Available online at www.sciencedirect.com
ScienceDirect
Journal of Integrative Agriculture
2013, 12(9): 1522-1529
September 2013
RESEARCH ARTICLE

MicroRNA Primary Transcripts and Promoter Elements Analysis in Soybean (*Glycine max* L. Merril.)

LI Jing[1*], LIU Yong-xin[1*], HAN Ying-peng[1], LI Yong-guang[1], GUO Mao-zu[2] and LI Wen-bin[1]

[1] *Key Laboratory of Soybean Biology, Ministry of Education/Northeast Agricultural University, Harbin 150030, P.R.China*
[2] *Department of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, P.R.China*

Abstract

The importance of microRNA (miRNA) at the post-transcriptional regulation level has recently been recognized in both animals and plants. In recent years, many studies focused on miRNA target identification and functional analysis. However, little is known about the transcription and regulation of miRNAs themselves. In this study, the transcription start sites (TSSs) for 11 miRNA primary transcripts of soybean from 11 miRNA loci (of 50 loci tested) were cloned by a 5′ rapid amplification of cDNA ends (5′ RACE) procedure using total RNA from 30-d-old seedlings. The features consistent with a RNA polymerase II mechanism of transcription were found among these miRNA loci. A position weight matrix algorithm was used to identify conserved motifs in miRNA core promoter regions. A canonical TATA box motif was identified upstream of the major start site at 8 (76%) of the mapped miRNA loci. Several *cis*-acting elements were predicted in the 2 kb 5′ to the TSSs. Potential spatial and temporal expression patterns of the miRNAs were found. The target genes for these miRNAs were also predicted and further elucidated for the potential function of the miRNAs. This research provides a molecular basis to explore regulatory mechanisms of miRNA expression, and a way to understand miRNA-mediated regulatory pathways and networks in soybean.

Key words: soybean, miRNA, primary transcript, RACE, TATA box, motif, promoter, *cis*-acting element

INTRODUCTION

The miRNAs, approximately 21 nucleotides (nt) long, are a large class of small non-coding RNAs (ncRNAs), which play important roles in post-transcriptional gene expression control. The miRNAs can negatively regulate gene expression through hybridization to completely or partially complementary sequences in target mRNAs. The double stranded (ds) mRNAs are targets for cleavage or inhibition of transport and translation (Voinnet 2009). Mature miRNAs are generated through multiple processing steps from longer primary transcripts (pri-miRNA) that contain imperfect fold back structures (Chen 2010). In animals, miRNA genes are transcribed by RNA polymerase II (Pol II) (Lee *et al*. 2004), yielding a pri-miRNA that is processed initially by the nuclear RNase III-like enzyme Drosha (Lee *et al*. 2003). The resulting pre-miRNA transcripts are transported to the cytoplasm and processed by dicer to yield mature miRNAs (Lee *et al*. 2002; Liu *et al*. 2009). Plant pri-miRNAs are mostly transcribed by RNA polymerase II (Pol II) from regions located between protein-coding genes. The RNA-binding protein DAWDLE (DDL) presumably stabilizes pri-miRNAs for their conversion in nuclear processing centers called D-bodies to stem-loop pre-miRNAs. This reaction entails the concerted action and physical interaction of the C_2H_2-zinc finger

Received 14 September, 2012 Accepted 12 December, 2012
LI Jing, Mobile: 13936282709, E-mail: lijing_101@126.com; Correspondence LI Wen-bin, Tel: +86-451-55190778, E-mail: wenbinli@neau.edu.cn
* These authors contributed equally to this study.

doi:10.1016/S2095-3119(13)60500-6

protein SERRATE (SE), the double-stranded RNA-binding protein HYPONASTIC LEAVES1 (HYL1), Dicer-like 1 (DCL1), and the nuclear cap-binding complex (CBC). Pre-miRNAs, or mature miRNAs produced by DCL1, are then exported to the cytoplasm possibly through the action of the plant exportin 5 ortholog HASTY and other unknown factors. Mature RNA duplexes excised from pre-miRNAs (miRNA/miRNA*, where miRNA is the guide strand and miRNA* is the degraded strand) are methylated by HEN1, a reaction that protects them from being degraded by the small RNA degrading nuclease (SDN) class of exonucleases. The guide miRNA strand is then incorporated into AGO proteins to carry out the silencing reactions (Voinnet 2009).

Computational and molecular cloning strategies had found 203 potential miRNA genes in soybean (*Glycine max* L. Merril.) genome by 2012 (Subramanian *et al*. 2008; Sunkar and Jagadeeswaran 2008; Zhang *et al*. 2008, 2009; Tuteja *et al*. 2009; Joshi *et al*. 2010; Liu *et al*. 2011; Song *et al*. 2011). Previously discovered miRNAs have been verified to play important roles in soybean development, seed yield, nitrogen fixation, and the responses to both abiotic and biotic stresses (Chen *et al*. 2009; Wang *et al*. 2009; Liu *et al*. 2010; Zeng *et al*. 2010; Song *et al*. 2011). Many studies focus on newly miRNA detections and target genes identification, but little was known about transcription and regulation for miRNAs. A crucial component in the analysis of a miRNA promoter region was the accurate identification of the transcription start site (TSS). In animals, the primary transcript was rapidly cleaved in the nucleus by the enzyme Drosha, and this presents a technical barrier for the large-scale experimental identification of TSSs. The primary transcripts that have been experimentally characterized in animal species have been observed to be on the order of 1-4 kb in length (Cai *et al*. 2004; Lee *et al*. 2004). The TSS may be as little as 50 nt or as much as 2.5 kb upstream of the first miRNA contained within the transcript, suggesting that promoter location cannot be inferred directly from miRNA location alone. In plants, only 53 primary transcripts were reported from *Arabidopsis* TSSs where distances range from 26- to 615-nt upstream of the pre-miRNA contained within the transcript. The fact that only a very limited number of published reports mention upstream regulatory sequences of miRNA genes may be largely attributed to this difficulty (Ohler *et al*. 2004). Therefore, understanding the mechanisms governing miRNA gene expression patterns will be necessary to understand miRNA-mediated regulatory pathways and networks.

In this study, a number of transcript start sites (TSSs) for soybean miRNA primary transcripts were identified using a 5′ rapid amplification of cDNA ends (RACE) procedure. Features associated with transcription initiation of these cloned TSSs in soybean miRNA genes were analyzed, revealing start sites, core promoters and other properties that were consistent with a pol II mechanism of transcription. A position weight matrix algorithm was used to identify conserved motifs in miRNA core promoter regions and a canonical TATA box motif was identified upstream of each of the major TSSs. Several *cis*-acting elements were predicted in 2 kb 5′ to the TSSs, and were used to infer the potential spatial and temporal expression patterns of these miRNAs. The target genes of these miRNAs were also predicted and used to further elucidate the potential function of these miRNAs. These data provide a molecular basis to explore regulatory mechanisms of miRNA expression, and a way to understand miRNA-mediated regulatory pathways and networks in plants.

RESULTS

Soybean miRNA primary transcripts

In order to clone the soybean miRNA primary transcripts and determine if they contained 5′-cap structures typical for RNA pol II transcripts, a series of 5′ RACE reactions were done for 50 known miRNAs of soybean. Used was total RNA from 30-d-old seedlings that was pretreated with calf intestine alkaline phosphatase (CIAP) plus tobacco acid pyrophosphatase (TAP). Only the transcripts containing a 5′ cap would be further ligated to adapters, and subsequently amplified by PCR. As controls, TAP(-) was used to detect whether incomplete RNA sufficient degraded, and M-MLV(-) was used to exclude false positive caused by DNA contamination. Products were developed from 11 miRNAs using locus-specific primer sets (including gma-

MIR166a, gma-MIR172a, gma-MIR172f, gma-MIR394a, gma-MIR394b, gma-MIR396a, gma-MIR1508a, gma-MIR1535, gma-MIR1536, gma-MIR2108a, and gma-MIR2108b) (Fig. 1).

All of their controls were negative, indicating that the products were derived from complete primary transcripts. The details of these 11 miRNA 5′ end transcript (pre-miRNA upstream to TSS) were listed in Table 1. The lengths between the start site of primary miRNA transcripts and the pre-miRNA start sites differed ranging from 75 to 2 627 bp, with a mean of 698 bp.

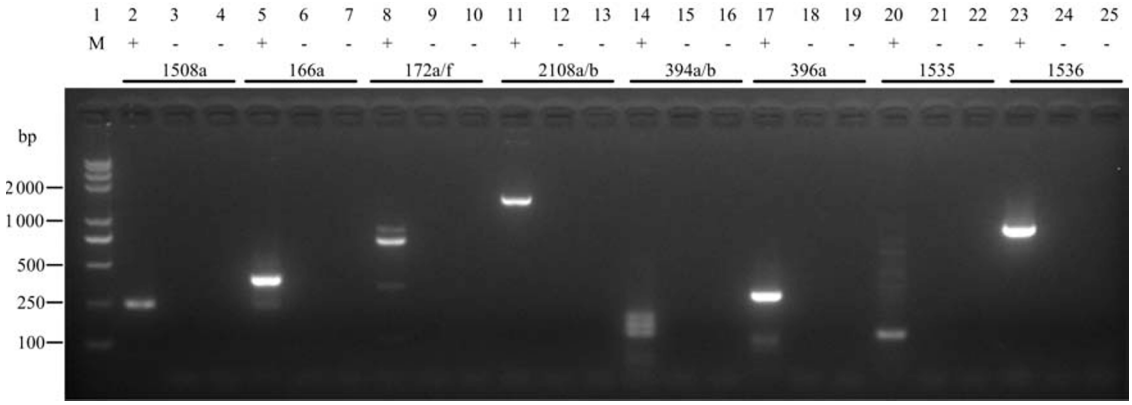

Fig. 1 RLM-5′ RACE reactions on eight of the eleven miRNA transcripts. M, the DNA marker. In each set of three lanes containing miRNAs the "+" represents RLM-5′ RACE reactions of each miRNAs; the first "-" in each miRNA represented the TAP(-) control to detect whether RNA was sufficiently degraded; the second "-" in each miRNA represents the M-MLV(-) control to exclude false positive caused by DNA contamination, and each lanes were resolved on 1% (w/v) agarose gel.

Table 1 The basic information of miRNA primary transcript of 5′ end (pre-miRNA upstream to TSS)

GB[1)]	Gene name	Chromosome	Chain	TSS	Pre-miRNA[2)]	Length (nt)[3)]
HQ735290	gma-MIR1508a	Gm16	+	32 903 556	32 903 737	181
HQ735291	gma-MIR166a	Gm16	-	1 912 942	1 912 715	227
JN225488	gma-MIR172f	Gm11	+	16 933 399	16 934 086	687
JN225489	gma-MIR2108a	Gm17	+	24 100 742	24 103 207	2 465
JN225490	gma-MIR2108b	Gm05	-	11 859 209	11 856 582	2 627
JN225491	gma-MIR172a	Gm12	-	6 111 101	6 110 862	239
JN225492	gma-MIR394a	Gm17	+	41 783 668	41 783 747	79
JN225493	gma-MIR394b	Gm14	-	48 985 056	48 984 981	75
JN225494	gma-MIR396a	Gm13	-	26 338 445	26 338 273	172
JN225495	gma-MIR1535	Gm19	+	43 786 764	43 786 813	49
JN225496	gma-MIR1536	Gm15	-	43 789 103	43 788 222	881

[1)] Accession number of GenBank.

[2)] Soybean precursor miRNA start site in chromosome.

[3)] The length between primary miRNA transcript start site and pre-miRNA start site.

Conserved miRNA promoter sequence cored motifs in soybean

Each of the 11 positive 5′ RACE products was further analyzed as a single fragment of a distinct size. All the primary transcripts were mapped on genome according to the location annotation of miRbase (http://www.miRbase.org) (Griffiths-Jones *et al.* 2008). The TSS motif (-4 to +5) base composition was analyzed and shown in Fig. 2. The vast majority (73%) of transcripts initiated with an adenosine, and 64% of the transcripts were preceded by a guanine. Most of positions flanking the TSS were dominated by thymine and adenosine.

To identify conserved motifs flanking the initiation sites at each mapped locus, a 60-bp genomic segment (-50 to +10 relative to the start site) was computationally analyzed using BioProspector (Liu *et al*. 2004). A 8-bp

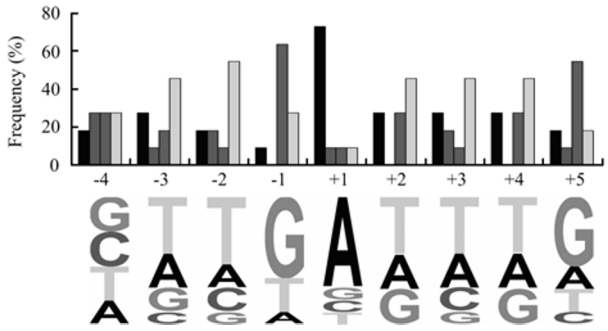

Fig. 2 Base composition at positions flanking miRNA transcription initiation sites.

TATA-box like sequence was identified as an only conserved motif in this region, and the position weight matrix (PWM) of this motif was showed in Table 2. This motif was detected upstream from eight (73%) of the mapped transcription start sites. In order to further evaluate these putative TATA-box like motifs, the PWM and 60-bp genomic segments were input into a second website program MotifMatcher (Ao *et al*. 2004). The results showed that ten of the 60-bp genomic segments were found to have putative TATA-box like motifs, and eight of them have high scores (scores>6) (Fig. 2).

To determine if the high frequency occurrence of the TATA box-like sequence was uniquely associated with this specific region, we examined the distribution of the TATA box like sequence in an extended upstream region (-200 to +50) using MotifMatcher (Fig. 3). The most high scored TATA box-like sequence centered at consensus position -24 from the start site and it was concluded that most or all of these motifs corresponded to authentic TATA boxes (Fig. 4).

Computational identification of *cis*-acting elements in miRNA promoter sequences and their functional analysis

cis-Acting elements can influence miRNA gene expression (Liu *et al*. 2008; Liu *et al*. 2010). To elucidate the spatial and temporal expression pattern and upstream regulation of the eleven soybean miRNAs, their promoter regions (from TSS to 1 000 bp 5′) were analyzed for the potential *cis*-acting elements using PlantCARE database (Lescot *et al*. 2002). The potential *cis*-acting elements in miRNA promoter regions were shown in Table 3. Several known stress-responsive elements were identified including abscisic acid response element (ABRE); anaerobic induction element (ARE); fungal elicitor element (Box-W1); MYB binding sites involved in drought inducement (MBS); heat stress responsive element (HSE); low temperature responsive element (LTR); defense stress responsive element (TC-rich repeats); and wound responsive element (WUN-motif). Other regulatory elements that were involved in the growth and regulation of response to circadian control, such as gibberellic acid (GA), ethylene, salicylic acid (SA) and methyl jasmonate (MeJA) were also found. In order to further explore the function of the miRNAs, the targets were identified using the web-based software psRNAtargets (Zhao and Dai 2011). The potential target of these miRNAs was listed in Table 4. The results indicated that miRNAs played important roles in regulation of transcription factors, metabolisms, stress responses, signaling transductions, cellular structural components, and other cellular process.

DISCUSSION

In this study 11 of 50 miRNAs were identified using the RNA from 30-d-old seedlings. Of the 11 miRNAs, two had been validated previously (Subramanian *et al*. 2008; Wang *et al*. 2009) and nine were newly validated.

The length range between pri-miRNA TSS and pre-miRNA start site (75-2 627 bp, mean 698 bp) in soybean was remarkably wider than in *Arabidopsis thaliana* (range from 26 to 615 bp, mean 208 bp). The soybean miRNA primary transcripts were longer in non-con-

Table 2 Position-weight matrix for conserved TATA box-like motif

Nucleotide	Position (5′→3′)							
	1	2	3	4	5	6	7	8
A	0.0145	0.9607	0.0145	0.9607	0.9607	0.9607	0.0145	0.8425
C	0.0141	0.0141	0.0141	0.0141	0.0141	0.0141	0.0141	0.0141
G	0.1247	0.0064	0.0064	0.0064	0.0064	0.0064	0.1247	0.1247
T	0.8467	0.0187	0.965	0.0187	0.0187	0.0187	0.8467	0.0187

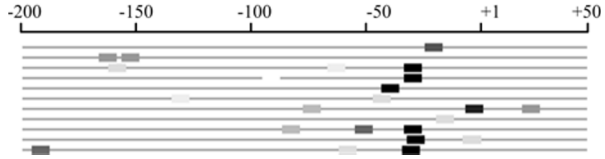

Fig. 3 Frequency of putative TATA box-like motifs within a 250-bp (-200 to +50 bp) context encompassing all mapped miRNA transcript initiation sites. Colors represent different profiles. Darker colors represent stronger matches to the consensus profile.

served miRNA, such as gma-MIR2108a, gma-MIR2108b and gma-MIR1536 (Xie *et al*. 2005).

A diverse set of putative miRNA transcripts from more than 30 plant species were identified from expressed sequence tag (EST) databases, suggesting that plant miRNA precursor transcripts are likely polyadenylated (Jones-Rhoades and Bartel 2004). These soybean miRNA primary transcript data were consistent with the previous reports, suggesting that plant miRNA genes are transcribed by RNA pol II mechanism (Lee *et al*. 2004; Xie *et al*. 2005).

According to the base composition characteristics of TSS motifs in Fig. 2, they were consistent with the transcription pattern by RNA pol II (Shahmuradov *et al*. 2003). This discovery was similar with the previous report for *A. thaliana.* However, the positions (-1 and +5) were dominated by guanine in soybean rather than dominated by cytosine in *Arabidopsis*.

The function of miR172 has been extensively studied. The earliest report showed that it likely acts in

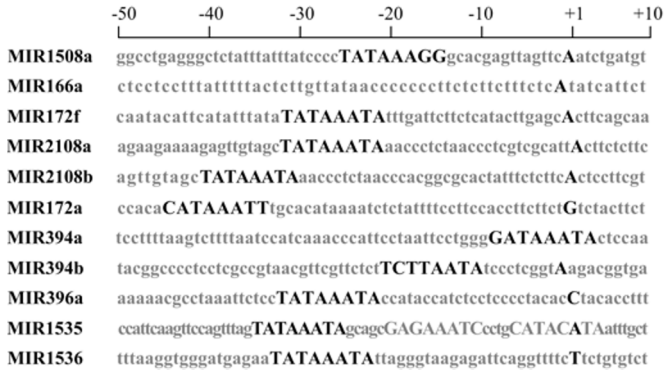

Fig. 4 Genomic sequences of 60 bp (-50 to +10 bp) flanking of transcript start sites from 11 miRNA loci. Putative TATA box-like motifs (bold) are indicated. MotifMatcher scores are given at the end of each sequence.

Table 3 Kinds and numbers of known stress-related elements in the upstream regions of miRNA genes

Element	Function	1508a	166a	172f	2108a	2108b	172a	394a	394b	396a	1535	1536
ABRE	Abscisic acid	1	1		2	1						
ARE	Anaerobic						1		1		1	
GC-motif	Anoxic											1
TGA-element	Auxin	1			1			1		1	1	
MSA-like	Cell cycle		1									
Circadian	Circadian control	1	1				1	1		1	1	
TC-rich repeats	Defense and stress		2			2		1			1	1
MBS	Drought	1			2		1			2		
GCN4_motif	Endosperm expression						1			1		1
Skn-1_motif	Endosperm expression		2	1		3	1	3		1	1	2
ERE	Ethylene	1						1	1	1	2	
Box-W1	Fungal elicitor									1		
GARE-motif	Gibberellin	1			1	2				1		
HSE	Heat stress			1						1	1	
LTR	Low temperature	1	1			1						
CGTCA-motif	MeJA	2	1		1	1	1	1		1	1	1
TGACG-motif	MeJA	1			1		1					
CAT-box	Meristem expression			1						1		
TCA-element	Salicylic acid			1	1	1	3	1		1	1	
WUN-motif	Wound							1				

Table 4 Potential targets of soybean miRNA[1)]

miRNA	Target protein	Target function	Target
miR166a	Class III HD-Zip protein	Transcription factor	TC230399, TC470337
	PHAVOLUTA-like HD-ZIP III	Transcription factor	BM309730
miR172a/f	AP2 transcription factor	Transcription factor	BE659941, FK339807, TC440496, TC457282, TC469666, TC480184, TC480543, TC483493, TC486044
	Superoxide dismutase [Cu-Zn]	Stress response	TC452799
miR394a/b	Wound induced protein	Stress response	TC484496, TC484496
	fatty acid CoA ligase protein	Metabolism	TC436065
	methyladenine glycosylase family protein	Metabolism	TC422001
	Enolase 2	Metabolism	AW186027
	Lipoxygenase	Metabolism	BE612202
miR396a	Cytochrome complex iron-sulfur subunit	Metabolism	TC491219, TC214790
	Growth-regulating factor	Transcription factor	TC231066, TC218495, TC216688
	Cysteine proteinase precursor	Metabolism	TC206710
	ATP binding /ATP-dependent helicase	Metabolism	TC220021
	Heat shock protein	Stress response	TC219546
	Calmodulin-binding family	Signaling transduction	BI700778
miR1508a	GMFP1 isoprenylated protein	Metabolism	TC426568
	Histone H1	Cellular component	TC459720, TC473921
	Gibberellin regulated protein	Stress response	TC486700
	Formin-like protein 9	Stress response	FK486749
	En/Spm-like transposon protein	Cellular process	TC461929
	DNA repair protein RadC	Cellular process	FK387437
	Generic methyltransferase	Cellular process	TC464464
	TGF-beta receptor	Signaling transduction	TC421314
	Syringolide-induced protein 19-1-5	Metabolism	AW306720
miR1536	Growth-on protein GRO11	Transcription factor	TC462825
	Patellin-3	Reproduction	TC421364
	auxin-independent	Transcription factor	TC475456
miR2108a/b	Elongation factor 1-alpha	Cellular process	AW351184, BU577594
	ATP-dependent RNA helicase-like protein	Cellular process	EH222473
	Patellin-4	Reproduction	TC468998

[1)] Proteins of unknown function were not listed.

cell-fate specification as a translational repressor of APETALA2 in *Arabidopsis* flower development (Chen 2004). The miR172 of soybean was first identified in nitrogen-fixed soybean nodules and was presumed to play a critical role on symbiotic nitrogen fixation (Subramanian *et al*. 2008; Wang *et al*. 2009). In this study, the miR172a promoter region was shown to have a potential anaerobic responsive element, and it might be induced by anaerobic environments. This result further suggested that the miR172a might take part in nodule nitrogen-fixing. The miR172a targets were predicted to mainly be an AP2 family transcription factor, consistent to the soybean sequencing results by Song *et al*. (2011).

The previous reports have demonstrated that miRNAs may be involved in a kind of negative feedback loops model, such as *miR159* gene in soybean and *miR160* gene in *Arabidopsis* (Megraw *et al*. 2006; Liu *et al*. 2010). In this study, four miRNAs were also found to have negative feedback loops model. The *miR394a* promoter region of soybean had wound responsive element and targeted in wound induced protein. The *miR396a* promoter region of soybean had auxin and heat shock responsive element and targeted in a growth regulating factor and heat shock protein. The *miR1508a* promoter region had gibberellin acid responsive element and targeted in gibberellin regulation protein. The *miR2108a* promoter region had auxin responsive element and targeted in elongation factor 1-α.

CONCLUSION

Potential spatial and temporal expression patterns of eleven soybean miRNAs were inferred. The target genes for these miRNAs were predicted and further elucidated for the potential function of the miRNAs. This study provides a molecular basis to explore regulatory mechanisms of miRNA expression in soybean. Understanding of miRNA-mediated regulatory pathways and networks in soybean has been advanced.

MATERIALS AND METHODS

5′ RACE mapping of miRNA transcripts

Twenty seeds of the soybean cultivar, Dongnong 50, were grown in a growth chamber for 30 d and the leaves, stems and roots were used for RNA isolation. Total RNA was extracted using RNAiso Plus (TaKaRa, Dalian, China). Primers for 50 miRNAs were assembled from published reports (Subramanian *et al*. 2008; Sunkar and Jagadeeswaran 2008; Tuteja *et al*. 2009; Zhang *et al*. 2008, 2009; Joshi *et al*. 2010; Song *et al*. 2011). The 5′ ends of miRNA transcripts were then mapped by 5′-Full RACE Kit (TaKaRa Dalian, China). First, calf intestine alkaline phosphatase (CIAP) was added for degrading incomplete mRNA, rRNA, tRNA, and DNA. Then tobacco acid pyrophosphatase (TAP) was added in order to decap the full-length mRNAs. The control of TAP(-) without the second step to test whether RNA sufficient dephosphorylation had occurred. The decapped mRNA was ligated with 5′ RACE adaptor, and complementary DNA (cDNA) was synthesized using reverse transcriptase M-MLV (RNase H^-) and random oligonucleotide hexamers as primers. In the second negative control, M-MLV(-) wasn't added in order to test for false positives caused by DNA contamination. Finally, a cDNA pool containing equal amounts of reaction products from each organ was used as template in 5′ RACE nested-PCR. The forward (5′ RACE outer and inner) primers were general, which was specific to the RNA adaptor. The reverse primers (gene-specific primers, GSP) were designed based on the sequence of the precursor miRNA (Kozomara and Griffiths-Jones 2010). The GSP specificity was tested using BLAST (word length=7) on the website phytozome http://www.phytozome.net/search.php?show=blast&method=Org_Gmax (Altschul *et al*. 1997; Schmutz *et al*. 2010). The 5′ RACE nested-PCR was performed by LA *Taq* (TaKaRa, Dalian, China) and the procedure followed the protocol of 5′-Full RACE Kit. The PCR products from a positive 5′ RACE reaction were gel purified and cloned in pMD®18-T vector (TaKaRa Dalian, China). A minimum of three clones were sequenced for each PCR product bands. The cloned vectors were transformed into *Escherichia coli* DH5α strain for replication and sequencing. Sequences corresponding to transcript 5′ ends were deposited at GenBank with accession numbers listed in Table 1.

Computational identification of conserved miRNA promoter sequence cored motifs

The genomic sequence was downloaded from website Phytozome (http://www.phytozome.org) (Schmutz *et al*. 2010). A 9-bp (-4 to +5) sequence flanking the TSS was used for base component analysis and the TSS motif matrix construction according to the method described previously (Shahmuradov *et al*. 2003). A 60-bp (-50 to +10) genomic sequence flanking the start site for each of the 11 miRNA loci (that was likely to contain the core promoter sequence) was analyzed using BioProspector (all parameter in default), a Gibbs sampling-based motif-finding program (Liu *et al*. 2004). Searches with a motif width of 6-8 nt were carried out. In all cases, TATA box-like sequences were identified as the only conserved motifs present. The presence of the conserved TATA box-like motif matrix (8-nt width) in each 60-bp genomic segment was checked using MotifMatcher with up to three matches per segment allowed (Ao *et al*. 2004). The algorithm gives a score for placement of each TATA box-like sequence detected. These are log-odds-based scores calculated as $\ln[P_{(\text{observed}|\text{PWM})}/P_{(\text{observed}|\text{background model})}]$, where the numerator is the probability of the observed sequence according to the position weight matrix (PWM). The represented motif and the denominator were the probability of the sequence according to a simple Markov chain constructed by examining frequencies of nucleotide occurrences throughout a background sequence set (Ao *et al*. 2004). A second search by MotifMatcher was done using an extended upstream region (-200 to +50) to analyze the distribution of the putative TATA box motif with the 8-nt motif matrix generated by BioProspector as a sample motif. Up to three matches to the TATA box-like motif were allowed.

Computational identification of *cis*-acting elements in miRNA promoter sequences and their functional analysis

The promoter regions (from TSS to 1 000 bp 5′) were analyzed for the potential *cis*-acting elements and motifs. The analysis of *cis*-acting elements for the miRNA promoters was performed by tools at the PlantCARE database (http://intra.psb.ugent.be:8080/PlantCARE), a database of plant *cis*-acting regularly elements (Lescot *et al*. 2002). In order to further explore the functions of these miRNAs, the targets were predicted using the newly web-based software psRNAtarget (Zhao and Dai 2011). The soybean target genes database used was the *Glycine max* (soybean) DFCI Gene Index (GMGI), version release 16, and the software default parameters were used.

Acknowledgements

This study was conducted in the Key Laboratory of Soybean Biology of Chinese Education Ministry, Agriculture Research System of China and the Key Laboratory of Northeastern Soybean Biology and Breeding/Genetics of Chinese Agriculture Ministry, financially supported by the National High-Tech R&D Program of China (2006AA10Z1F1), the National Core Soybean Genetic Engineering Project, China (2011ZX08004-002), the National Natural Science Foundation of China (60932008, 30971810), the National Basic Research Program of China (2009CB118400), the Ministry of Education Innovation Team of Soybean Molecular Design, China and the Innovation Team of the Education Bureau of Heilongjiang Province, China.

References

Altschul S F, Madden T L, Schaffer A A, Zhang J, Zhang Z, Miller W, Lipman D J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**, 3389-3402.

Ao W, Gaudet J, Kent W J, Muttumu S, Mango S E. 2004. Environmentally induced foregut remodeling by PHA-4/FoxA and DAF-12/NHR. *Science*, **305**, 1743-1746.

Cai X, Hagedorn C H, Cullen B R. 2004. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA-A Publication of the RNA Society*, **10**, 1957-1966.

Chen R, Hu Z, Zhang H. 2009. Identification of microRNAs in wild soybean (*Glycine soja*). *Journal of Integrative Plant Biology*, **51**, 1071-1079.

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

Chen X M. 2010. Small RNAs – secrets and surprises of the genome. *The Plant Journal*, **61**, 941-958.

Griffiths-Jones S, Saini H K, van Dongen S, Enright A J. 2008. miRBase: tools for microRNA genomics. *Nucleic Acids Research*, **36**, 154-158.

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

Joshi T, Yan Z, Libault M, Jeong D H, Park S, Green P J, Sherrier D J, Farmer A, May G, Meyers B C, *et al.* 2010. Prediction of novel miRNAs and associated target genes in *Glycine max*. *BMC Bioinformatics*, **11**, 14.

Kozomara A, Griffiths-Jones S. 2010. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Research*, **39**, 152-157.

Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, *et al.* 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature*, **425**, 415-419.

Lee Y, Jeon K, Lee J T, Kim S, Kim V N. 2002. MicroRNA maturation: stepwise processing and subcellular localization. *EMBO Journal*, **21**, 4663-4670.

Lee Y, Kim M, Han J, Yeom K H, Lee S, Baek S H, Kim V N. 2004. MicroRNA genes are transcribed by RNA polymerase II. *EMBO Journal*, **23**, 4051-4060.

Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, van de Peer Y, Rouze P, Rombauts S. 2002. PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Research*, **30**, 325-327.

Liu H H, Tian X, Li Y J, Wu C A, Zheng C C. 2008. Microarray-based analysis of stress-regulated microRNAs in *Arabidopsis thaliana*. *RNA-A Publication of the RNA Society*, **14**, 836-843.

Liu X, Luo G, Bai X, Wang X J. 2009. Bioinformatic analysis of microRNA biogenesis and function related proteins in eleven animal genomes. *Journal of Genetics and Genomics*, **36**, 591-601.

Liu Y, Wei L, Batzoglou S, Brutlag D L, Liu J S, Liu X S. 2004. A suite of web-based programs to search for transcriptional regulatory motifs. *Nucleic Acids Research*, **32**, 204-207.

Liu Y X, Chang W, Han Y P, Zou Q, Guo M Z, Li W B. 2011. In silico detection of novel microRNAs genes in soybean genome. *Agriculture Sciences in China*, **10**, 1336-1345.

Liu Y X, Han Y P, Chang W, Zou Q, Guo M Z, Li W B. 2010. Genomic analysis of microRNA promoters and their cis-acting elements in soybean. *Agriculture Sciences in China*, **9**, 1561-1570.

Megraw M, Baev V, Rusinov V, Jensen S T, Kalantidis K, Hatzigeorgiou A G. 2006. MicroRNA promoter element discovery in *Arabidopsis*. *RNA-A Publication of the RNA Society*, **12**, 1612-1619.

Ohler U, Yekta S, Lim L P, Bartel D P, Burge C B. 2004. Patterns of flanking sequence conservation and a characteristic upstream motif for microRNA gene identification. *RNA-A Publication of the RNA Society*, **10**, 1309-1322.

Schmutz J, Cannon S B, Schlueter J, Ma J, Mitros T, Nelson W, Hyten D L, Song Q, Thelen J J, Cheng J, *et al.* 2010. Genome sequence of the palaeopolyploid soybean. *Nature*, **463**, 178-183.

Shahmuradov I A, Gammerman A J, Hancock J M, Bramley P M, Solovyev V V. 2003. PlantProm: a database of plant promoter sequences. *Nucleic Acids Research*, **31**, 114-117.

Song Q X, Liu Y F, Hu X Y, Zhang W K, Ma B A, Chen S Y, Zhang J S. 2011. Identification of miRNAs and their target genes in developing soybean seeds by deep sequencing. *BMC Plant Biology*, **11**, 5.

Subramanian S, Fu Y, Sunkar R, Barbazuk W B, Zhu J K, Yu O. 2008. Novel and nodulation-regulated microRNAs in soybean roots. *BMC Genomics*, **9**, 160.

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

Tuteja J H, Zabala G, Varala K, Hudson M, Vodkin L O. 2009. Endogenous, tissue-specific short interfering RNAs silence the chalcone synthase gene family in glycine max seed coats. *The Plant Cell*, **21**, 3063-3077.

Voinnet O. 2009. Origin, biogenesis, and activity of plant microRNAs. *Cell*, **136**, 669-687.

Wang Y, Li P, Cao X, Wang X, Zhang A, Li X. 2009. Identification and expression analysis of miRNAs from nitrogen-fixing soybean nodules. *Biochemical and Biophysical Research Communications*, **378**, 799-803.

Xie Z, Allen E, Fahlgren N, Calamar A, Givan S A, Carrington J C. 2005. Expression of *Arabidopsis* MIRNA genes. *Plant Physiology*, **138**, 2145-2154.

Zeng H Q, Zhu Y Y, Huang S Q, Yang Z M. 2010. Analysis of phosphorus-deficient responsive miRNAs and cis-elements from soybean (*Glycine max* L.). *Journal of Plant Physiology*, **167**, 1289-1297.

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

Zhang Z, Yu J, Li D, Liu F, Zhou X, Wang T, Ling Y, Su Z. 2009. PMRD: plant microRNA database. *Nucleic Acids Research*, **38**, 806-813.

Zhao P X, Dai X B. 2011. PsRNATarget: a plant small RNA target analysis server. *Nucleic Acids Research*, **39**, 155-159.

(Managing editor WANG Ning)