RESEARCH PAPER

MicroRNAs in the shoot apical meristem of soybean

Chui E. Wong¹, Ying-Tao Zhao²,³, Xiue-Jie Wang², Larry Croft⁴,⁵,⁶, Zhong-Hua Wang⁴,⁵, Farzad Haerizadeh¹, John S. Mattick⁵,⁶, Mohan B. Singh¹, Bernard J. Carroll⁴,⁵ and Prem L. Bhalla¹,∗

¹ ARC Centre of Excellence for Integrative Legume Research, Faculty of Land and Food Resources, The University of Melbourne, Parkville, Victoria, 3010, Australia
² State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, 100101, China
³ Graduate University of the Chinese Academy of Sciences, Beijing, 100101, China
⁴ ARC Centre of Excellence for Integrative Legume Research, University of Queensland, Brisbane, Queensland 4072, Australia
⁵ School of Chemistry and Molecular Biosciences, University of Queensland, Brisbane, Queensland 4072, Australia
⁶ Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland 4072, Australia

To whom correspondence should be addressed. E-mail: premlb@unimelb.edu.au

Received 10 August 2010; Revised 4 November 2010; Accepted 30 November 2010

Abstract

Plant microRNAs (miRNAs) play crucial regulatory roles in various developmental processes. In this study, we characterize the miRNA profile of the shoot apical meristem (SAM) of an important legume crop, soybean, by integrating high-throughput sequencing data with miRNA microarray analysis. A total of 8423 non-redundant sRNAs were obtained from two libraries derived from micro-dissected SAM or mature leaf tissue. Sequence analysis allowed the identification of 32 conserved miRNA families as well as 8 putative novel miRNAs. Subsequent miRNA profiling with microarrays verified the expression of the majority of these conserved and novel miRNAs. It is noteworthy that several miRNAs* were expressed at a level similar to or higher than their corresponding mature miRNAs in SAM or mature leaf, suggesting a possible biological function for the star species. In situ hybridization analysis revealed a distinct spatial localization pattern for a conserved miRNA, miR166, and its star species suggesting that they serve different roles in regulating leaf development. Furthermore, localization studies showed that a novel soybean miRNA, miR4422a, was nuclear-localized. This study also indicated a novel expression pattern of miR390 in soybean. Our approach identified potential key regulators and provided vital spatial information towards understanding the regulatory circuits in the SAM of soybean during shoot development.

Key words: Development, meristem, microRNAs, soybean.

Introduction

The small RNA (sRNA)-based silencing system is a major facet of gene expression in eukaryotes and this has revolutionized our understanding of gene regulatory pathways. In plants, microRNAs [miRNAs; 21–22 nucleotides (nt)] are one of the best characterized classes of these small RNAs. They are produced from primary miRNA transcripts (pri-miRNAs) transcribed by RNA polymerase II (reviewed by Mallory et al., 2008). The pri-miRNAs form an imperfect hairpin structure that is subjected to DICER-LIKE1 (DCL1)-mediated cleavages to first produce a miRNA precursor (pre-miRNA) and subsequently a miRNA/miRNA* duplex. The strand of the miRNA duplex that has the least stable 5’ pairing is preferentially maintained as the mature miRNA strand and is incorporated into the RNA-induced silencing complex (RISC) while the passenger strand miRNA gets degraded. However, recent studies have indicated that much remains to be uncovered about the selectivity of the mature miRNA strand as there...
are increasing numbers of examples where the miRNA* species accumulates to a level that can be higher than the canonical miRNA in both animals and plants (Ruby et al., 2006; Pant et al., 2009). The RISC-incorporated miRNA subsequently guides the RISC to ‘select’ mRNAs containing a target site based on Watson–Crick complementarity, after which the RISC down-regulates the expression of the corresponding mRNA. Binding of a miRNA to its target mRNA results in its down-regulation through repression of its translation or induction of its cleavage (Brodersen et al., 2008). Recent work in moss has suggested that some miRNAs can also exert their effect on gene expression by influencing DNA methylation (Khraiwesh et al., 2010).

The field of plant miRNA discovery was initially dominated by the identification of conserved miRNAs that are expressed across diverse angiosperms and often represented by multiple loci in sequenced genomes (reviewed by Axtell and Bowman, 2008). For example, in the miRBASE version 10.1, there are only 39 plant miRNA families present in at least two phylogenetically distant plant species, 21 of which are expressed among diverse angiosperms (Griffiths-Jones et al., 2008). The recent development of high-throughput sequencing technology has facilitated the identification of less widely expressed and less conserved miRNAs that can be clade or species specific; this is exemplified by the report of at least 39 novel, non-conserved rice miRNA families expressed in grains (Zhu et al., 2008). While conserved miRNAs primarily regulate developmentally essential transcription factors (e.g. Aukerman and Sakai, 2003; Combier et al., 2006), non-conserved miRNAs target genes with various specialized functions (e.g. Abdel-Ghany and Pilon, 2008).

Recent work has revealed that miRNA-based gene regulation is essential in coordinating events taking place at the shoot apical meristem (SAM) (reviewed by Bhalla and Singh, 2006; Chuck et al., 2009). The SAM is located at the tip of the shoot and contains a population of stem cells that are capable of self-renewal and generating lateral organs that account for the continuous growth of plants throughout their lifetime. The number of stem cells in the SAM remains almost constant and the underlying regulatory mechanisms are dependent upon reciprocal signalling between stem cells and the niche (Singh and Bhalla, 2006; Tucker and Laux, 2007). Studies in Arabidopsis thaliana and rice have demonstrated that members of the AGONAUTE (AGO) family, the catalytic component of the RISC complex, are essential regulators of SAM as mutants display defective SAM or leaf polarity (Nagasaki et al., 2007; Noguera et al., 2007; Liu et al., 2009b). More specifically, it was shown that AGO10 represses miRNA165/166 (miR165/166) for proper SAM maintenance as well as the establishment of leaf polarity (Liu et al., 2009b). However, there is a lack of information on the expression profile of miRNAs in the SAM.

In this study, we performed high-throughput sequencing analysis of soybean sRNAs from either dissected SAM or mature leaf with the objective of discovering miRNAs that potentially play important roles in regulating the two functions of vegetative SAM: the maintenance of pluripotent stem cells and the initiation of leaf primordia. As the full repertoire of miRNAs that are specifically expressed in cells belonging to different spatial domains of a SAM is unlikely to be represented in sRNA libraries prepared using whole seedlings, we used micro-dissected SAM. We further integrated the large-scale sequencing data with microarray analysis of miRNA expression in micro-dissected SAM and mature leaf tissue. This has resulted in the identification of eight putative novel miRNAs with the expression of seven verified by subsequent microarray study. Further in situ hybridization analysis on selected candidates has implicated the nuclear-localization of one novel miRNA as well as a possible functional role for the star strand of miRNA in plant.

Materials and methods

Cloning and sequencing of soybean sRNAs

Soybean (Glycine max) cultivar Bragg was grown under greenhouse conditions located at the University of Melbourne, Australia. SAMs with leaf primordia were micro-dissected from 10-d-old soybean under the dissecting microscope at 40× magnification. Dissected samples were quickly frozen in liquid nitrogen and stored at –80°C until RNA extraction for the construction of an sRNA library. Mature leaves excluding the main vein were used in the making of the leaf sRNA library.

Subsequent RNA extraction and cDNA library synthesis were carried out by Vertis Biotechnologie AG (Freising, Germany). Briefly, sRNA species of <200 bases were extracted using the mirVana miRNA isolation kit from Ambion (Austin, TX, USA). The sRNAs were then separated on a denaturing 12.5% polyacrylamide gel and stained with SYBR Green II. miRNAs with a length of 15–30 bases were obtained by passive elution of the RNAs from the gel and then precipitated with ethanol and dissolved in water. For cDNA synthesis, the miRNAs were first poly(A)-tailed using yeast poly(A) polymerase followed by ligation of a RNA linker oligo to the 5’-phosphate of the miRNAs. First-strand cDNA synthesis was then performed using an oligo(dT)-linker primer and Moloney murine leukemia virus reverse transcriptase. The resulting cDNA was then PCR amplified for 15 cycles. As the combined length of the oligo(dT)-linker primer and the RNA linker oligo is 80 bases, cDNAs containing inserts of 15–30 bp should have a total length between 95 and 110 bp. This size fraction was obtained by separation on and subsequent elution from a 4% agarose gel. For cloning, the cDNA was subjected to a limited exonuclease treatment to generate 5’ overhangs at both ends of the cDNAs. The size-fractionated cDNA was then directionally ligated into the EcoRI and BamHI sites of the plasmid vector pBSIIsk+ and electroporated into T1-phage-resistant TransforMaxTM EC100TM (Epigenic) electrocompetent cells.

To produce the template for 454 pyrosequencing, three aliquots from each library were sampled and grown in Luria–Bertani medium with selection for 6 h (~12 doubling times). Plasmids were purified from each culture and resuspended in 30 μl of TE buffer. Then 8 μl of plasmid preparation was used as template in 20-μl, 35-cycle PCRs with the Mic5 (5’-AAGAGGACGAGA CAGAACGCAG-3’) and BSF (5’-GACTGGAAAGCGGCG CAGTGCCTG-3’) primers. Replicate PCRs were performed for each original library aliquot. The six PCRs from each library were then combined, purified using Qiagen MinElute PCR Purification Kit and resuspended in 30 μl of elution buffer (10 mM Tris–HCl pH 8.5) to produce the 454 template for sequencing. DNA yield was determined by agarose gel electrophoresis using λ DNA as a quantity standard, 1 μl of each 454 template was demonstrated to contain ~300 ng of DNA. Thus we produced 9 μg of template for
Sequence analysis and prediction of miRNA candidates

Sequencing reads were trimmed of adaptors, low-complexity regions and low-quality sequences before being used for mapping against soybean genome sequence downloaded from ftp://ftp.jgi-psf.org/pub/JGI_data/Glycine_max/assembly/. Repetitive regions of the soybean genome were identified using RepeatMasker software. sRNA sequences that mapped to the repeat region or to >20 loci in the genome were removed from further analysis. sRNAs that were homologous with known tRNA, rRNA, other non-coding RNAs downloaded from the Rfam database (Gardner et al., 2009), or to known protein databases (NCBI) were also excluded from further analysis. sRNAs with perfect matches to the genome sequence together with their precursor sequences were subjected to folding analysis using Mfold software (Zuker, 2003). Targets of novel miRNAs were predicted using a web-based application psRNATarget available at http://bioinfo3.noble.org/psRNA_Target/ with default settings against G. max DCFI Gene Index Release 12 (Zhang, 2005).

Microarray analysis of miRNA expression profile

Arrays were custom built on a medium density (4000-feature) modified oligonucleotide microfluidic platform by LC Sciences (Houston, TX, USA). Arrays were queried with Cy5-labelled short RNAs purified from ~8 µg of total RNA using Millipore centrifugal filters YM-100 and YM-3, which enrich for RNAs of between 10 and ~300 nt. Probe synthesis was performed in a µParaffo microfluidic chip at LC Sciences. Locked nucleic acid (LNA)-modified RNA probes were synthesized in situ using photogenerated acid coupled with conventional DMT chemistry (LNA modification ensures Tm normalization). A spacer segment of polyethylene glycol was used to extend the hybridizing segment away from the substrate. Purified sRNAs were 3’-extended with a poly(A) tail using poly(A) polymerase, followed by ligation of a Cy5 fluorophore-conjugated oligonucleotide tag to the poly(A) tail. Hybridization was accomplished overnight at 34°C using amicro-circulation pump (Atactic Technologies, Houston, TX, USA), using 100 µl of 6×SSPE buffer (0.90M NaCl, 60mM Na2HPO4, 6mM EDTA pH 6.8) containing 25% formalamide. Fluorescence was determined by laser scanning using an AxonGenePix 4000B Microarray Scanner (10µm pixel size; Molecular Devices, Inc., Sunnyvale, CA, USA). The image was processed using ArrayPro (Media Cybernetics, Bethesda, MD, USA) with a morphological filter open, 3×3 cross, two passes. Background signal was calculated from the average background probes, which contained no oligonucleotide probes. The signal was determined by spike-in normalization and LOWESS (locally-weighted regression).

In addition to control probes added by the manufacturer, positive control probes were added for known mature miRNAs, and 5S RNAs. Sets of negative control probes for sRNAs derived from degraded mRNAs (ACTIN and GADPH) were also added. These probes remained in the background signal range in all experiments. In other unpublished experiments, we have shown that single mismatches in probes produce amedian order of magnitude loss of signal strength, though the drop-off strength is strongly position dependent.

In situ hybridization of miRNAs and sRNA gel blot analysis

Fixation of soybean shoot apices and subsequent in situ hybridization was carried out according to Wong et al. (2009) with the exception that the overnight hybridization of sections with probes was carried out at a temperature that was 15°C below the Tm calculated for each probe. Probes used were modified LNA labelled with digoxigenin at the 3’ end purchased from Exiqon (Woburn, MA, USA). As for gel blot analysis of miRNA, total RNA was isolated from tissues using TRI reagent (Sigma) according to the manufacturer’s instructions and subsequent analysis was carried out according to Kim et al. (2010).

Results and discussion

Sequencing and analysis of soybean sRNAs

To identify miRNAs that play roles in regulating activities taking place in the SAM, one sRNA library was constructed from dissected soybean SAMs (Fig. 1) and another was constructed from mature leaf tissue. The libraries were sequenced by 454 Life Sciences using pyrosequencing technology, producing ~82000 reads. Following quality processing (see Materials and methods), sequences were aligned to the reference soybean genome sequence (downloaded from ftp://ftp.jgi-psf.org/pub/JGI_data/Glycine_max/assembly/) and those that matched the soybean genome sequence perfectly were subjected to further computational analysis after filtering for previously annotated features including messenger RNAs, tRNAs, tRNAs, other known non-coding RNAs, and protein-coding genes. This resulted in a total of 8423 unique sRNA sequences, representing 14950 clone reads (Table 1) and most of these sRNAs were 20–21 nt in length (Fig. 2).

Subsequently, we focused on sRNAs that were 20–23 nt in length (5463 sRNA sequences, representing 10166 clone reads) as most miRNAs discovered to date are of this size range. The flanking regions of these sRNAs matching the soybean genome were subjected to secondary structure prediction for each locus. Candidate miRNAs were then selected as those that fulfilled the hairpin structure criteria previously defined (Meyers et al., 2008), and representatives of these hairpin structures are shown in Fig. 3. There are a total of 32 known miRNA families comprising 62 members...

Fig. 1. Scanning electron micrograph of a soybean shoot apex from a 10-d-old plant (polar view). The tissues dissected for constructing the SAM sRNA library are indicated in red while the SAM used for microarray analysis is in yellow. LP, leaf primordia; TL, trifoliate leaflet; S, stipule. Scale bar 100 µm.
represented by our dataset with perfect matching or up to 2 nt variation while eight sequences are putative novel miRNA candidates (Supplementary Table S1 available at JXB online).

The detection of miRNA* sequences has been put forward as strong evidence for the DCL-mediated cleavage of the candidate miRNA during its biogenesis; hence the identification of sequences representing both the miRNA and miRNA* is one primary criterion for miRNA annotation (Meyers et al., 2008). We searched for corresponding miRNA* sequences for predicted miRNAs, and since the average frequency of miRNA* is ~10% of that of the miRNA (Rajagopalan et al., 2006), one putative novel miRNA, and 38 conserved miRNA candidates with corresponding miRNA* were detected in our libraries (Supplementary Table S1 at JXB online).

Expression analysis of miRNAs in SAM and mature leaf

We subsequently developed a soybean miRNA microarray using miRNA and several miRNA* sequences derived from this study as well as other soybean miRNAs available at miRBASE (Griffiths-Jones et al., 2008) (see Materials and methods). We utilized this array to compare the repertoire of miRNAs in the SAM (Fig. 1) and mature leaf as well as to verify the expression of novel miRNA candidates identified. A total of 31 miRNAs or 42 miRNAs were detected to be

**Table 1. Summary of sRNA sequencing and analysis**

<table>
<thead>
<tr>
<th></th>
<th>SAM</th>
<th>Leaf</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Raw reads</td>
<td>Non-redundant reads</td>
</tr>
<tr>
<td>Total sequence reads</td>
<td>30173</td>
<td>22231</td>
</tr>
<tr>
<td>sRNAs mapped to the G.max genome assembly</td>
<td>14791</td>
<td>8814</td>
</tr>
<tr>
<td>Removed sRNAs mapped to repeat regions</td>
<td>7721</td>
<td>3854</td>
</tr>
<tr>
<td>Removed sRNAs with &gt;20 loci</td>
<td>1393</td>
<td>1081</td>
</tr>
<tr>
<td>Removed structural RNAs</td>
<td>347</td>
<td>302</td>
</tr>
<tr>
<td>sRNAs with length 20–23 nt</td>
<td>4233</td>
<td>2622</td>
</tr>
<tr>
<td>sRNAs mapped to predicted hairpins</td>
<td>2163</td>
<td>243</td>
</tr>
<tr>
<td>20- to 23-nt sRNAs mapped to predicted hairpins</td>
<td>1931</td>
<td>183</td>
</tr>
<tr>
<td>Conserved miRNAs(^a)</td>
<td>758</td>
<td>42(28)</td>
</tr>
<tr>
<td>Non-conserved miRNAs(^a)</td>
<td>5</td>
<td>3(1)</td>
</tr>
</tbody>
</table>

\(^a\) Number of miRNAs with miRNA* detected in the libraries with up to 2 nt variants is given in parentheses.

**Fig. 2.** Size distribution of: (A) sequence reads and (B) non-redundant clones of filtered soybean sRNAs from SAM or leaf library.

**Fig. 3.** Novel soybean miRNAs identified by 454 sequencing analysis. Putative secondary structures for three novel miRNAs uncovered in this study are shown with red indicating the mature miRNA and blue indicating the miRNA* detected for gma-miR4994.
expressed in the SAM or leaf, respectively (Supplementary Table S2 at JXB online). The expression of a greater diversity of miRNAs in the leaf in comparison with the SAM likely reflects the structural complexity of the leaf tissues as well as multiple metabolic and developmental activities taking place in the leaves that require regulation by miRNAs.

Although there is a general trend of agreement between the frequency of miRNAs represented in the library and the intensity reading detected by microarray (Table 2, 3), there is a lack of concordance for some miRNAs (Supplementary Table S1 at JXB online) and this may result from insufficient depth of sequencing coverage or cloning bias. There are seven novel miRNAs with expression verified using microarray (Table 2). We consider these sequences as novel miRNAs.

Novel miRNAs

Our computational approach has predicted a total of eight putative novel miRNAs and subsequent microarray experiment verified the expression of seven of these novel miRNAs including one miRNA family with cloned miRNA*s (Table 2). To determine whether any of these novel miRNAs are specific to soybean, we interrogated the genome sequences of grape, Arabidopsis, rice, and poplar for miRNA genes homologous to these novel miRNAs; only one out of seven novel miRNAs appear to be conserved in the species examined (Table 2). This implies that six of the novel miRNAs are likely to be species or legume specific.

Subsequently, the spatial expression pattern was investigated for two novel miRNAs, miR4994 and miR4422. Unfortunately, no clear signal was detected for miR4994 (Fig. 4A). The failure to detect the expression of miR4994 by in situ hybridization is most likely due to its low expression level as evident from the microarray data (Table 2). Meanwhile, the expression of miR4422 is most fascinating as its expression is nuclear-localized (Fig.4B) and on closer inspection, it seems to be largely confined to spots in nucleoli of most cells in the central and peripheral zone while its expression becomes diffused in nuclei of cells in the rib zone (Fig.4C). When a similar section was stained with 4’,6-diamidino-2-phenylindole (DAPI), there was intense fluorescence associated with the genomic DNA in the nucleus but such signal was very weak in the nucleoli (Fig. 4D) in contrast to the miR4422 signals suggesting that the signal observed is specific to the miRNA. When northern blot analysis was carried out for the novel miRNA, we successfully detected the miRNA in both SAM and mature leaf tissue (Fig. 4E) consistent with the microarray data (Table 2). It is tempting to speculate that this miRNA may play a role in regulating gene expression in the nucleus, e.g. by regulating transcription of target transcripts that play roles in cell differentiation. Nuclear-localized miRNAs are yet to be documented in plants, but in animals there are increasing reports of nuclear-localized miRNAs playing essential nuclear roles and some miRNAs that may revisit the nucleus once exported to the cytoplasm (Hwang et al., 2007; Politz et al., 2006, 2009).

Table 2. Novel miRNAs predicted from the soybean sRNA sequence data miRNAs verified by microarray analysis are shown. Targets of novel miRNAs were predicted using a web-based application psRNATarget available at http://bioinfo3.noble.org/psRNATarget/ with default settings against G. max DCFI Gene Index Release 14 (Zhang, 2005).

<table>
<thead>
<tr>
<th>miRNA family</th>
<th>miRNA ID</th>
<th>Sequence</th>
<th>Loci</th>
<th>miRNA*</th>
<th>Frequency</th>
<th>Cy5 processed intensity signal</th>
<th>Predicted target</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR4993</td>
<td>gma-miR4993</td>
<td>GAGCGGCGGGCGGUGGAGGAUG</td>
<td>1 N 1 0 63 10</td>
<td></td>
<td></td>
<td></td>
<td>Acetolactate synthase 2, glycosyltransferase-like protein</td>
</tr>
<tr>
<td>miR4994</td>
<td>gma-miR4994</td>
<td>GGUUGUCUCAAGGAGCUACAC</td>
<td>1 Y 2 0 62 57</td>
<td></td>
<td></td>
<td></td>
<td>Tetratricopeptide-like helical, gibberellin 2-oxidase</td>
</tr>
<tr>
<td>miR4422</td>
<td>gma-miR4422a</td>
<td>AUAGCUUGGCGGGCGGACACCU</td>
<td>1 N 0 3 1900 487</td>
<td></td>
<td></td>
<td></td>
<td>Unknown protein</td>
</tr>
<tr>
<td>miR4995a</td>
<td>gma-miR4995</td>
<td>AGGCAGUGCUUUGGUAAAGG</td>
<td>1 N 0 1 8 123</td>
<td></td>
<td></td>
<td></td>
<td>ATP sulfurylase precursor</td>
</tr>
<tr>
<td>miR4424</td>
<td>gma-miR4424a</td>
<td>AGAACAAGGUGUUGUUUACCGG</td>
<td>1 N 0 1 65 4</td>
<td></td>
<td></td>
<td></td>
<td>MYB transcription factor</td>
</tr>
<tr>
<td>miR4996</td>
<td>gma-miR4996</td>
<td>UAGAACUGCCCAUGUUCUC</td>
<td>1 N 0 9 15 1184</td>
<td></td>
<td></td>
<td></td>
<td>Polyphenol oxidase, Mth19 protein</td>
</tr>
<tr>
<td>miR4997</td>
<td>gma-miR4997</td>
<td>GAUGCUCGAAGCGGCGAAGUGAGG</td>
<td>1 N 0 2 67 16</td>
<td></td>
<td></td>
<td></td>
<td>Hypothetical protein</td>
</tr>
</tbody>
</table>

* Novel miRNA likely to be conserved in other plant species.
Spatial expression of miR166 and miR166*

As the polarity of leaves is established in the SAM, it was not surprising to identify a number of miRNAs reported to regulate such activity that are expressed in the SAM. The miRNAs concerned include miR166 and miR390 (Table 3). While the expression of miR166 is consistent with its role in leaf development, the high expression of its miRNA* revealed by both deep sequencing and microarray.

**Fig. 4.** Characterization of novel miRNAs. (A) *In situ* hybridization analysis of gma-4994 with no signal being detected. (B) *In situ* hybridization analysis of gma-miR4422a with longitudinal sections of soybean shoot apices hybridized with complementary LNA probe. (C) Close-up of SAM from (A). (D) DAPI staining of a similar SAM section demonstrating the intense staining of the genomic DNA in the nucleus but lesser staining in the nucleoli in contrast to the signals associated with gma-miR4422a. (E) Expression analysis of gma-miR4422a using: lane 1, 20 μg of soybean leaf total RNA; lane 2, 10 μg of soybean SAM total RNA; lane 3, 20 μg of aphid total RNA as negative control. The expression analysis was carried out according to Kim *et al.* (2010). LP, leaf primordia.

**Table 3.** Conserved miRNAs identified from the deep sequencing of soybean sRNA libraries with expression verified by microarray analysis. Signals >50 are above background in the microarray analysis. Some miRNAs have clones of different lengths as a result of sequence heterogeneity at the RNA ends. The sequence of the most abundant clone is shown.

<table>
<thead>
<tr>
<th>miRNA family</th>
<th>miRNA ID</th>
<th>Loci</th>
<th>miRNA*</th>
<th>Frequency</th>
<th>Cy5 processed intensity signal</th>
<th>Predicted or known targets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SAM</td>
<td>Leaf</td>
<td>SAM</td>
</tr>
<tr>
<td>miR156</td>
<td>gma-miR156f</td>
<td>1</td>
<td>N</td>
<td>0</td>
<td>2</td>
<td>29</td>
</tr>
<tr>
<td>miR156</td>
<td>gma-miR156h</td>
<td>2</td>
<td>N</td>
<td>0</td>
<td>2</td>
<td>40</td>
</tr>
<tr>
<td>miR159</td>
<td>gma-miR159a</td>
<td>1</td>
<td>Y</td>
<td>256</td>
<td>310</td>
<td>18139</td>
</tr>
<tr>
<td>miR160</td>
<td>gma-miR160</td>
<td>4</td>
<td>Y</td>
<td>2</td>
<td>12</td>
<td>112</td>
</tr>
<tr>
<td>miR162</td>
<td>gma-miR162b</td>
<td>1</td>
<td>Y</td>
<td>1</td>
<td>2</td>
<td>13</td>
</tr>
</tbody>
</table>

SQUAMOSA PROMOTER BINDING PROTEIN-LIKE family of transcription factors (Wu and Poethig, 2006); SQUAMOSA PROMOTER BINDING PROTEIN-LIKE family of transcription factors (Wu and Poethig, 2006); MYB transcription factors (Achard *et al.*, 2004; Allen *et al.*, 2007); Auxin response factors (Mallory *et al.*, 2009); DCL1 (Xie *et al.*, 2003).
<table>
<thead>
<tr>
<th>miRNA family</th>
<th>miRNA ID</th>
<th>Sequence</th>
<th>Loci</th>
<th>miRNA*</th>
<th>Frequency</th>
<th>Cy5 processed intensity</th>
<th>Predicted or known targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR164</td>
<td>gma-miR164</td>
<td>UGGAGAAGCAGGCACGUGC</td>
<td>3 Y</td>
<td>0 4</td>
<td>577 2635</td>
<td>NAC transcription factor (Raman et al., 2008)</td>
<td></td>
</tr>
<tr>
<td>miR166</td>
<td>gma-miR166</td>
<td>UCGGACCAGGCUCUCAUCCCC</td>
<td>5 Y</td>
<td>10 10</td>
<td>416 720</td>
<td>HD-ZIP transcription factors (Juarez et al., 2004; Williams et al., 2005)</td>
<td></td>
</tr>
<tr>
<td>gma-miR166a/b*</td>
<td></td>
<td>GGAUUGUUGCGGUCGCAGG</td>
<td>1 Y</td>
<td>177 16</td>
<td>586 292</td>
<td>Unknown proteina</td>
<td></td>
</tr>
<tr>
<td>miR167</td>
<td>gma-miR167</td>
<td>UGGAGCUGGCAUGAUCCU</td>
<td>4 N</td>
<td>49 4</td>
<td>673 5987</td>
<td>Auxin response factors (Wu et al., 2006)</td>
<td></td>
</tr>
<tr>
<td>miR168</td>
<td>gma-miR168b</td>
<td>UCUCUUGGUCGUGGCGG</td>
<td>1 Y</td>
<td>35 47</td>
<td>1643 6360</td>
<td>ARGONAUTE1 (Vaucheret et al., 2004)</td>
<td></td>
</tr>
<tr>
<td>miR169</td>
<td>gma-miR169a</td>
<td>UGCCAAGGACUGUCCGAGG</td>
<td>3 Y</td>
<td>0 1</td>
<td>813b 8c</td>
<td>Nuclear factor Y (NF-Y) transcription factor (Li et al., 2008)</td>
<td></td>
</tr>
<tr>
<td>miR171</td>
<td>gma-miR171a/b</td>
<td>UGGAGCUGGCAUGAUCGCA</td>
<td>4 N</td>
<td>16 4</td>
<td>34 61</td>
<td>SCARECROW-LIKE transcription factors (Síné et al., 2009)</td>
<td></td>
</tr>
<tr>
<td>miR319</td>
<td>gma-miR319a</td>
<td>UUGCACUGAGGGAGCUCCC</td>
<td>2 Y</td>
<td>4 0</td>
<td>5320 11183</td>
<td>TCP transcription factor (Palatnik et al., 2003)</td>
<td></td>
</tr>
<tr>
<td>gma-miR319b</td>
<td></td>
<td>UUGCACUGAGGGGCCCCCUU</td>
<td>1 N</td>
<td>5 0</td>
<td>652 88</td>
<td>TCP1 transcription factor, boron transporter, transcription factor PCF6, eukaryotic translation initiation factor</td>
<td></td>
</tr>
<tr>
<td>miR482</td>
<td>gma-miR482b/c</td>
<td>UGCCAAGGACGUCGCCUCCU</td>
<td>1 Y</td>
<td>4 49</td>
<td>4 129</td>
<td>Cytochrome c oxidase, resistance-like proteina</td>
<td></td>
</tr>
<tr>
<td>gma-miR482b</td>
<td></td>
<td>UCUUCUAAAGGCACGCUCCUU</td>
<td>1 Y</td>
<td>0 8</td>
<td>211 1316</td>
<td>Disease resistance-like proteina</td>
<td></td>
</tr>
<tr>
<td>gma-miR482b*</td>
<td></td>
<td>UAGGGGGAUUGGGGAAGG</td>
<td>1 NA</td>
<td>3 9</td>
<td>455 6041</td>
<td>No target predicted with the default settinga</td>
<td></td>
</tr>
<tr>
<td>miR2118</td>
<td>gma-miR2118a/b</td>
<td>UUGGCGAUCACCCGCAUCCU</td>
<td>2 Y</td>
<td>211 922</td>
<td>2994 12554</td>
<td>Zinc finger protein HTR, disease resistance proteina</td>
<td></td>
</tr>
<tr>
<td>miR530</td>
<td>gma-miR530a</td>
<td>UGCAUUGACGUCGCCUU</td>
<td>1 Y</td>
<td>0 6</td>
<td>0 127</td>
<td>CONSTANS interacting protein 2a, nuclear transcription factor Ya</td>
<td></td>
</tr>
<tr>
<td>miR1507</td>
<td>gma-miR1507a-5p</td>
<td>AGAGGGUGUAUGGAGAGAG</td>
<td>1 N</td>
<td>9 3</td>
<td>241 235</td>
<td>WD40-like, actin depolymerizing factora</td>
<td></td>
</tr>
<tr>
<td>gma-miR1507c</td>
<td></td>
<td>CCUCAUUCACCCAGCAUCCU</td>
<td>1 Y</td>
<td>13 34</td>
<td>4 13</td>
<td>NBS-LRR resistance-like protein, MYB transcription factor, actin8</td>
<td></td>
</tr>
<tr>
<td>gma-miR1507c*</td>
<td></td>
<td>AGAGGGUGUAUGGAGAGAG</td>
<td>1 NA</td>
<td>18 0</td>
<td>94 105</td>
<td>AP2 domain transcription factor, heat shock protein, polygalacturonase-inhibiting protein9</td>
<td></td>
</tr>
<tr>
<td>miR1508</td>
<td>gma-miR1508a</td>
<td>UGAAAGGGAAUUGACGUGG</td>
<td>2 Y</td>
<td>0 8</td>
<td>612 19831</td>
<td>Calcium-dependent protein kinasea</td>
<td></td>
</tr>
<tr>
<td>miR1509</td>
<td>gma-miR1509a</td>
<td>UUAUCAAGAAUUACACGUGU</td>
<td>1 Y</td>
<td>1 6</td>
<td>1571 4088</td>
<td>Hypothetical protein8</td>
<td></td>
</tr>
<tr>
<td>miR1510</td>
<td>gma-miR1510a</td>
<td>UGUUGUUUGUACUUCUCCACC</td>
<td>2 Y</td>
<td>2 17</td>
<td>14 20</td>
<td>NBS-LRR type disease resistance protein, brassinosteroid receptor, hydrolasea</td>
<td></td>
</tr>
</tbody>
</table>
expression analysis is most intriguing (Table 3). The abundance of miR166* may imply regulatory roles since inhibitory activity of miRNA* species has been demonstrated in *Drosophila melanogaster* (Okamura et al., 2008). In soybean, the expression of miR166 was found below and on the abaxial side of the incipient leaf (Fig. 5A) in line with that of maize (Nogueira et al., 2007, 2009). Results from two independent experiments showed that for miR166*, faint signals were observed at the tip of leaf primordia and the peripheral region of the SAM with signals becoming stronger and punctate in older leaf primordia (Fig. 5B). There is thus a distinct expression pattern between miR166 and its star strand suggesting that they serve different roles in regulating leaf development. Further transgenic experiments that involve the silencing of either strand of miRNA will reveal the function of this miRNA*. Nevertheless, the expression of both star and mature strands of the candidate miRNA most likely reflects the diversification of miRNA function during evolution. Recent studies are leading to the discovery of bi-functional miRNAs where both miRNA and its corresponding star form target different genes belonging to the same developmental pathway. For example, miR9 and miR9* regulate human transcriptional repressor REST (RE1-silencing transcription factor) and its co-repressor CoREST, respectively (Packer et al., 2008).

miR159 expression in soybean

Among the conserved miRNAs that are expressed in either SAM or leaves, the most highly expressed miRNA in both tissues is miR159 (Table 3). The miR159 family has been recorded in all flowering plants examined to date (Axtell and Bartel, 2005). In Arabidopsis, miR159 regulates short-day photoperiod flowering time and anther development via...
modulation of the expression of several MYB transcription factors (Achard et al., 2004). Recent work has revealed that members of the miR159 family regulate the expression of MYB33 and MYB65 redundantly and that only the double mutant mir159ab displayed pleiotropic developmental defects including reduced apical dominance and curled leaves (Allen et al., 2007). A miRNA that is very closely related to miR159 is miR319. In Arabidopsis, it has been shown that the specificity of these closely related miRNAs are achieved in part by their different expression levels with miR319 expressed at a much lower level than miR159 (Palatnik et al., 2007). Though this trend of low expression is not conserved in soybean according to our microarray data (Table 3), further experiments are necessary to rule out any possibility that the signal detected on the array is due to miR159 cross-hybridizing with the miR319 probe.

To reveal the spatial expression patterns and to obtain further insights into the function of miR159 in soybean, we carried out in situ localization of miR159 in the soybean shoot apex. As shown in Fig. 5C, signals associated with the expression of miR159 were detected throughout the SAM and leaf primordia. This spatial expression pattern together with its high expression level in the leaf is consistent with the expression reported for miR159 in Arabidopsis (Allen et al., 2007), suggesting a conserved function of this miRNA in soybean. Future studies aimed at elucidating the targets of the miRNA in the SAM may reveal miR159-regulated pathways associated with SAM function.

**Other conserved miRNAs**

The microarray data showed that miR390, miR160, and miR169 are the most differentially expressed conserved miRNAs in the SAM, relative to expression in mature leaf tissue (Table 3). When the spatial expression pattern for miR390 was examined, signals were detected in the outer peripheral zone of the soybean SAM, in cells that will leave the meristem to form lateral organs, and in the adaxial side of the emerging leaf primordia (Fig. 5D). However, in older leaf primordia, the adaxial expression became striated extending from the lateral to the marginal domain of the leaf suggesting a novel role of miR390 in regulating leaf vein formation (inset of Fig. 5D).

Both miR390 and miR160 indirectly or directly target ARF (auxin responsive factor) transcripts (Wang et al., 2005; Nogueira et al., 2007). miR390 initially targets the non-coding TAS3 transcript for cleavage, which is subsequently amplified by RDR6 and processed by DCL4 to produce ta-siRNAs that target ARF2, ARF3, and ARF4. These ARFs have been implicated in the control of leaf polarity (Nogueira et al., 2007). miR160 directly targets ARF10, ARF16, and ARF17, and transgenic Arabidopsis expressing a mutant arf16 that is not recognized by miR160, displays severe defects in shoot development, including leaf abnormalities (Wang et al., 2005).

miR169 was abundant in the soybean SAM but undetectable using sRNA from mature leaves on microarrays.
This miRNA targets nuclear factor Y (NF-Y) transcription factors, and in Arabidopsis, it is down-regulated by drought stress through an abscisic acid-dependent pathway (Li et al., 2008). The functional significance of the abundant miR169 expression in the soybean SAM remains to be determined.

The expression pattern of miR396 as observed in our data is interesting (Table 3). While miR396 is highly expressed in mature soybean leaves, no detectable signal was observed in the SAM. This miRNA has been shown to target GROWTH REGULATING FACTOR (GRF1) (Liu et al., 2009a) and GRF1 was reported to be expressed strongly in Arabidopsis shoot tips and flower buds but weakly in mature stem and leaf tissues (Kim et al., 2003). In fact, our recent transcript profiling work comparing the expression profile of SAM with non-meristem tissues in soybean has identified a number of GRFs to be up-regulated in the SAM (GmaAffx.4788.2.1_at, GmaAffx.75820.1.S1_at, Gma.16796.1.S1_at, Gma16741.1.S1_at, Gma.11947.1.S1_at) with log2 fold change ranging from 2.9 to 4.9 (Haerizadeh et al., 2009). A closer inspection of the sequences reveals a highly conserved region of 20 nt among the identified GRF transcripts that seemed to be targeted by miR396 (Fig.6), implying an inverse correlation between the expression of miR396 and its targets. This inverse correlation between a miRNA and its target genes has recently been demonstrated in Arabidopsis (Rodriguez et al., 2010) and soybean (Joshi et al., 2010). The absence of miR396 in the SAM more than likely allows its targets to be highly expressed in the SAM and the developing leaf primordia to exert their functions. On the other hand, the abundant presence of miR396 in mature leaves may be critical to remove GRF transcripts, perhaps to stabilize gene expression patterns necessary for the proper functioning of leaves. Interestingly, recent transcriptome analysis of laser micro-dissected domains of maize SAM showed up-regulated expression of Zm-grf1 in the SAM periphery and leaf primordia of wild type, and down-regulation in leaf-arrested mutant SAMS (Brooks et al., 2009).

Our study reveals that miR164 is expressed at much higher levels in mature soybean leaves as compared with SAM (Table 3). The targets of miR164 are NAC-domain transcription factors that include CUP-SHAPED COTYLEDON1 (CUC1) and CUC2 (Mallory et al., 2004; Raman et al., 2008). In Arabidopsis, CUC1, CUC2, and CUC3 function redundantly in initiating the SAM and establishing organ boundaries (Aida et al., 1997). It has been demonstrated that miR164-directed control of CUC1 is critical for normal embryonic, vegetative, and floral development (Mallory et al., 2004). Our data on the expression of miR164 in soybean SAMs agree with the conclusion that proper dosage or localization of this small regulatory miRNA is required for the separation of adjacent embryonic, vegetative, and floral organs. In this instance, the presence of miR164 is necessary to modulate rather than fully down-regulate the expression of its target genes.

From the microarray data, one other highly expressed conserved miRNA in both SAM and leaf tissue is miR2118a/b which is consistent with the frequency of the miRNA clone detected from the sequencing (Table 3). Targets predicted for this miRNA include transcripts encoding a putative HTR (Hordeum repressor of transcription) zinc finger transcription repressor and a disease resistance protein. The HTR transcription repressor has been shown to repress expression from certain promoters in wheat, and proposed to mediate developmental and phytohormone-responsive gene expression (Raventos et al., 1998). If miR2118a/b expression in the SAM is to modulate the HTR level, then the putative miRNA is likely to play a crucial role in mediating key developmental activities within the SAM. Meanwhile, a number of other novel miRNAs are predicted to target transcripts that encode disease resistance proteins (Table 3). Since similar disease-related sequences have been shown to be repressed in SAM in comparison with non-meristem tissues (Wong et al., 2008; Liang et al., 2009), the more abundant expression of miR2118a/b in the SAM in particular may be down-regulating expression of their targets in the SAM.

In summary, integrating a sequencing approach with miRNA microarray expression profiling has allowed us to identify eight putative novel miRNAs in soybean. Subsequent in situ hybridization analysis has revealed a nuclear-localized miRNA. We also discovered a distinct expression pattern between a conserved miRNA and its star species implicating a novel biological function for miRNA* in plants. Moreover, profiling of miRNAs in the soybean SAM will pave the way towards understanding the regulatory circuits underlying developmental processes in the SAM.

Supplementary data

Supplementary data are available at JXB online.

Table S1. Conserved and novel miRNAs identified from the soybean sRNA sequence data.

Table S2. Microarray intensity data for soybean miRNAs.

Acknowledgements

We gratefully acknowledge financial support from the Australian Research Council in the form of the ARC Centre of Excellence grant (CE0348212) for this project.

References


