Analysis of promoters of microRNAs from a 
*Glycine max* degradome library*

Yi-qiang HAN1, Zheng HU2, Dian-feng ZHENG3, Ya-mei GAO†‡1

(1College of Life Science and Technology, Heilongjiang Bayi Agricultural University, Daqing 163319, China)
(2The National Key Facility for Crop Gene Resources and Genetic Improvement, Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China)
(3College of Agronomy, Heilongjiang Bayi Agricultural University, Daqing 163319, China)
†E-mail: gaoym800@126.com

Received July 01, 2013; Revision accepted Aug. 25, 2013; Crosschecked

Abstract: Objective: MicroRNAs (miRNAs) are genome-encoded, small non-coding RNAs that play important functions in development, biotic and abiotic stress responses, and other processes. Our aim was to explore the regulation of miRNA expression. Methods: we used bioinformatics methods to predict the core promoters of 440 miRNAs identified from a soybean (*Glycine max*) degradome library and to analyze *cis*-acting elements for 369 miRNAs. Results: The prediction results showed that 83.86% of the 440 miRNAs contained promoters in their upstream sequences, and 8.64% (38 loci) in their downstream sequences. The distributions of two core promoter elements, TATA-boxes and transcriptional start sites (TSSs), were similar. The *cis*-acting elements were examined to provide clues to the function and regulation of spatiotemporal expression of the miRNAs. Analysis of miRNA *cis*-elements and targets indicated a potential ARF- and GARF-mediated negative feedback loop for miRNA expression. Conclusions: The features of miRNAs from a *Glycine max* degradome library obtained here provide insights into the transcription regulation and functions of miRNAs in soybean.

Key words: *Glycine max*, MicroRNA (miRNA), Promoter, *cis*-acting element, Prediction

1 Introduction

MicroRNAs (miRNAs) are genome-encoded, ~21-23 nucleotide non-coding RNAs that negatively regulate gene expression by directing mRNA cleavage or interfering with translation (Carrington *et al.*, 2003). Studies have shown that miRNAs have the potential to regulate all physiological cell functions (Llave *et al.*, 2002). In plants, miRNAs regulate the expression of a large number of genes in diverse biological processes (Chen, 2004; Lu *et al.*, 2005; Zeng *et al.*, 2010; Shamimuzzaman *et al.*, 2012), as identified by computational prediction, microarray analysis, and high-throughput sequencing (Jones-Rhoades *et al.*, 2004; Liu *et al.*, 2008; Zhao *et al.*, 2010). Most miRNAs exist as independent transcription units, which are transcribed into long primary transcripts (pri-miRNAs) by RNA polymerase II, then cleaved to miRNA precursors (pre-miRNAs) (Bartel, 2004; Lee, 2004). MicroRNAs have broad regulatory effects throughout the plant. Therefore, understanding the mechanisms regulating miRNA gene expression remains an

†Corresponding author
* Project supported by the National Key Project for Basic Research of China (No. 31271652 and 31301708)

© Zhejiang University and Springer-Verlag Berlin Heidelberg 2013
important fundamental goal.

Class II promoters contain two parts, a core promoter and upstream elements. The core promoter consists of at least two elements: a TATA box, which begins at about the −30 bp position, and an initiator, which is centered on the transcription start site (TSS) (Liu et al., 2010). Promoter elements of pre-miRNAs are similar to those of protein-coding genes (Smale, 2001). Bioinformatic analysis of the core miRNA promoters in model plants, including Arabidopsis thaliana and Oryza sativa, confirmed that the promoters of most known miRNA genes are of the same type as those of protein-coding genes and are transcribed by RNA polymerase II (Xie et al., 2005; Zhou et al., 2007; Cui et al., 2009). TATA-less promoters were also identified in Arabidopsis and rice. The promoters of most miRNA genes are found within 500 bp upstream of the TSS. Although substantial work has examined the promoters of protein-coding genes, the nature of miRNA promoter elements is still one of the most interesting aspects of small RNA biology (Megraw et al., 2006).

Soybean (Glycine max) is one of the most economically important agricultural crops in the world, and one of the most important global sources of protein and oil, both for food and for livestock feed. Soybean was the first legume species with a published high-quality draft genome sequence (Schmutz et al., 2010). Because miRNAs play an important regulatory role in a wide variety of developmental and metabolic processes in plants, more and more soybean miRNAs have been identified (Wang et al., 2008; Zhang et al., 2008; Chen et al., 2009; Song et al., 2011; Wong et al., 2011; Kulcheski et al., 2011; Guo et al., 2011; Li et al., 2011; Zeng et al., 2012; Hu et al., 2013), but the features of the upstream sequences of these miRNAs have not been comprehensively analyzed. Core promoters and cis-acting elements of 82 soybean miRNAs from the miRBase database were analyzed by bioinformatics methods (Liu et al., 2010), but this analysis involved a relatively small number of miRNA genes. We have identified substantial numbers of miRNAs by deep sequencing of a Glycine max degradome library (Hu et al., 2013). In this study, the core promoters and cis-acting elements of those miRNAs were predicted by bioinformatics methods, to reveal the features of the core promoters of soybean miRNAs. The miRNA expression regulation pathway was also analyzed. These studies provide insights into the regulation of soybean miRNA expression.

2 Materials and methods

2.1 MicroRNA data and classification

A total of 440 miRNAs were identified from a soybean cDNA degradome library that was constructed from root, stem, leaf, and inflorescence tissues. The NCBI-GEO accession number of the Glycine max degradome sequences is GSE33380. All soybean miRNA sequences for prediction were divided into two groups based on their genomic background: “intronic miRNAs”, which overlap protein-coding genes, and “intergenic miRNAs”, which reside between protein-coding genes. The protein-coding genes included verified and hypothetical genes. The 440 soybean miRNAs included 34 intronic miRNAs and 406 intergenic miRNAs.

2.2 Identification of potential core promoters of soybean miRNAs

For promoter prediction, the upstream and downstream sequences of the 440 miRNA genes were downloaded from the soybean genome database (http://www.phytozome.net/index.php) (Goodstein et al., 2012). The sequences upstream of pre-miRNAs (hairpin precursors) of each of the miRNA genes were obtained according to Cui et al. (2009). The 200 bp sequence downstream of each miRNA gene was also used for promoter prediction. For 83.86% of the pre-miRNAs (hairpin precursors), the upstream sequence was 2000 bp or longer. Promoters were predicted by the plant promoter identification program TSSP (softberry.com), which is designed for predicting plant Pol II promoters (Shahmuradov et al., 2005). The
predictions were obtained at the default TSSP settings.

2.3 Analysis of cis-acting elements of soybean miRNAs

Potential promoter regions (from TSS to 800 bp upstream) were obtained to predict the potential cis-acting elements and motifs. If there were overlapping TSS in the selected region, only the sequence between the two TSSs was analyzed, to exclude redundancy (Liu et al., 2010). The PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html), a database of plant cis-acting regulatory elements (Lescot et al., 2002), was used to analyze the cis-acting elements of the miRNAs.

3 Results

3.1 The core promoters and enhancers of soybean miRNAs

Putative promoters were divided into TATA-box/TSS core promoters and enhancers, and counted independently. From 440 miRNAs, we identified 699 predicted core promoters, with 102 predicted enhancers in upstream sequences. There were 38 core promoters, but no enhancers, predicted in the downstream sequence.

Among the upstream sequences of the 440 miRNAs, 71 (16.14%) were found to contain no predicted promoters (Fig. 1A). Of the other 369 (83.86%) promoter-containing sequences, 127 (28.86%) were predicted to have only one promoter, 168 (38.18 %) contained two, and the rest, 74 (16.90%), contained three or more predicted promoters. Among the predicted enhancers in upstream sequences of the 440 miRNAs, 345 (78.41%) were found to contain no enhancers (Fig. 1B). Among the other 95 (21.59%)
enhancer-containing sequences, 88 (20.00%) were predicted to have only one enhancer, 7 (1.59%) contained two enhancers, and no sequences were predicted to contain three or more enhancers. Most pre-miRNAs examined contained no predicted enhancers.

In the downstream sequences of the 440 miRNAs, 402 (91.36%) were found to contain no predicted promoter (Fig. 1C) and 38 (8.64%) were predicted to have only one. None of the 440 miRNAs contained predicted enhancers in the downstream sequence (Fig. 1D).

Pre-miRNAs from intergenic regions generally had more core promoters than the intronic pre-miRNAs (Supplementary Table S1), and two or fewer core promoters were identified for each intronic miRNA. In the pre-miRNA downstream sequences, there were few core promoters, and never more than one.

### 3.2 Distribution of the predicted core promoters and enhancers of soybean miRNA genes

We analyzed the specific sequences of pre-miRNAs using the TSSP program, and counted the TSS, TATA-box and enhancers separately. For the predicted core promoters in the upstream sequences of the 440 miRNAs, a total of 699 TSSs were identified. A strong peak was observed near the start position of each transcript. For intergenic miRNAs, the distribution could be separated into three distinct regions (Fig. 2A). The first peak was found close to the pre-miRNA within the 1 kb upstream region, which contained 56.41% of the total predicted TSSs (663). A second broad region, from −1 kb to −1.6 kb, contained 27.60% of the predicted TSSs. The third region, from −1.6 kb to −2.0 kb, contained 15.99% of the predicted TSSs. For intronic miRNAs, there were also three peaks of TSSs (from 0 kb to −1.2 kb, −1.2 kb to −1.6 kb, and −1.6 kb to −2.0 kb).

The distribution of TATA-boxes was similar to the distribution of TSSs. The core promoters of four miRNAs (gm-cmiR160f, gm-cmiR319k, gm-cmiR13 and gm-cmiR4i) were found to be TATA-less. For
intergenic miRNAs, we identified three distinct regions (Fig. 2B). The first peak was found close to the pre-miRNA within the 0.6 kb upstream region, and contained 36.42% of the total (659) predicted TATA-boxes. A second broad region, from −0.6 kb to −1.2 kb, contained 28.68% of predicted TATA-boxes. The third region, from −1.1 kb to −2.0 kb, contained 34.90%. For intronic miRNAs, there were also three peaks of TATA-boxes (from 0 kb to −0.6 kb, −0.6 kb to −1.2 kb, and −1.2 kb to −2.0 kb).

In intergenic and intronic miRNAs, 102 enhancers were predicted. For intergenic miRNAs, most (74.51%) of the predicted enhancers were found within 1 kb upstream of pre-miRNAs. The remaining 15.69% occurred within -1 kb to -1.6 kb. The third region, from −1.6 kb to −2 kb, contained 9.80% of the predicted enhancers (Fig. 2C). For the predicted core promoters in downstream sequences of the 440 miRNAs, a total of 36 (8.18%) TSSs and TATA-boxes were identified.

3.3 Cis-acting elements in potential promoters and an miRNA expression regulation pathway related to hormone metabolism

To further elucidate the spatiotemporal expression pattern and the function of miRNAs from the Glycine max degradome library, sequences 800 bp upstream of the potential TSS of 369 miRNA genes were analyzed using the PlantCARE database, excluding 71 miRNAs that had no predicted TSS. Cis-acting elements and motifs of 355 miRNA promoters were predicted, excluding 14 miRNAs without sufficient sequence information. There were 126 types of cis-acting elements. Every miRNA contained a TATA-box core promoter element at around -30 bp from the TSS and a CAAT-box, a common cis-acting element, in the promoter and enhancer regions (Supplementary Table S2). In addition, 99.72% of the miRNAs contained a light responsive element (LRE). Cis-acting elements were grouped into several classes according to their functions, such as stress-responsive elements (SRE), cis-acting elements regulating plant growth (GRE), hormone regulated cis-acting elements (HRE), and tissue specific cis-acting elements (TSE) (Supplementary Table S3) as described by Liu et al. (2010). Other regulatory elements, such as AT-rich DNA, cis-acting regulatory elements essential for anaerobic induction, and elicitor-responsive elements, were not classified. Of the miRNA sequences examined, 86.48% contained an SRE, including anaerobic induction elements (AREs), defense stress-responsive elements (TC-rich repeats), ABA (abscisic acid)-response elements (ABREs), low-temperature responsive elements (LTRs), MYB binding sites involved in drought response (MBSS), salicylic acid responsiveness elements (SAEs), heat-stress responsive elements (HSEs), and light-responsive elements (LREs). Also, 35.49% of miRNAs contained a GRE, including a leaf morphology development element (HD-Zip 2), differentiation of the palisade mesophyll cells element (HD-Zip 1), meristem specific activation element (OCT, NON-box, doCT, CCGTCC-box, CAT-box), zein metabolism regulation element (O2-site), circadian control element (circadian), and cell cycle regulation elements (MSA-like). TSEs were identified in 72.39% of miRNA sequences examined, including nodule specific factor binding site (Nodule-site2), involved in the endosperm-specific negative expression element (AACA_motif), endosperm expression elements (GCN4_motif, Skn-1_motif), and cis-acting regulatory elements involved in root-specific expression (as1) and seed-specific regulation (RY-element). HRE elements were found in 70.70% of miRNA sequences examined, including those possibly involved in the regulation of the response to plant hormones, such as gibberellic acid (GA), auxin-responsive element, ethylene, salicylic acid, methyl jasmonate, and flavonoids.

Among HRE elements, numerous copies of the ARF (auxin response factor) elements are found in the putative promoter regions of MIR160d, MIR160f, MIR167a, and MIR167g. Numerous copies of GARF (gibberellin response factor) elements are more frequently found in the putative promoter regions of MIR160f, MIR160g, MIR160h, MIR167a, MIR167c, MIR167d, and MIR167i. In our miRNA target gene data, multiple ARF family members, including ARF17, ARF18, ARF6 and ARF8, were predicted to be directly targeted by miR160 and miR167. Based on the above data, we propose a possible ARF- and GARF-mediated negative feedback regulatory loop involving miR160, miR167 and certain ARF family members (Fig. 3).
ARF17 and ARF18 are post-transcriptionally regulated by miR160, and ARF6 and ARF8 are regulated by miR167 (Fig. 3A). In the negative feedback regulatory loop, miR160 and miR167 repress their target ARF.

Fig. 3 A. Certain ARF family members are directly regulated by miRNAs (miR160 and miR167). B. A possible negative feedback regulatory loop involving miR160, miR167 and certain ARF family members.

Numerous copies of the ARF recognition motif are found in the putative promoter regions of MIR160d, MIR160f, MIR167a, and MIR167g. Also, numerous copies of GA response elements are found in the putative promoter regions of MIR160f, MIR160g, MIR160h, MIR167a, MIR167c, MIR167d, and MIR167i. We speculate that GA regulates miRNAs to inhibit the formation of auxin.

ARFs affect miR160 and miR167 accumulation via binding to the AR regulation element of the MIR160 and MIR167 promoters. In addition to the feedback regulation of miR160 and miR167 by ARF, GARF can regulate miR160 and miR167 via binding to the GARF elements of the MIR160 and MIR167 promoters (Fig. 3B). According to the feedback regulation of miR160 and miR167 by ARF and GARF, we speculate that there is GA-auxin cross-talk in the pathway of the regulation of miRNA expression.

4 Discussion

The core promoters and cis-acting elements of 440 miRNAs from a Glycine max degradome library were analyzed using bioinformatics methods to reveal the general features of the promoters of a large number of soybean miRNAs, examine their cis-acting elements, and provide insights into the transcriptional regulation and functions of miRNAs in soybean. Recent studies of Arabidopsis and rice showed that most core promoters contain a TATA-box, and our results for soybean support this observation. Among 699 core promoters predicted, only 4 (0.57%) were TATA-less promoters. Also, 95 of 440 miRNAs (14.59%) contained enhancers. The promoters predicted by bioinformatic analysis will offer valuable information for the design of verification assays and for functional research on miRNAs.

The expression of genes at the transcriptional and post-transcriptional levels is often controlled by small cis-acting sequence elements in or near the regulated genes. Cis-acting elements have been analyzed extensively (Liu et al., 2008; Zhou et al., 2007; Liu et al., 2010). In this study, cis-acting elements of 369 soybean miRNAs were predicted and were classified into several classes according to their functions, such as SRE, GRE, HRE, and TSE. Among those cis-acting elements, HRE element sites were further analyzed to reveal the potential functional correlation between the transcription factors, miRNAs and the target genes of
the miRNAs. In *Arabidopsis*, the ARF element was found to be over-represented among binding motifs in miRNA promoters. MicroRNAs with putative ARF binding sites upstream may play a role in negative feedback loops that control their own expression (Megraw *et al.*, 2006; Jones-Rhoades *et al.*, 2006; Xie *et al.*, 2010). In soybean, we also identified ARF elements in some miRNA promoters, such as miR160 and miR167. Multiple ARF family members, including ARF17, ARF18, ARF6 and ARF8, were predicted to be directly targeted by miR160 and miR167. Although the target genes of soybean miRNAs have not been experimentally verified, targeting of ARFs by soybean miRNAs was indicated by experimental results showing that *Arabidopsis* miR160 and miR167 target ARFs (Wu *et al.*, 2006; Mallory *et al.*, 2005; Wang *et al.*, 2005; Liu *et al.*, 2007). According to the negative feedback model, we propose a possible ARF-mediated negative feedback regulatory loop in soybean, which indicates that miRNAs often act as regulators of regulators with many secondary effects on gene expression. We also identified many types of GARF-binding motifs in promoters of MIR160 and MIR167. Gou *et al.* (2010) found that gibberellins regulate lateral root formation in *Populus* through interacting with auxin, and auxin levels increased in GA-deficient and insensitive transgenic roots, suggesting crosstalk between GA signaling and other hormone pathways. GA may negatively regulate auxin, but the genes and pathways responsible have remained unclear. Here, we found GARF-binding motifs in the promoters of MIR160 and MIR167 and propose a GARF-mediated negative feedback regulatory loop for soybean miRNA expression, suggesting that gibberellins regulate the expression of miR160 and miR167 to interplay with auxin. The functions of these computationally identified cis-elements of soybean miRNA will require experimental validation.

Although few soybean miRNAs have been studied thoroughly, the multiple types of cis-acting elements we have identified and the predicted target genes of these soybean miRNAs both indicate that soybean miRNAs have diverse regulatory functions. For example, the gma-miR156 family has 71 types of cis-acting elements and regulates 27 target genes. Moreover, the more members in each miRNA family, the more genes the miRNAs target, and the more types of cis-acting elements are contained in the promoter of each miRNA.

In summary, we have performed a bioinformatic analysis of the core promoter sequences (TSS, TATA-box, and enhancer) of soybean pre-miRNAs (upstream and downstream sequences) from a *Glycine max* degradome library. Six hundred and ninety-nine promoters and 102 enhancers for 440 soybean pre-miRNAs were detected. Most miRNA genes were intergenic and there were more core promoters in intergenic genes than in intronic genes. Distribution analysis showed that the locations of the TATA-box and TSS were similar. There were 126 types of cis-acting elements identified in 355 miRNA promoters, including stress-responsive elements, cis-acting elements regulating plant growth, hormone-regulated cis-acting elements, and tissue specific cis-acting elements. In addition, we suggest an ARF- and GARF-mediated negative feedback regulatory loop for soybean miRNA expression. These findings improve our understanding of the specific sequences and motifs upstream of pre-miRNAs in soybean. They also provide a basis for further research on the functional role of miRNA and the regulation of miRNA expression in soybean.

**Compliance with ethics guidelines**

Yi-qiang HAN, Zheng HU, Dian-feng ZHENG, and Ya-mei GAO declare that they have no conflict of interest.

This article does not contain any studies with human or animal subjects performed by the any of the authors.

**References**


