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

Computational identification of miRNAs and their targets in *Phaseolus vulgaris*

J. Han, H. Xie, M.L. Kong, Q.P. Sun, R.Z. Li and J.B. Pan

Beijing Key Laboratory for Agricultural Application and New Technique, College of Plant Science and Technology, Beijing University of Agriculture, Beijing, China

Corresponding authors: J. Han / J.B. Pan
E-mail: hanjun422@yahoo.com.cn / bacpjb@yahoo.com.cn

Genet. Mol. Res. 13 (1): 310-322 (2014)
Received April 8, 2013
Accepted September 5, 2013
Published January 17, 2014
DOI http://dx.doi.org/10.4238/2014.January.17.16

ABSTRACT. MicroRNAs (miRNAs) are a class of non-coding small RNAs that negatively regulate gene expression at the post-transcriptional level. Although thousands of miRNAs have been identified in plants, limited information is available about miRNAs in *Phaseolus vulgaris*, despite it being an important food legume worldwide. The high conservation of plant miRNAs enables the identification of new miRNAs in *P. vulgaris* by homology analysis. Here, 1804 known and unique plant miRNAs from 37 plant species were blast-searched against expressed sequence tag and genomic survey sequence databases to identify novel miRNAs in *P. vulgaris*. All candidate sequences were screened by a series of miRNA filtering criteria. Finally, we identified 27 conserved miRNAs, belonging to 24 miRNA families. When compared against known miRNAs in *P. vulgaris*, we found that 24 of the 27 miRNAs were newly discovered. Further, we identified 92 potential target genes with known functions for these novel miRNAs. Most of these target genes were predicted to be involved in plant development, signal transduction, metabolic pathways, disease resistance, and environmental stress response. The identification of the novel miRNAs in *P. vulgaris* is anticipated to

provide baseline information for further research about the biological functions and evolution of miRNAs in *P. vulgaris*.

Key words: Computational identification; MicroRNAs; Target genes; *Phaseolus vulgaris*

INTRODUCTION

MicroRNAs (miRNAs) are a newly discovered class of endogenous, highly conserved, non-coding small RNAs. Most mature miRNAs are approximately 21 nucleotides (nt) in length, with their precursor being able to fold into a stem-loop structure (Bartel, 2004). miRNAs have been broadly reported to be located on non-coding regions of genomes from various species (Megraw and Hatzigeorgiou, 2010). The *MIR* genes of plants are first transcribed by polymerase II to produce primary miRNA (pri-miRNA), which is then processed by the enzyme Dicer like 1 (DCL1) into a product called precursor miRNA (pre-miRNA). This precursor has a stem-loop structure, and the mature miRNA sequence is located on one arm of the hairpin structure. The pre-miRNA is subsequently cleaved into the double-stranded mature miRNA: Mirna* duplex (Bartel, 2004). Finally, the single-stranded mature miRNA is assembled into RNA-induced silencing complex (RISC), to induce it to bind to the complementary target mRNA (Bartel, 2004). Increasing evidence has shown that miRNAs have multiple roles in plant biological processes by negatively regulating their target proteins, including plant organ development and differentiation, signal transduction, morphogenesis, and adaptive responses to diverse biotic and abiotic stresses (Mallory et al., 2005; Larue et al., 2009; Li et al., 2012; Sunkar et al., 2012).

Since the first plant miRNA was identified in *Arabidopsis*, their crucial roles in post-transcriptional gene regulation have received increasing attention. Tremendous effort has been made to identify new miRNAs in different plant species, resulting in the discovery of large numbers of miRNAs in plants. At present, approximately 5000 mature miRNAs have been discovered from plants, and deposited in the public available miRNA database (miRbase, Release August 19, 2012). Most of these miRNAs were identified from plants with sequenced genomes, including 675 from *Medicago truncatula*, 591 from *Oryza sativa*, 506 from *Glycine max*, 323 from *Populus trichocarpa*, 299 from *Arabidopsis thaliana*, 206 from *Sorghum bicolor*, 201 from *Arabidopsis lyrata*, 174 from *Zea mays*, 163 from *Vitis vinifera*, and 142 from *Brachypodium distachyon*. To date, only 8 miRNAs have been identified from *Phaseolus vulgaris*, and deposited in the current edition of the miRNA registry.

The common bean (*P. vulgaris* L.) is one of the most important food legumes worldwide, and it has been suggested that it should be used as a model species for studying legume crops (Broughton et al., 2003). Although its genome sequence is unknown, the public EST and GSS databases make it possible to identify miRNAs in *P. vulgaris* by comparative genome-based homology searching. In this study, 1804 known and unique plant miRNA sequences were employed to search the EST and GSS databases to predict the potential miRNAs in *P. vulgaris*. A total of 27 potential miRNAs were detected, 24 of which were newly discovered in *P. vulgaris*. Subsequently, we also blasted these newly predicted miRNA sequences against the *P. vulgaris* mRNA database. We found that 92 potential target genes with known functions were predicted for these newly identified miRNAs. This study is anticipated to facilitate re-

search focused on elucidating the regulatory roles of miRNAs in plant growth, organ development, metabolism, and other biological processes in *P. vulgaris*.

MATERIAL AND METHODS

Mature miRNA query sequences, *Phaseolus vulgaris* EST, and GSS

A total of 3580 previously known plant mature miRNA sequences from 37 plant species were downloaded from miRbase (http://www.mirbase.org/). After the removal of redundant sequences, the remaining unique miRNA sequences were employed as query sequences to blast against the *Phaseolus vulgaris* EST and GSS database, which were available online from the National Center for Biotechnology Information (NCBI) GenBank nucleotide databases (http://www.ncbi.nlm.nih.gov/).

Availability of software

BLASTN and BLASTX (http://www.ncbi.nlm.nih.gov/BLAST/) was used to search for *P. vulgaris* miRNA homologs and to analyze the protein-encoding sequences, respectively. Based on the ClustalW alignments by Codoncode Aligner, the repeated EST or GSS of the same genes were discarded. The Zuker RNA folding algorithm software MFold 3.5 (Zuker, 2003), which is freely available online (http://mfold.rna.albany.edu/?q=mfold), was used to generate the secondary structures of the RNAs. Target genes for the novel miRNAs were predicted by the web tool psRNATarget (Dai and Zhao, 2011) (http://bioinfo3.noble.org/psRNATarget/).

Procedure and screening criteria for miRNA identification

The procedure for the identification of potential miRNAs was shown in Figure 1. Previously known plant miRNAs were screened to remove redundant sequences, and the remaining unique miRNA sequences were used as queries to search for miRNA homologs in *P. vulgaris*. Searching parameters were set as follows: maximum target sequences for 1000, expect threshold for 10, the remaining parameters were default. All of the EST or GSS, with no more than 4 mismatches against the query sequences, were saved. The repeated sequences of the same genes were discarded by Codoncode Aligner. The remaining sequences were then used to conduct a BLASTX analysis to remove the protein-encoding sequences.

The stem-loop structures of candidate precursor miRNAs were generated by the software MFold 3.5 (Zuker, 2003), where the default parameters were used. The following criteria were employed to screen the potential *P. vulgaris* miRNAs or pre-miRNAs: 1) predicted mature miRNAs should have no more than 4 nucleotide mismatches with the known plant miRNAs, 2) pre-miRNA sequence within an EST or GSS could fold into a perfect or near perfect stem-loop hairpin secondary structure, 3) the mature miRNA sequence should be located on one arm of the hairpin structure, 4) no more than 6 mismatches between the potential mature miRNA sequence and the opposite miRNA* sequence were allowed, 5) no loops or breaks in miRNA sequences were allowed, 6) the secondary structures of predicted pre-miRNA should have higher MFEI (minimal folding-free energy index) and negative MFE

(minimal free energy) compared to other small RNAs, 7) MFE should be lower than -20 kcal/mol, while MFEI should be over 0.85, and 8) the content of A+U should be within 30 to 70% (Ambros et al., 2003).

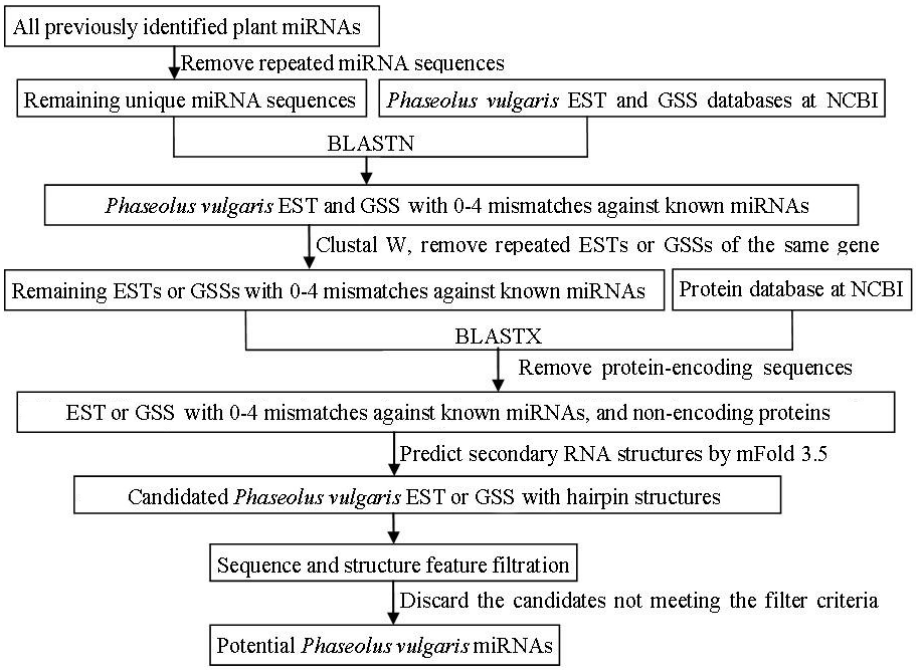

Figure 1. Procedure for the prediction of *Phaseolus vulgaris* potential miRNAs by searching EST and GSS homologs of previously known plant miRNAs.

Prediction of targets for novel miRNAs

Previous studies have demonstrated that most known plant miRNAs bind to their targets with perfect or nearly perfect sequence complementarity (Bartel, 2004). This phenomenon provides a powerful approach to identify plant miRNA targets by searching for miRNA homologs. In this study, targets of newly identified miRNAs were predicted by the web tool psRNATarget, using the *P. vulgaris* (Bean) DFCI Gene index (PHVGI) release 3.1 as the sequence library for the target search. The following criteria were used to identify miRNA targets: 1) no more than 4 mismatches were allowed between the mature miRNA and its potential target site; 2) no more than 1 mismatch was allowed at nucleotide positions 1-9; 3) no more than 2 consecutive mismatches were allowed; and 4) no mismatches were allowed at positions 10 and 11.

Phylogenetic analysis of the new miRNAs

Because most plant miRNAs and their precursor sequences are derived from the same gene families, they are strongly conserved and have high sequence identity, even between distantly related species. The evolutionary relationships between the precursor sequences of the

novel *P. vulgaris* miRNAs and their counterparts in the same family from other plant species were analyzed by MEGA 5.1 (http://www.megasoftware.net/mega.php).

RESULTS AND DISCUSSION

Identification of potential miRNAs

The miRNA homology-based EST and GSS blast analyses were conducted by comparing 1804 unique plant miRNAs with the *P. vulgaris* EST and GSS databases. After searching homology sequences, the redundant sequences of the same genes were removed, and then the protein-coding sequences was removed. This filtering, accordance with the screening criteria, left 828 EST and 765 GSS, which were selected as miRNA candidates. The secondary structures of the candidate miRNA sequences were predicted by MFold 3.5, and finally manually inspected according to the screening criteria described in the methods. Finally, a total of 27 potential *P. vulgaris* miRNAs, belonging to 24 miRNA families, were identified. Within the 27 predicted miRNAs, 12 miRNAs were identified from the EST and 15 miRNAs from the GSS (Table 1). Considering the known *P. vulgaris* miRNAs deposited in the miRNA database, all of the mature miRNA sequences identified in our study were compared with known *P. vulgaris* miRNAs. The results indicated that only 3 miRNAs had identical sequence with known *P. vulgaris* miRNAs, with the other 24 miRNAs being newly identified for this species (Table 1).

Table 1. Identified miRNAs in *Phaseolus vulgaris*.

miRNAs	GenBank ID	Source	miRNA mature sequences (5'→3')	Location	NM (nt)	LM (nt)	LP (nt)	(A+U)%	MFEs (kcal/mol)	MFEIs
miR156	EI480836	GSS	uugacagaagagagagagcaca	5	0	22	85	54.12	-43.60	1.12
miR167a	FE692413	EST	ugaagcugccagcaugaucua	5	0	21	86	50.00	-36.90	0.86
miR167b	FD792204	EST	ugaagcugccagcaugaucuga	5	0	22	80	56.25	-44.70	1.28
miR168	FE674450	EST	ucgcuuggugcaggucgggaa	5	0	21	114	42.11	-60.50	0.92
miR169	EI430521	GSS	ugagccaaggauggcuugccg	5	0	21	86	53.49	-57.60	1.44
miR171a	EI500954	GSS	ugauugagccgcgucaauauc	3	0	21	98	63.27	-45.80	1.27
miR171b	GW908487	EST	uaauugagccgcgucaauauc	3	1	21	81	58.02	-38.00	1.18
miR172	EI491452	GSS	aagacaucuccaagauucaga	3	4	21	91	67.03	-30.40	1.01
miR397	EI458339	GSS	ucauugagugcagcauugauc	3	2	21	186	59.14	-70.00	0.92
miR398	EI460202	GSS	uguguucucaggucaccccuu	3	0	21	112	60.71	-58.20	1.32
miR399	EG594328	EST	ugccaaaggagaguugcccug	3	0	21	75	54.67	-38.80	1.14
miR403	EI488011	GSS	uuagauucacgcacaaacuug	3	0	21	97	59.79	-42.16	1.08
miR408	FE687314	EST	augcacugccucuucccuggc	3	0	21	112	50.89	-60.80	1.11
miR414	EI477854	GSS	accuuuucaucaucaucauca	5	4	21	150	48.67	-79.30	1.03
miR447	EI444518	GSS	uuggggaagagaugauuuguuu	5	3	22	102	67.71	-31.40	0.95
miR1436	EI486878	GSS	acauuaugaaacggagggagu	3	2	21	169	66.27	-78.10	1.37
miR1514	FE709266	EST	uucauuuugaaaauaggcauug	5	0	22	127	66.93	-45.50	1.08
miR1888	EI490416	GSS	caaguuaagaauugagaagaa	3	3	21	108	68.37	-33.80	0.99
miR2118	HS104359	EST	uugccgauuccacccauuccua	3	0	22	92	57.61	-40.20	1.03
miR2595	EI433455	GSS	ucaauuuucuucuugaugucu	5	4	21	108	62.04	-35.40	0.86
miR2628	FE898012	EST	agagaaagaaagaugaguaa	3	4	20	116	61.21	-40.90	0.91
miR2672	EI468445	GSS	uuaaucgaccaaguugguaaaa	5	4	22	140	65.00	-50.40	1.03
miR2919	HS103295	EST	aaggggggggggggaauu	5	3	19	89	30.34	-71.70	1.16
miR5021	GW909603	EST	ugagaagaugaugaggaaau	5	4	20	202	59.90	-71.00	0.88
miR5205	EI470379	GSS	uauauauuuggggacggagguagu	3	4	24	87	64.37	-47.80	1.54
miR5658a	FE698270	EST	augaugaugaugaugaugaug	3	2	21	68	54.42	-27.60	0.89
miR5658b	EI503792	GSS	augaugaugaugaugaugucu	5	3	21	146	62.33	-52.70	0.96

The miRNAs newly identified in *P. vulgaris* are shown in red color. NM = No. of mismatch; LM = length of mature miRNAs; LP = length of precursors; MFEs = minimal folding free energies; MFEIs = minimal folding-free energy indexes.

The characteristics of the identified *P. vulgaris* miRNAs were studied to distinguish miRNAs from other small RNAs, such as miRNA mature sequences, length of mature miRNAs, miRNA location, length of precursors, MFEs, and MFEIs (Table 1). The identified *P. vulgaris* pre-miRNAs were between 68 and 202 nt in length, with an average of 111.37 ± 34.17 nt; however, the majority of the pre-miRNAs (24 of 27 or 88.89%) were 60-150 nt in length (Table 1). These results matched the lengths of most previously predicted pre-miRNAs in plant species (Zhang et al., 2007; Unver and Budak, 2009; Barozai et al., 2012; Neutelings et al., 2012). Although the identified *P. vulgaris* miRNA precursors were of a wide range of lengths, they could be folded into the typical miRNA secondary structures (Figure 2). The lengths of these identified mature miRNA sequences ranged from 19 to 24 nucleotides; however, most of the potential miRNAs (17 of 27 or 62.96%) were 21 nt in length (Table 1). This length was quite similar to other mature miRNAs already identified in other plants (Zhang et al., 2006a; Xie et al., 2010; Barozai et al., 2012; Neutelings et al., 2012).

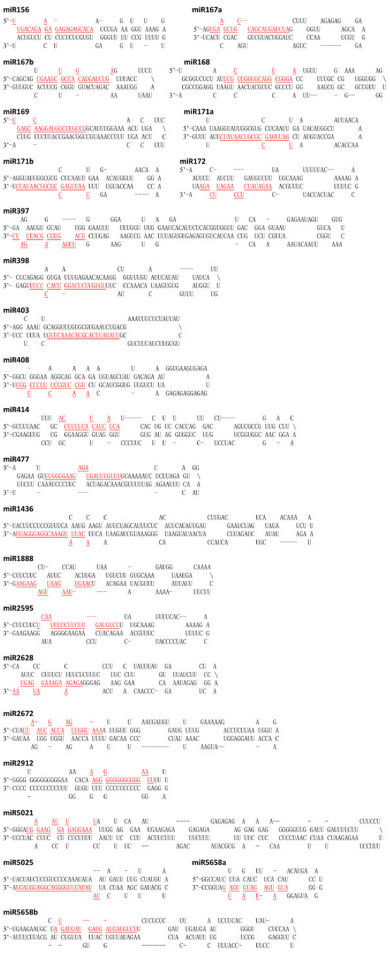

Figure 2. Predicted hairpin secondary structures of the newly identified *Phaseolus vulgaris* miRNA precursors; mature miRNA sequences are underlined.

The A+U contents of these predicted *P. vulgaris* pre-miRNA sequences ranged from 30.34 to 68.37%, with an average of 57.95 ± 8.53% (Table 1), which closely matched the results of previous studies (Ambros et al., 2003; Neutelings et al., 2012). MFE values are important for evaluating the stability of RNA secondary structures. In general, the lower the MFE, the more stable the secondary structure of an RNA sequence. The MFE values of the identified *P. vulgaris* miRNA precursors varied broadly from -79.30 kcal/mol to -27.60 kcal/mol, with an average of -49.38 ± 14.95 kcal/mol (Table 1). The MFEI of each potential miRNA precursor was calculated for the precise discrimination of the miRNA from other types of small RNAs (Adai et al., 2005; Zhang et al., 2005). We found that the precursors of these miRNAs had high MFEIs, ranging from 0.86 to1.54 (Table 1), with an average of 1.09 ± 0.18. These values were significantly higher compared to those reported for tRNAs (0.64), rRNAs (0.59), and mRNAs (0.62-0.66), indicating that these identified *P. vulgaris* miRNAs are probably true miRNAs (Zhang et al., 2006b).

Mature miRNA sequences have been reported to be evenly located on the 2 arms of the stem-loop hairpin structures of potential pre-miRNAs (Gorodkin et al., 2006). We also obtained the same result in our study for the 27 identified *P. vulgaris* miRNAs, of which 13 (48.15%) were found to be located on the 5'-arms of the stem-loop hairpin structures, while the other 14 (51.85%) were located on the 3'-arms (Table 1 and Figure 2).

Targets of newly predicted miRNAs

The prediction of targets for the identified miRNAs is anticipated to help us elucidate the important function and regulation of these novel miRNAs in *P. vulgaris*. Most plant miRNAs are perfectly or near-perfectly complementary with their targets (Schwab et al., 2005). For this reason, searching potential miRNA targets by blasting the mature miRNA sequences against the cDNA or EST database is considered a reliable technique. Nevertheless, several studies have demonstrated that many target genes have 1 to 4 nucleotide mismatches with mature miRNA sequences (Xie et al., 2010, 2011).

We identified a total of 92 potential targets with known functions for the 13 newly predicted miRNA families in *P. vulgaris*. The number of potential targets for every miRNA family ranged from 1 (miR398) to 23 (miR5658) (Table 2). However, 18 targets with unknown functions have not been listed here. Because a small number of sequences could be employed for target searching, no targets were found for miR168, miR172, miR403, miR408, miR1436, miR2672, miR2919, or miR5205 in *P. vulgaris*.

The 92 potential targets belonged to several gene families, and had multiple biological functions. A total of 24 predicted targets for new *P. vulgaris* miRNAs were transcription factors (Table 2). In addition to the transcription factors, their predicted targeted genes were involved in a broad range of biological processes, such as abiotic stress response, metabolism, transportation, disease resistance, and signal transduction.

Highly conserved *P. vulgaris* miRNAs also had conserved miRNA target sites on specific target genes, which was also observed in previous studies of other plants (Frazier et al., 2010; Xie et al., 2010, 2011). For example, it has been established that the plant-specific transcription factor, squamosa promoter binding protein, is involved in regulating changes during the early flower development and vegetative phase. In addition, it has been widely accepted that this transcription factor is a conserved target of the miR156 family in plants (Yin et al., 2008; Xie et al., 2011). In our study, 2 squamosa promoter binding protein-encoding genes were also identified as the targets of miR156 in *P. vulgaris* (Table 2). Some CCAAT-

Table 2. Predicted targets of the newly identified miRNAs in *Phaseolus vulgaris.*

miRNA families	Targeted protein	Functions of targetes	Targeted genes
miR156	Squamosa promoter binding protein	Transcription factor	FE706771 (1) TC26626 (1)
	ABC transporter E family member	Transporter	TC23933 (2)
	Aquaporin SIP1;2		TC22849 (2.5)
	Haloalkane dehalogenase 2-like	Metabolism	TC20482 (3)
miR167	Dioxygenase	Metabolism	TC20369 (2.5) TC27684 (2.5)
	Acyltransferase	Metabolism	TC26796 (3)
miR169	CCAAT-binding transcription factor	Transcription factor	TC20287 (1) FE700223 (1.5) TC28312 (2)TC25936 (2)
	Auxin Efflux Carrier	Signal transduction	CV537472 (3)
miR171	GRAS family transcription factor	Transcription factor	TC20655 (0.5)TC22093 (0.5)
	ATP-dependent chaperone ClpB		TC28957 (3)
	Ac transposase-like protein		FD794302 (2.5)
miR397	Laccase-17-like	Metabolism	CB541070 (0.5)
	Lipoxygenase	Metabolism	TC28581 (2.5)
	Glucosyltransferase	Metabolism	TC20381 (3)
miR398	Monogalactosyldiacylglycerol synthase	Metabolism	TC20478 (3)
	Pollen-specific protein SF3	Transcription factor	TC26066 (2)
miR414	Adipocyte plasma membrane protein	Signal transduction	TC20889 (1.5)
	Tubulin		FE691368 (2)
	TOC1	Transcription factor	TC21380 (2.5)
	Zinc finger family protein	Transcription factor	TC26868 (2.5) TC20427 (3) FE691610 (3)
	Histidine acid phosphatase	Metabolism	TC25565 (2.5)
	ABC transporter	Transporter	TC23456 (2.5)
	Transcription factor 25-like	Transcription factor	TC31171 (2.5)
	F-box/kelch-repeat protein	Transcription factor	CB540056 (3)
	HMG1/2-like protein	Transcription factor	CV543900 (3)
	Ethylene-responsive late embryogenesis	Stress response	TC30955 (3)
	Transcription initiation factor TFIID	Transcription factor	CV541321 (3)
	Cell number regulator 5-like	Development	CV543778 (3)
	Ribosomal protein S6 family protein		CV534067 (3)
	bZIP transcription factor	Transcription factor	TC31739 (3)
miR447	Tubulin A		TC20420 (3)TC21705 (3) TC31555 (3)
	Xyloglucan endotransglucosylase	Metabolism	TC31914 (3)
	Rab family protein		TC26265 (3)
	ATP-dependent protease La	Metabolism	FD792124 (3)
miR1888	U-box domain-containing protein		TC24440 (2)
	AT hook motif DNA-binding family	Transcription factor	FD790511 (3)
	Glycosyltransferase	Metabolism	TC25804 (3)
miR2595	GATA transcription factor MED5	Transcription factor	TC27879 (2)
	NDPH:quinone oxidoreductase	Metabolism	TC23497 (2.5)
	SURP motif-containing protein		TC29671 (3)
	Heat shock protein 70	Stress response	TC31270 (3)
	40S ribosomal protein S5		TC23413 (3)
	Transcription factor MYB86	Transcription factor	CV540785 (3)
miR2628	Jasmonate Zim-domain protein	Signal transduction	TC21766 (2)
	Anion-selective channel protein	Signal transduction	TC24615 (2)
	Lipase class 3 family protein	Metabolism	TC29044 (2.5)
	F-box protein PP2-A13-like	Transcription factor	TC25861 (2.5)
	ATP-dependent RNA helicase eIF4A		TC25757 (2.5)
	WD repeat-containing protein	Signal transduction	TC28514 (2.5)
	Gibberellin 20-oxidase	Signal transduction	TC21749 (3)
	Developmental protein FluG	Development	FE690630 (3)
	Leucine zipper protein ATHB-6	Transcription factor	TC22980 (3)
	MADS-box protein	Transcription factor	FE898584 (3)
miR5021	NOD3	Development	FG231724 (2.5)
	SPX domain-containing protein 3		TC27882 (2.5)
	SufE-like protein	Metabolism	TC21780 (3)
	Aryl-alcohol dehydrogenase	Metabolism	TC21619 (3)
	Uridine-cytidine kinase C-like	Metabolism	FE711956 (3) TC23026 (3)

Continued on next page

Table 2. Continued.

miRNA families	Targeted protein	Functions of targetes	Targeted genes
miR5658	Transcription factor bHLH130	Transcription factor	TC32102 (0)
	Chlorophyll a-b binding protein 3		TC22988 (0)
	Zinc finger (C3HC4 RING finger) protein	Transcription factor	CV536229 (0) EE743300 (0.5) TC23382 (1.5) TC29750 (1.5)
	Threonine-protein kinase EFR-like	Transcription factor	FE709880 (0)
	L-lactate dehydrogenase	Metabolism	TC24498 (0)
	Calcium-regulated forisome	Signal transduction	TC26079 (0)
	DNA-3-methyladenine glycosylase II	Metabolism	TC27351 (0.5)
	ATGP4		TC29192 (0.5)
	Squalene monooxygenase-like	Metabolism	EX304227 (1)
	CBL-interacting protein kinase 12		TC24696 (1.5)
	ERD7 protein	Stress response	TC32079 (1.5)
	Plasma membrane intrinsic protein	Metabolism	TC20291 (1.5)
	Galactomannan galactosyltransferase	Metabolism	CB541209 (0)
	Heat shock protein	Stress response	TC21029 (0.5)
	Transcription factor E(Y)2	Transcription factor	TC30210 (1) TC23663 (1)
	NBS-LRR resistance protein	Disease resistance	TC22558 (1)
	BTF3b-like transcription factor	Transcription factor	FD789516 (1)
	Zinc finger CCHC domain protein	Transcription factor	TC22646 (1.5)
	Leucine zipper protein HDZ1	Transcription factor	TC30819 (2)

box binding transcription factors have been reported as targets of the ath-miR169 family in *Arabidopsis* (Xie et al., 2005). *MtHAP2-1* is a new transcription factor of the CCAAT-binding family, and has been reported to be regulated by miR169 in *Medicago truncatula*, with miR169 being considered essential for nodule development in *M. truncatula* (Combier et al., 2006). We also found 4 CCAAT-binding transcription factor encoding targets for the miR169 family in *P. vulgaris* (Table 2). Recent research revealed that the miR171 family negatively regulates shoot branching and decreases primary root elongation by targeting GRAS gene family members *SCARECROW-LIKE6-II* (*SCL6-II*), *SCL6-III*, and *SCL6-IV* for cleavage in *Arabidopsis* (Wang et al., 2010). Interestingly, *MtNsp2*, which is targeted by the miR171 family in *M. truncatula*, encodes a GRAS transcription factor essential for nodule symbiosis signaling (Branscheid et al., 2011). Here, 2 GRAS transcription factors were predicted to be targeted by miR171 in *P. vulgaris* (Table2). Previous studies showed that several *AUXIN RESPONSE FACTORS* (*ARFs*) were targets for miR167 family in *Oryza sativa* (Liu et al., 2012). These *ARFs* play important roles in plant growth and development, by regulating the expression of auxin-responsive genes. Copper homeostasis is regulated by conserved miR398 in *Arabidopsis*, whereby miR398 directs the degradation of *Copper/zinc Superoxide Dismutase* mRNA when copper is limited (Yamasaki et al., 2007). Nevertheless, these conserved targets were not identified for the miR167 and miR398 families in *P. vulgaris*, mainly because the sequence information was incomplete.

The general transcription factor TFIID contains the TATA-binding protein (TBP) and a set of 13-14 TBP-associated factors (TAFs). TFIID is recruited by the activity of promoters, and possibly repressed genes (Cler et al., 2009). We predicted that miR414 directed the regulation of TFIID in *P. vulgaris*. Several other transcription factor groups were also detected as the targets of miR414 (Table 2), such as, bZIP transcription factor, which is involved in plant development and stress responses (Kaminaka et al., 2006). F-box/kelch-repeat protein controls the daytime expression pattern of the *CONSTANS* (*CO*) protein, and serves as an important regulator of plant flowering (Imaizumi et al., 2005). *TIMING OF CAB EXPRESSION 1* (TOC1) is the critical clock component, and is a transcriptional regulator that binds DNA. *TOC1* controls growth and development in *Arabidopsis* (Gendron et al., 2012).

MYB domain transcription factors play key regulatory roles in plant developmental processes, metabolism, and responses to biotic and abiotic stress response (Dubos et al., 2010). MYB transcription factors have been proven to be the targets of miR414 in potato and the model grass species *Brachypodium distachyon* (Unver and Budak, 2009; Xie et al., 2011). In contrast, in this study, transcription factor MYB86 was predicted to be the target of miR2595 in *P. vulgaris*. miR414 and miR5658 targeted zinc finger transcription factors are a superfamily of proteins involved in numerous activities of plant growth and development, and are also known to regulate resistance mechanisms for various biotic and abiotic stresses (Giri et al., 2011). The basic/helix-loop-helix (bHLH) transcription factors that control cell proliferation and cell lineage establishment are potential targets for *P. vulgaris* miR5658. miR2628 and miR5658 target the mRNA of homeobox-leucine zipper transcription factor, which is reported to regulate leaf polarity and vascular differentiation via the miR165/166 family (Jung and Park, 2007). Previous studies have revealed that miR444 family members target MADS-box transcription factors that are involved in diverse biological functions, including tolerance to abiotic stresses (Ding et al., 2011). Yet, in this study, MADS-box transcription factor was identified as the target of *P. vulgaris* miR2628.

Some of the newly identified *P. vulgaris* miRNAs also targeted genes involved in metabolism (Table 2), such as dioxygenase, acyltransferase, laccase, lipoxygenase, glucosyltransferase, monogalactosyldiacylglycerol synthase, L-lactate dehydrogenase, lipase, squalene monooxygenase, and galactomannan galactosyltransferase. We also found that several novel *P. vulgaris* miRNAs have complementary sites with the jasmonate zim-domain proteins, anion-selective channel protein, WD repeat-containing protein, gibberellin 20-oxidase, and calcium-regulated forisome (Table 2). Therefore, all of these genes might play important roles in signal transduction.

Many studies have shown that plant miRNAs are involved in biotic and abiotic stress (Xie et al., 2010; Gao et al., 2011; Sunkar et al., 2012). We also identified the stress-related proteins targeted by *P. vulgaris* miRNAs. For instance, miR2595 and miR5658 also targeted the heat shock proteins that are expressed in response to heat stress, which indicates the important role of these miRNAs during heat stress. Most plant resistance genes (R genes) encode nucleotide-binding site leucine-rich repeat (NBS-LRR) proteins (McHale et al., 2006). The NBS-LRR resistance protein was predicted to be the target of the *P. vulgaris* miR5658 family.

Phylogenetic analysis of the new miRNAs

Mature miRNA sequences, along with their corresponding precursor sequences, are highly conserved among different plant species (Zhang et al., 2006a). This phenomenon provides opportunities for the investigation of evolutionary relationships of miRNAs belonging to the same families in different plant species. To gain further insights about the evolutionary relationships of the newly identified miRNAs and their counterparts in other plant species, the phylogenetic relationships of several conserved *P. vulgaris* miRNAs with other members from the same families were analyzed by MEGA 5.1, including the miR156, miR167, miR168, miR169, miR171, miR398, and miR408 family. The precursor sequences of these miRNA families were obtained from the miRNA database (Release August 19, 2012, http://www.mirbase.org/).

Phylogenetic trees for 7 conserved miRNA families were constructed based on the pre-miRNA sequence comparisons. We concluded from the phylogenetic trees that the evolutionary relationships of *P. vulgaris* miRNAs with their counterparts from diverse plant species differed in different families. For instance, most members of the miR168 and miR408 families from legume plants were clustered into 1 group, indicating that these two families are highly conserved

in legume plants (Figure 3A and B). In comparison, the members of the miR156, miR167,

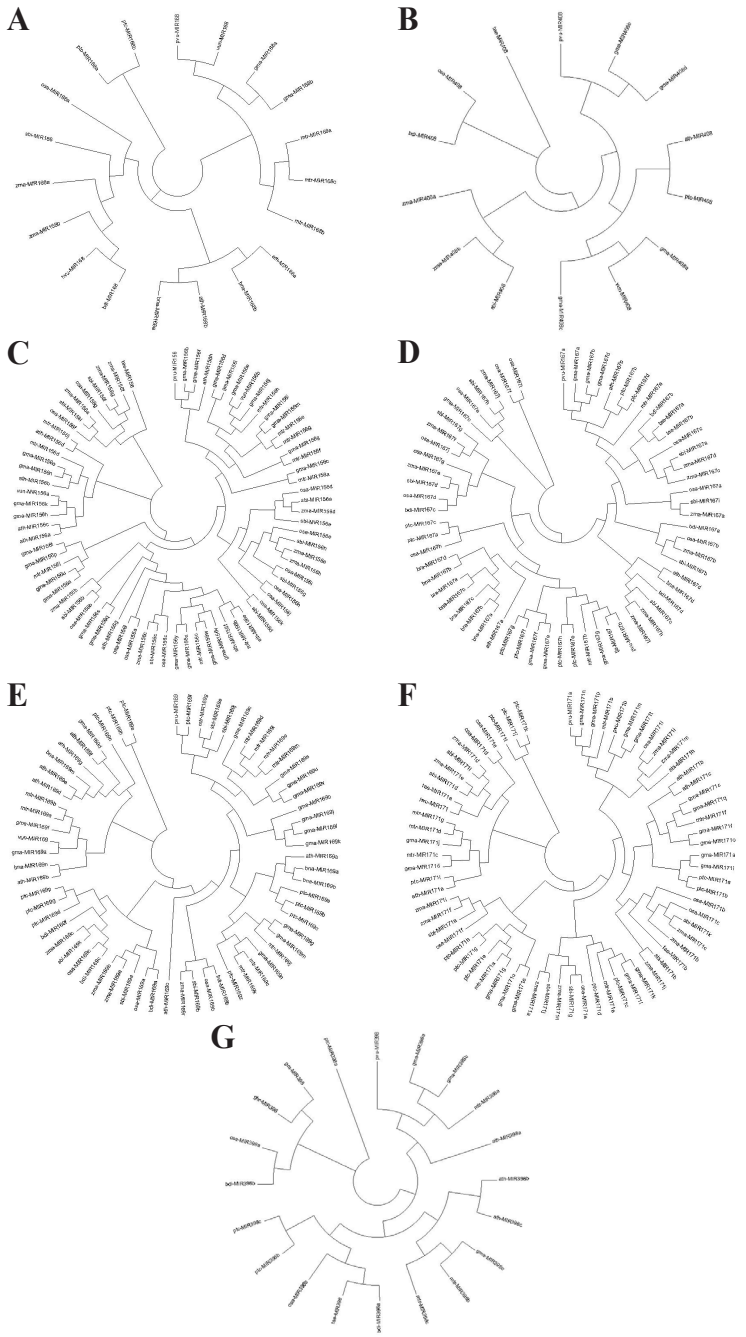

Figure 3. Phylogenetic trees obtained by aligning the precursor sequences of miRNAs in different families. **A.** MIR168 family; **B.** MIR408 family; **C.** MIR156 family; **D.** MIR167 family; **E.** MIR169 family; **F.** MIR171 family; **G.** MIR398 family.

miR169, miR171, and miR398 families from legume plants were clustered into more than 1 group, indicating that these miRNA families might have quickly diverged in legume plants. However, all of the *P. vulgaris* miRNAs from these families had some counterparts from legume plants on the same branches. For instance, pvu-MIR156 was the most closely related to gma-MIR156b and gma-MIR156f in the plant miR156 family (Figure 3C). In comparison, pvu-MIR167, gma-MIR167a, gma-MIR167b, gma-MIR167d, and mtr-MIR167a were clustered into 1 small branch (Figure 3D). pvu-MIR169 and mtr-MIR169g were located on 1 small branch (Figure 3E). pvu-MIR171a, pvu-MIR171b, gma-MIR171m, gma-MIR171n, gma-MIR171p, gma-MIR171t, and mtr-MIR171b formed a small group (Figure 3F). pvu-MIR398, gma-MIR398a, gma-MIR398b, and mtr-MIR398a formed a cluster on another small branch (Figure 3G). In addition, 2 members of the *P. vulgaris* miR171 family were closely related on the phylogenetic tree. Overall, these results indicate that different miRNAs might evolve at different rates in the plant kingdom, with some plant miRNA families being more conserved than the others.

ACKNOWLEDGMENTS

Research supported by the Scientific Research Project of Beijing Educational Committee, China (#PXM2011_014207_000037), and the National Natural Science Foundation of China (#31301312).

REFERENCES

Adai A, Johnson C, Mlotshwa S, Archer-Evans S, et al. (2005). Computational prediction of miRNAs in *Arabidopsis thaliana*. *Genome Res.* 15: 78-91.

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

Barozai MY, Baloch IA and Din M (2012). Identification of MicroRNAs and their targets in *Helianthus*. *Mol. Biol. Rep.* 39: 2523-2532.

Bartel DP (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116: 281-297.

Branscheid A, Devers EA, May P and Krajinski F (2011). Distribution pattern of small RNA and degradome reads provides information on miRNA gene structure and regulation. *Plant Signal. Behav.* 6: 1609-1611.

Broughton WJ, Hernández G, Blair M and Beebe S (2003). Beans (*Phaseolus* spp.) - model food legumes. *Plant Soil* 252: 55-128.

Cler E, Papai G, Schultz P and Davidson I (2009). Recent advances in understanding the structure and function of general transcription factor TFIID. *Cell Mol. Life Sci.* 66: 2123-2134.

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.

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

Ding Y, Chen Z and Zhu C (2011). Microarray-based analysis of cadmium-responsive microRNAs in rice (*Oryza sativa*). *J. Exp. Bot.* 62: 3563-3573.

Dubos C, Stracke R, Grotewold E, Weisshaar B, et al. (2010). MYB transcription factors in *Arabidopsis*. *Trends Plant Sci.* 15: 573-581.

Frazier TP, Xie F, Freistaedter A, Burklew CE, et al. (2010). Identification and characterization of microRNAs and their target genes in tobacco (*Nicotiana tabacum*). *Planta* 232: 1289-1308.

Gao P, Bai X, Yang L, Lv D, et al. (2011). osa-MIR393: a salinity- and alkaline stress-related microRNA gene. *Mol. Biol. Rep.* 38: 237-242.

Gendron JM, Pruneda-Paz JL, Doherty CJ, Gross AM, et al. (2012). *Arabidopsis* circadian clock protein, TOC1, is a DNA-binding transcription factor. *Proc. Natl. Acad. Sci. U. S. A.* 109: 3167-3172.

Giri J, Vij S, Dansana PK and Tyagi AK (2011). Rice A20/AN1 zinc-finger containing stress-associated proteins (SAP1/11) and a receptor-like cytoplasmic kinase (OsRLCK253) interact via A20 zinc-finger and confer abiotic stress tolerance in transgenic *Arabidopsis* plants. *New Phytol.* 191: 721-732.

Gorodkin J, Havgaard JH, Enstero M, Sawera M, et al. (2006). MicroRNA sequence motifs reveal asymmetry between the

stem arms. *Comput. Biol. Chem*. 30: 249-254.
Imaizumi T, Schultz TF, Harmon FG, Ho LA, et al. (2005). FKF1 F-box protein mediates cyclic degradation of a repressor of CONSTANS in *Arabidopsis*. *Science* 309: 293-297.
Jung JH and Park CM (2007). MIR166/165 genes exhibit dynamic expression patterns in regulating shoot apical meristem and floral development in *Arabidopsis*. *Planta* 225: 1327-1338.
Kaminaka H, Näke C, Epple P, Dittgen J, et al. (2006). bZIP10-LSD1 antagonism modulates basal defense and cell death in *Arabidopsis* following infection. *EMBO J*. 25: 4400-4411.
Larue CT, Wen J and Walker JC (2009). A microRNA-transcription factor module regulates lateral organ size and patterning in *Arabidopsis*. *Plant J.* 58: 450-463.
Li F, Pignatta D, Bendix C, Brunkard JO, et al. (2012). MicroRNA regulation of plant innate immune receptors. *Proc. Natl. Acad. Sci. U. S. A.* 109: 1790-1795.
Liu H, Jia SH, Shen DF, Liu J, et al. (2012). Four AUXIN RESPONSE FACTOR genes downregulated by microRNA167 are associated with growth and development in *Oryza sativa*. *Funct. Plant Biol.* 39: 744.
Mallory AC, Bartel DP and Bartel B (2005). MicroRNA-directed regulation of *Arabidopsis* AUXIN RESPONSE FACTOR17 is essential for proper development and modulates expression of early auxin response genes. *Plant Cell* 17: 1360-1375.
McHale L, Tan X, Koehl P and Michelmore RW (2006). Plant NBS-LRR proteins: adaptable guards. *Genome Biol.* 7: 212.
Megraw M and Hatzigeorgiou AG (2010). MicroRNA promoter analysis. *Methods Mol. Biol.* 592: 149-161.
Neutelings G, Fenart S, Lucau-Danila A and Hawkins S (2012). Identification and characterization of miRNAs and their potential targets in flax. *J. Plant Physiol.* 169: 1754-1766.
Schwab R, Palatnik JF, Riester M, Schommer C, et al. (2005). Specific effects of microRNAs on the plant transcriptome. *Dev. Cell* 8: 517-527.
Sunkar R, Li YF and Jagadeeswaran G (2012). Functions of microRNAs in plant stress responses. *Trends Plant Sci*. 17: 196-203.
Unver T and Budak H (2009). Conserved microRNAs and their targets in model grass species *Brachypodium distachyon*. *Planta* 230: 659-669.
Wang L, Mai YX, Zhang YC, Luo Q, et al. (2010). MicroRNA171c-targeted SCL6-II, SCL6-III, and SCL6-IV genes regulate shoot branching in *Arabidopsis*. *Mol. Plant* 3: 794-806.
Xie F, Frazier TP and Zhang B (2010). Identification and characterization of microRNAs and their targets in the bioenergy plant switchgrass (*Panicum virgatum*). *Planta* 232: 417-434.
Xie F, Frazier TP and Zhang B (2011). Identification, characterization and expression analysis of MicroRNAs and their targets in the potato (*Solanum tuberosum*). *Gene* 473: 8-22.
Xie Z, Allen E, Fahlgren N, Calamar A, et al. (2005). Expression of *Arabidopsis* MIRNA genes. *Plant Physiol.* 138: 2145-2154.
Yamasaki H, Abdel-Ghany SE, Cohu CM, Kobayashi Y, et al. (2007). Regulation of copper homeostasis by micro-RNA in *Arabidopsis*. *J. Biol. Chem.* 282: 16369-16378.
Yin Z, Li C, Han X and Shen F (2008). Identification of conserved microRNAs and their target genes in tomato (*Lycopersicon esculentum*). *Gene* 414: 60-66.
Zhang B, Pan X, Cannon CH, Cobb GP, et al. (2006a). Conservation and divergence of plant microRNA genes. *Plant J.* 46: 243-259.
Zhang B, Pan X, Cobb GP and Anderson TA (2006b). Plant microRNA: a small regulatory molecule with big impact. *Dev. Biol*. 289: 3-16.
Zhang B, Wang Q, Wang K, Pan X, et al. (2007). Identification of cotton microRNAs and their targets. *Gene* 397: 26-37.
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.