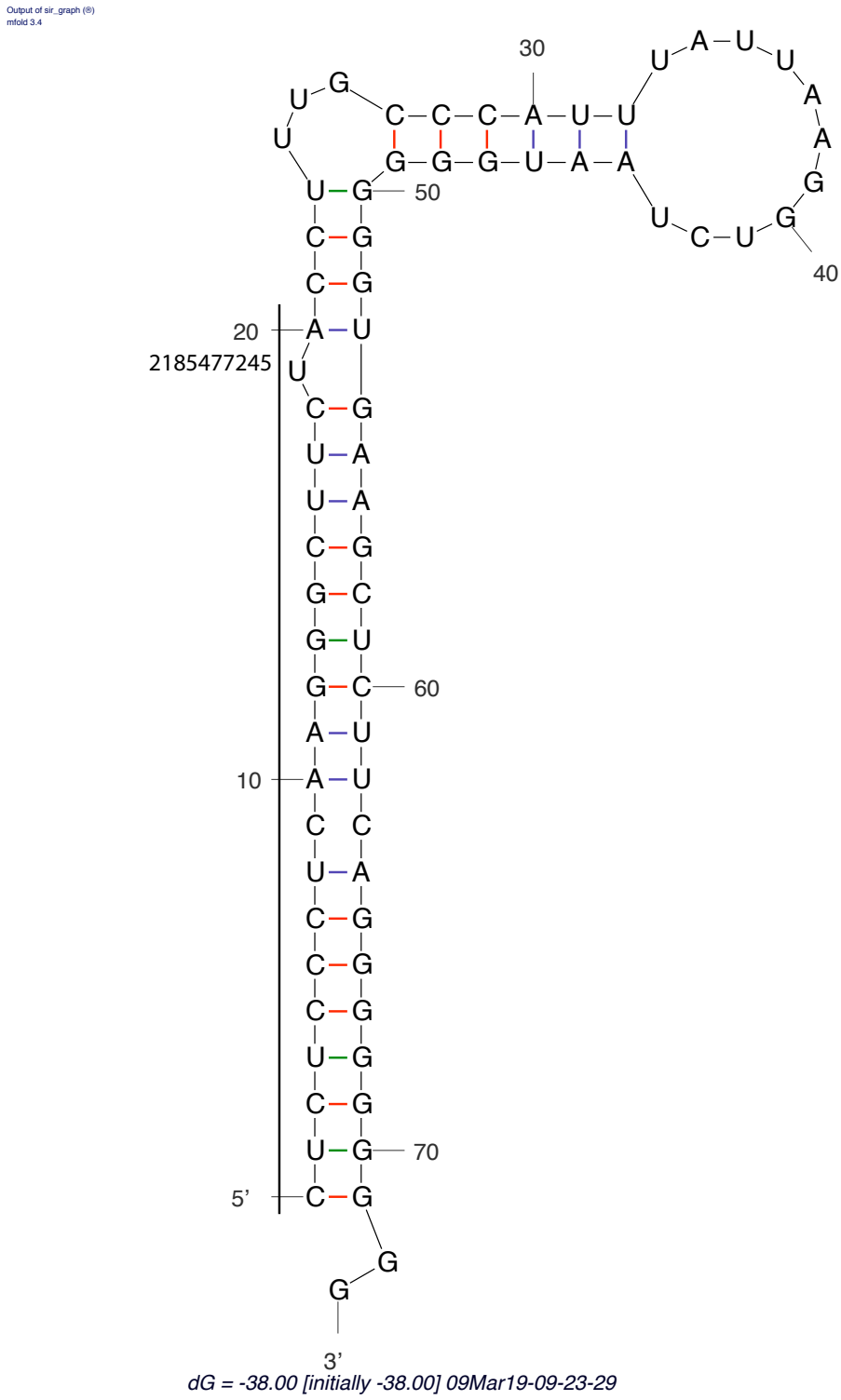

477
Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sative
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max
0 1 2 3 4 5 6 7 8 9
Number of miRNAs



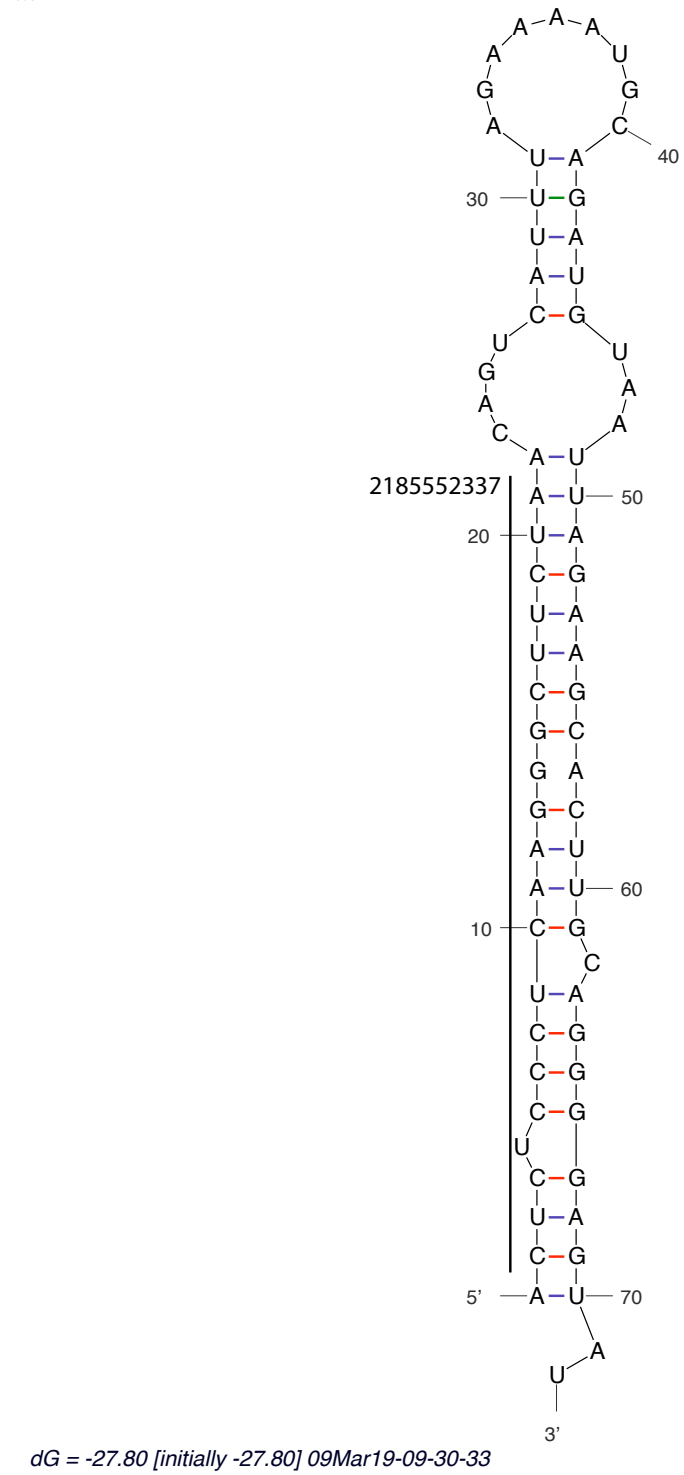
dG = -40.60 [initially -40.60] 09Mar19-10-24-15



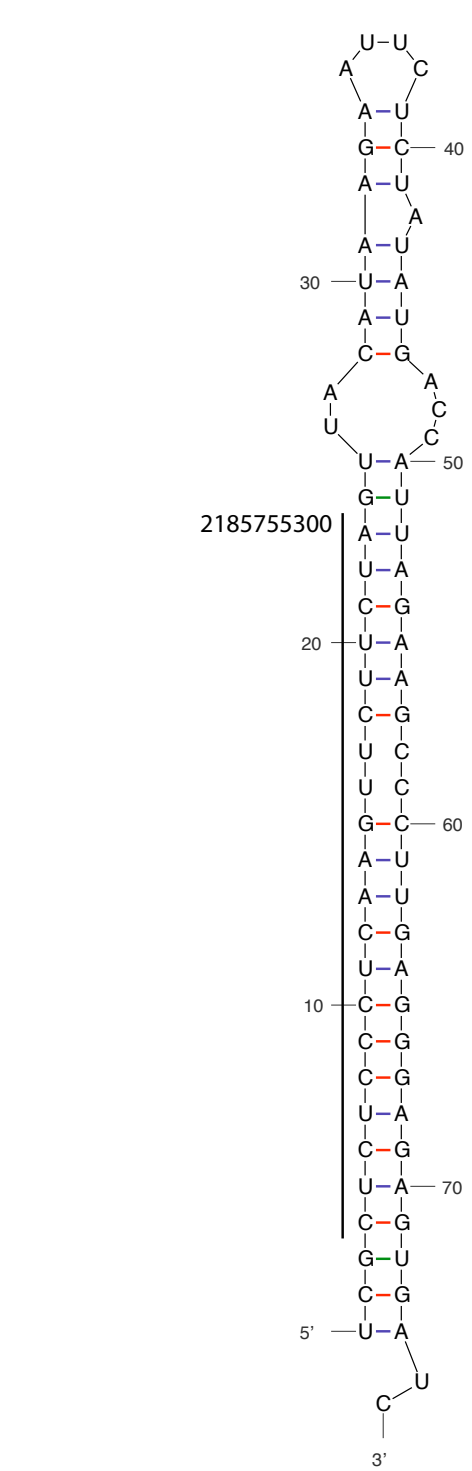
dG = -36.40 [initially -36.40] 09Mar19-10-29-38



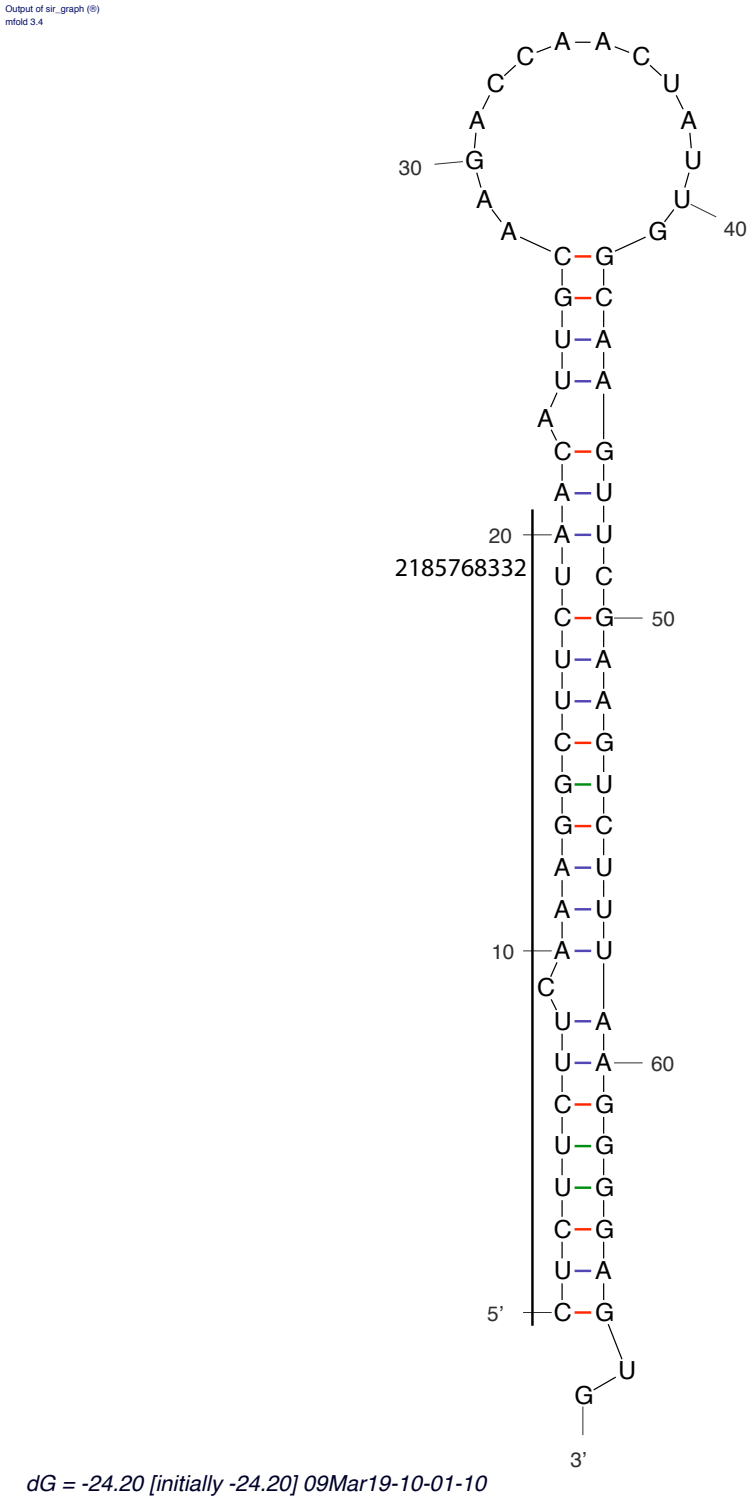
dG = -38.00 [initially -38.00] 09Mar19-09-23-29



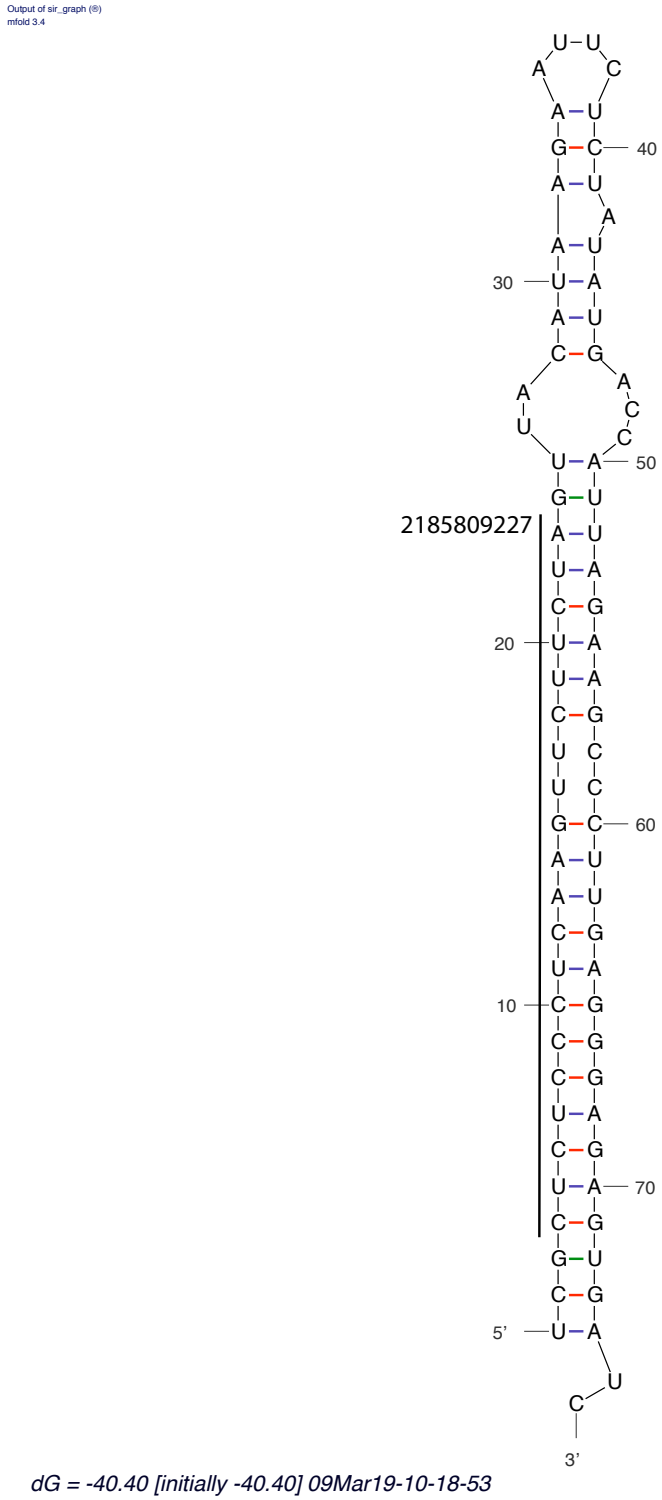
dG = -27.80 [initially -27.80] 09Mar19-09-30-33



dG = -40.40 [initially -40.40] 09Mar19-09-53-09

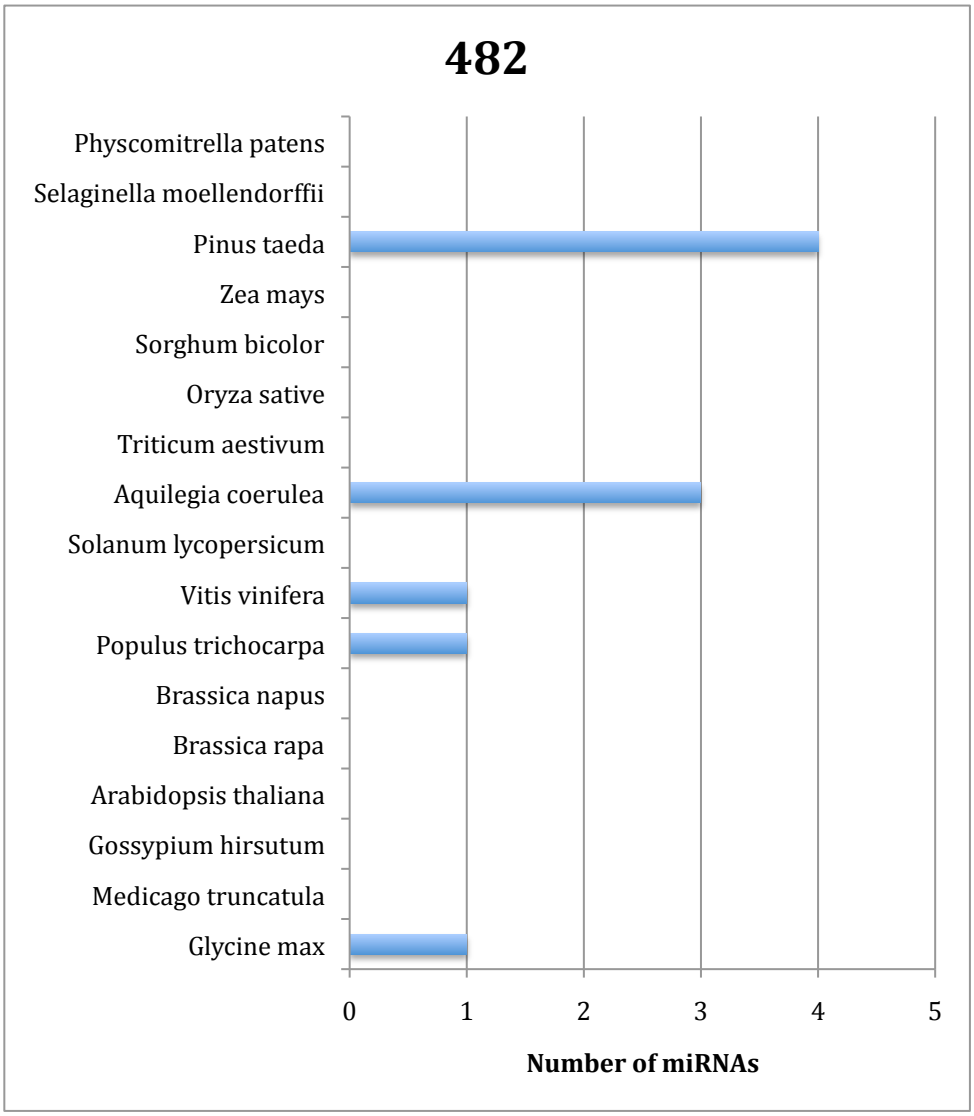


dG = -24.20 [initially -24.20] 09Mar19-10-01-10



dG = -40.40 [initially -40.40] 09Mar19-10-18-53

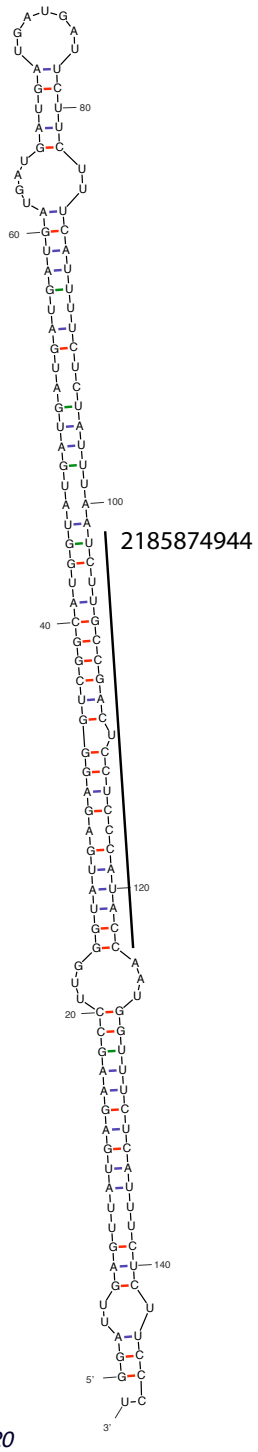
```
                                                                                20
2185586659mature  U C U U G C C G A C U C C U C C C A U A C C
2185874944mature  U C U U G C C G A C U C C U C C C A U A C C
2185892817mature  U C U U G C C G A C U C C U C C C A U A C C
                  U C U U G C C G A C U C C U C C C A U A C C
```



microRNA482

Output of sir_graph (©)
mfold_util_ng 4.1

Created Tue Apr 7 09:04:39 2009



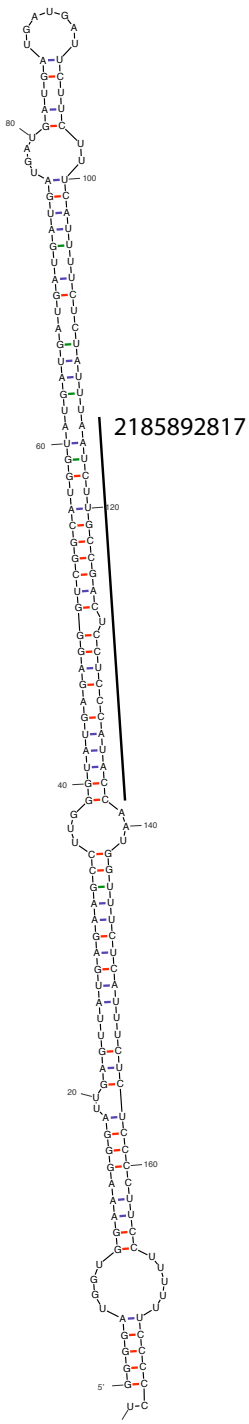
dG = -64.20 [initially -64.20] 09Apr07-09-04-20

microRNA482

microRNA482

Output of sir_graph (©)
mfold_util_ng 4.1

Created Tue Apr 7 08:39:40 2009

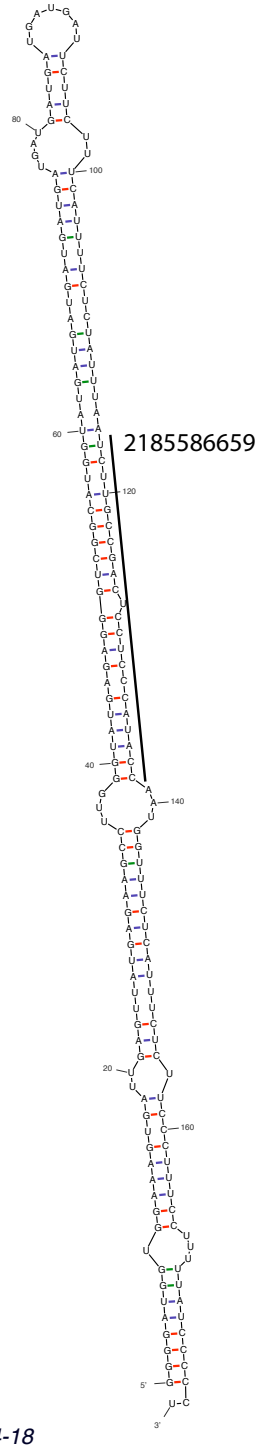


dG = -81.80 [initially -81.80] 09Apr07-08-39-28

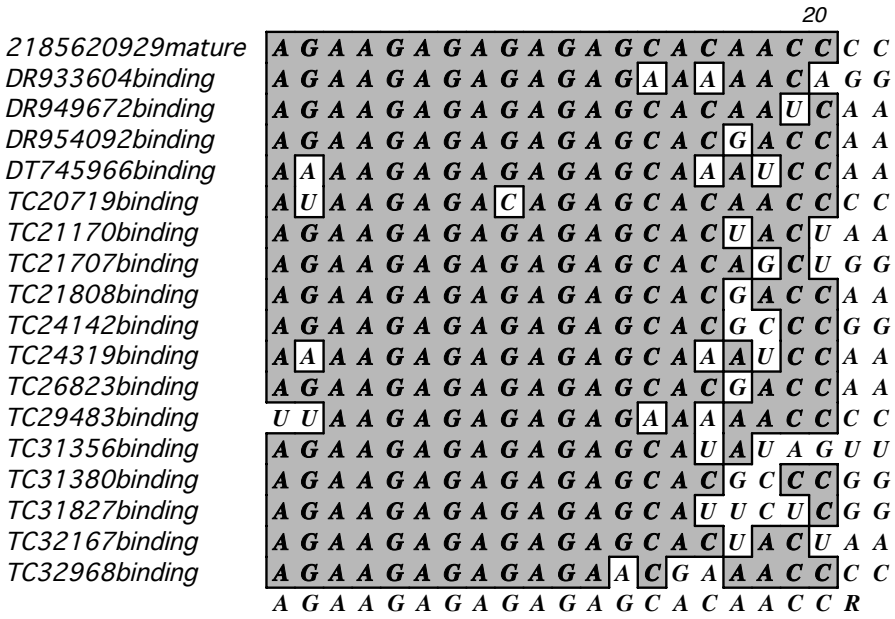
microRNA482

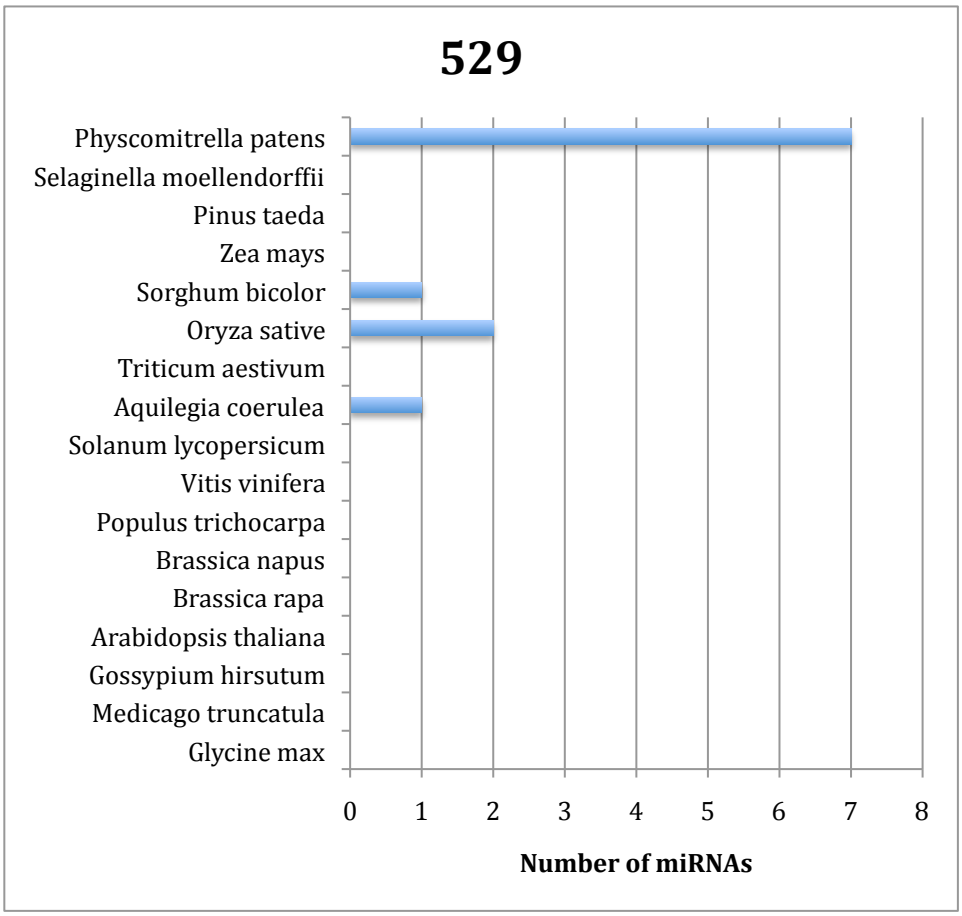
Output of sir_graph (©)
mfold_util_ng 4.1

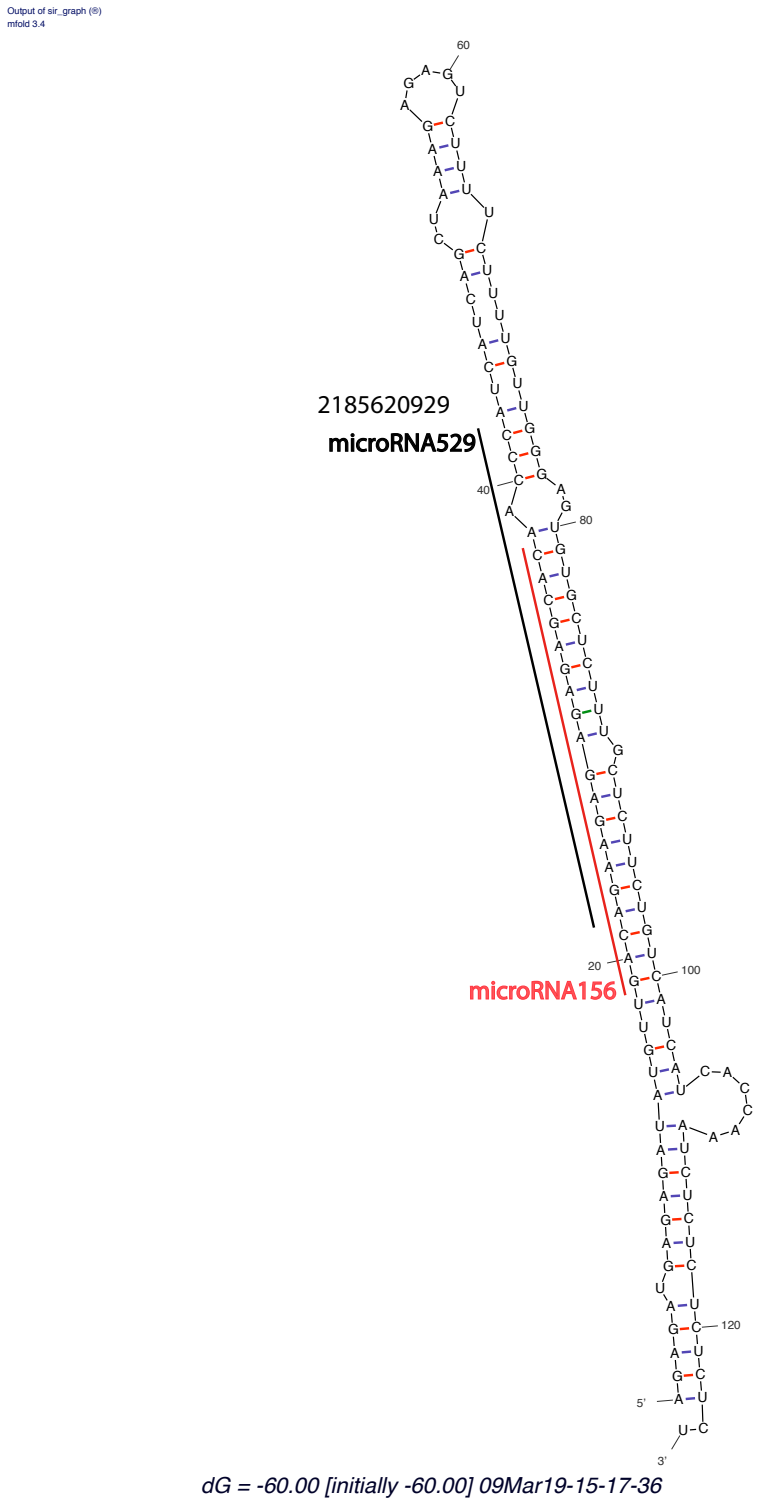
Created Tue Apr 7 09:24:39 2009



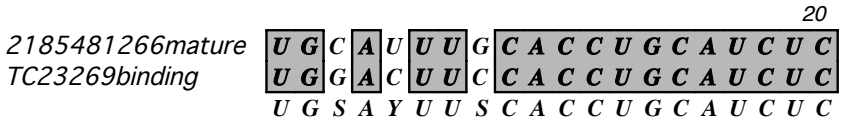
dG = -81.50 [initially -81.50] 09Apr07-09-24-18

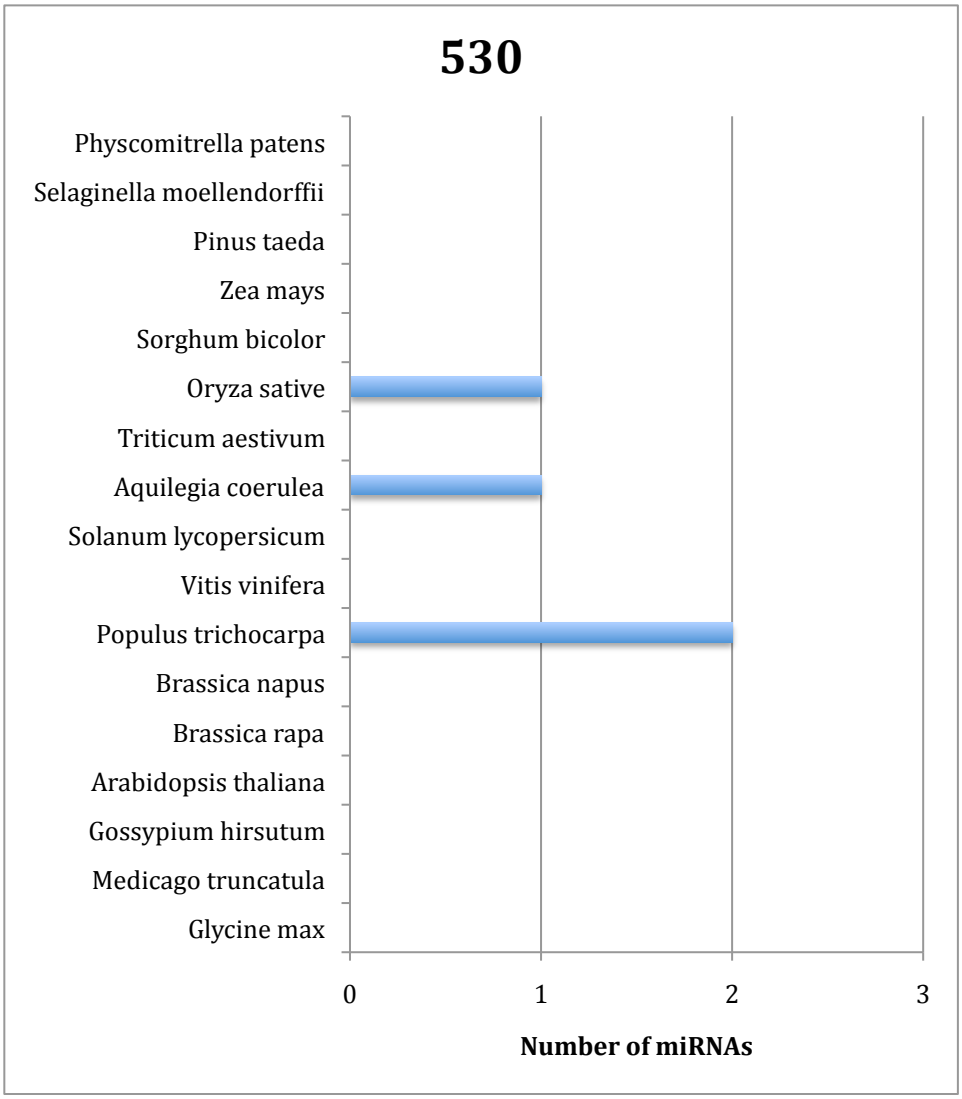

20
2185620929mature A G A A G A G A G A G A G C A C A A C C C C
DR933604binding A G A A G A G A G A G A G A A A A A C A G G
DR949672binding A G A A G A G A G A G A G C A C A A U C A A
DR954092binding A G A A G A G A G A G A G C A C G A C C A A
DT745966binding A A A A G A G A G A G A G C A A A U C C A A
TC20719binding A U A A G A G A C A G A G C A C A A C C C C
TC21170binding A G A A G A G A G A G A G C A C U A C U A A
TC21707binding A G A A G A G A G A G A G C A C A G C U G G
TC21808binding A G A A G A G A G A G A G C A C G A C C A A
TC24142binding A G A A G A G A G A G A G C A C G C C C G G
TC24319binding A A A A G A G A G A G A G C A A A U C C A A
TC26823binding A G A A G A G A G A G A G C A C G A C C A A
TC29483binding U U A A G A G A G A G A G A A A A A C C C C
TC31356binding A G A A G A G A G A G A G C A U A U A G U U
TC31380binding A G A A G A G A G A G A G C A C G C C C G G
TC31827binding A G A A G A G A G A G A G C A U U C U C G G
TC32167binding A G A A G A G A G A G A G C A C U A C U A A
TC32968binding A G A A G A G A G A G A A C G A A A C C C C
A G A A G A G A G A G A G C A C A A C C R


529
Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sative
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max
0
1
2
3
4
5
6
7
8
Number of miRNAs



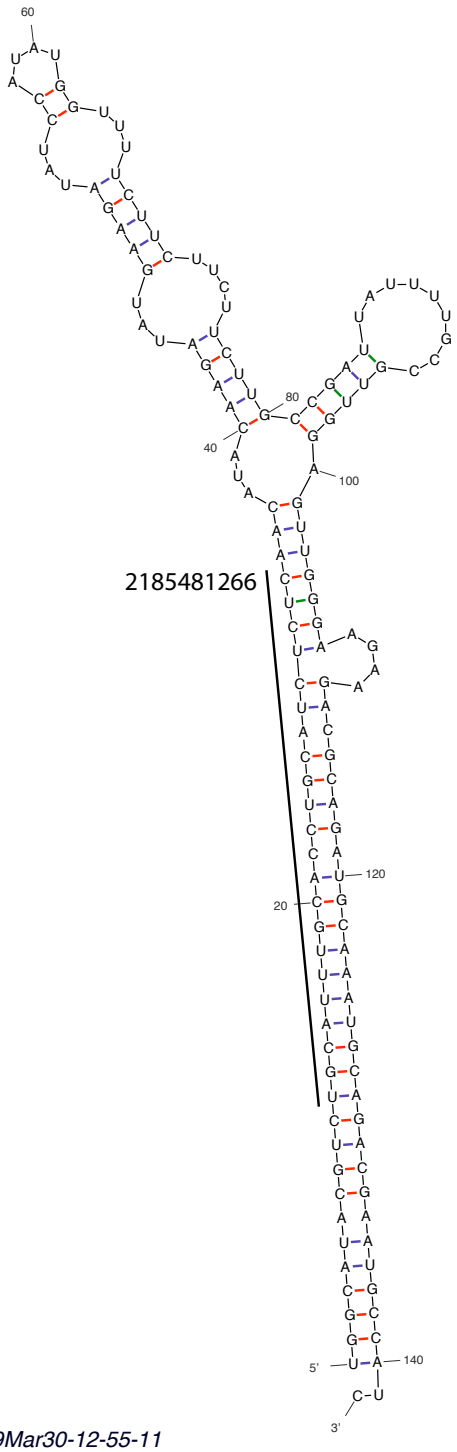
dG = -60.00 [initially -60.00] 09Mar19-15-17-36


20
2185481266mature
TC23269binding
U G C A U U G C A C C U G C A U C U C
U G G A C U C C A C C U G C A U C U C
U G S A Y U S C A C C U G C A U C U C


530
Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sative
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max
0
1
2
3
Number of miRNAs

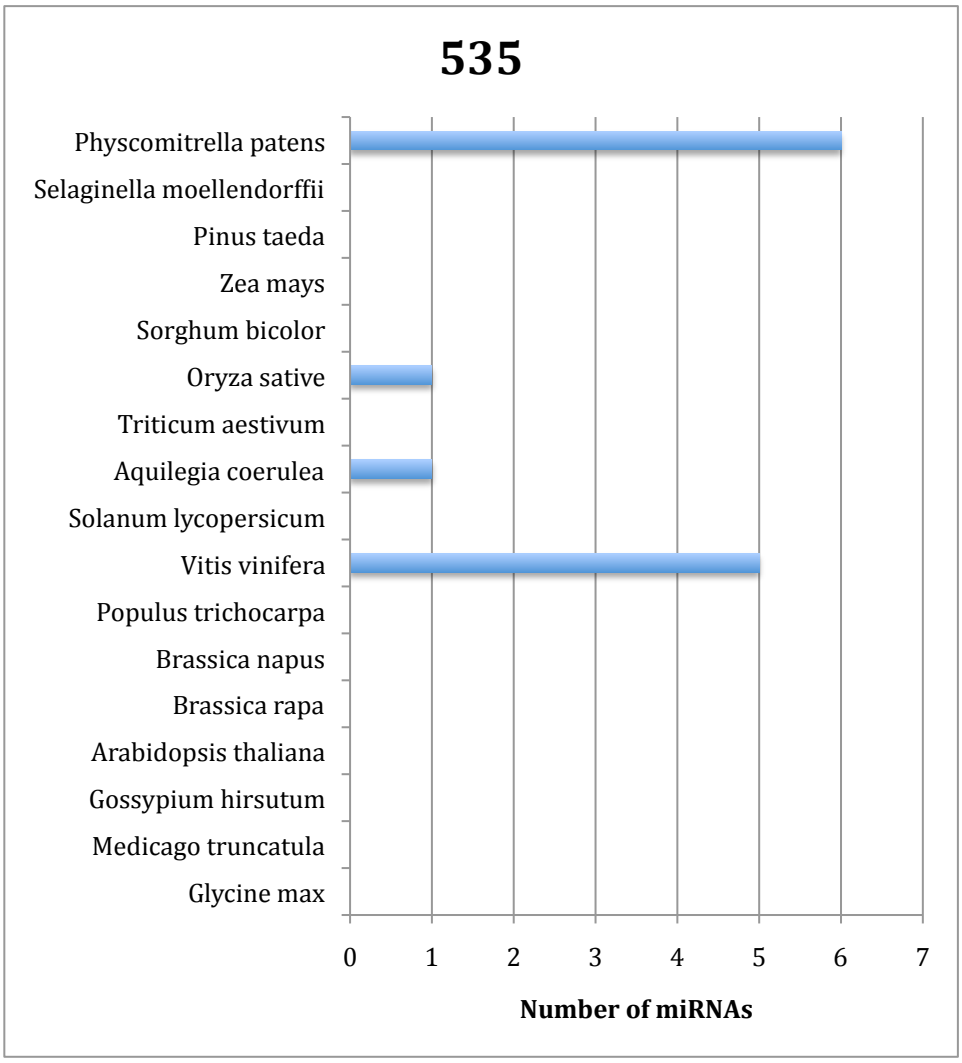
Output of sir_graph (©)
mfold_util_ng 4.1

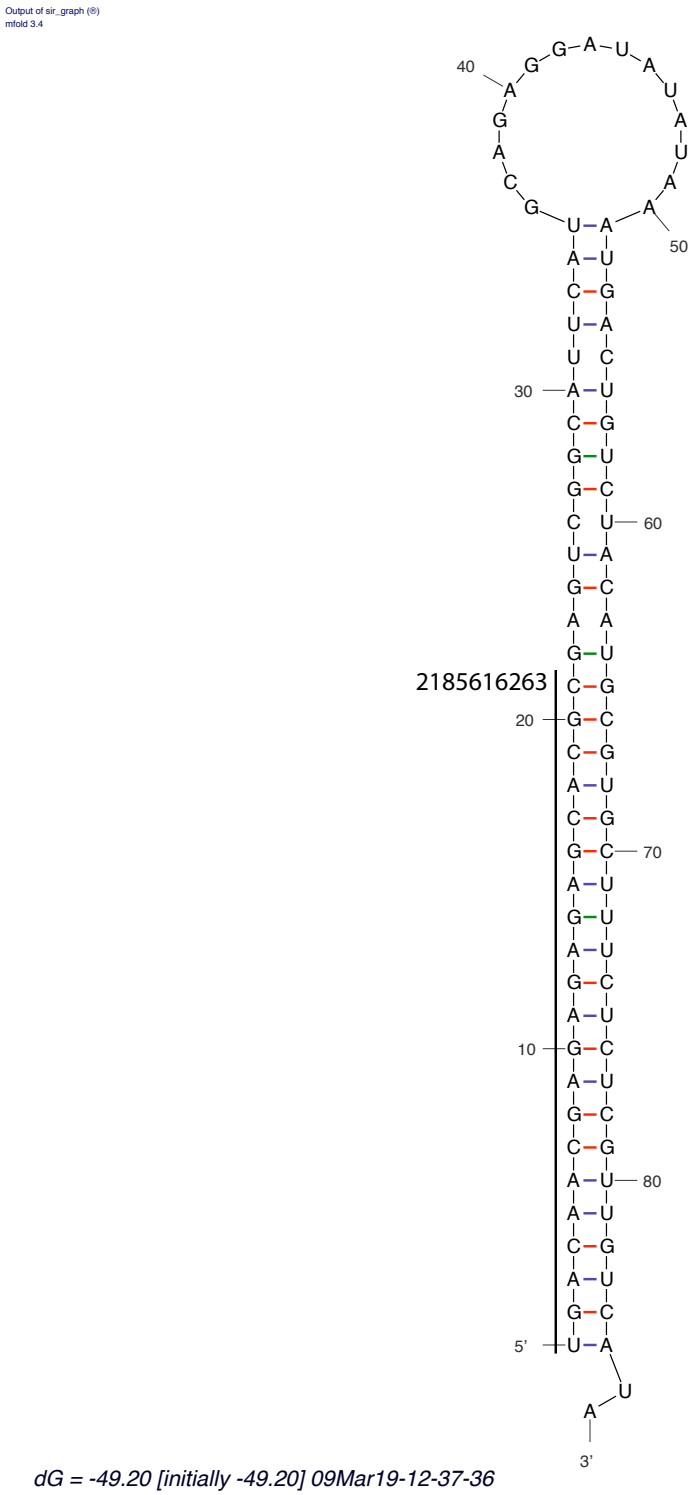
Created Mon Mar 30 12:55:21 2009



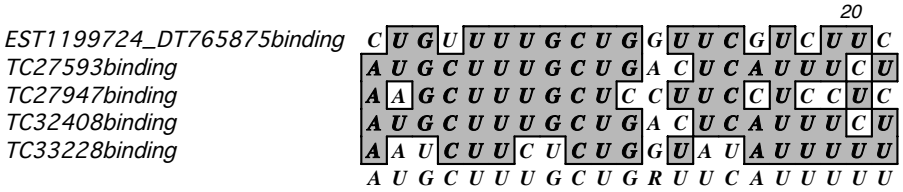
dG = -62.80 [initially -64.40] 09Mar30-12-55-11

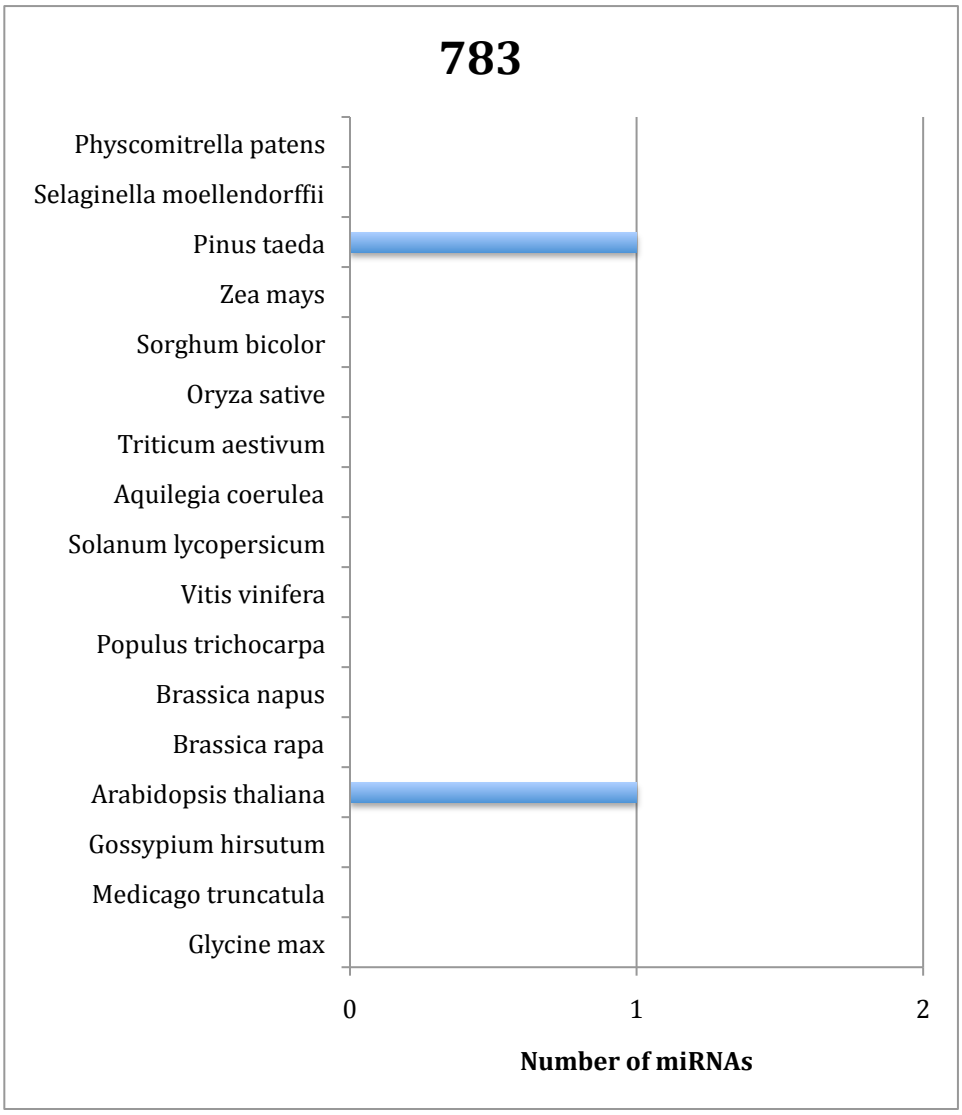
20
2185616263mature
U G A C A A C G A G A G A G A G C A C G C G
U G A C A A C G A G A G A G A G C A C G C G


535
Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sative
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max
0
1
2
3
4
5
6
7
Number of miRNAs



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


20
EST1199724_DT765875binding
TC27593binding
TC27947binding
TC32408binding
TC33228binding
A U G C U U U G C U G R U U C A U U U U U


783
Physcomitrella patens
Selaginella moellendorffii
Pinus taeda
Zea mays
Sorghum bicolor
Oryza sative
Triticum aestivum
Aquilegia coerulea
Solanum lycopersicum
Vitis vinifera
Populus trichocarpa
Brassica napus
Brassica rapa
Arabidopsis thaliana
Gossypium hirsutum
Medicago truncatula
Glycine max
0
1
2
Number of miRNAs