GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Prediction of potential novel microRNAs in soybean when in symbiosis

G.A. Barros-Carvalho[1,2], A.R. Paschoal[3], F.C. Marcelino-Guimarães[1] and M. Hungria[1]

[1]Centro Nacional de Pesquisa de Soja,
Empresa Brasileira de Pesquisa Agropecuária, Londrina, PR, Brasil
[2]Departamento de Bioquímica e Biotecnologia,
Universidade Estadual de Londrina, Londrina, PR, Brasil
[3]Universidade Tecnológica Federal do Paraná, Cornélio Procópio, PR, Brasil

Corresponding author: M. Hungria
E-mail: mariangela.hungria@embrapa.br

Genet. Mol. Res. 13 (4): 8519-8529 (2014)
Received December 11, 2013
Accepted March 31, 2014
Published October 20, 2014
DOI http://dx.doi.org/10.4238/2014.October.20.28

ABSTRACT. MicroRNAs (miRNAs) are small molecules, noncoding proteins that are involved in many biological processes, especially in plants; among these processes is nodulation in the legume. Biological nitrogen fixation is a key process, with critical importance to the soybean crop. This study aimed to identify the potential of novel miRNAs to act during the root nodulation process. We utilized a set of transcripts that were differentially expressed in soybean roots 10 days after inoculation with *Bradyrhizobium japonicum*, which were obtained in a previous study, and performed a set of computational analyses that led us to select new miRNAs potentially involved in nodulation. Among these analyses, the set of transcripts were submitted to an *in silico* annotation of noncoding RNAs, including a search of similarity against miRNA public databases, *ab initio* tools for miRNA identification, structural search against miRNA families, prediction of the secondary structure of miRNA precursors, and prediction of the sequences of mature miRNAs.

Subsequently, we applied filter procedures based on miRNA selections described in the literature (e.g., free energy value). In the next step, a manual curation inspection of the annotation was performed and the top candidates were selected and used for prediction of potential target genes, which were later checked manually in the database of the soybean genome. This prediction led us to the identification of 9 potential new miRNAs; among these, 4 were conserved in other plants. Moreover, we predicted their target genes might play important roles in the regulation of nodulation.

Key words: Computational prediction; *Glycine max*; MicroRNAs; Nodulation; Target genes

INTRODUCTION

MicroRNAs (miRNAs) are small molecules of single-stranded RNA, noncoding proteins, with an approximate size of 21-24 nucleotides (nt), that negatively regulate target genes at the post-transcriptional level in eukaryotes (Jung et al., 2009; Voinnet, 2009). There are differences between miRNAs of plants and animals, especially in terms of functional mechanisms and biogenesis (Jung et al., 2009). Unlike most animals, plants generally have high complementarity, if not perfect, between miRNAs and their targets, resulting in the regulation of gene expression by cleaving the coding regions of messenger RNAs (mRNAs) (Jung et al., 2009). miRNAs have great importance in biological systems, especially in plants, and regulate a variety of processes such as growth and development, and responses to biotic and abiotic stresses. For example, Kulcheski et al. (2011) identified new miRNAs in the soybean that are responsive to biotic and abiotic stresses. Moreover, miRNAs are involved in gene regulation during nodulation of legumes by rhizobia (Jung et al., 2009; Simon et al., 2009), as shown in *Medicago truncatula*, where the transcription factor MtHAP2 is regulated by mir-169 (Combier et al., 2006). Another important study conducted by Joshi et al. (2010) expanded our knowledge on miRNAs and showed that they can regulate a variety of processes such as growth and development in the soybean by identification of 87 new miRNAs in the different tissues of this legume (e.g., seed, flower, root, and nodule). Currently, the soybean [*Glycine max* (L). Merrill] is the most important legume crop worldwide, particularly as a rich source of protein for human and animal consumption (Hungria et al., 2006). It is well known that the process of biological nitrogen fixation (BNF) is very important for increasing the productivity of this legume and has economic and environmental advantages as a replacement for N-fertilizers (Hungria et al., 2006). Given the complexity of the nodulation process, the identification of novel miRNAs that participate in the complex process of BNF can highly contribute to our knowledge on the coordination of the process between symbionts. The first study to identify miRNAs in the nodulated soybean was conducted by Subramanian et al. (2008), who studied the role of miRNAs in the symbiosis of the soybean-*Bradyrhizobium* and identified 35 new miRNA families. The second study on miRNAs involved in soybean nodulation was published the following year by Wang et al. (2009), who identified 32 miRNA sequences. Despite the current knowledge of miRNAs involved in the regulation of nodulation, activity during the initial stages (e.g., the regulation of homeostasis and auxin signaling process) and nodule maturation (e.g., the nitrogen fixation process) (Subramanian et al., 2008; Wang et al.,

2009; Li et al., 2010), their functions are still not fully understood. Computational strategies have proven to be successful, highly effective, and important in identifying new miRNAs, as shown by the studies of Qiu et al. (2007), Xie et al. (2007), and Lu and Yang (2010), among others, where new miRNAs and target genes were identified by *in silico* analysis based on expressed sequence tags of *Gossypium hirsutum*, *Brassica napus*, and *Vigna unguiculata*, respectively. Our study performed an *in silico* analysis of miRNAs and their target genes present in a subtractive library of soybean roots inoculated with the *Bradyrhizobium japonicum* strain CPAC 15 obtained in a previous study (Barros de Carvalho et al., 2013). We aimed to identify new miRNAs that could be involved in regulation of nodulation in the early developmental stages of soybean roots.

MATERIAL AND METHODS

Obtaining the sequences for computational analysis

To search potential novel miRNAs involved in nodulation, we used a subtractive library dataset (available in browse, subtractive libraries: http://www.lge.ibi.unicamp.br/soybean), which contains 3776 sequences differentially expressed in soybean roots obtained at 10 days after inoculation (DAI) with *B. japonicum* strain CPAC 15 (non-nodulated x nodulated plants). This subtractive library has been integrated into the GENOSOJA project [see Barros de Carvalho et al. (2013) for more details].

Identification of miRNAs involved in nodulation

In order to predict potential miRNAs in the subtractive library, we submitted the sequence to an *in silico* annotation of noncoding RNAs. The steps of this annotation are summarized in **Figure S1**. The annotation process included a search of structure by using the INFERNAL software of the Rfam database data (Nawrocki et al., 2009), the BLAST program to search for similarity against miRNAs in public databases based on the NRDR database (Paschoal et al., 2012), and an *ab initio* miRNA prediction tool to infer predictions of biological features using only a computational model (Kadri et al., 2009). Based on the final report, sequences with the strongest evidence of miRNAs were selected and subjected to manual annotation against the miRBase database (version 18) (Griffiths-Jones et al., 2006) in which comparisons were made between the sequences of candidate pre-miRNAs and deposited stem-loop sequences (BLASTn).

To reinforce the hypothesis that these sequences of pre-miRNAs can be accurately processed in mature miRNAs, the candidates were evaluated according to criteria previously suggested (Ambros et al., 2003; Joshi et al., 2010; Kulcheski et al., 2011); the sequence of pre-miRNA should: 1) form a secondary structure (stem-loop), 2) have low free energy [i.e., -20 kcal/mol, according to Thakur et al. (2011)], 3) have a duplex (miRNA:miRNA*) that is inserted in the "arm" of the stem-loop structure, and 4) restrict pairing between the sense and antisense miRNA to $\leq$4 mismatches. The prediction of secondary structures of miRNA precursors was performed with the Mfold software using the default parameters; this analysis was based on the formation of the secondary structure and minimum free energy (ΔG) (Zuker, 2003).

The miRNA prediction was based on the computational identification of the sequence and position of the mature miRNA within the precursor predicted by the algorithm MaturePred (Xuan et al., 2011). Finally, the last step was to search for miRNA mature similarity, obtained by MaturePred against the miRBase database.

Prediction of miRNA targets

The prediction of potential target genes of miRNAs identified in this study was performed by using the psRNA Target Server software (Dai and Zhao, 2011). We used the sequences of mature miRNAs predicted to identify possible targets present in *G. max* (DFCI gene index release 16). The target sequences were considered to be positive when the score was <3.0 using the default parameters of the algorithm. The results of the analyses were verified manually using the Phytozome database (available at http://www.phytozome.net/) to identify the genes and their respective annotations.

RESULTS

With regard to the computational prediction of miRNAs, the sequences present in the root subtractive library were submitted to an *in silico* annotation of noncoding RNAs in order to identify candidates for miRNA precursors; 161 hypothetical sequences of miRNA precursors were identified based on evidence from the Rfam database provided by the INFERNAL software (Nawrocki et al., 2009); 19 were selected according to the best bit scores (Table 1). Among the candidate sequences, 11 were identified in the antisense strand (-) and 8 in the sense strand (+). This prediction also indicated the start and end of the pre-miRNA sequences, indicating their respective sizes and directions of the strands (Table 1). All candidates were termed gma-MIR-Cand (xx).

Subsequently, a manual search of the database was performed for deposited miRNA-precursor sequences. The alignment resulted in 6 candidates showing similarity to families of pre-miRNAs of the soybean, indicating that they might represent paralog sequences (Table 1).

The formation of the secondary structure stem-loop was obtained using Mfold (Table 2), a broadly used algorithm for computational prediction of miRNAs in plants (Zhang et al., 2005; Qiu et al., 2007; Subramanian et al., 2008; Wang et al., 2009; Kulcheski et al., 2011) based mainly on minimum estimates of ΔG (Zuker, 2003). The free energy of pre-miRNAs is generally lower than in that in other noncoding RNAs (ncRNAs) (Bonnet et al., 2004); therefore, an important parameter in predicting new miRNAs is the quantification of the stability of secondary structures (Mathews and Turner, 2006). The study conducted by Thakur et al. (2011) reported that there is great variability in the ΔG of pre-miRNA structures in plants, especially between -20 and -100 kcal/mol, which is most concentrated from approximately -30 to -70 kcal/mol. In our study, ΔG values ranged from -11.1 to -60.6 kcal/mol (Table 2); the best candidates presented ΔG values of $\leq$-20 kcal/mol, representing potential authentic precursors.

The processing of pre-miRNA produces mature miRNA that negatively regulates the mRNA of the target genes possibly expressed in root cells during nodulation through repression or degradation of the mRNA. In order to understand this regulatory sequence, we performed an *in silico* prediction with the MaturePred software to locate mature miRNAs from new precursor candidates, based mainly on pre-miRNAs of plants experimentally verified and deposited into the miRBase database. As shown in Table 2, most sequences of the mature miRNAs had sizes of 21 nucleotides. Furthermore, the locations of predicted candidates on the soybean genome included the intronic regions, intergenic regions, 5' and 3' untranslated regions (UTRs), and coding sequences (CDSs). Although most pre-miRNAs in plants are transcribed from the intronic and intergenic regions, they can also arise from CDSs and UTRs, as observed by Fahlgren et al. (2010), Kulcheski et al. (2011), and Nozawa et al. (2012).

Nine potential miRNAs were identified in our study (Table 2), and their secondary structures are shown in Figure 1.

Table 1. Sequence of pre-miRNAs revealed by the noncoding RNA annotation pipeline, based on the Rfam database and results of evaluation of candidate sequences in the miRBase database.

		Rfam						miRBase			
SSH sequence ID	Size[a]	Candidate miRNA	Bit score	Strand[b]	Start	End	Size[c]	Organism/family miRNA	Score	E-value	Evidence
Contig5263	1624	gma-MIR-Cand01	34.98	-	312	217	96	*Arabidopsis thaliana* (MIR414)	120	0.001	Possible ortholog
Contig1834	523	gma-MIR-Cand02	28.99	-	414	253	162	*Oryza sativa* (MIR812)	143	3E-05	Possible ortholog
Contig24974	601	gma-MIR-Cand03	27.53	+	464	544	81	*Arabidopsis thaliana* (MIR825)	88	0.33	Possible ortholog
SJ10-E1-R05-043 H05-UC.F	569	gma-MIR-Cand04	27.39	+	235	378	144	*Glycine max* (MIR1530)	105	0.032	Possible paralog
Contig16659	707	gma-MIR-Cand05	24.85	-	630	488	143	*Glycine max* (MIR1530)	138	6E-05	Possible paralog
Contig13869	540	gma-MIR-Cand06	23.04	-	372	279	94	*Oryza sativa* (MIR5543)	91	0.25	Possible ortholog
Contig26154	1100	gma-MIR-Cand07	22.02	+	9	86	78	*Oryza sativa* (MIR1874)	90	0.22	Possible ortholog
Contig8113	1092	gma-MIR-Cand08	21.79	-	211	51	161	*Medicago truncatula* (MIR2591)	126	7E-04	Possible ortholog
Contig24959	925	gma-MIR-Cand09	21.51	-	918	823	96	*Oryza sativa* (MIR2921)	94	0.15	Possible ortholog
Contig6190	668	gma-MIR-Cand10	20.98	+	551	626	76	*Glycine max* (MIR 1520)	99	0.036	Possible paralog
Contig29200	1466	gma-MIR-Cand11	20.95	-	95	14	82	*Medicago truncatula* (MIR2608)	114	0.002	Possible ortholog
SJ16-E1-L08-029 E05-UC.F	590	gma-MIR-Cand12	20.64	+	8	129	122	*Glycine max* (MIR1530)	109	0.012	Possible paralog
Contig7552	1058	gma-MIR-Cand13	20.58	+	130	257	128	*Glycine max* (MIR4393)	103	0.040	Possible paralog
Contig27883	2064	gma-MIR-Cand14	20.57	-	1059	1007	53	*Medicago truncatula* (MIR2609)	111	0.001	Possible ortholog
Contig24513	1248	gma-MIR-Cand15	20.37	+	45	158	114	*Glycine max* (MIR1522)	92	0.28	Possible paralog
Contig21304	1224	gma-MIR-Cand16	20.23	+	242	335	94	*Oryza sativa* (MIR396)	102	0.030	Possible ortholog
Glyma08g08970.1	859	gma-MIR-Cand17	20.20	-	92	6	87	*Arabidopsis thaliana* (MIR414)	169	7E-08	Possible ortholog
Contig19125	964	gma-MIR-Cand18	19.86	-	869	800	70	*Arabidopsis thaliana* (MIR403)	83	0.64	Possible ortholog
Contig8009	1370	gma-MIR-Cand19	19.42	-	1355	1229	127	*Sorghum bicolor* (MIR437)	110	0.010	Possible ortholog

[a]Size of the original sequence from the subtractive library. [b]DNA strand where the pre-miRNA was localized. [c]Size of the sequence of the predicted precursor.

Table 2. Mature miRNA identified based on the secondary structure prediction by Mfold and *in silico* prediction.

Mfold		MaturePred				Phytozome		
Candidate miRNA	Free energy (ΔG)	Mature miRNA	Size	Score	Arm	Ch	Start - End	Region
gma-MIR-Cand01*	-37.30 kcal/mol	UGAUCGUGAUCGUGAUCGUGC	21 nt	0.649064	3'	Gm15	7382273-7382293	CDS
gma-MIR-Cand02*	-29.10 kcal/mol	UAGUAUAAAAAAGUAUUAAAU	21 nt	0.653008	5'	Gm17	2217418-2217438	Intronic
gma-MIR-Cand03*	-25.50 kcal/mol	UUCUAUAGAAGUCAAGUGCAC	21 nt	0.582132	3'	Gm20	819015-819035	Intronic
gma-MIR-Cand04*	-32.20 kcal/mol	UCUAUCUAUCUAUCUAUCUAU	21 nt	0.712194	5'	Gm17	2225498-2225518	Intronic
gma-MIR-Cand05	-22.10 kcal/mol	UUUAUUUCUAAAUAUUUAUUG	21 nt	0.579147	5'	Gm06	41501044-41501064	3' UTR
gma-MIR-Cand06	-16.90 kcal/mol	ACAAAGUGGUUAAUUAACAAC	21 nt	0.546925	5'	Gm16	20464796-20464816	CDS/intergenic
gma-MIR-Cand07*	-60.60 kcal/mol	AUCCAGAAGGAGUCAACUCUC	21 nt	0.698796	3'	Gm10	47394262-47394282	CDS
gma-MIR-Cand08	-20.80 kcal/mol	AUGAUAAAUCUAUAAGGCAAA	21 nt	0.44891	3'	Gm04	49212058-49212078	5' UTR
gma-MIR-Cand09	-15.80 kcal/mol	AUAAGUUUUAAGCUUGAACAC	21 nt	0.516254	3'	Gm20	34162445-34162465	Intergenic
gma-MIR-Cand10	-11.90 kcal/mol	AAUUCGUUAUUAUCUAUC	18 nt	0.528889	3'	Gm04	4131636-4131653	3' UTR
gma-MIR-Cand11*	-23.20 kcal/mol	UCUCUCUCUCUCUUUGUUUUU	21 nt	0.522843	5'	Gm12	20305882-20305902	5' UTR
gma-MIR-Cand12	-11.40 kcal/mol	GUUUUCUGUAUAAAUUUCAUA	21 nt	0.683337	5'	Gm03	36784285-36784305	Intronic
gma-MIR-Cand13*	-20.50 kcal/mol	UGUUUUUAUAUGAAAAGGUGA	21 nt	0.520922	5'	Gm04	6683683-6683703	Intronic
gma-MIR-Cand14	-11.70 kcal/mol	UUUUAGUGACGAAAUAUUUCA	21 nt	0.745488	5'	Gm10	9075172-9075192	Intergenic
gma-MIR-Cand15	-11.10 kcal/mol	CUUCUCAUUUAUGAAUUAUGA	21 nt	0.46095	3'	Gm20	37227792-37227812	Intronic
gma-MIR-Cand16*	-56.20 kcal/mol	AGGUGAAGGACUACGCCGAUU	21 nt	0.741033	3'	Gm03	36536344-36536364	CDS
gma-MIR-Cand17*	-32.80 kcal/mol	UGAUGGUGAUCAUGAUCAUGA	21 nt	0.622788	5'	Gm08	6375449-6375469	CDS
gma-MIR-Cand18	-17.00 kcal/mol	ACACCAAUGAGGAACAAUGUA	21 nt	0.569925	3'	Gm09	40325686-40325706	CDS
gma-MIR-Cand19	-13.30 kcal/mol	AAUUUAAAAUUUAGCCAUUUA	21 nt	0.687659	3'	Gm04	4458909-4458929	3' UTR

CDS = coding sequence. UTR = unstranslated region. *Potential soybean miRNAs predicted *in silico*.

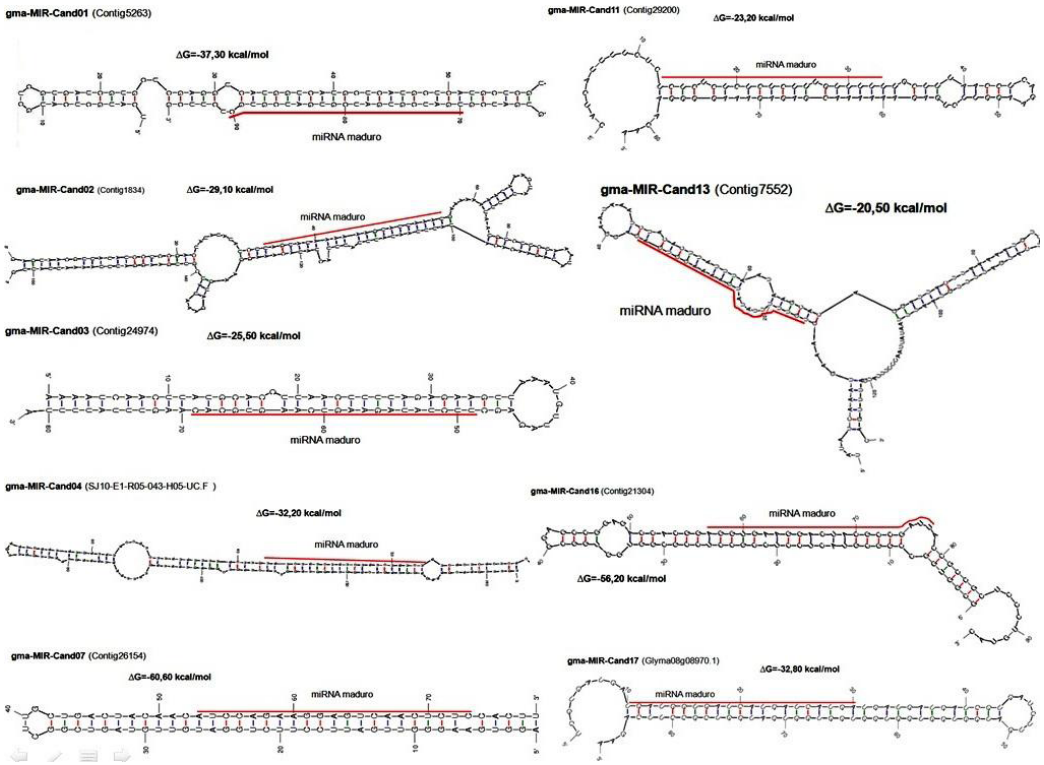

Figure 1. Secondary structures of pre-miRNAs and mature miRNAs predicted *in silico* in soybean inoculated with *Bradyrhizobium japonicum.*

It should be noted that the sequence of the mature miRNA is located within the "arm"of the stem-loop structure, and that the presence of mismatches in pairing miRNA:miRNA* was ≤4. Importantly, the actual sizes of the precursors may be slightly larger or smaller than those presented here. The secondary structures of the other pre-miRNAs are presented in **Figure S2**.

DISCUSSION

The soybean is the most important legume crop, mainly as a protein source for human and animal consumption. Our knowledge on the involvement of miRNAs in the process of nodulation and nitrogen fixation in the soybean is still very limited; thus, to understand the importance of these regulatory elements and the complexity in BNF, efforts have been spent on identifying new miRNAs involved in this process, thus increasing our knowledge base of this important biological process.

In this study, we performed a computational prediction initially using transcripts expressed in the soybean roots inoculated with *B. japonicum* (Barros de Carvalho et al., 2013). Among the sequences of pre-miRNAs revealed by the ncRNA annotation pipeline (Table 1), we can see that the candidates gma-MIR-Cand04, gma-MIR-Cand05, and gma-MIR-Cand12 aligned with the precursor MIR1530, whereas gma-MIR-Cand10 aligned

with MIR1520, and gma-MIR-Cand15 aligned with MIR1522. These families are involved in the early stages of nodulation in the soybean at 3 h after inoculation with *B. japonicum*, as identified by Subramanian et al. (2008). Another soybean family miRNA showing similarity with a candidate precursor was MIR4393 (gma-MIR-Cand13), identified by Joshi et al. (2010) in a study involving 4 different organs of the soybean (i.e., seed, flower, root, and nodule).

The family MIR396 from *Oryza sativa* was shown to be conserved among several plant species, including the soybean, *Arabidopsis thaliana*, *Sorghum bicolor*, *Saccharum officinarum*, *Zea mays*, *Populus trichocarpa*, and *M. truncatula*. In the soybean, it was identified in the early stages of nodulation in response to stresses (Subramanian et al., 2008; Kulcheski et al., 2011). The *M. truncatula* families MIR2591, MIR2608, and MIR2609, which showed similarities to the candidate precursors, were also identified as differentially regulated during nodulation in the current study (Lelandais-Brière et al., 2009). There were no previous reports of involvement in nodulation for the other candidates (Table 1). Through these alignments, we observed that the sequences obtained in our study had similarities to previously identified precursor sequences (Table 1).

In the search for mature sequences of these miRNA precursors, according to the parameters established, this study indicated 9 potential miRNAs involved in nodulation. Among these 9 miRNAs successfully predicted, 4 had similarities to miRNAs identified in other plants such as *M. truncatula* (Devers et al., 2011; Zhai et al., 2011), *O. sativa* (Xue et al., 2009), and *A. thaliana* (Breakfield et al., 2012). It is interesting that involvement of these miRNAs has been identified in the biological processes of other plant species, such as symbiosis and defense response and development, which are also released in the soybean during nodulation (Table 3). Many molecular events are triggered in a coordinated manner for successful symbiosis between a host plant and a symbiont, thus leading to morphological and physiological changes in the host plant (for more details see Oldroyd et al., 2011).

Table 3. Similarity search in the miRBase database using the predicted mature miRNA sequences.

Candidate miRNA	Mature miRNA sequence	miRBase			
		Organism	Mature sequence ID	Score	Condition
gma-MIR-Cand01*	UGAUCGUGAUCGUGAUCGUGC	-	-	-	-
gma-MIR-Cand02*	UAGUAUAAAAAAGUAUUAAAU	-	-	-	-
gma-MIR-Cand03*	UUCUAUAGAAGUCAAGUGCAC	-	-	-	-
gma-MIR-Cand04*	**UCUAUCUAUCUAUCUAUCUAU**	*Medicago truncatula*	mtr-miR5290	58	Symbiosis
gma-MIR-Cand05	UUUAUUUCUAAAUAUUUAUUG	-	-	-	-
gma-MIR-Cand06	ACAAAGUGGUUAAUUAACAAC	-	-	-	-
gma-MIR-Cand07*	**AUCCAGAAGGAGUCAACUCUC**	*Medicago truncatula*	mtr-miR5225b	62	Defense response
gma-MIR-Cand08	AUGAUAAAUCUAUAAGGCAAA	-	-	-	-
gma-MIR-Cand09	AUAAGUUUUAAGCUUGAACAC	-	-	-	-
gma-MIR-Cand10	AAUUCGUUAUUAUCUAUC	-	-	-	-
gma-MIR-Cand11*	**UCUCUCUCUCUCUUUGUUUUU**	*Oryza sativa*	osa-miR535-5p	61	Defense response
gma-MIR-Cand12	**GUUUUCUGUAUAAAUUUCAUA**	*Oryza sativa*	osa-miR2106	62	Seed development
gma-MIR-Cand13*	UGUUUUUAUAUGAAAAGGUGA	-	-	-	-
gma-MIR-Cand14	UUUUAGUGACGAAAUAUUUCA	-	-	-	-
gma-MIR-Cand15	CUUCUCAUUUAUGAAUUAUGA	-	-	-	-
gma-MIR-Cand16*	AGGUGAAGGACUACGCCGAUU	-	-	-	-
gma-MIR-Cand17*	**UGAUGGUGAUCAUGAUCAUGA**	*Arabidopsis thaliana*	ath-miR5658	68	Root development
gma-MIR-Cand18	ACACCAAUGAGGAACAAUGUA	-	-	-	-
gma-MIR-Cand19	AAUUUAAAAUUUAGCCAUUUA	-	-	-	-

Bold = sequences with similarity to miRNA have been identified in other plants.

Therefore, it appears that the miRNAs obtained by this computational identification may be involved in the regulatory process of nodulation, thus representing new soybean miRNAs.

We should also mention that our analyses are in agreement with the requirements described by Meyers et al. (2008). These authors described the main criteria for the annotation of miRNAs in plants and included 1) a prediction of the secondary structure of the transcript precursor, which forms the hairpin structure characteristic, as observed in our study, where the miRNA:miRNA* duplex is derived from opposite stem-arms; and 2) the presence of mismatches in pairing miRNA:miRNA* was ≤4, and the presence of asymmetric bulges were minimal within the miRNA:miRNA* duplex (Figure 1). In addition, we observed the conservation of some miRNAs among plant species based on the similarity search in the miRBase database (**Figure S3**).

In addition to identifying 9 potential miRNAs in this study, we investigated the gene targets that have been being regulated. (Table 4). gma-MIR-Cand01 and gma-MIR-Cand04 appear to regulate splicing factors that act in the processing of mRNA (Lopato et al., 1996). gma-MIR-Cand016 and gma-MIR-Cand017 also appear to regulate target genes involved in gene expression. The bZIP transcription factor (TF) seems to be the target of gma-MIR-Cand017; this TF negatively regulates nodulation (Nishimura et al., 2002) and is present in plant-defense responses (Dröge-Laser et al., 1997). Other transcription factors have been shown to be regulated by miRNAs during nodulation, such as MtHAP2, which is controlled by miR169; HD-ZIPIII, regulated by miR166; and NAC1, regulated by miR164; all were identified during nodulation in *M. truncatula* (Combier et al., 2006; Boualem et al., 2008; D'haeseleer et al., 2011). In the soybean, Subramanian et al. (2008) also showed that the TFs are targets of miRNAs during nodulation, as confirmed recently by Joshi et al. (2010). Moreover, herein we showed that a bZIP can be regulated by gma-MIR-Cand17.

Table 4. Predicted soybean targets for the candidate miRNA involved in nodulation.

miRNA ID	Locus target	Target description	Function
gma-MIR-Cand01	Glyma13g11090	LTP family	Structural constituent of cell wall
	Glyma04g40740	Alternative splicing factor ASF/SF2	Nucleic acid binding
gma-MIR-Cand02	Glyma04g02900	Predicted hydrolase/acyltransferase	Catabolism
gma-MIR-Cand03	Glyma06g05020	Serine carboxypeptidase	Proteolysis and peptidolysis
gma-MIR-Cand04	Glyma07g35050	Splicing factor	Nucleic acid binding
	Glyma06g05880	YGGT family protein	Membrane transport
gma-MIR-Cand07	Glyma20g27950	Ubiquitin	Protein binding
gma-MIR-Cand11	Glyma08g17120	CTP synthase	Pyrimidine nucleotide biosynthesis
gma-MIR-Cand13	Glyma13g19640	NAD dependent epimerase	Cellular metabolism
gma-MIR-Cand16	Glyma18g23590	OB fold nucleic acid binding protein	Nucleic acid binding
gma-MIR-Cand17	Glyma03g40730	bZIP transcription factor	Regulation of transcription

The other identified targets regulated by miRNAs identified in our study act in the modulation of metabolic changes of the plant, also observed at 3 DAI with *B. japonicum* in the soybean (Subramanian et al., 2008), where miRNAs were shown to be regulatory elements of a variety of targets, including proteases, water channels, and metabolic enzymes.

Thus, the results obtained in our study contribute to our understanding of the role of miRNAs in the regulation of nodulation in the soybean. The results add strong evidence that miRNAs facilitate and coordinate symbiotic interactions, from nodule organogenesis to nodule functioning. For example, miR482, miR1512, and miR1515 were previously demonstrated to

influence the number of nodules on soybean roots inoculated with *B. japonicum* (Li et al., 2010).

Considering that our analyses meet the main criteria for the identification of novel miRNAs, we can indicate that the 9 miRNAs identified in our study may orchestrate the process of nodulation in soybeans 10 days after inoculation with the nitrogen fixing bacterium *B. japonicum* strain CPAC 15.

We also emphasize the importance of computational predictions for the identification of novel miRNAs, showing that these tools can help us to expand our knowledge on the presence of miRNAs in different conditions. For this, the library transcripts deposited in public databases can be used. This type of study can facilitate the choice of miRNAs for experimental validation.

ACKNOWLEDGMENTS

Research partially supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) projects Genosoja and Repensa (#562008/2010-1). G.A. Barros-Carvalho acknowledges the MSc fellowship from CAPES (Coordenação de Aperfeiçoamento do Pessoal de Nível Superior), and M. Hungria acknowledges the research fellowship from CNPq (#300547/2010-2). Approved for publication by the Editorial Board of Embrapa Soja as manuscript XX/2014.

Supplementary material

REFERENCES

Ambros V, Bartel B, Bartel DP, Burge CB, et al. (2003). A uniform system for microRNA annotation. *RNA* 9: 277-279.

Barros de Carvalho GA, Batista JS, Marcelino-Guimarães FC, Costa do Nascimento LC, et al. (2013). Transcriptional analysis of genes involved in nodulation in soybean roots inoculated with *Bradyrhizobium japonicum* strain CPAC 15. *BMC Genomics* 14: 153.

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

Boualem A, Laporte P, Jovanovic M, Laffont C, et al. (2008). MicroRNA166 controls root and nodule development in *Medicago truncatula*. *Plant J.* 54: 876-887.

Breakfield NW, Corcoran DL, Petricka JJ, Shen J, et al. (2012). High-resolution experimental and computational profiling of tissue-specific known and novel miRNAs in *Arabidopsis*. *Genome Res.* 22: 163-176.

Combier JP, Frugier F, de Billy F, Boualem A, et al. (2006). MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in *Medicago truncatula*. *Genes Dev.* 20: 3084-3088.

D'haeseleer K, Den HG, Laffont C, Plet J, et al. (2011). Transcriptional and post-transcriptional regulation of a NAC1 transcription factor in *Medicago truncatula* roots. *New Phytol.* 191: 647-661.

Dai X and Zhao PX (2011). psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res.* 39: W155-W159.

Devers EA, Branscheid A, May P and Krajinski F (2011). Stars and symbiosis: microRNA- and microRNA*-mediated transcript cleavage involved in arbuscular mycorrhizal symbiosis. *Plant Physiol.* 156: 1990-2010.

Dröge-Laser W, Kaiser A, Lindsay WP, Halkier BA, et al. (1997). Rapid stimulation of a soybean protein-serine kinase that phosphorylates a novel bZIP DNA-binding protein, G/HBF-1, during the induction of early transcription-dependent defenses. *EMBO J.* 16: 726-738.

Fahlgren N, Jogdeo S, Kasschau KD, Sullivan CM, et al. (2010). MicroRNA gene evolution in *Arabidopsis lyrata* and *Arabidopsis thaliana*. *Plant Cell* 22: 1074-1089.

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

Hungria M, Campo RJ, Mendes IC and Graham PH (2006). Contribution of Biological Nitrogen Fixation to the N Nutrition of Grain Crops in the Tropics: the Success of Soybean (*Glycine max* L. Merr.) in South America. In: Nitrogen Nutrition and Sustainable Plant Productivity (Singh RP, Shankar N and Jaiwal PK, eds.). Studium Press, Houston, 43-93.

Joshi T, Yan Z, Libault M, Jeong DH, et al. (2010). Prediction of novel miRNAs and associated target genes in *Glycine* max. *BMC Bioinformatics* 11 (Suppl 1): S14.

Jung JH, Seo PJ and Park CM (2009). MicroRNA biogenesis and function in higher plants. *Plant Biotech. Rep.* 3: 111-126.

Kadri S, Hinman V and Benos PV (2009). HHMMiR: efficient de novo prediction of microRNAs using hierarchical hidden Markov models. *BMC Bioinformatics* 10 (Suppl 1): S35.

Kulcheski FR, de Oliveira LF, Molina LG, Almerão MP, et al. (2011). Identification of novel soybean microRNAs involved in abiotic and biotic stresses. *BMC Genomics* 12: 307.

Lelandais-Brière C, Naya L, Sallet E, Calenge F, et al. (2009). Genome-wide *Medicago truncatula* small RNA analysis revealed novel microRNAs and isoforms differentially regulated in roots and nodules. *Plant Cell* 21: 2780-2796.

Li H, Deng Y, Wu T, Subramanian S, et al. (2010). Misexpression of miR482, miR1512, and miR1515 increases soybean nodulation. *Plant Physiol.* 153: 1759-1770.

Lopato S, Mayeda A, Krainer AR and Barta A (1996). Pre-mRNA splicing in plants: characterization of Ser/Arg splicing factors. *Proc. Natl. Acad. Sci. U. S. A.* 93: 3074-3079.

Lu Y and Yang X (2010). Computational identification of novel microRNAs and their targets in *Vigna unguiculata*. *Comp. Funct. Genomics* 2010: 128297

Mathews DH and Turner DH (2006). Prediction of RNA secondary structure by free energy minimization. *Curr. Opin. Struct. Biol.* 16: 270-278.

Meyers BC, Axtell MJ, Bartel B, Bartel DP, et al. (2008). Criteria for annotation of plant microRNAs. *Plant Cell* 20: 3186-3190.

Nawrocki EP, Kolbe DL and Eddy SR (2009). Infernal 1.0: inference of RNA alignments. *Bioinformatics* 25: 1335-1337.

Nishimura R, Ohmori M and Kawaguchi M (2002). The novel symbiotic phenotype of enhanced-nodulating mutant of Lotus japonicus: astray mutant is an early nodulating mutant with wider nodulation zone. *Plant Cell Physiol.* 43: 853-859.

Nozawa M, Miura S and Nei M (2012). Origins and evolution of microRNA genes in plant species. *Genome Biol. Evol.* 4: 230-239.

Oldroyd GE, Murray JD, Poole PS and Downie JA (2011). The rules of engagement in the legume-rhizobial symbiosis. *Annu. Rev. Genet.* 45: 119-144.

Paschoal AR, Maracaja-Coutinho V, Setubal JC, Simões ZL, et al. (2012). Non-coding transcription characterization and annotation: a guide and web resource for non-coding RNA databases. *RNA Biol.* 9: 274-282.

Qiu CX, Xie FL, Zhu YY, Guo K, et al. (2007). Computational identification of microRNAs and their targets in *Gossypium hirsutum* expressed sequence tags. *Gene* 395: 49-61.

Simon SA, Meyers BC and Sherrier DJ (2009). MicroRNAs in the rhizobia legume symbiosis. *Plant Physiol.* 151: 1002-1008.

Subramanian S, Fu Y, Sunkar R, Barbazuk WB, et al. (2008). Novel and nodulation-regulated microRNAs in soybean roots. *BMC Genomics.* 9: 160.

Thakur V, Wanchana S, Xu M, Bruskiewich R, et al. (2011). Characterization of statistical features for plant microRNA prediction. *BMC Genomics* 12: 108.

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

Wang Y, Li P, Cao X, Wang X, et al. (2009). Identification and expression analysis of miRNAs from nitrogen-fixing soybean nodules. *Biochem. Biophys. Res. Commun.* 378: 799-803.

Xie FL, Huang SQ, Guo K, Xiang AL, et al. (2007). Computational identification of novel microRNAs and targets in *Brassica napus*. *FEBS Lett.* 581: 1464-1474.

Xuan P, Guo M, Huang Y, Li W, et al. (2011). MaturePred: efficient identification of microRNAs within novel plant pre-miRNAs. *PLoS One* 6: e27422.

Xue LJ, Zhang JJ and Xue HW (2009). Characterization and expression profiles of miRNAs in rice seeds. *Nucleic Acids Res.* 37: 916-930.

Zhai J, Jeong DH, De Paoli E, Park S, et al. (2011). MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. *Genes Dev.* 25: 2540-2553.

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

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