Planta (2008) 229:161–182
DOI 10.1007/s00425-008-0818-x

ORIGINAL ARTICLE

Identification of soybean microRNAs and their targets

Baohong Zhang · Xiaoping Pan · Edmund J. Stellwag

Received: 11 July 2008 / Accepted: 4 September 2008 / Published online: 25 September 2008

Abstract The microRNAs (miRNAs) are a newly identified class of small non-protein-coding regulatory RNA. Using comparative genomics, we identified 69 miRNAs belonging to 33 families in the domesticated soybean (*Glycine max*) as well as five miRNAs in the soybean wild species *Glycine soja* and *Glycine clandestine*. TaqMan® MicroRNA Assay analyses demonstrated that these miRNAs were differentially expressed in soybean tissues, with certain classes expressed preferentially in both a spatio-temporal and a tissue-specific manner. Detailed sequence analyses revealed that soybean pre-miRNAs vary in length from 44 to 259 nt with an average of 106 ± 45 nt, harbor mature miRNAs that differ in their physical location within the pre-miRNAs, and encode more than a single mature miRNA. Comparative sequence analyses of soybean miRNA sequences showed that uracil is the dominant base in the first position at the 5′ end of the mature miRNAs while cytosine is dominant at the 19th position, which is indicative that these two bases may have an important functional role in miRNA biogenesis and/or miRNA-mediated gene regulation. Soybeans were unique among plants in the frequency of occurrence of miRNA clusters. For the first time, antisense miRNAs were identified in plants. The five antisense miRNAs and their sense partners from soybean belonged to three miRNA families (miR-157, miR-162 and miR-396). Antisense miRNAs were also identified in soybean wild species. Mature antisense miRNA products appeared to have 1–3 nucleotide changes compared to their sense partners, which suggests that both strands of a miRNA gene can produce functional mature miRNAs and that antisense transcripts may differ functionally from their sense partners. Based on previously established in silico methods, we predicted 152 miRNA-targeted mRNAs, which included a large percentage of mRNAs that encode transcription factors that regulate plant growth and development as well as a lesser percentage of mRNAs that encode environmental signal transduction proteins and central metabolic processes.

Keywords Antisense microRNA · Comparative genomics · Expressed sequence tag · MicroRNA cluster · Posttranscriptional gene regulation · Soybean

Abbreviations

3′ UTR	3′ Untranslated region
AMFE	Adjusted MFE
ap2	Apetal 2
ARFs	Auxin response factors
CBF	CCAAT-binding transcription factor
EST	Expressed sequence tag
GSS	Genome survey sequence
GIF1	GRF1-interacting factor 1
HSP	Heat shock protein
miRNA	MicroRNA
MFEI	Minimal folding free energy index
pre-miRNAs	miRNA precusor
NCBI	National Center for Biotechnology Information

Electronic supplementary material The online version of this article (doi:10.1007/s00425-008-0818-x) contains supplementary material, which is available to authorized users.

B. Zhang (✉) · E. J. Stellwag
Department of Biology, East Carolina University, Greenville,
NC 27858, USA
e-mail: zhangb@ecu.edu

X. Pan
Department of Chemistry, Western Illinois University, Macomb,
IL 61455, USA

NAP	Nucleosome assembly protein
PCR	Polymerase chain reaction
qRT-PCR	Quantitative real time PCR
RISC	RNA-induced silencing complex
SBP	Squamosa-promoter binding protein-like protein

Introduction

Soybean (*Glycine max*) is one of most important agricultural crops in the USA and around the world. Because soybean seeds contain a high percentage of protein (40%) and oil (20%), soybean is considered the most nutritious crop and soybean seeds are processed into a variety of food products, such as soybean milk and tofu. Recently, soybean has been adopted as a potential source of biofuels. The widespread agricultural use of soybeans and the demand for increased production will require the development of cultivars with higher yields and improved resistance to environmental stressors. Thus, there are growing needs to modify the soybean to increase its yield and resistance to different environmental stresses. Although progress has been made, several critical problems remain, including the disease resistance and the need for increased yield. Newly discovered microRNAs (miRNAs) may play important roles in soybean development, nitrogen fixation, and the response to abiotic and biotic stresses.

The miRNAs are a class of small regulatory RNAs, which negatively regulate gene expression at the posttranscriptional levels by binding target mRNAs for mRNA cleavage or inhibition of mRNA translation (Zhang et al. 2006c). Many investigations have shown that miRNAs play an important role in a variety of biological and metabolic processes in plants and animals (Carrington and Ambros 2003; Ambros and Chen 2007; Zhang et al. 2007a). In plants, miRNAs function to control tissue (leaf, root, stem, and flower) differentiation and development, phase switching from vegetative growth to reproductive growth, signal transduction, and the response to biotic and abiotic stress (e.g., salinity, drought, and pathogens) (Chen 2005; Zhang et al. 2006c). Since the first miRNAs were discovered in plants in 2002 (Park et al. 2002; Reinhart et al. 2002), several hundred miRNAs have been identified in plants by computational and experimental approaches (Zhang et al. 2006e). A catalog of plant miRNAs includes 184 from *Arabidopsis thaliana*, 269 from rice, 234 from *Populus trichocarpa* (Griffiths-Jones 2004; Griffiths-Jones et al. 2006), and 188 from maize (Zhang et al. 2006a). However, very little is known about miRNAs in soybean despite its agricultural and economic significance (Zhang et al. 2005, 2006b; Subramanian et al. 2008; Sunkar and Jagadeeswaran 2008).

Comparative genomics across vastly divergent taxa has shown that many miRNAs are highly evolutionarily conserved from species to species, ranging from moss to high flowering eudicot species in the plant kingdom (Floyd and Bowman 2004; Zhang et al. 2006b) and from worms to humans in the animal kingdom (Pasquinelli et al. 2000, 2003; Altuvia et al. 2005). The extensive evolutionary conservation of miRNA provides a powerful approach to their identification using comparative genomics. Using this strategy, we recently developed an expressed sequence tag (EST) and a genome survey sequence (GSS) approach to identify miRNAs (Zhang et al. 2005; Pan et al. 2007). By using this approach, we have successfully identified more than 600 miRNAs in 71 plant species, including several important crops, such as wheat, tomato, tobacco, cotton and maize (Zhang et al. 2005, 2006a, b, 2007b). This approach has also been employed by other scientists to identify miRNAs in other plant species (Guo et al. 2007; Xie et al. 2007; Gleave et al. 2008). There are several significant advantages of using EST analysis for identifying miRNAs: (1) EST analysis can be employed to identify conserved miRNAs not only in model species, whose genomes have been published, but also in species for which only EST sequences have been determined; (2) EST analysis provides direct evidence for miRNA expression that cannot be inferred from genomic sequence surveys since EST are derived from transcribed sequences (mRNA) (Adams et al. 1991; Matukumalli et al. 2004); (3) miRNA identification using EST analysis can be conducted without specialized software using the BLASTn search algorithm and so is readily available for widespread use (Altschul et al. 1997). Although several computational programs have been developed for predicting miRNAs, all these programs are based on genome sequences and require that these programs run individually on a computer; there is no any clue on their expression of miRNAs predicted by these programs (Zhang et al. 2006e). Thus, the difficulty related to genome-based miRNA prediction is remedied by EST-based analyses. Based on these three advantages, EST analysis will significantly enhance our ability to identify miRNAs and to investigate miRNA structure, function and evolution. Except identifying more than the 600 miRNAs in 71 plant species, we employed EST analysis to analyze the evolutionary relationships of miRNAs. Our results demonstrated that many miRNAs are highly conserved evolutionarily across all major lineages of plants, including mosses, gymnosperms, monocots and eudicots. We further concluded that regulation of gene expression by miRNAs appears to have existed during the earliest stages of plant evolution and has been constrained (functionally) for more than 425 million years (Zhang et al. 2006b).

Currently, 394,370 soybean ESTs have been deposited in National Center for Biotechnology Information (NCBI)

GenBank EST database (based on data collected on 8 February 2008). This provides a valuable resource for the identification of potential miRNAs in soybean. The goal of this study is to identify soybean miRNAs and their potential targets. To achieve this goal, we first compared all of these EST sequences with the 184 known *Arabidopsis thaliana* miRNAs to identify potential miRNA homologs in soybean; then, selected putative soybean miRNAs were validated by quantitative real time PCR (qRT-PCR) using miRNA specific primers and probes (Chen et al. 2005). Based on these newly identified soybean miRNAs, we also predicted the potential miRNA targets in soybeans by BLASTn search.

Materials and methods

Reference set of miRNAs

To identify potential soybean miRNAs, a total of 184 known *A. thaliana* miRNAs were defined as a reference set of miRNA sequences. *A. thaliana* miRNAs were used as reference miRNAs because *A. thaliana* and soybeans are eudicots and a large number of *A. thaliana* miRNAs have been deposited in the publicly available miRNA database. The 184 *A. thaliana* mature miRNAs and their precursor sequences were downloaded from the miRNA database (miRBase Sequence Database, http://microrna.sanger.ac.uk/sequences/; release 10.1, December 2007). Although some of these *A. thaliana* miRNAs were initially identified by computational approaches, a majority of them have been validated by experimental approaches including direct cloning, PCR, and/or Northern blotting (Griffiths-Jones 2004; Griffiths-Jones et al. 2006).

Soybean ESTs, cDNAs, and mRNAs

Soybean EST, mRNA, and cDNA sequences were obtained from the GenBank nucleotide databases at NCBI and the soybean nucleotide databases from the Institute for Genome Research (TIGR) at http://www.tigr.org. A total of 394,370 soybean ESTs were deposited in the NCBI EST database and all of these ESTs were screened against the 184 known *A. thaliana* miRNAs.

Identifying potential soybean miRNAs using EST-based comparative genomics

Soybean miRNAs were identified according to our previously published method (Zhang et al. 2005, 2006a, b, 2007b). There are two important parameters in EST analysis; one is conservation of sequences, another is the second hairpin stem-loop structure of the potential pre-miRNAs. Figure S1 summarizes the general procedure for identifying conserved miRNAs in soybeans using EST-based comparative genomics. Briefly, the mature sequences of known *A. thaliana* miRNAs were subjected to a BLASTn search against all of the publicly available EST databases in NCBI using BLASTn 2.2.9 (1 May 2004) (Altschul et al. 1997). To improve the BLASTn search, the Blast parameter settings were adjusted as follows: expect values were set at 1,000 to increase the number of potential hits; the default word-match size between the query and database sequences was set at seven; and the number of descriptions and alignments was raised to 1,000. If the searches reveal partial sequence similarity to an *A. thaliana* mature miRNA sequence, the non-aligned regions were manually inspected and compared to determine the number of matching nucleotides to assess their potential as miRNA candidates. Those EST sequences that closely matched (no more than 4 mismatches, including insertion and/or deletion nucleotides) the previously known *A. thaliana* mature miRNAs were included in the set of miRNA candidates used for additional characterization based on the following criteria. The entire EST sequence (containing the conserved miRNA sequence) was selected to predict the secondary structures and to screen for miRNA precursor sequences. Although we also did BLASTn searches using miRNA precursor sequences, the mature sequences were the primary source of sequences for BLASTn searches against the EST databases of NBCI GenBank because only mature miRNAs, rather than miRNA precursors, are highly conserved in plants (Zhang et al. 2006b).

Expressed sequence tag sequences with four or fewer mismatches and/or indels (deletion/insertion nucleotides) compared to the previously identified *A. thaliana* miRNAs, were further compared with each other to eliminate redundancies. Then, the secondary structures of the non-redundant sequences were generated using the Zuker folding algorithm, as implemented through the web-based computational software MFOLD 3.2 (Mathews et al. 1999; Zuker 2003). MFOLD 3.2 is publicly available at http://www.bioinfo.rpi.edu/applications/mfold/rna/form1.cgi. The software default parameters were used to predict the secondary structures of the selected sequences. All MFOLD outputs including free energy (ΔG kcal/mol), the number of nucleotides (A, G, C and U), location of the matching regions, and the number of arms per structure were recorded. The minimal folding free energy index (MFEI) for each sequence was calculated as previously described (Zhang et al. 2006d). In previous studies, we found that miRNA precursor sequences have significantly higher MFEI than other non-coding or coding RNAs, and the candidate RNA sequences are more likely to be miRNAs when the MFEI is greater than 0.85 (Zhang et al. 2006d). To avoid mistakenly designating other types of RNAs as miRNA candidates, MFEI was also considered when predicting secondary structures.

In this study, an RNA sequence was considered a miRNA candidate only if it fit all of the following criteria: (1) predicted mature miRNAs had no more than four nucleotide substitutions compared with *A. thaliana* mature miRNAs; (2) the RNA sequence can fold into an appropriate stem-loop hairpin secondary structure; (3) the mature miRNA could be localized in one arm of the hairpin structure; (4) no more than 6 mismatches between the predicted mature miRNA sequence and its opposite miRNA* sequence in the secondary structure; (5) no loop or break in the miRNA or miRNA* sequences; (6) predicted secondary structure had high MFEI and negative MFE. Overall, the application of these criteria for inclusion of RNAs as miRNAs reduced the number of RNAs analyzed, minimized the likelihood that non-miRNAs would be included in subsequent analyses, and significantly reduced the total number of predicted false miRNAs.

Expression of soybean miRNAs

Two approaches were employed to confirm and establish the expression of miRNAs in soybean. They are qRT-PCR analysis and EST analysis. All identified soybean miRNA precursor sequences were used to perform BLASTn searches of the NCBI EST database, and BLASTn results were recorded and analyzed. The potential tissues in which we expected miRNA expression for miRNA candidates were determined based on the tissue sources reported for each EST in the NCBI database.

Isolation of total RNA

Soybean (*Glycine max* cv. NC Raleigh) seeds were kindly provided by Dr. Joseph Burton (ARS, USDA, Raleigh, NC, USA). The soybean seeds were cultivated in the Greenhouse Facility of East Carolina University. Total RNA was isolated from 4-week-old soybean seedlings using mirVana^{TM} miRNA Isolation Kit (Ambion, Austin, TX, USA) according to the manufacturer's protocol. Briefly, 0.1 g seedling tissues were harvested, weighed and immediately immersed into 1 mL Lysis/Binding Buffer in a microcentrifuge tube on ice. Collected samples were immediately homogenized using the Fisher Scientific PowerGen* Model 125 Homogenizer (Pittsburgh, PA, USA). Then, 300 μL of homogenized tissue was transferred into a centrifuge tube and 30 μL miRNA homogenate additive was added and mixed well by vortexing 20 s and left on ice for 10 min. Following the addition of 300 μL acid-phenol:chloroform (5:1, v/v; pH 4.5) and thorough vortexing for 60 s, the mixture was centrifuged for 5 min at 10,000*g* at room temperature. The upper aqueous phase was removed and 1.25 volumes of 100% ethanol (at room temperature) was added, mixed and allowed to incubate for 1 min on ice. The ethanol-treated samples were passed through the filter cartridge, which was washed three times with 500–700 μL of wash solution, and the purified RNA was eluted from the filter cartridge using 100 μL of elution buffer. The eluted RNA was stored at −20°C prior to analysis. The quality and the quantity of the total RNAs were measured using NanoDrop ND-1000 (NanoDrop Technologies, Wilmington, DE, USA).

To confirm soybean miRNAs and analyze the expression of miRNAs in soybean tissues, qRT-PCR was employed by using the Applied Biosystems TaqMan® microRNA Assays Protocol (Foster City, CA, USA). A two-step assay was performed in TaqMan-based real-time quantification of miRNAs. The first step involved a reverse-transcription reaction in which a stem-loop RT primer was used to reverse-transcribe mature miRNAs to cDNAs. The second step involved real-time PCR, in which the expression level of miRNAs was monitored and quantified using qRT-PCR that includes miRNA-specific forward primer, reverse primer and FAM dye-labeled TaqMan probes (Chen et al. 2005).

Reverse-transcription reaction

The miRNA reverse-transcription reactions contained 150 ng of total RNAs, 0.25 mM each of dNTPs, 3.33 U/μL MultiScribe reverse transcriptase, 1× reverse-transcription buffer, and 0.25 U/μL RNase inhibitor. The total volume of reverse-transcription reactions was adjusted to 15 μL using nuclease-free water. The miRNA reverse-transcription reactions were performed using an Eppendorf Mastercycler (Eppendorf North America, Westbury, NY, USA). The temperature program was 30 min at 16°C, 30 min at 42°C, 5 min at 85°C and then held at 4°C. All reverse-transcription reactions, including no-template controls and RT minus controls, were performed in duplicate.

Real-time reaction

Real-time PCR reactions were performed using TaqMan® microRNA Assays kit (Applied Biosystems) on an Applied Biosystems 7300 Sequence Detection System. Twenty μL PCR reaction mixtures were prepared and each contained 1 μL 20× TaqMan MicroRNA Assay primers and probes, 10 μL 2× TaqMan Universal PCR Master Mix, 1 μL of product from reverse-transcription reaction (after fivefold dilution), and 8 μL nuclease-free water. The reactions were incubated in a 96-well plate at 95°C for 10 min, followed by 45 cycles of 95°C for 15 s and 60°C for 60 s. After the completion of the real-time reactions, the threshold was manually set and the threshold cycle (C_T) was recorded. The C_T is defined as the fractional cycle number at which the fluorescence passes the fixed threshold (Chen et al. 2005). All reactions were conducted in triplicate.

Potential miRNA targets

There is ample documentation that the mechanism of miRNA-mediated gene regulation requires perfect or near-perfect complementarity between the miRNAs and their targeted mRNAs for directly cleaving mRNAs or repressing protein translation (Rhoades et al. 2002; Zhang et al. 2006c). A BLASTn search based on the complementarity between miRNAs and their targets has become a powerful approach for identifying plant miRNA targets. To date, a majority of miRNA targets in plants have been predicted based on BLASTn searches of databases followed by confirmation using one or several experimental approaches, including Northern blotting, qRT-PCR and 5′ rapid amplification of cDNA ends (5’RACE). In this study, we also used the BLASTn search to predict miRNA targets in the soybean. The procedure was similar to that described above for predicting soybean miRNA homologs. A modification was that we used the identified soybean miRNAs to do a BLASTn search in the protein-coding gene databases instead of the EST database. We searched the potential miRNA targets using the identified soybean miRNAs against the GenBank protein-coding nucleotide databases using BLASTn searches and the soybean nucleotide databases from TIGR using miRU (Zhang 2005). The parameters that were used in BLAST searches for miRNA targets, include total numbers of mismatched nucleotides between miRNAs and the potential targets and the alignment structures. The conservation of a target site in other plant species was also considered for identifying miRNA targets and eliminating false positives. The total number of allowed mismatches at complementary sites between miRNA sequences and potential mRNA targets were limited to no more than four (no mismatch between positions 10 and 11, no more than 1 mismatch between positions 1 and 9, and no more than 2 mismatches at other positions), and no gaps were allowed at the complementary sites. Because the proteome of soybean has not been fully annotated, we performed BLASTn searches using the predicted miRNAs against protein-coding nucleotide database as well as EST databases; in the later case, after identifying potential ESTs, we used the identified EST sequences to do a homology search against the protein-coding mRNA database in other plant species and decided the potential targeted genes based on the degree of similarity of protein-coding mRNAs among plant species.

Results

Identification of soybean miRNAs

After BLASTn searches of the NCBI EST databases using the 184 miRNAs from *A. thaliana* as probes and further screening based on analysis of the secondary structures of putative miRNA from the MFOLD 3.2 results, a total of 69 miRNAs were identified from a total of 394,370 soybean EST sequences (Table 1, Fig. 1 and Suppl. Fig. S2). These results provide evidence that about 0.0175% soybean ESTs contained potential miRNAs in the total pool of transcribed RNAs. This number is higher than the previously reported 0.010% for other plant species (Zhang et al. 2006b). There are two reasons that the soybean miRNA percentages are higher than those previously reported for other plant species; one is that several new miRNA families have been identified and the total number of plant miRNAs has increased since the previous study; another reason is that we modified the BLASTn search to include indels in addition to nucleotide substitutions as a measure of miRNA variation. In this study, we also identified five miRNAs from ESTs deposited in the NCBI EST database for *G. soja* (3 miRNA from 18,511 ESTs) and *G. clandestine* (2 miRNA from 911 ESTs). *G. soja* and *G. clandestine* are two important wild species of soybean.

Mature miRNAs can be located within either arm of the secondary hairpin stem-loop structures. Of the 74 soybean miRNAs identified, 38 (51.35%) are located in 3′ arm of the stem-loop hairpin structures while 36 (48.65%) are in 5′ arm. This property of soybean miRNAs is similar to those of other plant species in which mature miRNAs are typically confined to the stem-loop hairpin region.

The 69 miRNAs identified in soybean were classified into 33 miRNA families. The large number of miRNAs identified in soybean suggests that miRNAs are common in soybean and that miRNAs are highly conserved from *A. thaliana* to soybean. However, the abundance of miRNAs in each miRNA family differs (Fig. 2). Of the 69 miRNAs, 7 miRNAs belong to miR-166 family; 6 belong to miR-157 and miR-169, respectively; miR-172, miR-396 and miR-171 contain 4 members; and only one miRNA was identified from among each of the other miRNA families. The distribution of miRNAs among the various miRNA families in soybean is similar to other plant species, such as *A. thaliana*, rice, maize and cotton (Sunkar et al. 2005; Zhang et al. 2006a, b, 2007b). This uneven distribution of miRNAs in different family indicates that different miRNAs may have different evolutionary history and play a different role in plant development and growth.

The diversity of soybean miRNAs is also observed in the length of pre-miRNA sequences (Fig. 3). The length of soybean pre-miRNAs varies from 44 to 259 with an average of 105.7 ± 45.4 nt and with a median of 92 nt. More than 50% of pre-miRNAs are between 79 and 112 nt in length. This distribution of pre-miRNA lengths is similar to those reported for *Arabidopsis*, rice, cotton and maize (Sunkar et al. 2005; Zhang et al. 2006a, b, 2007b).

Table 1 Soybean miRNAs identified by comparative genomics and secondary structures

miR	Species	Query miRNAs	Mature sequence	MN	ML	Arm	PL	A (%)	C (%)	G (%)	U (%)	(A + U) (%)	G + C (%)	G/C	U/A	AMFE	MFE	MFEI	EST
gma-miR156a	Glycine max	ath-miR156	UGACAGAAGAGAGUGAGCAC	0	20	5′	83	24.10	18.07	28.92	28.92	53.01	46.99	1.60	1.20	50.72	42.1	1.08	BE807821
gma-miR156b	Glycine max	ath-miR156	UGACAGAAGAGAG*A*GAGCAC	1	20	3′	46	30.43	17.39	26.09	26.09	56.52	43.48	1.50	0.86	13.91	6.4	0.32	CX708501
gma-miR157a	Glycine max	ath-miR157a, b,c	UUGACAGAAGAUAGAGAGCAC	0	21	3′	155	28.39	21.29	23.23	27.10	55.48	44.52	1.09	0.95	25.87	40.1	0.58	BI971210
gma-miR157b	Glycine max	ath-miR157a, b,c	UUGACAGAAGAUAGAGAGCAC	0	21	5′	85	28.24	18.82	23.53	29.41	57.65	42.35	1.25	1.04	46.82	39.8	1.11	BE210632
gma-miR157c	Glycine max	ath-miR157a, b,c	UGACAGAAGA*C*UAGAGAGCAC	1	21	5′	85	29.41	23.53	18.82	28.24	57.65	42.35	0.80	0.96	49.65	42.2	1.17	BE210632
gma-miR157d	Glycine max	ath-miR157a, b,c	UUGACAGAAGAUAGAGAGCAC	0	21	5′	112	31.25	16.07	18.75	33.93	65.18	34.82	1.17	1.09	40.18	45	1.15	AW756919
gma-miR157e	Glycine max	ath-miR157a, b,c	UGACAGAAG*U*AUAGAGAGCAC	1	21	5′	112	33.93	18.75	16.07	31.25	65.18	34.82	0.86	0.92	37.50	42	1.08	AW756919
gma-miR157f	Glycine max	ath-miR157d	UUGACAGAAGA*G*AGAGAGCAC	1	21	5′	85	22.35	22.35	28.24	27.06	49.41	50.59	1.26	1.21	50.71	43.1	1.00	BG650023
gma-miR159a	Glycine max	ath-miR159a	UUUGGAUUGAAGGGAGCUCUA	0	21	3′	176	21.02	19.32	24.43	35.23	56.25	43.75	1.26	1.68	46.99	82.7	1.07	BM893181
gma-miR159b	Glycine max	ath-miR159a	UUUGGAUUGAAGGGAG*A*UCCU	1	21	3′	259	25.87	19.31	26.64	28.19	54.05	45.95	1.38	1.09	29.96	77.6	0.65	BQ299420
gma-miR160a	Glycine max	ath-miR160a,b,c	UGCCUGGCUCCCUGUAUGCCA	0	21	5′	80	22.50	23.75	27.50	26.25	48.75	51.25	1.16	1.17	56.13	44.9	1.10	CA801322
gma-miR160b	Glycine max	ath-miR160a,b,c	UGCCUGGCUCCCUGUAUGCCA	0	21	5′	44	20.45	29.55	25.00	25.00	45.45	54.55	0.85	1.22	29.09	12.8	0.53	BG882856
gma-miR160c	Glycine max	ath-miR160a,b,c	UGCCUGGCUCCCUGUAUGCC*U*	1	21	3′	107	25.23	14.95	28.04	31.78	57.01	42.99	1.88	1.26	31.78	34	0.74	BM887596
gma-miR162b	Glycine max	ath-miR162a,b	UCGAU*G*AACC*G*CUGCAUCCAG	2	21	3′	84	27.38	19.05	26.19	27.38	54.76	45.24	1.38	1.00	47.02	39.5	1.04	CF807240
gma-miR162a	Glycine max	ath-miR162a,b	UCGAUAAACCUCUGCAUCCAG	0	21	3′	84	26.19	26.19	19.05	28.57	54.76	45.24	0.73	1.09	44.64	37.5	0.99	CF807240
gma-miR163	Glycine max	ath-miR163	AAGAGGACUU*U*GAACUU*G*UGA	2	21	5′	135	28.15	13.33	25.19	33.33	61.48	38.52	1.89	1.18	22.07	29.8	0.57	EV274047
gma-miR166a	Glycine max	ath-miR166	UCGGACCAGGCUUCAUUCCCC	0	21	3′	86	22.09	22.09	31.40	24.42	46.51	53.49	1.42	1.11	44.07	37.9	0.82	EV280596
gma-miR166b	Glycine max	ath-miR166	UCGGACCAGGCUUCAUUCCCC	0	21	3′	130	31.54	20.77	24.62	23.08	54.62	45.38	1.19	0.73	35.46	46.1	0.78	EV280596
gma-miR166c	Glycine max	ath-miR166	UCGGACCAGGCUUCAUUCCCC	0	21	3′	107	18.69	22.43	26.17	32.71	51.40	48.60	1.17	1.75	44.77	47.9	0.92	BI893541
gma-miR166d	Glycine max	ath-miR166	UCGGACCAGGCUUCAUUCCC*U*	1	21	3′	136	30.15	21.32	16.91	31.62	61.76	38.24	0.79	1.05	33.09	45	0.87	EV267001
gma-miR166e	Glycine max	ath-miR166	UCGGACCAGGCUUCAUUCCC*G*	1	21	3′	156	21.79	16.67	23.08	38.46	60.26	39.74	1.38	1.76	37.24	58.1	0.94	BI972515
gma-miR166f	Glycine max	ath-miR166	UCGGACCAGGCUUCAUUCCCC	0	21	3′	86	22.09	22.09	30.23	25.58	47.67	52.33	1.37	1.16	44.30	38.1	0.85	EV266011
gma-miR166 g	Glycine max	ath-miR166	UCGGACCAGGCUUCAUUCCCC	0	21	3′	128	29.69	18.75	27.34	24.22	53.91	46.09	1.46	0.82	38.05	48.7	0.83	EV266011
gma-miR167a	Glycine max	ath-miR167a,b	UGAAGCUGCCAGCAUGAUCUA	0	21	5′	82	21.95	23.17	28.05	26.83	48.78	51.22	1.21	1.22	46.46	38.1	0.91	BI095235
gma-miR167b	Glycine max	ath-miR167d	UGAAGCUGCCAGCAUGAUCUG	0	21	5′	64	25.00	20.31	23.44	31.25	56.25	43.75	1.15	1.25	56.56	36.2	1.29	BM892909
gma-miR168	Glycine max	ath-miR168a,b	UCGCUUGGUGCAGGUCGGGAA	0	21	5′	87	16.09	27.59	35.63	20.69	36.78	63.22	1.29	1.29	46.90	40.8	0.74	BE661028
gma-miR169a	Glycine max	ath-miR169a	CAGCCAAGGAUGACUUGCCGA	0	21	5′	66	25.76	19.70	15.15	39.39	65.15	34.85	0.77	1.53	11.21	7.4	0.32	AW201497
gma-miR169b	Glycine max	ath-miR169b,c	CAGCCAAGGAUGACUUGCCGG	0	21	5′	82	15.85	24.39	25.61	34.15	50.00	50.00	1.05	2.15	43.54	35.7	0.87	CA953278
gso-miR169a	Glycine soja	ath-miR169b,c	CAGCCAAGGAUGACUUGCCGG	0	21	5′	103	29.13	18.45	21.36	31.07	60.19	39.81	1.16	1.07	50.10	51.6	1.26	BF598910
gma-miR169c	Glycine max	ath-miR169b,c	CAGCCAAGGAUGACUUGCCGG	0	21	5′	111	30.63	17.12	19.82	32.43	63.06	36.94	1.16	1.06	50.81	56.4	1.38	AW596073

Table 1 continued

miR	Species	Query miRNAs	Mature sequence	MN	ML	Arm	PL	A (%)	C (%)	G (%)	U (%)	(A + U) (%)	G + C (%)	G/C	U/A	AMFE	MFE	MFEI	EST
gma-miR169d	Glycine max	ath-miR169d, e,f,g	*GC*AGCCAAGGAUGACUUGCCG	2	21	5′	84	15.48	23.81	26.19	34.52	50.00	50.00	1.10	2.23	42.50	35.7	0.85	CA953278
gso-miR169b	Glycine soja	ath-miR169d, e,f,g	*GC*AGCCAAGGAUGACUUGCCG	2	21	5′	105	28.57	18.10	21.90	31.43	60.00	40.00	1.21	1.10	49.14	51.6	1.23	BM524615
gma-miR169e	Glycine max	ath-miR169d, e,f,g	UGAGCCAAGGAUGA*G*UUGCC*U*	2	21	3′	254	41.34	20.08	16.54	22.05	63.39	36.61	0.82	0.53	17.68	44.9	0.48	AW310292
gso-miR169 g*	Glycine soja	ath-miR169 g*	UCGGCAAGUUG*G*CCUUGGCU	1	20	3′	101	28.71	17.82	21.78	31.68	60.40	39.60	1.22	1.10	50.10	50.6	1.27	BM524615
gma-miR169 g*	Glycine max	ath-miR169 g*	UCGGCAAGUUG*G*CCUUGGCU	1	20	3′	109	30.28	16.51	20.18	33.03	63.30	36.70	1.22	1.09	50.83	55.4	1.39	AW596073
gma-miR171a	Glycine max	ath-miR171a	U*C*AUUGAGCCG*U*GCCAAUAUC	2	21	3′	79	25.32	17.72	21.52	35.44	60.76	39.24	1.21	1.40	49.37	39	1.26	CA937914
gma-miR171b	Glycine max	ath-miR171a	UAAUUGAGCCGCGUCAAUAUC	2	21	3′	79	30.38	17.72	22.78	29.11	59.49	40.51	1.29	0.96	41.14	32.5	1.02	BM892213
gma-miR171c	Glycine max	ath-miR171b,c	UUGAGCCGUGCCAAUAUCACG	0	21	3′	85	24.71	18.82	23.53	32.94	57.65	42.35	1.25	1.33	48.24	41	1.14	CA937914
gma-miR171d	Glycine max	ath-miR171b,c	UUGAGCCG*C*GCCAAUAUCAC*U*	2	21	3′	88	31.82	19.32	25.00	23.86	55.68	44.32	1.29	0.75	43.41	38.2	0.98	CO979466
gma-miR172a	Glycine max	ath-miR172a,b	AGAAUCUUGAUGAUGCUGCAU	0	21	3′	120	25.83	19.17	28.33	26.67	52.50	47.50	1.48	1.03	41.75	50.1	0.88	BU084569
gma-miR172b	Glycine max	ath-miR172a,b	AGAAUCUUGAUGAUGCUGCAU	0	21	3′	114	27.19	18.42	24.56	29.82	57.02	42.98	1.33	1.10	45.18	51.5	1.05	BI320499
gma-miR172c	Glycine max	ath-miR172c,d	AGAAUC*C*UGAUGAUGCUGCAG	1	21	5′	89	22.47	22.47	31.46	23.60	46.07	53.93	1.40	1.05	25.17	22.4	0.47	CO981166
gma-miR172d	Glycine max	ath-miR172e	GGAAUC*C*UGAUGAUGCUGCA*G*	2	21	5′	141	23.40	23.40	27.66	25.53	48.94	51.06	1.18	1.09	37.73	53.2	0.74	EH223710
gma-miR319a	Glycine max	ath-miR319a,b	UUGGACUGAAGGGAGCUCCCU	0	21	3′	170	30.00	18.82	23.53	27.65	57.65	42.35	1.25	0.92	48.71	82.8	1.15	BQ630517
gma-miR319b	Glycine max	ath-miR319c	UUGGACUGAA*A*GGAGCUCCUU	1	21	3′	184	28.80	19.02	21.20	30.98	59.78	40.22	1.11	1.08	42.17	77.6	1.05	BG237979
gma-miR319c	Glycine max	ath-miR319c	UUGGA*G*UGAAGGGAGCUCC*AG*	3	21	3′	168	23.81	20.24	25.60	30.36	54.17	45.83	1.26	1.28	46.31	77.8	1.01	BQ453148
gma-miR394a	Glycine max	ath-miR394a,b	UUGGCAUUCUGUCCACCUCC	0	20	5′	79	17.72	29.11	21.52	31.65	49.37	50.63	0.74	1.79	40.25	31.8	0.80	BU082875
gma-miR394b	Glycine max	ath-miR394a,b	UUGGCAUUCUGUCCACCUCC	0	20	5′	65	18.46	30.77	20.00	30.77	49.23	50.77	0.65	1.67	39.54	25.7	0.78	AW099182
gma-miR395	Glycine max	ath-miR395a, d,e	*A*UGAAGUGUUUGGGGGAACUC	1	21	3′	67	19.40	16.42	29.85	34.33	53.73	46.27	1.82	1.77	53.58	35.9	1.16	AW596801
gma-miR396a	Glycine max	ath-miR396a	UUCCACAGCUUUCUUGAACUG	0	21	5′	101	22.77	25.74	18.81	32.67	55.45	44.55	0.73	1.43	41.49	41.9	0.93	CA853579
gma-miR396b	Glycine max	ath-miR396a	U*C*CCACAGCUUU*A*UUGAAC*C*G	3	21	5′	101	32.67	18.81	25.74	22.77	55.45	44.55	1.37	0.70	38.02	38.4	0.85	CA853579
gma-miR396c	Glycine max	ath-miR396b	UUCCACAGCUUUCUUGAACUU	0	21	5′	87	25.29	26.44	16.09	32.18	57.47	42.53	0.61	1.27	40.23	35	0.95	BM521827
gma-miR396d	Glycine max	ath-miR396b	U*C*CCACAGCUUUCUUGA*G*CUU	2	21	5′	87	32.18	16.09	26.44	25.29	57.47	42.53	1.64	0.79	45.52	39.6	1.07	BM521827
gcl-miR396a	Glycine clandestina	ath-miR396a	U*C*CCACAGCUUU*A*UUGAAC*C*G	3	21	3′	100	32.00	20.00	28.00	20.00	52.00	48.00	1.40	0.63	39.80	39.8	0.83	BG838673
gcl-miR396b	Glycine clandestina	ath-miR396b	UUCCACAGCUUUCUUGAACAG	1	21	3′	100	20.00	28.00	20.00	32.00	52.00	48.00	0.71	1.60	41.00	41	0.85	BG838673
gma-miR398a	Glycine max	ath-miR398a	UGUGUUCUCAGGUCACCCCUU	0	21	3′	94	24.47	18.09	24.47	32.98	57.45	42.55	1.35	1.35	46.81	44	1.10	CB063312
gma-miR398b	Glycine max	ath-miR398b,c	UGUGUUCUCAGGUC*A*CCCCUG	1	21	3′	98	24.49	24.49	21.43	29.59	54.08	45.92	0.88	1.21	48.78	47.8	1.06	BM732696
gma-miR399	Glycine max	ath-miR399d	UGCCAAAGGAG*U*UUUGC*AA*G	3	20	5′	48	33.33	16.67	22.92	27.08	60.42	39.58	1.38	0.81	23.13	11.1	0.58	AW234053
gma-miR408a	Glycine max	ath-miR408	AUGCACUGCCUCUUCCCUGGC	0	21	3′	119	31.09	18.49	32.77	17.65	48.74	51.26	1.77	0.57	33.92	40.36	0.66	EV270125
gma-miR408b	Glycine max	ath-miR408	AUGCACUGCCUCUUCCCUGGC	0	21	3′	112	28.57	19.64	29.46	22.32	50.89	49.11	1.50	0.78	42.32	47.4	0.86	AI856618
gma-miR414	Glycine max	ath-miR414	*A*UAUCUUCAUCAUC*C*UCGUCA	2	21	5′	211	24.17	27.96	15.64	32.23	56.40	43.60	0.56	1.33	19.19	40.5	0.44	EH221170
gma-miR415	Glycine max	ath-miR415	AACAGAG*G*AGAAACAGAA*A*A*G*	3	21	5′	61	37.70	11.48	22.95	27.87	65.57	34.43	2.00	0.74	11.80	7.2	0.34	EV282889

Table 1 continued

miR	Species	Query miRNAs	Mature sequence	MN	ML	Arm	PL	A (%)	C (%)	G (%)	U (%)	(A + U) (%)	G + C (%)	G/C	U/A	AMFE	MFE	MFEI	EST
gma-miR426	Glycine max	ath-miR426	*AG*UUGGAAAUUU*C*UCCUUAC	3	20	3′	71	33.80	15.49	26.76	23.94	57.75	42.25	1.73	0.71	31.27	22.2	0.74	BE347331
gma-miR447	Glycine max	ath-miR447a,b	UA*U*GGAC*C*AGAUGUUUUGUUG	3	21	5′	59	28.81	15.25	22.03	33.90	62.71	37.29	1.44	1.18	14.92	8.8	0.40	EV280751
gma-miR779	Glycine max	ath-miR779.2	UGAUUGGAAAUUUC*G*UU*G*ACU	3	21	5′	61	16.39	14.75	21.31	47.54	63.93	36.07	1.44	2.90	24.10	14.7	0.67	CF920966
gma-miR781	Glycine max	ath-miR781	UUAGAGUUUUCUGAAUA*CAGA*	3	21	3′	70	24.29	12.86	25.71	37.14	61.43	38.57	2.00	1.53	34.00	23.8	0.88	CA799476
gma-miR824	Glycine max	ath-miR824	UAGACCAUUUGUGAGAA*CAGA*	2	21	5′	66	21.21	18.18	15.15	45.45	66.67	33.33	0.83	2.14	15.30	10.1	0.46	CX711529
gma-miR825	Glycine max	ath-miR825	UUCUCAAGAAGGUG*G*AUGA*UG*	3	21	5′	220	26.36	20.00	22.27	31.36	57.73	42.27	1.11	1.19	27.32	60.1	0.65	BM732778
gma-miR830	Glycine max	ath-miR830	UAACUAUU*C*UGAGAAGAAA*U*A	3	21	5′	110	35.45	13.64	16.36	34.55	70.00	30.00	1.20	0.97	24.09	26.5	0.80	CD402807
gma-miR854	Glycine max	ath-miR854a, b,c,d	GAAGAGGAUA*G*GGAGGAGGAG	2	21	5′	77	24.68	18.18	31.17	25.97	50.65	49.35	1.71	1.05	38.05	29.3	0.77	EV275048
gma-miR860	Glycine max	ath-miR860	*G*CAAU*G*GAUUGGACUAUG*GGC*	4	21	3′	96	32.29	22.92	17.71	27.08	59.38	40.63	0.77	0.84	21.46	20.6	0.53	BM091672
gma-miR862	Glycine max	ath-miR862-5p	*U*CCAAUAGGU*G*GAGCAUGUG	3	20	3′	204	20.10	23.04	25.98	30.88	50.98	49.02	1.13	1.54	32.30	65.9	0.66	EV276797
gma-miR865	Glycine max	ath-miR865-5p	GU*U*AAUUUGGAUCUAAUUAA	3	20	3′	79	21.52	12.66	29.11	36.71	58.23	41.77	2.30	1.71	18.99	15	0.45	BM521823
gma-miR869	Glycine max	ath-miR869.1	*C*AUGGUUCAAU*G*CUGGUGUUA	3	21	5′	51	15.69	11.76	35.29	37.25	52.94	47.06	3.00	2.38	28.04	14.3	0.60	EV268005

MN number of mismatched nucleotides, *ML* length of mature miRNAs, *PL* length of pre-miRNAs, *Arm* mature miRNA location in the secondary stem-loop structures of pre-miRNA sequences, *EST* the EST sequence where the miRNA is deviated from, here only give one representative EST if there are more ESTs containing the miRNAs

Italic letters show nucleotide substitutions compared with the previously published *Arabidopsis* query miRNAs

Characteristics of soybean miRNAs

Although there are few obvious similarities in the nucleotides at specific positions along the mature miRNA sequences identified in this study, uracil constitutes about 70% of the bases at the 5′ end of the mature miRNAs while cytosine represents greater than 60% of the bases at position 19 (Fig. 4). These results are comparable to those obtained in comparisons to cotton miRNA in which 53 (71.62%) and 49 (66.22%) of 74 cotton miRNAs have U and C at the 1st and the 19th positions from the 5′ end, respectively. We previously reported that uracil is the predominant nucleotide at the 5′ end of mature miRNA sequences, and proposed that uracil may play an important role in miRNA biogenesis through recognition of targeted miRNA precursors by the RNA-induced silencing complex (RISC) (Zhang et al. 2006b). However, a further comparison of the nucleotide distribution at each position in all the mature plant miRNAs reported to date showed that cytosine, like uracil is the predominate nucleotide at position 19 (61% of cases). Why is cytosine the dominant nucleotide at position 19 in mature miRNA? One possible reason is that cytosine at this location may be important for targeting RISC or Dicer cleavage to specific sites on pre-miRNAs.

The percentage composition of four nucleotides (A, C, G and U) in soybean pre-miRNAs is not even nor are the G/C or A/U ratios (Table 2). Uracil is dominant in soybean pre-miRNA sequences and comprises 29.94 ± 5.33% of total nucleotide composition followed by adenine, guanine and cytosine (19.93 ± 4.24%). While the expected ratio of G/C and U/A bases in fully double-stranded RNA would be 1, we found ratios of 1.27 ± 0.41 and 1.22 ± 0.43, respectively. This suggests soybean pre-miRNAs contain about 20% U and G more than A and C, respectively, which could be a simple manifestation of the unequal distribution of nucleotides within single-stranded regions of the pre-miRNAs.

Minimal folding free energy (MFE) is one important characteristic that determines RNA and DNA secondary structure. The lower the MFEs, the higher the thermodynamic stability of the corresponding sequences; so the sequences with lower MFEs can form stable secondary stem-loop structures. Several studies have demonstrated that pre-miRNAs have high negative MFE (Bonnet et al. 2004b; Zhang et al. 2006d). In this study, we observed that soybean pre-miRNAs have high negative MFEs ranging from −6.4 to −82.8 kcal/mol with an average of −39.60 ± 17.26 kcal/mol. However, MFEs are strongly and positively corrected with their RNA/DNA sequence length. The longer the RNA sequences, the more freedom (the lower the MFEs) the sequences have to form stable secondary structures. To normalize the potential effect of nucleotide length on MFE calculations we developed a gma-miR156a

```
       -    -|        au   a ug  auu       a
ugacaga agag agugagcac  gcu g  gu   uguaug g
acugucu ucuc ucacucgug  ugg c  ua   acauac g
       a    u^        cg   g gu  ---       g
```

gma-miR157b

```
u          -          u  u  -   a   -|  gaa
 ugacagaaga uagagagcac ga ga uga aug cau   a
 acugucuucu aucucucgug uu cu acu uac gua   g
u          g          u  c  c   c   g^  agg
```

gma-miR160a

```
   c       c                  c    c   a
ugc uggcucc uguaugccauuuguagag ucau ga g
acg accgagg guaugcgguagguguuuc agua cu c
   a       a                  c    a   a
```

gma-miR162a

```
          c    c       uc   c  aa  gu-| uu
cuggaugcag gguu aucgauc  uuc ug  uc   ug  \
gaccuacguc ccaa uagcuag  aag ac  ag   ac  u
          u    a       cu   u  ca  acu^ aa
```

gma-miR167b

```
      u    g         ag      uuc
ugaagc gcca caugaucug  uuuacc   \
acuucg cggu guacuagac  gaaugg   u
      u    -         aa      uua
```

gma-miR168

```
   c          u     a       uc   -   aa--| ga
ucg uuggugcagg cggga ccgguuu  gcg cgg    ug  n
agu aacuacguuc gcccu gguuaag  cgc gcc    gc  g
   c          c     a       c-   g   gcug^ ga
```

gma-miR169b

```
 c       g u            uuauu   --| cu   g
g agccaag a gacuugccggcg     auu ug   cau u
u ucgguuu u uugaacggccgu     ugg ac   gua u
 a       g g            uccuu   cc^ uc   c
```

gma-miR169d

```
c       g              uuauu   --| cu   g
 agccaag augacuugccggcg     auu  ug  cau u
 ucgguuu uguugaacggccgu     ugg  ac  gua u
a       g              uccuu   cc^ uc   c
```

gma-miR171a

```
|       uc         a        au        u
gauauugg  cgguucaau agaaagca  gcucaaaa \
cuauaacc  gccgaguua uuuuuugu  uggguuuu g
^       gu         c        cc        u
```

gma-miR395

```
|                    g   uu    agu
gaguuccucugaacgcuucau uga  ggcu   \
cucaaggggguuugugaagua guu  ccgg   u
^                    -   u-    aua
```

gma-miR396c

```
|                a     uau  aucuuau   u  cca
uuccacagcuuucuuga cuucu   gc       auc cu   c
agggugucgaaagaacu gaaga   cg       uag ga   c
^                c     ucc  aauuu--   -  ccu
```

gma-miR398b

```
     u  u       c        ac-  auu     ---|       u
caggg cg ccugaga cacaugaa   ag   caaaa   uacaagca a
guccc gc ggacucu guguacuu   uc   guuuu   guguucgu u
     c  u       u        gac  agu     cca^       u
```

gma-miR399

```
    -|     guu   ca    aa
ugcc aaagga   uug  agcu  \
gugg uuuucu   agc  ucga  a
    a^     ac-   a-    aa
```

gma-miR779

```
       aa  uu-|   u     uuuuuu
ugauugg  au  ccuu ccuuc      u
acugauu  ua   ggaa gggag      u
       gg  uuu^   u     ccuucg
```

gma-miR869

```
|    u   u    g         g
ucuca ggu caau cugguguua a
agggu ccg guua gguuguggu g
^    u   -    a         u
```

gcl-miR396a

```
 c          a      c   aaa  a  u  ggga--|  aga    c
u ccacagcuuu uugaac gca   ca gg gu       gca   gugc a
a ggugucgaaa aacuug cgu   gu uc ca       cgu   uacg u
 a          g      a   ag-  a  u  aggaar^  acg    g
```

gso-miR169a

```
c|       au           uu   caaguga     ca            a
 agccaagg  gacuugccggca  agc          augag  ucauauauauau u
 ucgguucc  uugaacggcugu  uug          uacuc  aguauauauaua a
a^       gg           uc   -------     --            u
```

gso-miR169g*

```
|       au           uu   caaguga     ca            a
agccaagg  gacuugccggca  agc          augag  ucauauauauau u
ucgguucc  uugaacggcugu  uug          uacuc  aguauauauaua a
^       gg           uc   -------     --            u
```

gma-miR159a

```
           ga       uu    aucu    ----   ug   u     g     c      uc      uu     -| ucaaua
uggagcuccuu  aguccaa  gagg    uacu    ggg  aau gagcu cuuag uaugga  ccacag  cuacc ca      a
aucucgaggga  uuagguu  uucc    auga    ucc  uua uucga ggguc auaccu  gguguu  gaugg gu      g
           ag       uc    aauu    uauu   gu   c     g     u      uc      cu     u^ uucgu
```

gma-miR172a

```
                    a ug--    c-| ag     --  g   g  uugau   auc
auguagcaucaucaagauuc ca    caag  gc  guggu  gg ugg ac     gca   \
uacgucguaguaguucuaag gu    guuu  ug  uaccg  cc acc ug     cgu   u
                    a  caag    uc^ ga     aa  g   g  u----   gaa
```

gma-miR319c

```
            g        gu  g       augau       c  ua  g    uu    gau   uc       uu   aac
uuggaguucccu cacuccaa  cu aaaggau     gguaaac uc  cu cuag  caug   acc  ugacuuc  aac   \
gaccucgaggga gugagguu  gg uuuucua     ccauuug ag  ga gguc  guac   ugg  acugaag  uug   a
            a        au  -       guu--       -  ua  g    cc    guu   ga       c-   cgu
```

gma-miR408a

```
   -|    c      g   c  a  g     u       aaaugguaaagugagaauga
gcu gggaa aggcag gca ga ug agcua caacaga                    \
cgg cccuu uccguc cgu cu gc ucggu guugucu                    a
   u^    c      a   a  c  a     -       agagagagagagagagagga
```

Fig. 1 Predicted hairpin secondary structures of the selected soybean miRNAs identified in this study. Mature miRNA sequences are *shaded*. miRNA precursors may be slightly longer than the sequences shown in this figure

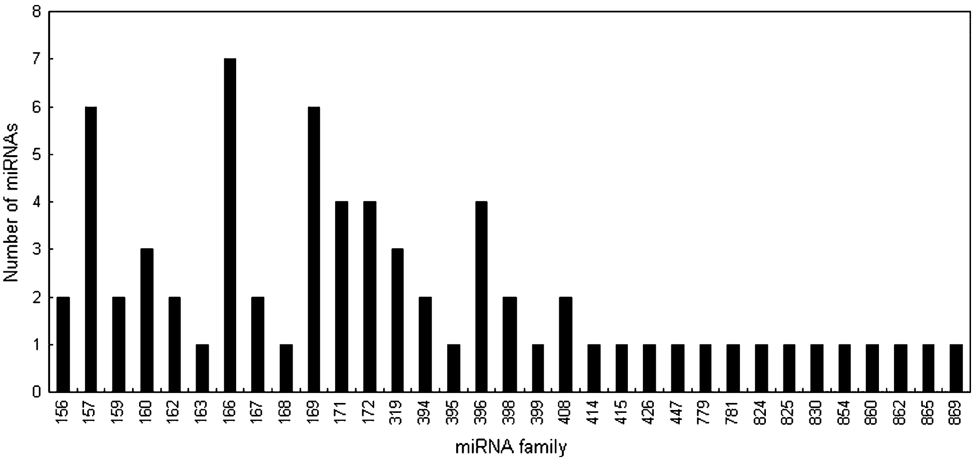


Fig. 2 Size of miRNA families in soybean

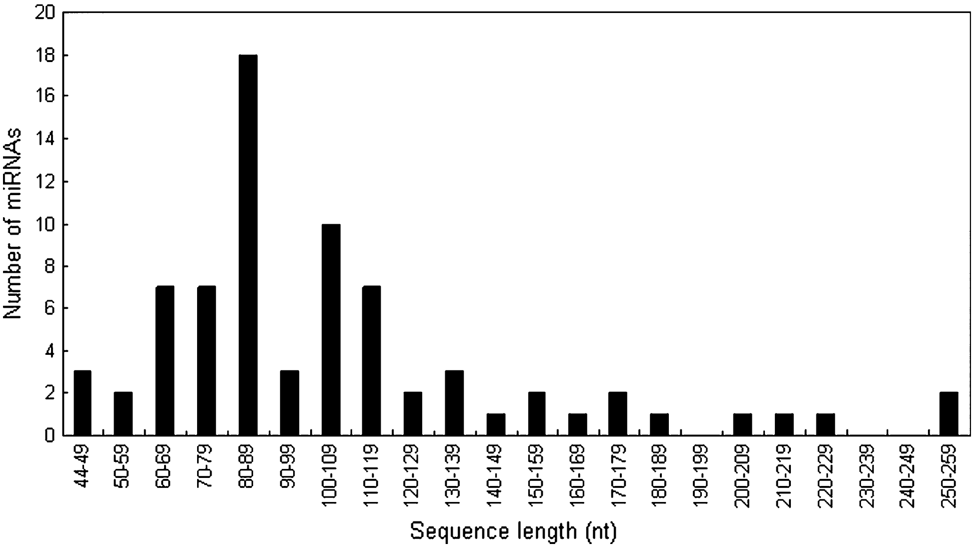


Fig. 3 Size distribution of pre-miRNAs in soybean

modification of the MFE calculation that we refer to as the adjusted MFE (AMFE). The adjusted MFE represents the MFE of an RNA sequence with 100 nt in length. Here, AMFEs of soybean pre-miRNAs ranged from −11.21 to −56.56 kcal/mol with an average of −37.84 ± 11.54 kcal/mol.

Although pre-miRNAs have high negative MFEs and AMFEs, several studies demonstrated that there is no significant difference between pre-miRNAs and other non-coding/coding RNAs (Zhang et al. 2006d). To better distinguish miRNAs from other RNAs, we developed a new criterion of miRNAs, called the minimal folding free energy index (MFEI), which combines three important parameters of RNAs: MFE, sequence length and G and C nucleotide content. A recent study demonstrated that the MFEI of miRNA precursor sequences was significantly higher than that of other RNAs including other small RNAs and mRNAs; a candidate RNA sequence is more likely to be an miRNA when the MFEI is greater than 0.85 (Zhang et al. 2006d). Now, MFEI is being adopted as an important criterion for distinguishing miRNAs from the other RNAs. In this study, the average MFEIs of the 74 soybean pre-miRNAs was 0.86; a majority of the identified soybean miRNAs have an MFEI value higher than 0.85.

The miRNA clusters in soybean

Five miRNA clusters were identified in soybean. These miRNA clusters are gma-miR166a-miR-166b cluster, gma-miR166f-miR-166 g cluster, gma-miR169c-miR-169g* cluster, gma-miR169b-miR-169d cluster, and the gma-miR171a-miR-171d cluster, which contains 10 different miRNAs. Each of these miRNA clusters was found in at least one EST sequence. More interestingly, we also found that the miR169c-miR169g* cluster exists in the wild soybean species *G. soja* and this cluster is highly conserved between *G. max* and *G. soja*. The clustered miRNAs represent 16% of the total of the identified miRNAs in

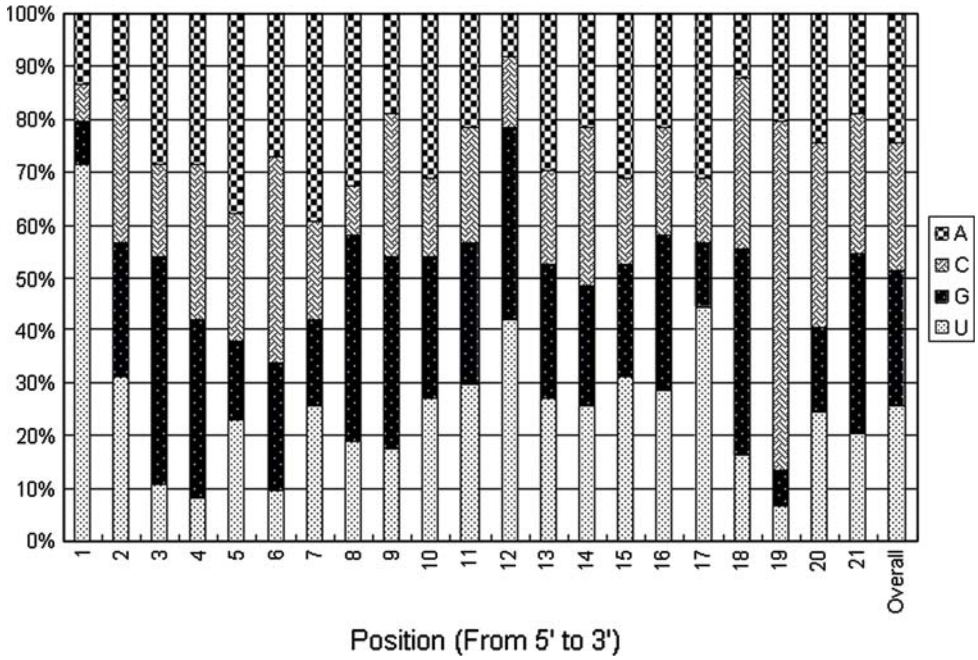


Fig. 4 Position-specific nucleotide preferences in soybean mature miRNAs. The percentage distribution of individual nucleotides at each position numbered 1–21 are designated as A, *checkered boxes*; C, *cross woven lines*; G, *black bars with white stippling*; and U, *white bars with black stippling*

Table 2 Major characteristics of the identified soybean pre-miRNAs

Characteristic	Minimal	Maximum	Median	Average	Stand derivation
Sequence length (nt)	44	259	92	105.7	45.4
A (%)	15.48	41.34	25.80	26.06	5.45
C (%)	11.48	30.77	19.11	19.93	4.24
G (%)	15.15	35.63	24.45	24.07	4.72
U (%)	17.65	47.54	30.57	29.94	5.33
G + C (%)	30.00	63.22	43.68	44.00	5.89
A + U (%)	36.78	70.00	56.33	56.00	5.89
G/C	0.56	3.00	1.25	1.27	0.41
U/A	0.53	2.90	1.11	1.22	0.43
MFE (−kcal/mol)	6.4	82.8	40.0	39.6	17.29
AMFE (−kcal/mol)	11.21	56.56	40.63	37.84	11.54
MFEI	0.32	1.39	0.87	0.86	0.26

soybean, suggesting that miRNA clusters are relatively common in soybean species. To the best of our knowledge, this is the first report of miR-166 and miR-171 family clusters in plants. In one of our previous studies, we discovered a small miR-169 cluster in cotton that consisted of two miR-166 precursors spaced 155 nt apart and oriented in the same direction (Zhang et al. 2007b). In this study, the same miR-169 cluster was also identified in the soybean (Fig. 5). This suggests that the miR-169 cluster is conserved even among distantly related angiosperm. Of the six soybean miRNA clusters identified in this study, some were encoded within the same precursor sequences whereas others were located on discrete precursors separated by genomic DNA. For example, miR169c and miR-169g* are located within different arms of the secondary hairpin structure of a same miRNA-169 precursor. This intra-premiRNA cluster was also observed in the wild species *G. soja*. Of the remaining 4 miRNA clusters, miRNAs were encoded within different precursors, separated by as few as several to as many as 168 nt. In the cases of those miRNAs that were encoded by different precursors, the mature miRNA encoding region was always located within the same arm of the secondary hairpin structure.

Sense- and antisense-strand miRNAs

Recent studies have shown that miRNAs are transcribed and processed from sense and antisense transcripts derived from the same genomic loci in both invertebrates and vertebrates, including fruit fly and human (Bender 2008; Stark et al. 2008; Tyler et al. 2008). However, to the best of our knowledge, no report has yet demonstrated that antisense miRNAs are also transcribed and processed from the same genomic loci in plant species. In this study, we observed that five pairs of soybean miRNAs are bidirectionally transcribed and processed from the same soybean genomic loci, generating both sense and antisense miRNAs (Fig. 6). The five pairs of soybean miRNAs are miR-157b and miR-157c, miR-157d and miR-157e, miR-162a and miR-162b, miR-396a and miR-396b, and miR-396c and miR-396d. These five pairs of transcripts belong to three distinct miRNA families. Although the EST database for wild soybean species is more limited than that of *G. max*,

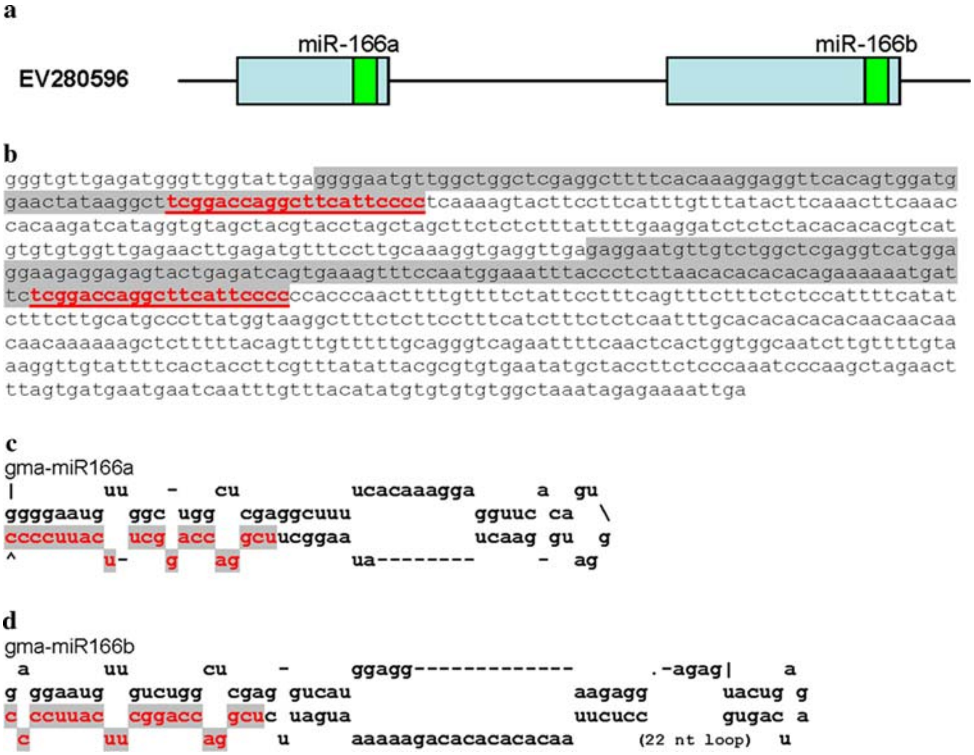

a
miR-166a
miR-166b
EV280596
b
gggtgttgagatgggttggtattgaggggaatgttggctggctcgaggcttttcacaaaggaggttcacagtggatg
gaactataaggcttcggaccaggcttcattcccctcaaaagtacttccttcatttgtttatacttcaaacttcaaac
cacaagatcataggtgtagctacgtacctagctagcttctctctttattttgaaggatctctctacacacacgtcat
gtgtgtggttgagaacttgagatgtttccttgcaaaggtgaggttgagaggaatgttgtctggctcgaggtcatgga
ggaagaggagagtactgagatcagtgaaagtttccaatggaaatttaccctcttaacacacacacagaaaaaatgat
tctcggaccaggcttcattccccccacccaactttgttttctattcctttcagtttctttctctccattttcatat
ctttcttgcatgcccttatggtaaggctttctcttcctttcatctttctctcaatttgcacacacacacaacaacaa
caacaaaaaagctcttttacagtttgtttttgcagggtcagaattttcaactcactggtggcaatcttgttttgta
aaggttgtattttcactaccttcgtttatattacgcgtgtgaatatgctaccttctcccaaatcccaagctagaact
ttagtgatgaatgaatcaatttgtttacatatgtgtgtgtggctaaatagagaaaattga
c
gma-miR166a
| uu - cu ucacaaagga a gu
ggggaaug ggc ugg cgaggcuuu gguuc ca \
ccccuuac ucg acc gcuucggaa ucaag gu g
^ u- g ag ua-------- - ag
d
gma-miR166b
a uu cu - ggagg------------- .-agag| a
g ggaaug gucugg cgag gucau aagagg uacug g
c ccuuac cggacc gcuc uagua uucucc gugac a
c uu ag u aaaaagacacacacacaa (22 nt loop) u

Fig. 5 MiRNA-166a-166b cluster in soybean EST EV280596. **a** Schematic diagram of the organization of the cluster. **b** EST sequence containing the miRNAs encoded within the cluster. *Shadowed sequences* represent pre-miRNAs; *underlined sequences* represent the mature miRNAs. **c** Predicted secondary structure of miR-166a. **d** Predicted secondary structure of miR-166b

we also documented one pair (miR-396a and miR-396b) of these sense-antisense miRNAs in the wild soybean species, *G. clandestine*. We also observed that the *G. clandestine* pair has several nucleotide substitutions relative to the *G. max* pair, indicative that these miRNAs have evolved distinctly in these two lineages and may be widely distributed among soybean and other plant species (Fig. 7).

In animals, antisense miRNAs have at least one nucleotide difference in their seed regions compared to their partners (Bender 2008; Stark et al. 2008; Tyler et al. 2008), suggesting that antisense miRNAs may play a different function in the regulation of target genes. In our study, we observed the same phenomenon in plants. Although the sense and antisense miRNAs are transcribed from a same miRNA locus at a same genome location, the mature products of sense and antisense miRNAs are not identical. All identified pairs of sense/antisense miRNAs have 1–3 nucleotide differences in relation to their anti-sense partners (Table 3). Further, a majority of the nucleotide changes occurred within the miRNA seed region. For example, sense and antisense miR-157 and miR-162 have one nucleotide difference at position 10 or 11, which is required for a specific mRNA target (Schwab et al. 2005). This suggests that antisense miRNAs may target different genes or function through a different mechanism in the plant kingdom. These antisense miRNAs may contribute to the functional diversification of miRNA genes in plant growth and development.

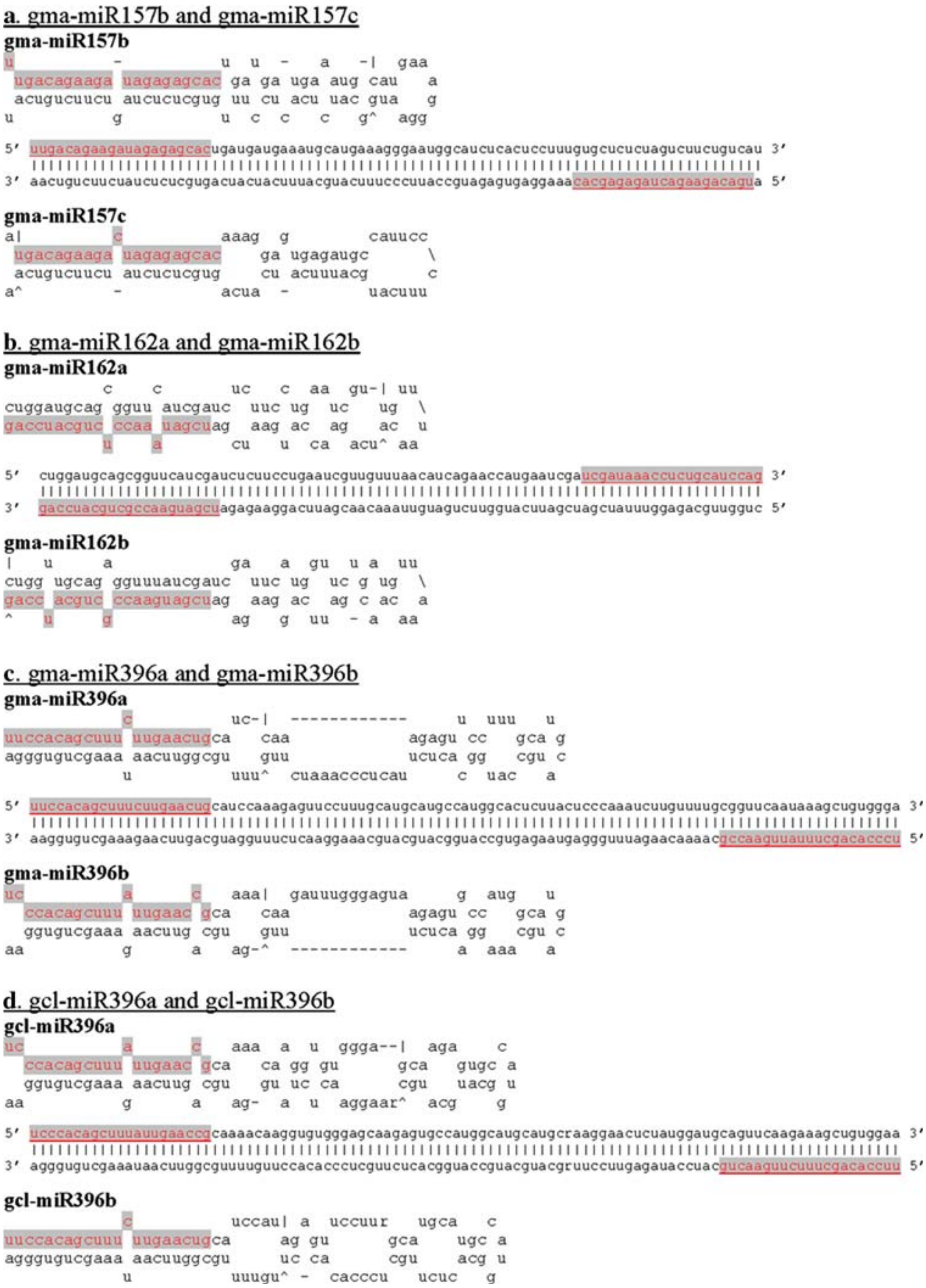
Expression of soybean miRNAs in public EST database

ESTs are partial sequences derived from transcripts. Although ESTs in available databases are not representative of transcribed DNA in all plant tissues or culture conditions, this bias could influence conclusions concerning the expression pattern of a specific gene, such as a miRNA, which is found in the EST database. However, mining EST databases in a systematic way could provide evidence that miRNAs found in a specific EST is expressed in a specific tissue. After mining 394,370 soybean ESTs in NCBI databases, we found that at least one EST contains the identified miRNA precursor sequences in different soybean tissues, including leaf, flower, root, stem, hypocotyl, cotyledon, seedling and somatic embryo (Table 4). This suggests that these miRNAs are expressed in a specific soybean tissue.

RT-PCR assay of putative soybean miRNAs

Stem-loop RT-PCR is a reliable method for detecting the expression of mature miRNAs, and it can distinguish miRNAs that vary in sequence from one another by as little as a single base pair (Chen et al. 2005). In this study, we used the unique primers, designed by the Applied Biosystems, to detect specific soybean mature miRNAs identified using in silico EST analysis. The miRNAs used in expression studies included miR-156, miR-157, miR-159,

Fig. 6 Sense and antisense miRNAs and their corresponding secondary structures


a. gma-miR157b and gma-miR157c
gma-miR157b
u - u u - a -| gaa
ugacagaaga uagagagcac ga ga uga aug cau a
acugucuucu aucucucgug uu cu acu uac gua g
u g u c c c g^ agg
5′ uugacagaagauagagagcacugaugaugaaaugcaugaaagggaauggcaucucacuccuuugugcucucuagucuucugucau 3′
3′ aacugucuucuaucucucgugacuacuacuuuacguacuuucccuuaccguagagugaggaaacacgagagaucagaagacagua 5′
gma-miR157c
a| c aaag g cauucc
ugacagaaga uagagagcac ga ugagaugc \
acugucuucu aucucucgug cu acuuuacg c
a^ - acua - uacuuu
b. gma-miR162a and gma-miR162b
gma-miR162a
c c uc c aa gu-| uu
cuggaugcag gguu aucgauc uuc ug uc ug \
gaccuacguc ccaa uagcuag aag ac ag ac u
u a cu u ca acu^ aa
5′ cuggaugcagcgguucaucgaucucuuccugaaucguuguuuaacaucagaaccaugaaucgaucgauaaaccucugcauccag 3′
3′ gaccuacgucgccaaguagcuagagaaggacuuagcaacaaauuguagucuugguacuuagcuagcuauuuggagacguugguc 5′
gma-miR162b
| u a ga a gu u a uu
cugg ugcag gguuuaucgauc uuc ug uc g ug \
gacc acguc ccaaguagcuag aag ac ag c ac a
^ u g ag g uu - a aa
c. gma-miR396a and gma-miR396b
gma-miR396a
c uc-| ------------ u uuu u
uuccacagcuuu uugaacugca caa agagu cc gca g
agggugucgaaa aacuuggcgu guu ucuca gg cgu c
u uuu^ cuaaacccucau c uac a
5′ uuccacagcuuucuugaacugcauccaaagaguuccuuugcaugcaugccauggcacucuuacucccaaaucuuguuuugcgguucaauaaagcugugga 3′
3′ aaggugucgaaagaacuugacguagguuucucaaggaaacguacguacgguaccgugagaaugaggguuuagaacaaaacgccaaguuauuucgacacccu 5′
gma-miR396b
uc a c aaa| gauuugggagua g aug u
ccacagcuuu uugaac gca caa agagu cc gca g
ggugucgaaa aacuug cgu guu ucuca gg cgu c
aa g a ag-^ ------------ a aaa a
d. gcl-miR396a and gcl-miR396b
gcl-miR396a
uc a c aaa a u ggga--| aga c
ccacagcuuu uugaac gca ca gg gu gca gugc a
ggugucgaaa aacuug cgu gu uc ca cgu uacg u
aa g a ag- a u aggaar^ acg g
5′ ucccacagcuuuauugaaccgcaaaacaaggugugggagcaagagugccauggcaugcaugcraaggaacucuauggaugcaguucaagaaagcuguggaa 3′
3′ agggugucgaaauaacuuggcguuuuguuccacacccucguucucacgguaccguacguacgruuccuugagauaccuacgucaaguucuuucgacaccuu 5′
gcl-miR396b
c uccau| a uccuur ugca c
uuccacagcuuu uugaacugca ag gu gca ugc a
agggugucgaaa aacuuggcgu uc ca cgu acg u
u uuugu^ - cacccu ucuc g

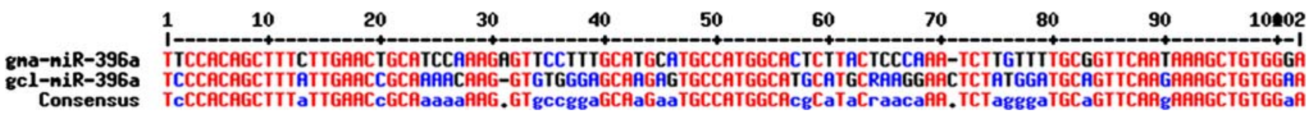

1 10 20 30 40 50 60 70 80 90 10102
gma-miR-396a TTCCACAGCTTTCTTGAACTGCATCCAAAGAGTTCCTTTGCATGCATGCCATGGCACTCTTACTCCCAAA-TCTTGTTTTGCGGTTCAATAAAGCTGTGGGA
gcl-miR-396a TCCCACAGCTTTATTGAACCGCAAAACAAG-GTGTGGGAGCAAGAGTGCCATGGCATGCATGCRAAGGAACTCTATGGATGCAGTTCAAGAAAGCTGTGGAA
Consensus TcCCACAGCTTTaTTGAACcGCAaaaaAAG.GTgccggaGCAaGaaTGCCATGGCAcgCaTaCraacaAA.TCTagggaTGCaGTTCAAgAAAGCTGTGGaA

Fig. 7 Alignment of soybean miR-396a with its corresponding antisense miR-396a* obtained from the cultivated soybean species *Glycine max* and the wild species *Glycine clandestine*. miR-396a and miR-396 are derived from the same genetic locus, see details in Fig. 6. This figure shows that miR-396 is highly conserved in soybean and its wild species. This suggests that antisense is also conserved in plant kingdom

miR-166, miR-169, miR-172, and miR-396. The qRT-PCR analyses demonstrated that all miRNAs were expressed in soybean seedlings. Based on the threshold cycle (C_T), we also observed the expressed level of a specific miRNA. A difference of one C_T unit represents a two-fold difference in the amount of expression. Analyzing the results from

Table 3 Sense miRNAs and their antisense partners

miRNA pair	miRNA	Species	Mature sequence	Length (nt)
1	gma-miR157b	Glycine max	UUGACAGAAGAUAGAGAGCAC	21
	gma-miR157c	Glycine max	UGACAGAAGA*C*UAGAGAGCAC	21
2	gma-miR157d	Glycine max	UUGACAGAAGAUAGAGAGCAC	21
	gma-miR157e	Glycine max	UGACAGAAG*U*AUAGAGAGCAC	21
3	gma-miR162a	Glycine max	UCGAUAAACCUCUGCAUCCAG	21
	gma-miR162b	Glycine max	UCGAU*G*AACC*G*CUGCAUCCAG	21
4	gma-miR396a	Glycine max	UUCCACAGCUUUCUUGAACUG	21
	gma-miR396b	Glycine max	U*C*CCACAGCUUU*A*UUGAAC*C*G	21
5	gma-miR396c	Glycine max	UUCCACAGCUUUCUUGAACUU	21
	gma-miR396d	Glycine max	U*C*CCACAGCUUUCUUGA*G*CUU	21
6	gcl-miR396a	Glycine clandestina	UCCCACAGCUUUAUUGAACCG	21
	gcl-miR396b	Glycine clandestina	U*U*CCACAGCUUU*C*UUGAAC*A*G	21

Italic letters indicate the nucleotide substitutions compared with the miRNAs transcribed from the complementary strand

Table 4 Number of ESTs containing a specific miRNA precursor

miRNAs	Total number of EST	Flower	Root	Leaf	Stem	Hypocotyl	Cotyledon	Seedling	Seed	Somatic embryo	Mixed tissues
156	4	2	1							1	
157	12	1	3				2	3	1		2
159	7			6			1				
160	4			1			1	1	1		
162	4					2		2			
163	4				1				1		2
166	21		4	1			1	7		2	6
167	7						1	2	2		2
168	3	1							1		1
169	21		2				12	6			1
171	5						3				2
172	19	1	6	1			1	5			5
319	8		1			6					1
394	3					1		1	1		
395	1						1				
396	12		1	2	4			1		2	2
398	3						2	1			
399	1		1								
408	9		1	4				2			2
414	12	1	3	2				1	1	2	2
415	8		3								5
426	8	1	1	1		1			1		3
447	4		1					2	1		
779	1		1								
781	1							1			
824	3		3								
825	7					1		3		1	2
830	5		1					3			1
854	7							1	2	1	3
860	12	1	3	1	1			2	1	2	1
862	8	1				1		1			5
865	2									2	
869	3							1	1		1

qRT-PCR, we observed that the expression level of miRNAs differ from each other in soybean seedlings. For example, miR-156 is expressed much higher level than miR-172. The C_T for miR-156 is 17.92 ± 0.11 awhile that is 22.88 ± 0.11 for miR-172, suggesting that the expression level of miR-156 is about 32 fold higher than the expression level of miR-172 (Fig. 8). The expression patterns observed for miR-156 and 172 appear to be related to their different functions. In *Arabidopsis* and rice, miR-156 regulates leaf development by targeting Squamosa-promoter binding protein-like protein (SBP) transcription factors (Rhoades et al. 2002; Schwab et al. 2005). Overexpression of miR-156 resulted in enhanced leaf initiation and produced bushier plants in *Arabidopsis* (Schwab et al. 2005). By comparison, miR-172 controls flower development and phase change from vegetative growth to reproductive growth by inhibiting the protein translation of an A class gene, apetal 2 (*ap2*) transcriptional factor (Aukerman and Sakai 2003; Chen 2004), which controls the timing of flower development and morphology (Lohmann and Weigel 2002). During plant vegetative growth, the expression level of miR-172 is very low but increases significantly immediately prior to flowering and reaches peak expression levels during the flowering period (Aukerman and Sakai 2003; Chen 2004). Overexpression of miR-172 inhibits the translation of the *ap2* and *ap2*-like genes, which results in early flowering and disruption of floral organ identity (Aukerman and Sakai 2003; Chen 2004). Thus, it is not difficult to understand that miR-172 is expressed at lower level in soybean seedling compared to miR-156. This also suggests that miR-172 may also regulate soybean flower development and phase change from vegetative growth to reproductive growth and points to the possibility of using miRNAs to modulate phase change to influence soybean yield.

Soybean miRNA targets

miRNAs regulate gene expression posttranscriptionally. miRNAs bind to the targeted mRNAs within the 3′ untranslated region (3′ UTR) or coding region of transcribed mRNAs and promote mRNA cleavage or translation repression. Usually, there are no more than four mismatches between miRNAs and their targeted mRNAs in plants (Rhoades et al. 2002; Schwab et al. 2005). The sequence relationship between miRNAs and their mRNA targets has been used successfully to identify miRNA targets in plants by performing BLASTn searches using

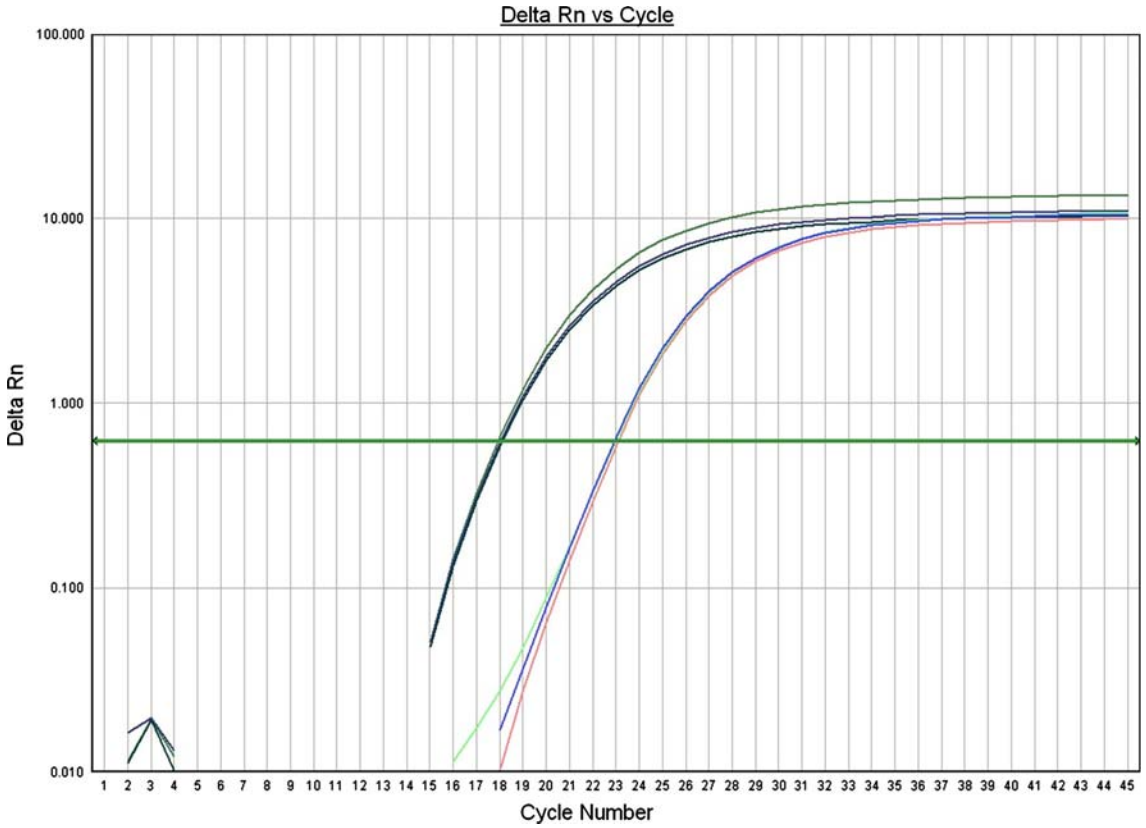


Fig. 8 Amplification plot of soybean miR-156 (*left*) and miR-172 (*right*). The same amount of cDNA was added to each qRT-PCR analysis. Each miRNAs-dependent reaction was repeated three times

putative miRNAs as query sequences to search NCBI mRNA databases or other mRNA databases.

BLASTn results conducted as described in Methods revealed a total of 152 potential miRNA targets in soybean protein-coding sequence databases. At least one targeted mRNA was identified for each of the soybean miRNA families except miR-394, miR-399, miR-426, miR-862 and miR-865 (Table 5, Fig. 9). These 152 potential miRNA targets belong to several gene families with diverse biological functions. Among the pool of mRNA targets, the majority are transcriptional factors, whereas others are associated with plant metabolism and response to environmental stress. These results are similar to those reported previously for other plant species, such as *Arabidopsis*, rice and corn (Rhoades et al. 2002; Bonnet et al. 2004a; Zhang et al. 2006a).

Transcriptional factors are an important component in transcriptional process. Transcriptional factors usually bind to a specific DNA sequence and control genetic information transfer from DNA to RNA. In this study, among the pool of mRNA targets, we found that nearly 40% of the miRNA families appear to target transcriptional factor-encoding mRNAs. Based on functional studies conducted in *Arabidopsis* and rice, it has been demonstrated that a number of the miRNAs we identified in soybean act on transcription factors that regulate plant development (Rhoades et al. 2002; Zhang et al. 2006c). These plant development regulators include miR-171, miR-172, miR-156/157, and miR-166, which control diverse developmental functions ranging from control of floral morphology and flowering time (miR-172) to leaf and root development (miR-156/157 and miR-171) (Zhang et al. 2006c). AP2 is one of the class A gene transcriptional factors that play an important role in floral morphology and flowering time. Results from *Arabidopsis* and maize show that miR-172 targets both AP2 in *Arabidopsis* (Aukerman and Sakai 2003; Chen 2004) and AP2-like gene glossy15 (gl15) in maize (Lauter et al. 2005; Zhang et al. 2006a). Similar, in this study, we found that miR-172 targets the *ap2* gene in soybean, which suggests that the function of miR-172 is highly conserved among plants. The scarecrow-like (SCL) (GRAS domain) family is a class of plant-specific transcription factors, which control a wide range of plant developmental process. Our study demonstrated that one member of SCL family, SCL6, is a potential target of miR-171. Consistent with our interpretation, functional studies in *Arabidopsis* have shown that miR-171 targets SCL6 expression predominantly in inflorescence and floral tissues (Llave et al. 2002; Reinhart et al. 2002). This suggests that miR-171 may play a role in soybean flower development. Both HD-ZIP III and SQUAMOSA promoter binding protein-like protein (SBP) transcription factors are important for leaf development. Functional studies have shown that miR-166 and miR-156/157 control leaf development via targeting these two classes of transcriptional factors in *Arabidopsis*, rice and maize (Juarez et al. 2004; Mallory et al. 2004). Here, we also identified that these two classes of transcriptional factors are potential targets of soybean miR-156/157 and miR-165/166 families. Several other soybean miRNAs also target specific transcriptional factors, such as miR-169 targets CCAAT-binding (CBF) transcription factor; miR-447 targets TCP transcriptional factor, one important transcriptional factor controlling leaf development. Auxin response factors (ARFs) are a class of transcriptional factors, which play a role in plant signaling transduction and root development. In this study we found miR-160 perfectly match to ARF 10 mRNAs and this binding site is highly evolutionary conserved in several plant species, including *Arabidopsis*, rice, maize and soybean.

Soybean miRNAs may also target transcriptional activators. We found that miR-168 and miR-169 target GRF1-interacting factor 1 (GIF1) and HSP2, respectively. In *Arabidopsis*, GIF1 is a transcriptional coactivator, which regulates the growth and shape of leaves and petals (Kim and Kende 2004). To our best knowledge, we have not yet seen a report indicating that miRNA is also a target transcriptional activator. This will enhance our knowledge of miRNA-mediated gene regulation.

Another major class of soybean miRNA targets are genes that control biological processes and that respond to environmental condition. Nucleosomes are the fundamental units in eukaryotic chromatin. Nucleosome assembly is a complicated biological process, and it required nucleosome assembly protein (NAP). This study demonstrated that miR-414 perfectly complementary with soybean NAP-1 mRNA, and the binding site is conserved from species to species in plant kingdom, including *Arabidopsis*. This suggests that miR-414 may play an important role in regulating nucleosome assembly.

Several studies demonstrated that several miRNA are involved in response to a various environmental stress, including cold, salinity, drought and nutritional deficiency (Jones-Rhoades and Bartel 2004; Sunkar and Zhu 2004; Lu et al. 2005; Zhang et al. 2005; Chiou 2007). Here, we found that miR-159, miR-398, miR-414 potentially target GST, SOD and cytochrome C reductase mRNAs, which all play an important role in plant responses to different biotic and abiotic environmental stresses.

Discussion

Although miRNAs have been extensively studied in the past several years, no systematic study has been performed on soybean, one of the most important crops around world.

Table 5 Potential targets of the identified miRNAs in soybean

miRNA	Targeted protein	Target function	Targeted genes or EST homologs of genes in other plant species[a]
156/157	Squamosa-promoter binding protein-like protein (SBP)	Transcription factor	TC209333 (1,1,Y)
	Conserved protein with CBS domain	Metabolism	TC210466 (1,2), TC211412 (1,2)
	Unknown		TC220074 (0,0), TC214166 (1,1), TC214342 (1,1), BE329522 (1,2), BI470983 (1,2,Y), TC220887 (1,2), TC221679 (1,2), TC232376 (1,2,Y), TC232420 (1,2), TC220034 (2,2), AW307452 (2,3), TC208172 (2.5,2)
159	GST	Metabolism and stress response	TC217088 (3,4)
	Unknown		AW705254 (2.5,2), CO984960 (1.5,3), CO984960 (1.5,4)
160	Auxin response factor 10	Signaling transduction	BM887596 (0,0,Y), TC208983 (0,0,Y)
	Unknown		BG651292 (0.5,1), BI427336 (0.5,1), TC213894 (0.5,1)
162	60S ribosomal protein L5		TC214109 (3,4), TC214179 (3,4), TC214267 (3,4), TC214287 (3,4)
	Unknown		TC222322 (0,0), AW459538 (3,4)
163	Unknown		BE800773 (2,4), BI971358 (2.5,2),TC209353 (2.5,3)
166	Class III HD-Zip protein	Transcription factor	TC230399 (1.5,3,Y)
	T01 homeodomain transcription factor (ATHB-14)	Transcription factor	BM309730 (2,4,Y)
	PHAVOLUTA-like HD-ZIPIII protein	Transcription factor	TC221756 (2,4,Y)
167	Auxin response factor 8	Transcription factor	TC207358 (2.5,3,Y), TC228776 (2.5,3,Y)
	Ribulose-1,5-bisphosphate carboxylase small subunit rbcS1	Metabolism	TC210116 (3,2)
	Unknown		CD411229 (0,0), TC221608 (0,0), TC218614 (3,4), TC203512 (3,4)
168	GRF1-interacting factor 1 (GIF1)	Transcriptional coactivator	TC226538 (3,4)
	Unknown	TC23387 (1.5,3,Y)	BG882680 (3,3)
169	RAPB protein	Transcription factor	
	CCAAT-box transcription factor complex WHAP12	Transcription factor	TC220009 (0,1,Y), TC232566 (1.5, 2)
	HAP2	Transcriptional activator	TC234410 (2,2,Y)
	Putative zinc fingers with GTPase activating proteins	Metabolism	TC217446 (2,4)
	Unknown		TC221767 (0,0), CO985073 (2.5,4,Y), BQ611496 (2.5,4)
171	Scarecrow-like 6 (SCL6)	Transcription factor	TC217781 (1,2,Y)
	Unknown		CO979466 (0,0), TC233437 (2,2,Y), CA937914 (3,4)
172	PHAP2B protein	Transcription factor	TC205405 (0,0,Y), TC205406 (0,0,Y), TC205407 (0,0,Y), TC205409 (0,0,Y), TC228902 (0,0,Y), TC228901 (0,1)
	APETALA2 protein	Transcription factor	BE659941 (0,0), TC228507 (3,3,Y)
	Transcription factor AHAP2	Transcription factor	TC217856 (0,0), TC217857 (0,0), TC217858 (0,0), TC217856 (0,1), TC217857 (0,1), TC217858 (0,1)
	A receptor protein kinase	Metabolism	BM142828 (2.5,4,Y), TC229891(2.5,4,Y)
	Unknown		TC208557 (1,2), TC221559 (3,3), BI320499 (0.5,2), BU084569 (0.5,2)
319	Unknown		CO984960 (1,1), BQ453148 (1.5,3), BE475558 (2,4), TC232126 (2,4)
395	ATP sulfurylase	Metabolism	TC204560 (1,2,Y), TC218491 (1,2,Y),
	Unknown		TC223520 (2.5,4), TC229234 (3,4)

Table 5 continued

miRNA	Targeted protein	Target function	Targeted genes or EST homologs of genes in other plant species[a]
396	Cysteine proteinase precursor	Metabolism	TC214755 (3,4,Y), TC214756 (3,4,Y), TC206710 (2.5,4,Y)
	Leucine-rich receptor-like protein kinase	Metabolism	BU760718 (2.5,4,Y)
	Unknown		TC223232 (0,0), TC232810 (2,4), BE657787 (2.5,4), TC231667 (1.5,3)
398	Superoxide dismutase	Stress response	TC204404 (2.5,2), TC204403 (2.5,3)
	Unknown		BG155619 (2,2)
408	Basic blue copper protein		TC204673 (2.5,4,Y), TC204674 (2.5,4,Y), TC225573 (0.5,2,Y), BF424264 (2.5,4,Y), TC225574 (0.5,2,Y),
	Uclacyanin 3-like protein		TC214395 (2,3,Y), TC214067 (2,3,Y)
	Unknown		CD391647 (0,0)
414	Nucleosome assembly protein 1 (SNAP-1)		TC215449 (0,0,Y), TC215450 (0,0,Y), TC215451 (0.5,2), TC215452 (0.5,2,Y), BU762452 (1,3)
	Nucleolar histone deacetylase HD2-P39		TC227389 (1.5,4)
	Nucleic acid binding protein-like		TC206439 (2,4,Y)
	Ubiquinol-cytochrome C reductase complex 17 kDa protein (Mitochondrial hinge protein)		TC207432 (2.5,4,Y)
	Unknown		TC204794 (1,3), TC230134 (1.5,3), TC231897 (1.5,3), BM891815 (2,3), TC219323 (2,3), TC219848 (2,3), TC219849 (2,3), TC228030 (2,3), TC204848 (2,4,Y), TC203448 (2,4), TC210126 (2,4), TC211437 (2,4), TC225279 (2,4), TC206810 (2.5,3), TC233319 (2,4), TC206811 (2.5,3), TC217772 (2.5,3,Y)
415	Golgi SNARE 12 protein (AtGOS12) (Golgi SNAP receptor complex member 1–2)		TC217990 (2,2)
	Unknown		TC216645 (2,3), BI975069 (3,2)
447	TCP family transcription factor-like	Transcriptional factor	TC205922 (3,3)
779	Unknown		BM954493 (3,4), CO982525 (3,4), TC235136 (3,4)
781	Unknown		BM887508 (3,4), TC221634 (3,4)
824	Vacuolar H + -ATPase B subunit	Metabolism	TC214816 (2.5,4)
825	Fasciclin-like AGP 11		TC217021 (3,3)
	GPAA1-like protein		BU760697 (3,4)
	Unknown		BI971529 (3,3)
830	Zinc finger protein	Transcription factor	TC214744 (3,4)
	Unknown		TC224967 (1,2), TC213605 (3,4),
854	Unknown		TC207726 (1.5,2), BU082412 (1.5,3), TC222865 (2,4)
860	Unknown		TC222745 (2.5,3), CD391541 (3,4)
869			BE657248 (0,0), TC210551 (2.5,4)

[a] The numbers in brackets stand for the score and mismatch number of miRU. Y means conserved in *Arabidopsis*

Recently, two groups identified several miRNAs from soybean by computational and direct cloning approaches (Zhang et al. 2005, 2006b; Subramanian et al. 2008; Sunkar and Jagadeeswaran 2008), but soybean miRNAs still remains largely unknown. This study not only systemically identified 69 miRNAs from 33 families in the domestic soybean and five miRNAs in the soybean wild species *G. soja* and *G. clandestine*, but also revealed several novel miRNA features through comparative genome analysis. These results demonstrate that miRNAs are common in soybean. The potential miRNA targeted genes are involved in the development, metabolic processes, and stress response. This is similar to other plant species (Rhoades et al. 2002; Bonnet et al. 2004a; Zhang et al. 2006a).

Fig. 9 Predicted miRNA targets and their complementary sites within defined mRNAs. Each *bottom strand* shows the miRNA sequence, and each *top strand* shows the corresponding complementary site within specific mRNA targets. Watson–Crick pairing (*vertical dashes*) and G:U wobble pairing (*circles*) are indicated

```
SBP (TC209333) (5′-3′) 373 gugcucucucucuucugucaa 393
                           ||||||||| |||||||||||
miR-157a (3′-5′)           CACGAGAGAUAGAAGACAGUU

ARF10 (BM887596) (5′-3′) 319 aggcauacagggagccaggca 339
                             |||||||||||||||||||||
miR-160c (3′-5′)             UCCGUAUGUCCCUCGGUCCGU

PHAVOLUTA-like HD-ZIPIII protein (TC221756) (5’-3’) 87 uugggaugaagccugguccgg 107
                                                       ||°|||||||||||||||||°
miR-166 (3’-5’)                                        CCCCUUACUUCGGACCAGGCU

WHAP12 (BC220009)(5′-3′) 786 aggcaacucauccuuggcucg 806
                             ||||||||||||||||||||°
miR-169e (3′-5′)             UCCGUUGAGUAGGAACCGAGU

SCL6 (TC217781)(5′-3′) 477   ugggauauuggcgcggcucaa 497
                              | ||||||||||||||||||
miR-169e (3′-5′)             UCACUAUAACCGCGCCGAGUU

APETALA2 (BE659941(5′-3′) 354   cugcagcaucaucaggauucc 374
                                |||||||||||||||||||||
miR-172 (3′-5′)                 GACGUCGUAGUAGUCCUAAGG

Basic blue copper protein (TC225573)(5′-3′) 80  cucagggaagaggcagugcau 100
                                                |||||||||||||||||||||
miR-408 (3′-5′)                                 CGGUCCCUUCUCCGUCACGUA

SNAP-1 (TC215449)(5′-3′) 1048 ugacgaggaugaugaagauau 1068
                              |||||||||||||||||||||
miR-414 (3′-5′)               ACUGCUCCUACUACUUCUAUA
```

More importantly, the occurrence of antisense miRNAs in plants was firstly reported in this study. Several recent studies observed that miRNAs can be transcribed from both sense and antisense strands of DNA in animals (Bender 2008; Stark et al. 2008; Tyler et al. 2008). However, plant antisense miRNAs has not been reported. The discovery of antisense miRNAs provides a new insight to miRNAs biogenesis and function. Five pairs of antisense miRNAs and their corresponding sense miRNAs have been identified in this study. These antisense miRNAs are also conserved in wild species of soybean. This suggests that antisense miRNAs may widely exist in plants. A majority of miRNA genes are transcribed by RNA polymerase (Pol II) following a common processing pathway (Lee et al. 2004). The miRNA gene is transcribed into a long primary miRNA sequence (pri-miRNA), which undertakes a series of subsequent processing events to generate a miRNA precursor (pre-miRNA) and finally a mature miRNA. When considering that sense and antisense miRNAs are complementary strands within the same genomic loci, several questions regarding the transcription of antisense miRNAs arise. How are these sense miRNAs and their partner antisense miRNAs transcribed? Are both sense and antisense miRNAs transcribed simultaneously or at different times? Are sense and antisense miRNAs transcribed following the same or independent mechanisms? If both sense and antisense miRNA are transcribed simultaneously, then what happens when the two transcriptional complexes encounter one another, i.e. how would the RNA Pol II complexes bypass each other when they meet?

The second major novel finding is that the occurrence of miRNA clusters is much higher in soybean than that in other plant species. It is common that animal miRNAs are clustered, although the significance of miRNA clustering is uncertain (Tanzer and Stadler 2004; Altuvia et al. 2005). However, miRNA clustering appears to be less common among plants. Only few miRNA clusters have been found in plants (Jones-Rhoades and Bartel 2004; Talmor-Neiman et al. 2006; Zhang et al. 2006b, 2007b). For example, Talmor-Neiman et al. (2006) observed that two miRNAs (miR-1219a and miR-1219b) were located within approximately 200 bp of each other. Surprisingly, we identified five miRNA clusters in soybean. The frequency is much higher than that in other plant species. Of the five miRNA clusters, miR-166 and miR-171 family clusters have not been reported earlier. It should be noted that these miRNA clusters were not observed in model plant species, such as *A. thaliana* and rice, suggesting that selected clustered plant miRNAs may evolve in a lineage-specific manner. In animals, many miRNA clusters are formed via miRNA duplication (Tanzer and Stadler 2004). Plant miRNA clusters may have been generated by a similar mechanism. Supporting evidence includes some miRNA clusters in *Arabidopsis* appear to have evolved via miRNA duplication (Allen et al. 2004; Maher et al. 2006). However, the mechanisms that drive miRNA clustering and miRNA cluster evolution are unclear. The study of clustered animal miRNAs has shown that clustered miRNAs have similar gene expression patterns and are transcribed together in a polycistronic manner. This indicates that common regulatory control may be a significant force in the maintenance of miRNA clustering (Tanzer and Stadler 2004; Altuvia et al. 2005). Additionally, there are as much as 40 miRNAs clustered together in animals whereas plant miRNA clusters only contained a small number of miRNAs. This difference suggests that the mechanisms driving miRNA evolution in animals and plants may be different.

Another interesting finding is that the domination of U and C bases at the first and 19th positions from the 5′ ends of the mature miRNAs. Although it has been reported that U is dormant at the first position (Zhang et al. 2006b), no study has reported that C base dominates the 19th position. It is unclear why U and C are dominated at these two positions. Further study on this phenomenon may allow a better understanding of the mechanism of miRNA biogenesis and function.

Conclusions

Our results from comparative genome-based in silico screening of soybean EST databases using *Arabidopsis* miRNAs and quantitative real time PCR (qRT-PCR) provide evidence for 69 miRNAs belonging to 33 families, with an additional five miRNAs identified in two wild soybean species. Based on sequence comparisons among the soybean miRNAs, we conclude that their precursors (pre-miRNAs) vary in length from 44 to 259 with an average of 106 ± 45 nt and that these precursors include previously unidentified clustered miRNAs as well as sense and antisense miRNAs, which have not been observed in plants. Further, comparative sequence analyses of the distribution of individual bases at each nucleotide position within mature soybean miRNAs revealed that uracil is the dominant nucleotide at position one (5′ most) while cytosine is the dominant nucleotide in position 19, which suggests that these two nucleotides may play an important role in miRNA biogenesis and/or miRNA-mediated gene regulation. miRNA-specific qRT-PCR analyses of RNA samples prepared from soybean seedlings revealed that miRNAs are differentially expressed both quantitatively and qualitatively in soybean tissues. Identification of putative miRNA-targeted mRNAs indicate that the targeted mRNAs include a large percentage (nearly 40%) that are transcription factors as well as sequences that encode genes that play a role in signal transduction and stress response. Overall, our results showed the extensive evolutionary conservation of miRNAs in plant species with an apparent significant increase in the number of clustered and antisense miRNAs in soybeans compared to previously studied plants.

Acknowledgments This work was supported by East Carolina University New Faculty Research Startup Funds Program (BZ), Western Illinois University New Faculty Research Startup Funds Program (XP), and a Science and Engineering Grant from DuPont (BZ). We would like to thank Dr. Joseph Burton of ARS, USDA for kindly providing soybean seeds. We also appreciate Dr. Brian McGonigle of DuPont for his support and encouragement.

References

Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, Kerlavage AR, Mccombie WR, Venter JC (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. Science 252:1651–1656

Allen E, Xie ZX, Gustafson AM, Sung GH, Spatafora JW, Carrington JC (2004) Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. Nat Genet 36:1282–1290

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Altuvia Y, Landgraf P, Lithwick G, Elefant N, Pfeffer S, Aravin A, Brownstein MJ, Tuschl T, Margalit H (2005) Clustering and conservation patterns of human microRNAs. Nucleic Acids Res 33:2697–2706

Ambros V, Chen XM (2007) The regulation of genes and genomes by small RNAs. Development 134:1635–1641

Aukerman MJ, 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

Bender W (2008) MicroRNAs in the *Drosophila bithorax* complex. Genes Dev 22:14–19

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

Bonnet E, Wuyts J, Rouze P, Van de Peer Y (2004b) Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. Bioinformatics 20:2911–2917

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

Chen CF, Ridzon DA, Broomer AJ, Zhou ZH, Lee DH, Nguyen JT, Barbisin M, Xu NL, Mahuvakar VR, Andersen MR, Lao KQ, Livak KJ, Guegler KJ (2005) Real-time quantification of microRNAs by stem-loop RT-PCR. Nucleic Acids Res 33:e179

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

Chen XM (2005) microRNA biogenesis and function in plants. FEBS Lett 579:5923–5931

Chiou TJ (2007) The role of microRNAs in sensing nutrient stress. Plant Cell Environ 30:323–332

Floyd SK, Bowman JL (2004) Gene regulation: ancient microRNA target sequences in plants. Nature 428:485–486

Gleave AP, Ampomah-Dwamena C, Berthold S, Dejnoprat S, Karunairetnam S, Nain B, Wang Y-Y, Crowhurst RN, MacDiarmid RM (2008) Identification and characterisation of primary microRNAs from apple (*Malus domestica* cv. Royal Gala) expressed sequence tags. Tree Genet Genomes 4:343–358

Griffiths-Jones S (2004) The microRNA registry. Nucleic Acids Res 32:D109–D111

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

Guo Q, Xiang AL, Yang Q, Yang ZM (2007) Bioinformatic identification of microRNAs and their target genes from *Solanum tuberosum* expressed sequence tags. Chin Sci Bull 52:2380–2389

Jones-Rhoades MW, Bartel DP (2004) Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Mol Cell 14:787–799

Juarez MT, Kui JS, Thomas J, Heller BA, Timmermans MCP (2004) MicroRNA-mediated repression of rolled leaf1 specifies maize leaf polarity. Nature 428:84–88

Kim JH, Kende H (2004) A transcriptional coactivator, AtGIF1, is involved in regulating leaf growth and morphology in *Arabidopsis*. Proc Natl Acad Sci USA 101:13374–13379

Lauter N, Kampani A, Carlson S, Goebel M, Moose SP (2005) MicroRNA172 down-regulates glossy15 to promote vegetative phase change in maize. Proc Natl Acad Sci USA 102:9412–9417

Lee Y, Kim M, Han JJ, Yeom KH, Lee S, Baek SH, Kim VN (2004) MicroRNA genes are transcribed by RNA polymerase II. EMBO J 23:4051–4060

Llave C, Xie ZX, Kasschau KD, Carrington JC (2002) Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. Science 297:2053–2056

Lohmann JU, Weigel D (2002) Building beauty: the genetic control of floral patterning. Dev Cell 2:135–142

Lu SF, Sun YH, Shi R, Clark C, Li LG, Chiang VL (2005) Novel and mechanical stress-responsive microRNAs in *Populus trichocarpa* that are absent from *Arabidopsis*. Plant Cell 17:2186–2203

Maher C, Stein L, Ware D (2006) Evolution of *Arabidopsis* microRNA families through duplication events. Genome Res 16:510–519

Mallory AC, Reinhart BJ, Jones-Rhoades MW, Tang GL, Zamore PD, Barton MK, Bartel DP (2004) MicroRNA control of PHABULOSA in leaf development: importance of pairing to the microRNA 5′ region. EMBO J 23:3356–3364

Mathews DH, Sabina J, Zuker M, Turner DH (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J Mol Biol 288:911–940

Matukumalli LK, Grefenstette JJ, Sonstegard TS, vanTassell CP (2004) EST-PAGE: managing and analyzing EST data. Bioinformatics 20:286–288

Pan XP, Zhang BH, SanFrancisco M, Cobb GP (2007) Characterizing viral microRNAs and its application on identifying new microRNAs in viruses. J Cell Physiol 211:10–18

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

Pasquinelli AE, McCoy A, Jimenez E, Salo E, Ruvkun G, Martindale MQ, Baguna J (2003) Expression of the 22 nucleotide *let-7* heterochronic RNA throughout the Metazoa: a role in life history evolution? Evol Dev 5:372–378

Pasquinelli AE, Reinhart BJ, Slack F, Martindale MQ, Kuroda MI, Maller B, Hayward DC, Ball EE, Degnan B, Muller P, Spring J, Srinivasan A, Fishman M, Finnerty J, Corbo J, Levine M, Leahy P, Davidson E, Ruvkun G (2000) Conservation of the sequence and temporal expression of *let-7* heterochronic regulatory RNA. Nature 408:86–89

Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP (2002) MicroRNAs in plants. Genes Dev 16:1616–1626

Rhoades MW, Reinhart BJ, Lim LP, Burge CB, Bartel B, Bartel DP (2002) Prediction of plant microRNA targets. Cell 110:513–520

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

Stark A, Bushati N, Jan CH, Kheradpour P, Hodges E, Brennecke J, Bartel DP, Cohen SM, Kellis M (2008) A single Hox locus in Drosophila produces functional microRNAs from opposite DNA strands. Genes Dev 22:8–13

Subramanian S, Fu Y, Sunkar R, Barbazuk WB, Zhu JK, Yu O (2008) Novel and nodulation-regulated microRNAs in soybean roots. BMC Genomics 9:160

Sunkar R, Girke T, Jain PK, Zhu JK (2005) Cloning and characterization of MicroRNAs from rice. Plant Cell 17:1397–1411

Sunkar R, Jagadeeswaran G (2008) *In silico* identification of conserved microRNAs in large number of diverse plant species. BMC Plant Biol 8:37

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

Talmor-Neiman M, Stav R, Frank W, Voss B, Arazi T (2006) Novel micro-RNAs and intermediates of micro-RNA biogenesis from moss. Plant J 47:25–37

Tanzer A, Stadler PF (2004) Molecular evolution of a microRNA cluster. J Mol Biol 339:327–335

Tyler DM, Okamura K, Chung WJ, Hagen JW, Berezikov E, Hannon GJ, Lai EC (2008) Functionally distinct regulatory RNAs generated by bidirectional transcription and processing of microRNA loci. Genes Dev 22:26–36

Xie FL, Huang SQ, Guo K, Xiang AL, Zhu YY, Nie L, Yang ZM (2007) Computational identification of novel microRNAs and targets in *Brassica napus*. FEBS Lett 581:1464–1474

Zhang BH, Pan XP, Anderson TA (2006a) Identification of 188 conserved maize microRNAs and their targets. FEBS Lett 580:3753–3762

Zhang BH, Pan XP, Cannon CH, Cobb GP, Anderson TA (2006b) Conservation and divergence of plant microRNA genes. Plant J 46:243–259

Zhang BH, Pan XP, Cobb GP, Anderson TA (2006c) Plant microRNA: a small regulatory molecule with big impact. Dev Biol 289:3–16

Zhang BH, Pan XP, Cox SB, Cobb GP, Anderson TA (2006d) Evidence that miRNAs are different from other RNAs. Cell Mol Life Sci 63:246–254

Zhang BH, Pan XP, Wang QL, Cobb GP, Anderson TA (2005) Identification and characterization of new plant microRNAs using EST analysis. Cell Res 15:336–360

Zhang BH, Pan XP, Wang QL, Cobb GP, Anderson TA (2006e) Computational identification of microRNAs and their targets. Comput Biol Chem 30:395–407

Zhang BH, Wang QL, Pan XP (2007a) MicroRNAs and their regulatory roles in animals and plants. J Cell Physiol 210:279–289

Zhang BH, Wang QL, Wang KB, Pan XP, Liu F, Guo TL, Cobb GP, Anderson TA (2007b) Identification of cotton microRNAs and their targets. Gene 397:26–37

Zhang YJ (2005) MiRU: an automated plant miRNA target prediction server. Nucleic Acids Res 33:W701–W704

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