농 학 석 사 학 위 논 문

고추 microRNA와 표적유전자의 대량 발굴

Identification of microRNAs and their target genes in chili pepper (*Capsicum annuum*)

2013년 02월

서울대학교 대학원

농생명공학부 응용생명화학 전공

황 동 규

A THESIS FOR THE DEGREE OF MASTER OF SCIENCE

Identification of microRNAs and their target genes in chili pepper (*Capsicum annuum*)

Advisor: Chanseok Shin

A thesis submitted to the faculty of the Seoul National University Graduate School in partial fulfillment of the requirement for the degree of Master of Science in the Department of Agricultural Biotechnology.

By

Dong-Gyu Hwang

Major in Applied Life Chemistry
Department of Agricultural Biotechnology
Seoul National University Graduate School

February 2013
Approved by Major Advisor

농 학 석 사 학 위 논 문

고추 microRNA와 표적유전자의 대량 발굴

Identification of microRNAs and their target genes in chili pepper (*Capsicum annuum*)

지도교수 신 찬 석

이 논문을 석사학위논문으로 제출함.

2013년 2월

서 울 대 학 교 대 학 원
농생명공학부 응용생명화학 전공
황 동 규

황동규의 논문을 석사학위논문으로 인준함.

2013년 2월

위 원 장 김 수 언 (인)
부 위 원 장 신 찬 석 (인)
위 원 이 상 기 (인)

ABSTRACT

Identification of pepper microRNAs and their targets in chili pepper (*Capsicum annuum*)

Dong-Gyu Hwang
Major in Applied Life Chemistry
Department of Agricultural Biotechnology
The Graduate School
Seoul National University

In this study, microRNAs (miRNAs) in pepper were identified extensively from 10 different libraries using high-throughput sequencing technology. Based on bioinformatics pipeline, 128 conserved miRNAs and 50 novel miRNAs were identified, which belong to 29 and 35 families, respectively. Moreover, northern blot analysis was used to further validate the expression of

identified miRNAs and to analyze their tissue-specific or developmental stage-specific expression patterns. Subsequently, miRNA target genes were searched by computational research, revealing 331 and 57 potential targets of conserved and novel miRNAs, respectively. Many of them were experimentally validated using 5' RNA ligase-mediated rapid amplification of cDNA ends (5'RLM-RACE). To conclude, this study fulfilled transcriptome-wide identification of pepper miRNAs and their target genes, which also provides a broad base of informative support for understanding the functional roles of miRNAs in Solanaceae, including pepper.

Key words: MicroRNA, *Capsicum annuum*, High-throughput sequencing, Northern blot anlaysis, 5′RLM-RACE, miRNA-directed target cleavage

***Student number*: 2011-21329**

CONTENTS

PAGE

ABSTRACT.....1

CONTENTS.....3

LIST OF TABLES AND FIGURES.....4

INTRODUCTION.....5

MATERIAL AND METHODS.....8

RESULTS.....13

DISCUSSION.....29

TABLES.....36

FIGURES AND LEGENDS.....62

REFERENCE.....76

ABSTRACT IN KOREAN.....89

ACKNOWLEDGEMENT.....90

LIST OF TABLES AND FIGURES

Table 1. Distribution of small RNAs in pepper from 10 different libraries

Table 2. Conserved miRNAs in pepper

Table 3. Novel miRNAs in pepper

Table 4. Predicted targets of conserved miRNAs

Table 5. Predicted targets of novel miRNAs

Table 6. Probe sequences used for northern blot analysis

Table 7. Gene sepcific primers used for 5' RACE

Table 8. 5'RACE results of novel miRNA targets

Figure 1. Length distribution and 5’ end analysis of small RNAs in pepper

Figure 2. Representative hairpin precursors of conserved miRNA families

Figure 3. Representative hairpin precursors of novel miRNA families

Figure 4. Expression patterns of conserved miRNAs in pepper

Figure 5. Expression patterns of novel miRNAs in pepper

Figure 6. Validation of conserved miRNA-directed target cleavage in pepper

Figure 7. Validation of novel miRNA-directed target cleavage in pepper

INTRODUCTION

MicroRNAs (miRNAs) are non-translated RNAs of about 21 nucleotides that play regulatory roles in animals and plants. They negatively regulates gene expression at the post-transcriptional level through mRNA cleavage or translational repression, depeneding on sequence complementarity between miRNA and target mRNA. It has been reported that miRNAs regulate many biological processes in plants, containing development, stress adaptation, metabolism, and hormone signaling [1,2,3]. In general, plant miRNAs have high complementarity with their target mRNAs by which cleavage mechanism occurs, and their binding sites on target mRNAs are mainly found in coding regions [1,3], inducing significantly vigorous effect on gene regulation. Recent studies also revealed that plant miRNAs often direct translational repression via a slicer-independent mechanism [4,5], and therefore post-transcriptional gene silencing by plant miRNAs involves a combination of slicing and translation inhibition.

Plant miRNA genes are placed in intergenic regions, generally having their own transcriptional units which are transcribed by RNA polymerase II as primary miRNAs. Subsequently, they are cleaved into precursor miRNA (pre-miRNAs) by the RNaseIII-type enzyme DICER-LIKE1 (DCL1), and further cleaved into miRNA:miRNA* duplexes, again by DCL1, in the nucleus. These

miRNA duplexes are stabilized by 2'-O-methylation that is catalyzed by Hua Enhancer 1 [6] and transported from the nucleus to the cytoplasm by HASTY [7]. One of the strands is finally incorporated into the RNA-induced silencing complex (RISC), where mainly target cleavage mechanism occurs

As common methods for miRNA identification in plants, two major approaches, experimental and bioinformatics approaches, have been used up to date. Experimental approaches have included forward genetics, direct cloning, and next generation high-throughput sequencing. High-throughput sequencing technology shows a good possibility for small RNA identification and has become available and affordable in general. By means of high-throughput sequencing a great number of miRNAs have been identified and registered on an online database (http://www.mirbase.org, release 18.0, November 2011). It currently consists of 21,643 mature miRNAs from 168 species, including 4,677 mature miRNAs from 52 plant species. The majority of miRNAs studied so far have been obtained from only a few model plant species, such as *Arabidopsis thaliana* (328), *Oryza sativa* (661), *Glycine max* (395) and *Medicago truncatula* (674). The Solanaceae family is one of the largest families in the plant kingdom, consisting of more than 3,000 species [11]. However, annotated miRNAs in Solanaceae are still very limited [8,9,10]. Pepper (*Capsicum annuum*), which belongs to the Solanaceae family, is one of the most economically important

crops cultivated worldwide, especially in South and Central America and Asia [12]. It is crucial to understand the function of pepper miRNAs, considering the importance of pepper for food, health products and medicines. The study of the miRNAs in pepper has previously been reported using an *in silico* approach [13]. However, research using high-throughput sequencing approaches has not been performed on pepper so far.

In this study, the high-throughput sequencing technology was employed to sequence and identify pepper miRNAs by taking advantage of an ongoing pepper sequencing project. [14]. This resulted out 29 and 35 of conserved and novel miRNA families that include 128 and 50 members, respectively, from 10 different libraries. Expression patterns of identified miRNAs were also analyzed using small RNA northern blot analysis in order to further validate their expression and to test tissue- or developmental stage-specific expression patterns. Additionally, pepper miRNA targets were predicted and analyzed, many of which were experimentally validated using 5' RLM-RACE analysis [15]. While most targets of conserved miRNA were validated most of non-conserved miRNAs were weakly expressed and tend to lack experimentally validated targets. On the whole, this work serves to expand the knowledge of pepper miRNAs and to give a better comparison between pepper miRNAs and those found in other plants, especially in Solanaceae

MATERIALS AND METHODS

Construction of small RNA library

A small RNA sequencing library was constructed using the Small RNA Sample Prep Kit (Illumina, CA, USA), according to the manufacturer's instructions. Briefly, 5μg of total RNA from pepper tissue samples (CM334) including, leaf, stem, root, flower, fruit-1 (6 DAP; DAP for days after pollination), fruit-2 (16 DAP), fruit-MG (36 DAP; MG for mature green), fruit-B (38 DAP), fruit-B5 (43 DAP) and fruit-B10 (48 DAP) were ligated to a 3' adaptor and a 5' adaptor sequentially and then converted to cDNA by RT-PCR. The resulting cDNAs were then amplified by PCR, gel-purified and submitted for Illumina/Solexa sequencing. The GEO accession number for series is GSE 41654.

MicroRNA identification

A miRNA prediction pipeline was written by Python scripting language. High-quality small RNA reads were obtained from raw reads through filtering out poor quality reads and removing adaptor sequences using FAXTX toolkit [16]. These clean sequences were then queried against non-coding RNAs (rRNA,

tRNA, snRNA, snoRNA) from the Rfam database (http://www.sanger.ac.uk/software/Rfam) and the tomato genome database, ITAG 2.3 Release (http://solgenomics.net/organism/Solanum_lycopersicum/genome). Any small RNA read matches to these sequences were excluded from further analysis. The reads between 18-26 nucleotides in length were selected and aligned with a draft contig sequence of an ongoing pepper genome project [14] using MicroRazerS program [17]. The perfect matched sequences were selectively chosen and mapped to the maximum of 25 loci per sequence.

To obtain the precursor sequences, potential miRNA sequences (reads ≥ 50) were extended upstream and downstream of 100 to 500nt with a step size of 100nt. Each putative precursor sequence was folded using RNAfold from Vienna RNA software package [18], and the potential miRNA* sequences were selected with mismatch ratio of 0.3 or less. The region of these putative precursor sequences with addition of 15nt marginal sequences were re-folded using RNAfold [18] to check whether miRNA/miRNA* duplex was suited to primary criteria for annotation of plant miRNAs [19]. All of the small RNA reads in the selected putative precursor region were mapped to examine the strand bias whether the reads of sense strand were accounted for 90% of total reads. The miRNA candidates were essentially grouped into families by mature sequence

similarity and/or loci. By the similarity search with miRBase release 18.0 (http://www.mirbase.org), all members of miRNA candidate families were classified to either the known miRNAs or the novel miRNA candidates. The normalizing factors were calculated using the DESeq library [20] in the R statistical software package (R Development Core Team, 2009).

MicroRNA target prediction

The putative target sites of miRNAs were identified by aligning mature miRNA sequences with a draft genome sequence [14] using TargetFinder, (http://carringtonlab.org /resources/targetfinder). miRNA targets were computationally predicted essentially as described [21,22,23]. Briefly, potential targets from FASTA searches were scored using a position-dependent, mispair penalty system. Penalties were assessed for mismatches, bulges, and gaps (+1 per position) and G:U pairs (+0.5 per position). Penalties were doubled if the mismatch, bulge, gap, or G:U pair occurred at positions 2 to 13 relative to the 5' end of the miRNAs. Only one single-nucleotide bulge or single-nucleotide gap was allowed, and targets with penalty scores of four or less were considered to be putative miRNA targets.

Northern blot analysis

Total RNA was extracted from different tissues using TriReagent (Ambion). A total amount of 20μg of total RNA from leaf, stem, root, inflorescence, fruit and seedling was individually separated in a 15% UREA polyacrylamide gel, electrophoretically transferred to Hybond-NX membrane (GE healthcare), and chemically cross-linked via 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) [24]. For labeling reaction of probes, 2μl of 10μM oligo, 2μl of 10X T4 PNK buffer (Takara), 2.5μl of [γ-^{32}P] ATP, >7000Ci/mmole (~150μCi/μl), 12.5μl of dH_2O and 1μl of T4 Poly Nucleotide Kinase (Takara) were added in a 20μl reaction for 1 hour at 37 °C. The labeled probes were further purified from unincorporated labels with PERFORMA Spin Columns (Edge Bio) according to the manufacturer's instructions. Probe sequences used for northern blot analysis were shown in Table 6. Hybridization and washing procedures were performed as described [25]. The membranes were exposed to a phosphorimager, and signals were analyzed using BAS-2500 (Fuji).

MicroRNA target validation assays

For miRNA target validation, gene-specific 5' RNA ligase-mediated rapid amplification of cDNA ends (5' RLM-RACE) was performed using the

GeneRacer Kit (Invitrogen). A amount of 5μg of total RNA from mixed seedling, root and flower tissues were ligated to 0.25μg of the GeneRacer RNA oligo adapter (5'–CGACUGGAGCACGAGGACACUGA CAUGGACUGAAGGAGUAGAAA–3'). The combination of oligo(dT) and random hexamers were then used to prime the first strand of cDNA synthesis in a reverse transcription reaction. The resulting cDNA was PCR-amplified with the GeneRacer 5' primer (5'–CGACTGGAGCAC GAGGACACTGA–3') and each respective gene-specific primer (shown in Table 7). The PCR product was further amplified by nested PCR using the GeneRacer 5' nested primer (5'–GGACACTGACATGGACTGAAGGAGTA–3') and each respective gene-specific primer (shown in Table 7). The final PCR product was gel-purified and finally cloned into pGEM-T Easy vector (Promega) for sequencing.

RESULTS

Small RNA analysis in pepper

To obtain endogenous small RNAs in pepper, high-throughput sequencing was used to generate small RNA sequences from leaf, stem, root, flower, fruit-1, fruit-2 , fruit-MG, fruit-B, fruit-B5 and fruit-B10, which yielded total reads from 38,168,067 (fruit-B5 library) to 87,761,235 (leaf library) (Table 1; This was performed by Jae Yun Lim). After removing adaptor sequences and filtering out low quality tags, clean reads from 37,640,821 (fruit-B5 library) to 74,310,576 (leaf library) were obtained, which were 18-26 nt in length, representing from 11,019,173 (fruit-B5 library) to 19,478,268 (fruit-B library) of unique sequences (Table 1). After further removal of tRNAs, rRNAs, snRNAs and snoRNAs, total reads from 19,591,941 (fruit-B5 library) to 40,888,746 (flower library) remained.

These small RNA sequences were mapped to draft pepper genome sequences to determine whether they could be considered candidate miRNAs. These candidate miRNA were further selected based on strict criteria for annotation of plant miRNA [19], including the presence of miRNA* sequences. Consequently, conserved and novel miRNAs were identified and represented in Tables 2 and 3, respectively. Total reads of conserved miRNAs varied from

544,977 (fruit-B library) to 3,390,712 (root library), whereas those of novel miRNAs varied from 76,608 (fruit-B5 library) to 256,460 (leaf library) (Table 1), suggesting that novel miRNAs are weakly expressed in general. More detailed information regarding the statistics of small RNA sequences of ten different libraries were all shown in Table 1. Furthermore, the length distribution and 5' end analysis were conducted for the genome-aligned small RNAs (Figure 1; This was performed by Jae Yun Lim). The majority of small RNAs were 24nt long and accounted for 46.91% of total small RNAs, followed by 21nt (14.16%), 22nt (13.99%) and 23nt (13.65%). Most of these small RNAs had a 5' terminal U or A, which is indicative of canonical small RNAs [26].

Identification of conserved miRNAs in pepper

With the purpose of identifying conserved miRNAs in pepper, genome-aligned small RNA sequences were BLASTN searched against currently known miRNAs in the miRbase (Release 18), allowing one or two mismatches between sequences. The BLASTN searches identified 128 conserved miRNAs corresponding to 29 miRNA families (Table 2; This was carried out in collaboration with Jae Yun Lim). The number of hairpin loci varied from 1 to 10, and the can-miR166 and can-miR171 families were the largest group having 12

members in the identified conserved families. The can-miR395 and can-miR172 families were the third (10 members) and fourth (9 members) groups, respectively (Table 2). Examples of representative hairpin structures are shown in Figure 2.

The sequencing frequencies of miRNAs from ten different libraries might be used to estimate the tissue- or developmental stage-dependent expression patterns and their possible roles. As a result, Illumina/Solexa sequencing revealed that the majority of conserved miRNAs showed tissue-specific or developmental stage-specific expression. For instance, can-miR156d-g was most highly expressed in root, but very lowly expressed in leaf, stem, flower and fruit. can-miR164a-b was expressed more abundantly in red fruits (fruit-B, B5 and B10) than in green fruits (fruit-1, 2 and MG) and other tissues. can-miR156a-c especially showed clear developmental stage-dependent expression patterns; the expression level gradually increased from fruit-1 (early stage of green fruit) to fruit-B10 (late stage of red fruit). The expression level of can-miR168 family was higher in leaf and lower in other tissues. can-miR390 family was expressed lowly in both green and red fruits, compared to other tissues. However, the expression level of a few miRNAs was similarly high (e.g. can-miR159 and can-miR166) or low (e.g. can-miR171f-g and can-miR172e) in all tissues.

The sequencing frequencies of miRNA families varied from 1 to 2,481,041, indicating expression varies significantly among different miRNA families. The can-miR166i-j (up to 2,481,041 reads in root), can-miR166a-h (up to 588,387 reads in fruit-2), and can-miR482f (up to 216,948 reads in flower) were among the most frequent in the libraries whereas several miRNAs, including can-miR169h, can-miR172e, can-miR398a-b, can-miR399h and can-miR399a-e, were detected in less than 100 reads.

Identification of novel miRNAs in pepper

For the identification of novel miRNAs, the genome-aligned small RNAs, exclusive of conserved miRNAs, were analyzed according to several criteria; first, only the miRNAs of which precursors were folded into stem-loop structures in hairpin prediction were selected for novel miRNA candidates. Precursors of these novel miRNAs had negative folding free energies ranging from -355.30 to -19.80 according to RNAfold. The average free energy of these novel miRNAs was about -104.85, much lower that of other plant miRNA precursors (-59.5 kcal/mole in *A. thaliana* and -71.0 kcal/mol in *O. sativa*). Next, base-paring criteria [19] between the miRNA and the other arm of the hairpin were applied; (1) mismatched miRNA bases are four or fewer, (2) asymmetric

bulges are two or less in size, and (3) one or less in frequency. In addition, it is investigated that these novel miRNA candidates contained both miRNA and miRNA* sequences in the sequencing libraries since the presence of miRNA* is strong evidence of precise biogenesis [19]. Finally, miRNAs showing high expression level with total frequency 1,000 or higher were chosen to be studied for more strict identification of novel miRNAs. As a result, 50 novel miRNAs that belongs to 35 families were identified (Table 3; This was carried out in collaboration with Jae Yun Lim). The largest family was can-miR-n019 with four members, and most of the novel miRNAs were mapped to a single locus in the pepper genome, contrary to multiple loci of conserved families (Table 3). Some of the representative secondary structures of their precursors were shown in Figure 3.

Classification of a large number of miRNAs from high-throughput sequencing data is usually faced with difficulty in some cases where multiple miRNAs accumulate from the same precursor. For example, there is a case where miRNA and miRNA* species are sequenced approximately at the same level, as is the case for ath-miR832 [27]. In the present study, likewise, can-miR-n009-5p and can-miR-n009-3p accumulated to approximately equal levels in most tissues examined, except the leaf (Table 3). Northern blot analysis revealed that can-miR-n009-3p, initially regarded as a star strand, was abundantly expressed in all

tissues (Figure 5B), implicating its functional activity.

The sequencing frequencies of these novel miRNAs also showed that their expression had clear tissue-specific or developmental stage-specific patterns. For example, can-miR-n001 was sequenced in more than 1,000 reads in stem, root and flower, but not sequenced in leaf. can-miR-n002 was most frequently sequenced in leaf (70,358 reads; the highest frequency of novel miRNAs in the libraries), moderately in stem and root and lowly in flower and fruit. can-miR-n026 had similar expression pattern; abundant expression in leaf, moderate expression in stem and root, and low expression in flower and fruit. In contrast, can-miR-n006 was expressed abundantly in fruit and expressed rarely in leaf, stem, root and flower. However, several novel miRNAs had similar expression pattern in all tissues (e.g. can-miR-n004), similar to some of the conserved miRNA families such as can-miR159, 160, 166, 171 and 172.

Expression patterns of miRNAs in pepper

It has been reported that a large number of miRNAs in plants have tissue-specific and developmental stage-specific expression patterns [25,28,29,30], some of which might have a wide range of crucial roles during development and stress adaptation [3,31]. Since it has been reported that there

were biases inherent in next-generation sequencing (NGS) technologies [32], northern blot analysis was expected to reveal more accurate *in vivo* expression patterns. Furthermore, detection of miRNAs by northern blot could provide more direct evidence for their true expression. Therefore northern blot analysis was performed to observe tissue-specific or developmental stage-specific expression patterns in different tissues; leaf (L), stem (S), root (R), inflorescence (I), fruit (F) and seedling (Se).

Among the 29 conserved families, 19 miRNA families were tested and 15 were detected as discrete bands (Figure 4). Four miRNA families (can-miR162, can-miR390, can-miR408, and can-miR477) were not detected by northern blot analysis. Expression levels of these miRNAs appeared not high enough to be detected by northern blot analysis [33]. As a result, tissue-specific and developmental stage-specific expression patterns were observed for some of conserved miRNAs. For instance, can-miR156 was expressed abundantly in the seedling, moderately in the root, inflorescence and leaf, but weakly in the stem and fruit (Figure 4A). Similarly, tissue-specific or developmental stage-specific expressions were observed in can-miR164, can-miR166, can-miR167, can-miR394, can-miR395, can-miR396, can-miR398, and can-miR530 (Figure 4C, D, E, I, J, K, L and N). The expression level of can-miR164 was high in the stem, root, inflorescence and fruit, and low in the leaf and seedling (Figure 4C). can-

miR166 and can-miR396 were expressed lowly in the leaf and inflorescence, respectively (Figure 4D, K). can-miR167 was expressed highly in the leaf, inflorescence and fruit and lowly in the stem, root and seedling (Figure 4E). The expression level of can-miR394 was higher in stem than in other tissues (Figure 4I). can-miR398 and can-miR-530 were accumulated in leaf (Figure 4L and N), and can-miR395 was exclusively expressed in the inflorescence (Figure 4J). However, can-miR159, can-miR168, can-miR171, can-miR319, and can-miR403 had no tissue-specific or developmental stage-specific expression patterns (Figure 4B, F, G, H, and M) They appeared to be expressed ubiquitously in all tissues. Some miRNAs, including can-miR167, can-miR319 and can-miR403 appeared as a doublet or triplet on northern blot analysis in all or part of the tissues.

The expression profiles obtained by northern blot analysis usually agreed with the sequencing data, as in the cases of can-miR159, can-miR164, can-miR167, can-miR168, can-miR171, can-miR394, can-miR395, can-miR396, can-miR403 and can-miR530 (Figure 4B, C, E, F, G, I, J, K, M and N, Table 2). However, there were discrepancies between northern blot analysis and sequencing frequencies although it was restricted to one or two tissues. Specifically, can-miR156 and can-miR398 were expressed more abundantly in leaf than stem although they were sequenced much less in leaf than stem (Figure

4A and L, Table 2). This reverse expression pattern between northern blot and sequencing data was observed for can-miR166 expression in leaf and root (Figure 4D, Table 2). In addition, can-miR319 expression in leaf was as abundant as in other tissues despite its much lower sequenced frequency than other tissues (Figure 4C and H, Table 2).

Subsequently, the expression patterns of novel miRNA families were analyzed by northern blot analysis. On the whole, 19 out of 35 novel miRNA families were subjected to northern blot analysis, and all of them (19) were detected. Representative results of the northern blot analysis for 15 novel miRNAs were shown in Figure 5. Overall, the expression level of many miRNAs were similarly high or low in all tissues; can-miR-n003, can-miR-n013, can-miR-n016, can-miR-n027, can-miR-n030, can-miR-n032 and can-miR-n033 (Figure 5E, M, C, O, D, G and P). Although these seven miRNAs did not show tissue-specific or developmental stage-specific expression patterns, the sequencing frequencies from different libraries suggested that four miRNAs (can-miR-n016, n027, n030 and n032) showed apparent tissue or developmental stage-dependent expression patterns (Table 3). In contrast, eight other miRNAs had tissue-specific or developmental stage-specific expression patterns. Specifically, can-miR-n002 was expressed abundantly in leaf, root and seedling, and weakly in stem, inflorescence and fruit (Figure 5A). can-miR-n026 was

expressed ubiquitously in all of the tissues, especially high in fruit (Figure 5F). can-miR-n004 was more highly expressed in stem, root and inflorescence than leaf, fruit and seedling (Figure 5H). can-miR-n005 was expressed most highly in root, moderately in stem and seedling, and lowly in leaf (Figure 5I). can-miR-n006 was abundantly expressed in leaf, stem, root and seedling, but not detected in inflorescence and fruit (Figure 5J). The expression level of can-miR-n007 was higher in stem and root, and lower in leaf, inflorescence, fruit and seedling (Figure 5K). The expression of can-miR-n010 was higher in leaf, stem, root and inflorescence, and lower in fruit and seedling (Figure 5L). can-miR-n017 was expressed moderately in stem and inflorescence, and slightly in leaf, root, fruit and seedling (Figure 5N). In the case of can-miR-n013 and can-miR-n030, a doublet was observed in all or part of the tissues (Figure 5M and D). Most of evolutionarily young miRNAs, such as species-specific or non-conserved miRNAs, have been reported to be expressed weakly [21,27,34,35,36] whereas highly conserved miRNAs are expressed at a higher level [27,36,37]. This was observed for several novel miRNAs, which showed weak expression in northern blot analysis although they were selected by strict criteria, including a high frequency of 1,000 or higher. In contrast, some of the novel miRNAs were expressed at a high level in northern blot analysis as well (Figure 5A, H, I, K, L, F and P), suggesting that they could play more role in pepper-specific biology

than the other weakly expressed novel miRNAs. In some cases, there were inconsistencies between expression levels obtained by Illumina/Solexa sequencing and northern blot analysis. As mentioned above, it is possible that sequencing technology could cause biases for certain sequences. Another possibility is that the probes used for northern blot analysis might have captured heterogeneous miRNAs.

Predicted targets of miRNAs in pepper

On the ground that plant miRNAs have a perfect or nearly-perfect match to their target mRNAs, FASTA searches were performed to predict targets of conserved miRNAs, and the potential targets were chosen by a position-dependent penalty scoring system in which predicted targets with penalty scores of four or less were selected. As a result, 331 potential target genes were identified from 26 out of 29 conserved miRNA families (Table 4; This was carried out in collaboration with Jae Yun Lim).

In general, most of the targets predicted for conserved miRNAs in pepper were transcription factors (Figure 4). For instance, predicted targets such as SQUAMOSA promoter binding proteins (SBP), auxin response factors (ARF), growth regulating factors (GRF), no apical meristem (NAM) and MYB family

belong to transcription factors (Table 4). Therefore, similar to other plants [3], pepper miRNAs appear to play a significant role in plant development and growth.

The functions of miRNAs in pepper could be predicted based on the functions known in other plants; most of the predicted targets of conserved miRNAs in pepper also had a conserved function with miRNA targets in other plants. The miRNA families having conserved targets were: can-miR156, can-miR159, can-miR160, can-miR164, can-miR167, can-miR171, can-miR172, can-miR319, can-miR394, can-miR395, can-miR396, and can-miR482 (Table 4). For example, SBP transcription factors, known to regulate flowering time in plants, were previously reported to contain complementary sequence of miR156 in other plants [38]. The target prediction results suggested that SBP transcription factors can be considered as the targets of can-miR156 as well. MYB transcription factors, which are known to regulate meristem formation and seed development [38,39,40,41], are conserved targets of miR159/319 families. The homologous transcripts of MYB genes were identified in pepper, which were considered as the target of can-miR159 and can-miR319 families. ARF, which are known to play important roles in plant development as activators or repressors of auxin-responsive transcription [42], are conserved targets of miR160/167 families in *Arabidopsis* [38,39]. The prediction results also indicate

that ARFs are presumed to be targeted by the can-miR160 and can-miR167 families. The miR482-mediated regulation of NBS-LRR disease resistance proteins, recognizing specific pathogen effectors and triggering resistance responses, are conserved in several plant species [43,44,45]. The target prediction results suggest that NBS-LRR disease resistance proteins in pepper were also presumed to be targeted by can-miR482 families. Similarly, many other targets, including GRF, F-box, TCP transcription factors, and NAM are potential targets of can-miR396, can-miR394, can-miR319, and can-miR164, respectively.

Among 35 novel miRNA families, a total of 57 potential target genes were identified from 19 novel miRNA families (Table 5; This was carried out in collaboration with Jae Yun Lim). Unlike conserved miRNAs, the targets of novel miRNAs were not enriched in transcription factors; only a target, WRKY transcription factor, was predicted. The largest group of predicted targets was F-box family proteins, predicted to be targeted by can-miR-n002 and can-miR-n005, both of which showed similar expression patterns in terms of high expression levels in root and seeding (see Figure 5). The second largest group were the following two groups; i) disease resistance-related genes such as verticillium wilt disease resistance proteins, late blight resistance proteins and NBS-LRR type disease resistance proteins. ii) protein kinase genes. Among the

protein kinase group, leucine-rich repeat protein kinases were known to be critical components of PTI (PAMP-triggered immunity), one of the plant immune systems triggered by pathogen-associated molecules [46,47]. Therefore, miRNAs predicted to target these kinases also appeared to be involved in immune system of pepper through cooperating with other miRNAs related to disease resistance protein. The fourth largest group was GDSL-lipase like chlorogenate-dependent caffeoyltransferase precursor known to have multifunctional properties and be related in lipid metabolism [48].

Validation of miRNA-directed target cleavage

It is necessary to experimentally validate whether computationally predicted targets are actually regulated by miRNAs in pepper by means of miRNA-directed cleavage of these targets because plant miRNAs regulate their target genes mainly by cleaving them [3,15]. Therefore, 5' RACE, universally used for detecting miRNA-directed target cleavage, was carried out for some of the predicted targets of conserved miRNAs; SBP, ARF, NAM, F-box, MYB, Argonuate, TCP, and sulphate transporter (Figure 6; This experiment was performed by June Hyun Park, Donghyn Kim, and Yourim Choi). One of the SBP transcription factor was targeted only by can-miR156d-g (Figure 6A),

whereas another SBP transcription factor was targeted both by can-miR156a-c and can-miR156d-g (Figure 6B). In addition, two cleavage sites of the SBP transcription factor were mapped owing to a single length difference between can-miR156a-c and can-miR156d-g (Figure 6B). Likewise, the miRNA-directed target cleavages of ARF, NAM, F-box, MYB, Argonuate, TCP, and sulphate transporter were validated as well (Figure 6C, D, E, F, G, H, I, J and K).

5'RACE analysis was also carried out for many of the novel miRNA targets. Of the 57 predicted targets of novel miRNAs, 26 targets were tested; 5 of them were validated, but most of them (21) could not be validated (Table 8). Overall, most of the predicted targets of novel miRNAs had high-penalty scores (49 out of 58, 86% had penalty scores of 3 or 4) (i.e., low sequence complementarity to miRNAs), especially for targets which could not be validated. On the contrary, the validated targets generally had lower penalty scores (i.e., higher sequence complementarity to miRNAs) (Table 8). The novel miRNA-directed target cleavage was validated for F-box family proteins and GDSL lipase-like chlorogenate-dependent caffeoyltransferase precursor (Figure 7). F-box family proteins were known to play a role in protein degradation as one component of SCF complex [49] and also associated with regulation of cell cycle and signal transduction [50]. Among the predicted novel miRNA targets, F-box family proteins were the largest group (Table 5) and were predicted to be

targeted by can-miR-n002 and can-miR-n005, both of which had a similar expression pattern; abundant expression in root and seedling (Figure 5). Additionally, F-box family proteins share higher sequence complementarity (i.e. low penalty score) to the can-miR-002 and can-miR-005. Consequently, positive results of four F-box family proteins were obtained (Figure 7A, B, C and D). GDSL lipase-like chlorogenate-dependent caffeoyltransferase precursor, another validated target, had a single nucleotide bulge on can-miR-n026, but the miRNA-mediated cleavage was still observed (Figure 7E).

In summary, it was experimentally confirmed that conserved and novel miRNAs in pepper negatively regulate the expression of many transcription factors (SBP, ARF, MYB, NAM, F-box and TCP), small RNA biogenesis protein (AGO), sulphate transporter and GDSL lipase-like chlorogenate-dependent caffeoyltransferase precursor by the miRNA-directed cleavage mechanism (Figure 6 and 7).

DISCUSSION

The functional study and database construction of plant miRNAs from non-model species are still in their infancy. However, many recent studies have demonstrated that plant miRNAs are closely associated with a variety of functional roles including developmental phase transition, morphogenetic processes, and so forth. Pepper is one of the most economically significant crops cultivated worldwide, but systematic study of miRNAs in pepper has been reported scarcely. A previous study of pepper miRNAs using an *in silico* approach [13] was able to identify a very limited number of miRNAs. Many previous studies employed expressed sequence tag (EST) analysis approaches to identify miRNAs from other species [51,52,53,54], especially for species whose genomes were poorly understood [2]. Since a limited number of ESTs were available in the database a great proportion of protein-coding genes were missed, which makes effective computational prediction of miRNAs and their target genes difficult. In this study, using draft genome sequences, a large number of miRNAs were identified in pepper extensively from 10 different libraries. This work therefore provides the first reliable draft of the pepper miRNA transcriptome. Identified miRNAs in pepper also were successfully differentiated from other small RNAs by extensive investigation of pre-miRNA fold-back

structures. From the 10 different libraries, many pepper miRNAs exhibited tissue-specific expression, as is commonly inferred [55]. These patterns were further validated and observed in the northern blot analysis as well.

Conserved miRNAs in pepper

ath-miR156, which is involved in floral development and phase change by targeting SBP transcription factors, is one of the most abundantly expressed miRNAs in *Arabidopsis* [23,56,57,58]. Expression of can-miR156 reached its highest level at the seedling stage as previously reported in other plants [59,60,61], suggesting it regulates juvenile-to-adult phase transition in pepper. Previous studies suggested that over-expression of ath-miR156 affected phase transition from vegetative growth to reproductive growth [62,63,64] including a decrease in apical dominance [62,65,66], which is in full agreement with expression analysis in pepper where can-miR156 was barely expressed in stem and fruit (Figure 4A). The 5' RACE assay confirmed that some members of the SBP transcription factor families are regulated by the miRNA-directed cleavage mechanism, depending on sequence specificity of can-miR156 (Figure 6A and B).

Similarly to can-miR156, can-miR164 was temporally regulated because

its expression was low in the seedling and high in adult tissues except for the leaf (Figure 4C). Expression of ath-miR164 negatively regulates cell death by repressing the NAC transcription factor *ORESARA1* (*ORE1*), a positive regulator of cell death [67,68]. In addition, decreased expression of ath-miR164 leads to increased expression of *ORE1* with leaf age [68], and this correlates with pepper expression analysis where the expression of can-miR164 was barely detectable in adult leaves (Figure 4C).

ath-miR166, which targets the HD-ZIP gene family, is conserved in many land plants [69]. In *Arabidopsis*, over-expression of ath-miR166 resulted in seedling arrest and diverse phenotypic changes, especially in floral structure and in shoot apical meristem (SAM) formation [70,71]. It was found that can-miR166 was highly expressed in flower followed by stem and root, and weakly in the seedling and leaf, which was closely associated with their over-expression consequences (Figure 4D).

In *Arabidopsis*, ath-miR159 and ath-miR319 share 17 identical nucleotides in their 21nt mature miRNA sequences, and they both play significant roles in plant development, morphogenesis and reproduction [72,73]. Computational prediction suggests that ath-miR159 and ath-miR319 might potentially regulate both MYB and TCP transcription factors [72,73] owing to their sequence similarity. However, when studied *in vivo*, over-expression of ath-

miR159 specifically down-regulated *MYB* mRNA expression, and over-expression of ath-miR319 specifically down-regulated *TCP* mRNA expression in *Arabidopsis* [74]. Similarly, in pepper, can-miR159 and can-miR319 are seemingly related due to their similarity in mature miRNA sequences. The 5' RACE assay confirmed that can-miR159 and can-miR319 specifically regulate MYB transcription factors and TCP transcription factors, respectively (Figure 6C, D and I). This result is consistent with those reported for *Arabidopsis*. Expression of miR159 is abundant in all tissues and widespread over the whole plant [59]. Similarly, it was found that expression of can-miR159 is stably expressed in all tissues in pepper as well (Figure 4B). However, miR319 is known to be expressed at much lower level and restricted to certain tissues and specific developmental stages in other plants [59], which seems to be different from can-miR319 whose expression is consistently high in all of the tissues tested (Figure 4H).

A previous study demonstrated that miR395 was induced under sulfate starvation conditions [75]. Several genes, including sulfate transporter and ATP sulfurylases (APS), both of which are involved in the sulfur assimilation pathway, are regulated by miR395 in *Arabidopsis* [76,77]. From the target prediction analysis, can-miR395 was also predicted to target sulfate transporter and APS (Table 4). The can-miR395-directed cleavage of sulfate transporter was tested by

the 5' RACE assay. It revealed that cleavage in the position corresponding to 9th and 10th nucleotides within their complementary site was clearly detected while no other cleavage product was found near around (Figure 6J). Although miRNAs are generally known to induce cleavage of their target mRNA between 10th and 11th nucleotides, non-canonical cleavage sites have also been reported in plants [23,75,78] and in mammals [79]. Interestingly, a recent study of miR395 in *Arabidopsis* reported the miR395-directed cleavage of *APS3* mRNA in the position corresponding to 9th and 10th nucleotides in their complementary site [76], which is similar to 5' RACE analysis in pepper.

Novel miRNAs in pepper

Many earlier studies repeatedly reported that non-conserved miRNAs are generally less abundant, more divergent, processed less precisely, and tend to lack of experimentally verifiable targets, suggesting that most are neutrally evolving [80,81,82] . Therefore, identification and characterization of novel miRNAs in pepper might be challenging, as there were many things to consider.

The variation in mature miRNA sequences, especially that of 5' terminal nucleotides, often showed differential Argonaut protein association [83,84], and hence the mature miRNA sequence itself is now considered to be an important

determinant of proper sorting into functional Argonaut complex. Therefore, the imprecise excision of the mature miRNAs from the stem-loop precursor could produce a functionally unstable miRNAs. For these reasons, very stringent standards were employed for the identification of novel miRNA candidates, as followed. 1) one of the proofs of miRNA biogenesis; the presence of the miRNA star sequences, forming a miRNA/miRNA* duplex with two nucleotides, 3' overhang; 2) the presence of reliable stem-loop structures with proper folding free-energy; 3) minimal in size and frequency of asymmetric bulge, along with four or fewer mismatched miRNA bases within miRNA/miRNA* duplex; 4) highly expressed small RNAs with frequency of 1,000 or higher in at least one or more from 10 different libraries examined.

As a result, a total of 50 highly reliable novel miRNAs candidates from various pepper tissues. These novel miRNAs had many predicted targets; consequently a 5' RACE assay was performed to reveal miRNAs involved in pepper-specific biology, but most of them could not be validated, which are consistent with previous studies. It has been reported that most of the non-conserved miRNAs had no experimentally supported targets [9,85], in contrast to the high validation rate of conserved miRNA targets . There are several possible explanations for this observation. 1) The novel miRNA targets that gave negative results upon 5' RACE analysis, other than technical failures, could be false

positive prediction owing to the limited genome annotation currently available in pepper, which was the major hindrance to high confidence prediction of miRNA targets. 2) The non-conserved miRNAs and their target genes are not co-expressed in the same tissue during development [80]. 3) The abundance of novel miRNAs is substantially lower than that of conserved miRNAs [21,27,34,35,36], as is observed as well. In addition, most of the predicted targets of novel miRNAs had high-penalty score; low sequence complementarity. 4) The non-conserved miRNAs might function mainly through translational inhibition [4,5], which is also supported by the fact that the level of non-conserved miRNA target transcripts were largely unchanged in miRNA biogenesis mutants [21]. 5) Some of the putative novel miRNAs are not *bona fide* miRNAs; although one of the major proofs of biogenesis where the perfect miRNA star sequences were found, DCL1 dependency of these novel miRNAs was not demonstrated because a *dcl1* mutant is not yet available for pepper. To conclude, while none of these possibilities can be ruled out at this point, it is currently difficult to explain the lack of experimental verifiable targets of the majority of non-conserved miRNAs; the other possible regulatory mechanisms behind it still remain to be elucidated.

TABLES

Table 1. Distribution of small RNAs in pepper from 10 different libraries

		Redundant		Non-redundant	
		Reads	Matching pepper genome	Reads	Matching pepper genome
Leaf	Raw reads	87,761,235			
	Adapter removed	74,310,576		14,057,223	
	rRNA/tRNA removed	38,704,582	24,558,036	11,709,178	8,086,024
	Match known miRNAs ≥1	2,274,251	2,167,895	3,905(62)	796(58)
	Match known miRNAs ≥50	2,273,930	2,167,618	3,823(41)	724(39)
	Conserved miRNA from predicted hairpins with abundance ≥1		2,197,192		352(33)
	Conserved miRNA from predicted hairpins with abundance ≥50		2,195,901		266(31)
	New miRNA from predicted hairpins with abundance ≥1		261,257		541(161)
	New miRNA from predicted hairpins with abundance ≥50		256,460		286(96)
Stem	Raw reads	52,765,696			
	Adapter removed	51,913,006		16,192,994	
	rRNA/tRNA removed	32,233,648	25,845,658	13,530,265	9,934,033
	Match known miRNAs ≥1	1,830,323	1,753,729	4,264(63)	805(58)
	Match known miRNAs ≥50	1,830,033	1,753,411	4,167(40)	711(37)
	Conserved miRNA from predicted hairpins with abundance ≥1		1,749,505		366(33)
	Conserved miRNA from predicted hairpins with abundance ≥50		1,747,201		232(31)
	New miRNA from predicted hairpins with abundance ≥1		212,480		561(164)
	New miRNA from predicted hairpins with abundance ≥50		207,500		305(118)
Root	Raw reads	50,434,157			
	Adapter removed	50,072,353		15,224,050	
	rRNA/tRNA removed	30,969,423	25,681,254	12,874,799	9,776,038
	Match known miRNAs ≥1	3,498,172	3,404,672	3,681(62)	833(57)
	Match known miRNAs ≥50	3,497,889	3,404,377	3,591(40)	751(38)
	Conserved miRNA from predicted hairpins with abundance ≥1		3,393,030		353(33)

	Conserved miRNA from predicted hairpins with abundance ≥50		3,390,712		225(32)
	New miRNA from predicted hairpins with abundance ≥1		170,849		554(165)
	New miRNA from predicted hairpins with abundance ≥50		164,805		243(106)
Flower	Raw reads	61,873,756			
	Adapter removed	61,476,055		16,877,140	
	rRNA/tRNA removed	40,888,746	28,328,301	15,028,664	11,552,651
	Match known miRNAs ≥1	2,316,757	2,258,587	3,135(62)	714(55)
	Match known miRNAs ≥50	2,316,430	2,258,373	3,003(39)	623(38)
	Conserved miRNA from predicted hairpins with abundance ≥1		2,281,533		364(33)
	Conserved miRNA from predicted hairpins with abundance ≥50		2,278,748		190(31)
	New miRNA from predicted hairpins with abundance ≥1		120,484		587(167)
	New miRNA from predicted hairpins with abundance ≥50		114,722		281(107)
Fruit-1	Raw reads	62,988,553			
	Adapter removed	62,293,585		19,478,268	
	rRNA/tRNA removed	39,270,026	31,930,008	16,737,872	12,375,524
	Match known miRNAs ≥1	1,687,948	1,614,810	3,391(71)	677(57)
	Match known miRNAs ≥50	1,687,612	1,614,436	3,270(36)	546(34)
	Conserved miRNA from predicted hairpins with abundance ≥1		1,602,497		333(33)
	Conserved miRNA from predicted hairpins with abundance ≥50		1,599,969		157(29)
	New miRNA from predicted hairpins with abundance ≥1		115,256		557(170)
	New miRNA from predicted hairpins with abundance ≥50		109,329		216(89)
Fruit-2	Raw reads	43,032,671			
	Adapter removed	42,289,844		17,537,387	
	rRNA/tRNA removed	31,504,608	25,518,067	15,580,091	11,752,392
	Match known miRNAs ≥1	1,243,288	1,193,441	3,098(62)	659(57)
	Match known miRNAs ≥50	1,242,853	1,193,103	2,959(36)	554(35)
	Conserved miRNA from predicted hairpins with abundance ≥1		1,184,731		348(33)
	Conserved miRNA from predicted hairpins with abundance ≥50		1,181,695		146(29)
	New miRNA from predicted hairpins with abundance ≥1		115,541		592(172)
	New miRNA from predicted hairpins with abundance ≥50		109,397		284(117)
Fruit-MG	Raw reads	46,725,193			
	Adapter removed	46,424,679		12,849,765	
	rRNA/tRNA removed	22,000,286	17,766,991	11,296,813	8,585,158
	Match known miRNAs ≥1	730,015	701,563	2,472(59)	580(50)
	Match known miRNAs ≥50	729,629	701,281	2,328(34)	480(33)

	Conserved miRNA from predicted hairpins with abundance ≥1		702,823		335(33)
	Conserved miRNA from predicted hairpins with abundance ≥50		699,794		118(26)
	New miRNA from predicted hairpins with abundance ≥1		91,716		566(167)
	New miRNA from predicted hairpins with abundance ≥50		85,099		224(87)
Fruit-B	Raw reads	47,761,557			
	Adapter removed	46,847,384		15,359,015	
	rRNA/tRNA removed	26,156,391	21,046,683	13,820,419	10,470,918
	Match known miRNAs ≥1	569,133	545,286	2,343(59)	596(55)
	Match known miRNAs ≥50	568,863	545,024	2,237(36)	501(34)
	Conserved miRNA from predicted hairpins with abundance ≥1		547,838		340(33)
	Conserved miRNA from predicted hairpins with abundance ≥50		544,977		123(29)
	New miRNA from predicted hairpins with abundance ≥1		102,788		569(171)
	New miRNA from predicted hairpins with abundance ≥50		96,081		235(90)
Fruit-B5	Raw reads	38,168,067			
	Adapter removed	37,640,821		11,019,173	
	rRNA/tRNA removed	19,591,941	15,809,606	10,046,174	7,616,358
	Match known miRNAs ≥1	742,857	714,548	2,422(58)	566(51)
	Match known miRNAs ≥50	742,549	714,256	2,330(33)	479(31)
	Conserved miRNA from predicted hairpins with abundance ≥1		731,273		325(33)
	Conserved miRNA from predicted hairpins with abundance ≥50		728,288		111(26)
	New miRNA from predicted hairpins with abundance ≥1		83,461		561(167)
	New miRNA from predicted hairpins with abundance ≥50		76,608		203(86)
Fruit-B10	Raw reads	51,427,930			
	Adapter removed	50,933,647		13,459,462	
	rRNA/tRNA removed	25,053,214	20,260,032	12,366,889	9,400,255
	Match known miRNAs ≥1	897,174	863,929	2,716(62)	627(58)
	Match known miRNAs ≥50	896,827	863,631	2,589(34)	522(32)
	Conserved miRNA from predicted hairpins with abundance ≥1		885,773		328(33)
	Conserved miRNA from predicted hairpins with abundance ≥50		883,130		136(27)
	New miRNA from predicted hairpins with abundance ≥1		107,247		566(170)
	New miRNA from predicted hairpins with abundance ≥50		102,035		268(101)

The number in parenthesis denotes the unique sequence. This was performed by Jae Yun Lim.

Table 2. Conserved miRNAs in pepper

Conserved miRNA	Sequence (5'→3')	Reads of miRNA sequences from 10 different libraries										Hp #
		Leaf	Stem	Root	Flower	Fruit-1	Fruit-2	Fruit-MG	Fruit-B1	Fruit-B5	Fruit-B10	
can-miR156a-c	UUGACAGAAGAUAGAGAGCAC	698	2407	7589	2815	32	225	1210	2715	6494	23305	3
can-miR156d-g	UGACAGAAGAGAGUGAGCAC	10	94	1166	32	4	20	52	53	91	152	4
can-miR156h	UUGACAGAAGAGAGUGAGCAU	84	176	72	74	53	89	40	52	65	105	1
can-miR159a-c	UUUGGAUUGAAGGGAGCUCUA	117204	61460	54057	64159	38667	29312	29700	56716	76148	101967	3
can-miR160	UGCCUGGCUCCCUGUAUGCCA	2724	4284	938	856	365	517	741	371	783	894	1
can-miR162a-e	UCGAUAAACCUCUGCAUCCAG	1484	4525	10282	3839	2061	4702	2101	2360	2832	2573	5
can-miR162f	UCGAUAAACCUCUGCAUCCGG	671	1368	4541	3319	3433	5796	3231	2933	3119	1747	1
can-miR164a-b	UGGAGAAGCAGGGCACGUGCA	95	381	118	293	699	458	3642	2902	18082	16181	2
can-miR164c	UGGAGAAGCAGGGCGCGUGCA	114	169	4	4	4	4	11	2	11	8	1
can-miR166a-h	UCGGACCAGGCUUCAUUCCCC	268932	307399	211712	468628	450083	588387	335556	220638	423083	315156	8
can-miR166i-j	UCUCGGACCAGGCUUCAUUCC	286713	355950	2481041	916101	245209	233061	244844	196621	333460	296253	2
can-miR166k	UCGAACCAGGCUUCAUUCCCC	133	304	164	269	360	469	276	217	345	271	1
can-miR166l	UCGGACCAGGCUUCAUUCCUC	109232	42342	15520	34410	38463	12901	12758	9383	17920	23174	1
can-miR167a-c	UGAAGCUGCCAGCAUGAUCUA	1921	249	548	1845	1940	3059	1819	1455	1666	1849	3
can-miR167d	UGAAGCUGCCAGCAUGAUCUGG	6379	249	2550	185	17	92	262	235	460	790	1

can-miR168a-b	UCGCUUGGUGCAGGUCGGGAC	3495	6790	8629	2794	1529	5060	2596	2144	3729	4012	2
can-miR168c	UCGCUUGGUGCAGGUCGGGAA	2998	2496	2447	955	720	1150	425	297	460	389	1
can-miR169a-g	UAGCCAAGGAUGACUUGCCU	89	765	3070	58	8	38	23	36	13	3	7
can-miR169h	UAUCGGCAGGUCAUCCUUGGC	7	17	56	14	4	2	2	0	2	1	1
can-miR171a-e	UGAUUGAGCCGUGCCAAUAUC	781	724	454	462	703	380	190	255	244	251	5
can-miR171f-g	UGAUUGAGCCGUGUCAAUAUC	179	73	3	83	20	9	12	24	10	10	2
can-miR171h	UGAGCCGAACCAAUAUCACUC	183	446	4003	813	994	404	172	104	85	86	1
can-miR171i	UUGAGCCGCGCCAAUAUCACU	79	263	243	8	30	27	11	8	8	5	1
can-miR171j	UUGAGCCGCGCCAAUAUCAUU	8	61	253	3	8	4	4	4	2	3	1
can-miR171k	UUGAGCCGUGCCAAUAUCACGU	337	124	104	47	55	15	11	46	80	81	1
can-miR171l	UUGAGCCGCGCCAAUAUCACG	1475	293	56	358	244	226	817	619	782	806	1
can-miR172a-d	AGAAUCUUGAUGAUGCUGCAU	472	324	53	210	52	81	110	161	206	155	4
can-miR172e	UGAAUCUUGAUGAUGCUGCAU	65	67	43	47	13	42	8	13	26	37	1
can-miR172f	AGAAUCUUGAUGAUGCUGCAG	1	4	2	285	20	12	14	62	241	475	1
can-miR172g	GGAAUCUUGAUGAUGCUGC	9	44	1	260	44	48	181	190	426	468	1
can-miR172h	AGAAUCUUGAUGCUGCUGCAU	3199	336	2	139	72	23	4	0	6	0	1
can-miR172i	GGAAUCUUGAUGAUGCUGCAG	7	85	5	545	56	174	383	583	1112	1569	1
can-miR319a	CUUGGACUGAAGGGAGCUCCC	588	64598	13611	41664	1719	730	2920	2704	3384	6198	1

can-miR319b-c	UUGGACUGAAGGGAGCUCCCU	178	143377	16249	5076	711	776	2872	2502	3285	3443	2
can-miR319d	UCUUGGACUGAAGGGUUCCCU	7	135	134	6142	14	26	122	241	616	1017	1
can-miR319e	UUGGACUGAAGGGAGCUCCC	669	50859	20520	25530	3858	1126	4486	4004	6198	7384	1
can-miR319f	UUGGACUGAAGGGAGCUCCUU	341	5936	2364	2375	500	448	816	916	1120	1228	1
can-miR319g	UUGGACUGAAGGGAGCUCCU	181	2819	783	1183	252	203	339	431	349	478	1
can-miR390a-b	AAGCUCAGGAGGGAUAGCGCC	258	471	1712	784	81	19	35	30	35	44	2
can-miR390c	AAGCUCAGGAGGGAUAGCACC	605	971	1241	125	42	364	64	75	93	162	1
can-miR393a-b	UCCAAAGGGAUCGCAUUGAUCC	3005	354	125	87	50	135	90	77	129	106	2
can-miR394a	UUGGCAUUCUGUCCACCUCC	1083	1717	406	581	184	71	107	114	99	122	1
can-miR394b	UUGGCAUUCUGUCUACCUCC	1481	930	78	542	47	151	290	179	132	111	1
can-miR395a-j	CUGAAGUGUUUGGGGGAACUC	124	54	117	150	10	28	39	74	45	39	10
can-miR396a	UUCCACAGCUUUCUUGAACUG	101458	18914	39283	8223	4475	20644	14973	12070	12491	10479	1
can-miR396b	UUCCACAGCUUUCUUGAACUU	16864	8481	7252	3095	1994	12496	1556	1324	2224	3435	1
can-miR396c	UUCCACAGCUUUCUUGAACUA	6618	3489	1041	180	147	863	145	138	134	141	1
can-miR397a	UCAUCUACGCUGCACUCAAUC	640	9	8	29	10	10	9	2	6	6	1
can-miR397b	AUUGAGUGCAGCGUUGAUGAC	57518	1047	2264	1621	757	1449	891	488	566	519	1
can-miR398a-b	UGUGUUCUCAGGUCACCCCUU	10	84	22	23	6	3	12	5	10	18	2
can-miR398c	UAUGUUCUCAGGUCACCCCUA	857	142	239	402	27	69	75	84	136	170	1

can-miR398d	UGUGUUCUCAGGUCGCCCCUG	67733	6350	4468	23607	30511	51585	19279	8530	7322	3941	1
can-miR399a-e	UGCCAAAGGAGAAUUGCCCUG	62	0	21	5	10	18	21	36	22	19	5
can-miR399f-g	UGCCAAAGGAGAGUUGCCCUG	423	20	91	131	278	188	258	465	303	326	2
can-miR399h	CGCCAAAGGAGAGCUGCCCUA	18	0	1	6	11	8	12	44	22	13	1
can-miR403	UUAGAUUCACGCACAAACUCG	8057	11621	19143	14241	8650	7548	6154	5391	6186	6813	1
can-miR408a	UGCACUGCCUCUUCCCUGGCU	76418	4108	2605	5224	4684	4980	3144	1334	1704	1715	1
can-miR408b	UGCACAGCCUCUUCCCUGGCU	17105	141	14	990	1026	1386	203	99	77	91	1
can-miR477	CCUCUCCCUCAAGGGCUUCUU	4	234	8115	6	0	101	16	63	19	11	1
can-miR482a	UUGCCGAUUCCGUCCAUACCGC	8254	4153	4415	1718	2583	2457	2387	1508	3290	3251	1
can-miR482b	UUUCCAAUUCCACCCAUUCCUA	2112	4529	1615	1543	3907	3278	1151	972	1147	1327	1
can-miR482c	UCUUGCCUACACCGCCCAUGCC	17638	23433	36831	7553	9871	17531	5334	4450	6747	8394	1
can-miR482d	UCUUACCGAUACCUCCCAUUCC	2016	3022	3637	1146	1413	752	500	452	815	879	1
can-miR482e	CUACCAACUCCACCCAUUCCUG	7	4	53	0	1	0	46	43	420	1660	1
can-miR482f	UCUUUCCUACUCCUCCCAUACC	166081	118677	105171	216948	173152	66770	46358	40194	66479	66781	1
can-miR482g	UUUCCUAUUCCACCCAUGCCAA	943	1235	1612	906	868	742	434	386	716	1013	1
can-miR530	UGCAUUUGCACCUGCACCUGU	12407	242	709	80	42	14	21	8	13	9	1
can-miR827	UUAGAUGAACAUCAACAAACA	4936	940	44	297	79	202	160	234	65	124	1
can-miR1446a-	UGAACUCUCUCCCUCAAUGGCU	329	10	1195	35	3	2	6	6	5	28	2

b												
can-miR4376	ACGCAGGAGAGA UGAUGCUGGA	2029	2249	301	90	168	113	49	36	42	21	1
can-miR4414a	AGCUGAUGACUC GUUGAUUCU	0	8	0	188	6	3	0	1	0	0	1
can-miR4414b	AGCUGCUGAAUC AUUGGUUCG	2	84	33	1716	10	5	49	95	81	40	1

MG stands for 'mature green'. **Hp#(Hairpin#)** indicates the number of hairpin locus. This was carried out in collaboration with Jae Yun Lim.

Table 3. Novel miRNAs in pepper

Novel miRNA	Sequence (5' → 3')	Reads of miRNA sequences from 10 different libraries										Hp #
		Leaf	Stem	Root	Flower	Fruit-1	Fruit-2	Fruit-MG	Fruit-B1	Fruit-B5	Fruit-B10	
can-miR-n001	UUUCUGUUUUGAUAGUAGGCCU	0	1025	976	1284	244	564	435	699	780	660	1
can-miR-n002a-c	UUGCAAACACACCUGAAUCGU	70358	30781	34831	323	55	239	309	298	405	660	3
can-miR-n003a	UGUAGUUGUAGCCAUUCUAUU	1379	1313	2138	2418	242	452	104	118	101	94	1
can-miR-n003b-d	UAGAGUGGCCACAACUAGAUG	1456	1132	1494	1256	860	619	214	171	201	142	3
can-miR-n004	UAAGAUCGAGCACAAGUUGUU	2798	2787	3656	3161	2170	2828	2507	2575	3389	3656	1
can-miR-n005	UUCGAUACGCACCUGAAUCGCC	76	9330	24545	268	5	76	218	112	145	120	1
can-miR-n006	CAACAAUCAUCCUUUGGGCUUU	57	308	223	856	655	15403	17379	19725	17193	12432	1
can-miR-n007	UUGAACCUUCAGGUGAGUUGC	640	896	2212	405	199	352	294	304	448	495	1
can-miR-n008	CGGCAUGAGAGAAAAUAUUGAGAA	0	0	4	4	0	420	953	1589	1433	4056	1
can-miR-n009a,b-5p	CCUGAACUGAACAAUACGAUC	7529	1197	1233	1059	1626	3154	1023	1197	452	557	2
can-miR-n009a,b-3p	UGGUAUUGUUCCGUUCAGGGA	2816	1287	780	575	391	1187	1250	1071	1302	842	2
can-miR-n010	UUAUGAGAUAAGUUCAACACG	1777	1179	1361	642	179	583	274	356	364	378	1
can-miR-n011	AGCAUCGAUUACAAUGACAAAAAG	29	76	79	170	74	410	431	698	1404	1284	1
can-miR-n012	UUCGGCUCAUCGUUGUUGCAGACG	300	775	492	400	683	874	849	794	970	533	1
can-miR-n013	UUUUAGCAAGAGUUGUUUUCC	671	578	912	618	696	357	145	225	199	136	1

can-miR-n014	CUGAAGUUCGUUACUGUUGUC	238	294	306	61	35	458	933	1622	945	744	1
can-miR-n015	UUCAGGCCUGAGAAACGAAAAACU	2397	11777	7357	12552	4329	4502	6435	9528	13191	11762	1
can-miR-n016a-b	AUCCAAUACAUCGUCCACAGC	8946	3982	958	1942	1952	1393	488	304	472	459	2
can-miR-n017	CUAAGCAACGCUAUGUCGAGU	13469	12610	5845	7831	7249	7142	4029	3330	3874	3730	1
can-miR-n018	AAUUGGACUGGCGCACAUCGGGAG	414	1597	1071	667	3192	2254	1715	1311	1923	2085	1
can-miR-n019a-d	UAAUAACUAGUAGUUGAGUGAU	0	3	1	1604	1	3	7	5	16	25	4
can-miR-n020	GGGGAUGUAGCUCAGAUGGUAGA	18719	7224	4696	1702	6942	1877	1842	3305	5541	3730	1
can-miR-n021	AUGGAUGAACAAUGCUGAAACAUU	173	1288	453	374	296	477	238	339	187	182	1
can-miR-n022a-c	UGAACAAGUAGACACAUGCGUCCU	307	237	217	252	199	228	240	280	479	2654	3
can-miR-n023	AUGACUUACUUUGACUUGGCACA	824	1424	611	808	819	752	915	999	935	824	1
can-miR-n024	UCAACUGCAAACCUGUAAGCCU	2392	5907	3492	11515	5334	5767	8755	9367	9857	9612	1
can-miR-n025	AGGAAGGAACUCCACGUCAUUGCU	57	12	12	124	331	3906	424	501	147	101	1
can-miR-n026	UGUCACAAUGAACUCCAUCCCA	11593	7823	4814	2852	1659	3427	2623	2672	3802	3961	1
can-miR-n027	AUGAAAGUUGUCGAUGCCAAC	6596	6076	10157	2132	2774	3073	1358	1096	1434	1273	1
can-miR-n028a-b	UCAUCCGAGAUCGUUUCGCUGA	1273	5468	1195	5568	4510	1967	3181	3131	3836	2915	2
can-miR-n029	UUAGAGUGAGCUCAACAGAGU	49	18	27	9	4	104	741	1605	1682	3318	1
can-miR-n030	UUCGAGGACCGUCAGUAGCAUA	1512	1	0	3	1	0	0	0	0	1	1
can-miR-n031	UUCCCAGUCCAGGCAUUCCAAC	2140	909	1063	1128	811	882	682	624	1000	1142	1

can-miR-n032	UGUUCCUGUAGAUAAGCCACU	524	319	374	360	213	733	1286	2260	2377	2543	1
can-miR-n033	UAGAGAAAGCAUGGCUUCAGGU	5101	7725	6752	4808	2785	2836	2713	3261	5975	5194	1
can-miR-n034	ACGGAUGAACUCGCAGAAGGACGA	59	8	372	125	100	928	1574	981	662	656	1
can-miR-n035	CUUCGAACUAACUCCUUCUUGACA	0	24	23	9	15	1049	812	1495	927	824	1

[a] These miRNAs are partially conserved in Solanaceae**. MG** stands for 'mature green'. **Hp#(Hairpin#)** indicates the number of hairpin locus. This was carried out in collaboration with Jae Yun Lim.

Table 4. Predicted targets of conserved miRNAs

Conserved miRNA	S	Predicted targets	Reference species
can-miR156a-c	3.5	acyltransferase-like protein	Nicotiana tabacum
can-miR156a-c	4	LIGULELESS1 protein, putative	Ricinus communis
can-miR156a-c	3	LIGULELESS1 protein, putative	Ricinus communis
can-miR156a-c	2	transcription factor squamosa promoter binding protein-like	Eucalyptus grandis
can-miR156a-c	2	LIGULELESS1 protein, putative	Ricinus communis
can-miR156a-c	2	promoter-binding protein SPL9	Vitis vinifera
can-miR156d-g	2	LIGULELESS1 protein, putative	Ricinus communis
can-miR156d-g	1	LIGULELESS1 protein, putative	Ricinus communis
can-miR156d-g	1	LIGULELESS1 protein, putative	Ricinus communis
can-miR156d-g	3	DNA topoisomerase family protein	Arabidopsis thaliana
can-miR156d-g	1	squamosa promoter-binding protein	Citrus trifoliata
can-miR156d-g	1	promoter-binding protein SPL9	Vitis vinifera
can-miR156h	2	LIGULELESS1 protein	Ricinus communis
can-miR156h	1.5	LIGULELESS1 protein	Ricinus communis
can-miR156h	1.5	LIGULELESS1 protein	Ricinus communis
can-miR156h	1.5	LIGULELESS1 protein	Ricinus communis
can-miR156h	1.5	LIGULELESS1 protein	Ricinus communis
can-miR159a-c	4	myb-related transcription factor	Solanum lycopersicum
can-miR159a-c	3.5	calcium-binding EF hand family protein	Arabidopsis lyrata subsp. lyrata
can-miR159a-c	4	GRAS family transcription factor	Populus trichocarpa
can-miR159a-c	3.5	r2r3-myb transcription factor, putative	Ricinus communis
can-miR159a-c	4	putative glycosyltransferase	Solanum aculeatissimum
can-miR159a-c	2.5	GAMyb-like1	Solanum lycopersicum
can-miR159a-c	3	GAMYB-like2	Solanum lycopersicum
can-miR159a-c	4	NHL repeat-containing protein	Arabidopsis thaliana
can-miR159a-c	4	polyprotein	Glycine max
can-miR159a-c	3.5	r2r3-myb transcription factor, putative	Ricinus communis
can-miR160	0.5	putative auxin response factor ARF16	Malus x domestica
can-miR160	1	auxin response factor 10	Solanum lycopersicum
can-miR160	1	Auxin response factor, putative	Ricinus communis
can-miR160	1.5	auxin response factor 17	Solanum lycopersicum

can-miR164a-b	4	UNE6 (unfertilized embryo sac 6)	Arabidopsis thaliana
can-miR164a-b	4	vesicle-associated membrane protein-related	Arabidopsis thaliana
can-miR164a-b	4	NAC domain-containing protein 21/22, putative	Ricinus communis
can-miR164a-b	3	NO APICAL MERISTEM	Solanum lycopersicum
can-miR164a-b	4	NAC domain-containing protein 21/22, putative	Ricinus communis
can-miR164a-b	3.5	lipase-related	Arabidopsis thaliana
can-miR164a-b	4	Putative late blight resistance protein, identical	Solanum demissum
can-miR164a-b	4	DP-glucuronate decarboxylase 1	Nicotiana tabacum
can-miR164c	3.5	NO APICAL MERISTEM	Solanum lycopersicum
can-miR164c	3.5	zinc finger (C3HC4-type RING finger) family protein	Arabidopsis thaliana
can-miR164c	4	lipase-related	Arabidopsis thaliana
can-miR166i-j	4	Cyclic nucleotide-gated ion channel, putative	Ricinus communis
can-miR167a-c	2.5	multidrug resistance pump, putative	Ricinus communis
can-miR167d	3.5	DNA binding protein, putative	Ricinus communis
can-miR167d	4	multidrug resistance pump, putative	Ricinus communis
can-miR168a-b	4	oxidoreductase family protein	Arabidopsis lyrata subsp. lyrata
can-miR168a-b	4	AGO1	Nicotiana benthamiana
can-miR168c	3	AGO1	Nicotiana benthamiana
can-miR169a-g	4	replication factor A 1, rfa1, putative	Ricinus communis
can-miR169a-g	4	beta-mannosidase enzyme	Solanum lycopersicum
can-miR171a-e	1	GRAS family transcription factor	Populus trichocarpa
can-miR171a-e	1	GRAS family transcription factor	Populus trichocarpa
can-miR171a-e	1	GRAS family transcription factor	Populus trichocarpa
can-miR171f-g	3.5	transmembrane receptor	Arabidopsis thaliana
can-miR171f-g	1.5	GRAS family transcription factor	Populus trichocarpa
can-miR171f-g	1.5	GRAS family transcription factor	Populus trichocarpa
can-miR171f-g	1.5	GRAS family transcription factor	Populus trichocarpa
can-miR171h	0	GRAS family transcription factor	Populus trichocarpa
can-miR171h	3	protein serine/threonine kinase	Arabidopsis thaliana
can-miR171i	4	Tyrosine recombinase XerC	Limnobacter sp. MED105
can-miR171i	4	GRAS family transcription factor	Populus trichocarpa
can-miR171i	1	GRAS family transcription factor	Populus trichocarpa
can-miR171i	1	GRAS family transcription factor	Populus trichocarpa
can-miR171i	3	GRAS family transcription factor	Populus trichocarpa
can-miR171i	4	riboflavin kinase/fmn adenylyltransferase, putative	Ricinus communis

can-miR171i	3	zinc finger protein	Nicotiana benthamiana
can-miR171j	1	GRAS family transcription factor	Populus trichocarpa
can-miR171j	3.5	GRAS family transcription factor	Populus trichocarpa
can-miR171j	3.5	zinc finger protein	Nicotiana benthamiana
can-miR171k	3.5	HIPL1 protein precursor, putative	Ricinus communis
can-miR171k	1.5	GRAS family transcription factor	Populus trichocarpa
can-miR171k	3	GRAS family transcription factor	Populus trichocarpa
can-miR171k	3	GRAS family transcription factor	Populus trichocarpa
can-miR171l	3.5	GRAS family transcription factor	Populus trichocarpa
can-miR171l	2	GRAS family transcription factor	Populus trichocarpa
can-miR171l	2	GRAS family transcription factor	Populus trichocarpa
can-miR171l	4	GRAS family transcription factor	Populus trichocarpa
can-miR171l	3.5	zinc finger protein	Nicotiana benthamiana
can-miR172a-d	3	Helix-loop-helix DNA-binding	Medicago truncatula
can-miR172a-d	4	DNA-repair protein UVH3, putative	Ricinus communis
can-miR172a-d	3.5	serine/threonine protein kinase, putative	Ricinus communis
can-miR172a-d	3.5	exostosin family protein	Arabidopsis thaliana
can-miR172a-d	2	apetala 2-like protein	Nicotiana tabacum
can-miR172a-d	2	apetala 2-like protein	Nicotiana tabacum
can-miR172a-d	4	symbiotic ammonium transporter	Glycine max
can-miR172a-d	3	transcription factor APETALA2	Citrus trifoliata
can-miR172a-d	3	AP2 domain-containing transcription factor	Populus trichocarpa
can-miR172e	4	NADH-ubiquinone oxidoreductase B8 subunit, putative	Arabidopsis thaliana
can-miR172e	4	methylase family protein	Arabidopsis thaliana
can-miR172e	4	ATP binding protein, putative	Ricinus communis
can-miR172e	4	Helix-loop-helix DNA-binding	Medicago truncatula
can-miR172e	4	ubiquitin-protein ligase	Cucumis melo subsp. melo
can-miR172e	4	DNA-repair protein UVH3, putative	Ricinus communis
can-miR172e	3	apetala 2-like protein	Nicotiana tabacum
can-miR172e	3	apetala 2-like protein	Nicotiana tabacum
can-miR172e	3	transcription factor APETALA2	Vitis vinifera
can-miR172e	3	transcription factor APETALA2	Vitis vinifera
can-miR172f	4	Putative mudrA protein - maize transposon MuDR	Oryza sativa
can-miR172f	4	methylase family protein	Arabidopsis thaliana
can-miR172f	4	Helix-loop-helix DNA-binding	Medicago truncatula

can-miR172f	4	MO25 protein -related	Brassica oleracea
can-miR172f	4	rginine decarboxylase	Solanum lycopersicum
can-miR172f	4	WD-40 repeat protein	Solanum lycopersicum
can-miR172f	4	S-locus receptor kinase	Olea europaea
can-miR172f	4	serine/threonine protein kinase, putative	Ricinus communis
can-miR172f	4	xostosin family protein	Arabidopsis thaliana
can-miR172f	3.5	DEAD/DEAH box helicase, putative	Arabidopsis thaliana
can-miR172f	1	AP2 domain-containing transcription factor	Populus trichocarpa
can-miR172f	1	apetala 2-like protein	Nicotiana tabacum
can-miR172f	2	transcription factor APETALA2	Vitis vinifera
can-miR172f	2	transcription factor APETALA2	Vitis vinifera
can-miR172g	3	phospholipase PLDa1	Solanum lycopersicum
can-miR172g	4	phospholipase PLDa1	Solanum lycopersicum
can-miR172g	4	BC-type transport-like protein	Arabidopsis thaliana
can-miR172g	2.5	protein kinase family protein	Arabidopsis thaliana
can-miR172g	4	RNA splicing factor, transesterification mechanism	Arabidopsis thaliana
can-miR172g	3.5	ATP binding protein, putative	Ricinus communis
can-miR172g	2.5	Helix-loop-helix DNA-binding	Medicago truncatula
can-miR172g	4	pre-mRNA-processing factor 39	Arabidopsis thaliana
can-miR172g	4	protein binding / zinc ion binding	Arabidopsis thaliana
can-miR172g	4	pentatricopeptide (PPR) repeat-containing protein	Arabidopsis thaliana
can-miR172g	3.5	pentatricopeptide (PPR) repeat-containing protein	Arabidopsis thaliana
can-miR172g	3.5	polyadenylate-binding protein, putative	Ricinus communis
can-miR172g	4	ATP binding protein, putative	Ricinus communis
can-miR172g	4	3'-5'-exoribonuclease/ RNA binding	Arabidopsis thaliana
can-miR172g	4	acuolar sorting protein, putative	Ricinus communis
can-miR172g	4	MCM protein-like protein	Nicotiana tabacum
can-miR172g	4	ubiquitin-protein ligase	Cucumis melo subsp. melo
can-miR172g	3	DNA-repair protein UVH3, putative	Ricinus communis
can-miR172g	4	LMBR1 integral membrane family protein	Arabidopsis thaliana
can-miR172g	3.5	RWP-RK domain-containing protein	Arabidopsis lyrata subsp. lyrata
can-miR172g	3.5	rginine decarboxylase	Solanum lycopersicum
can-miR172g	4	DNA binding	Arabidopsis thaliana
can-miR172g	4	LBD21 (LOB DOMAIN-CONTAINING PROTEIN 21)	Arabidopsis thaliana
can-miR172g	4	cytochrome P450	Coptis japonica var. dissecta

can-miR172g	3	P2 domain-containing transcription factor	Populus trichocarpa
can-miR172g	2.5	exostosin family protein	Arabidopsis thaliana
can-miR172g	4	aminophospholipid ATPase	Populus trichocarpa
can-miR172g	4	PTF1 (PLASTID TRANSCRIPTION FACTOR 1); transcription factor	Arabidopsis thaliana
can-miR172g	1.5	apetala 2-like protein	Nicotiana tabacum
can-miR172g	1.5	apetala 2-like protein	Nicotiana tabacum
can-miR172g	4	hydroxysteroid dehydrogenase, putative	Ricinus communis
can-miR172g	2.5	transcription factor Myb	Capsicum annuum
can-miR172g	1	transcription factor APETALA2	Vitis vinifera
can-miR172g	1	transcription factor APETALA2	Vitis vinifera
can-miR172h	3.5	axonemal dynein light chain, putative	Ricinus communis
can-miR172h	4	vacuolar protein sorting-associated protein, putative	Ricinus communis
can-miR172h	4	copper-binding family protein	Arabidopsis thaliana
can-miR172h	4	cytochrome P450, putative	Ricinus communis
can-miR172h	4	blight resistance protein RGA	Capsicum annuum
can-miR172h	4	early tobacco anther 1	Nicotiana tabacum
can-miR172h	3.5	EMB1047/FTSH12	Arabidopsis lyrata subsp. lyrata
can-miR172h	2.5	secreted glycoprotein 3	Ipomoea trifida
can-miR172h	3.5	UDP-glucosyltransferase, putative	Ricinus communis
can-miR172h	4	apetala 2-like protein	Nicotiana tabacum
can-miR172h	4	apetala 2-like protein	Nicotiana tabacum
can-miR172h	3	Negative regulator of the PHO system, putative	Ricinus communis
can-miR172h	4	phytochrome B	Solanum tuberosum
can-miR172i	3.5	phospholipase PLDa1	Solanum lycopersicum
can-miR172i	4	methylase family protein	Arabidopsis thaliana
can-miR172i	4	DNA-repair protein UVH3, putative	Ricinus communis
can-miR172i	1.5	apetala 2-like protein	Nicotiana tabacum
can-miR172i	1.5	apetala 2-like protein	Nicotiana tabacum
can-miR172i	1	transcription factor APETALA2	Vitis vinifera
can-miR172i	1	transcription factor APETALA2	Vitis vinifera
can-miR319a	3	phosphorylase family protein	Arabidopsis thaliana
can-miR319a	4	GAMyb-like1	Solanum lycopersicum
can-miR319b-c	2	phosphorylase family protein	Arabidopsis thaliana
can-miR319b-c	4	GAMyb-like1	Solanum lycopersicum
can-miR319b-c	2	anceolate	Capsicum annuum

can-miR319b-c	4	deetiolated 1-like protein	Solanum tuberosum
can-miR319e	3	GAMyb-like1	Solanum lycopersicum
can-miR319e	4	GAMYB-like2	Solanum lycopersicum
can-miR319e	4	lanceolate	Capsicum annuum
can-miR319e	2	phosphorylase family protein	Arabidopsis thaliana
can-miR319e	4	NHL repeat-containing protein	Arabidopsis thaliana
can-miR319e	4	lanceolate	Capsicum annuum
can-miR319e	3.5	putative transcription factor	Solanum tuberosum
can-miR319e	3.5	putative transcription factor	Solanum tuberosum
can-miR319f	1	phosphorylase family protein	Arabidopsis thaliana
can-miR319f	4	GAMyb-like1	Solanum lycopersicum
can-miR319f	3.5	lanceolate	Capsicum annuum
can-miR319f	4	cytochrome P450	Ipomoea nil
can-miR319f	4	amino acid transporter, putative	Ricinus communis
can-miR319f	3.5	transcription factor, putative	Ricinus communis
can-miR319f	3.5	lanceolate	Capsicum annuum
can-miR319f	3.5	putative transcription factor	Solanum tuberosum
can-miR319f	3.5	putative transcription factor	Solanum tuberosum
can-miR319g	3.5	erine-threonine protein kinase, plant-type, putative	Ricinus communis
can-miR319g	3.5	erine-threonine protein kinase, plant-type, putative	Ricinus communis
can-miR319g	1	phosphorylase family protein	Arabidopsis thaliana
can-miR319g	3	GAMyb-like1	Solanum lycopersicum
can-miR319g	4	GAMyb-like1	Solanum lycopersicum
can-miR319g	3	lanceolate	Capsicum annuum
can-miR319g	4	cytochrome P450	Ipomoea nil
can-miR319g	3.5	amino acid transporter, putative	Ricinus communis
can-miR319g	4	NHL repeat-containing protein	Arabidopsis thaliana
can-miR319g	3.5	transcription factor, putative	Ricinus communis
can-miR319g	3	lanceolate	Capsicum annuum
can-miR319g	4	TCP family transcription factor	Cyclamen persicum
can-miR319g	2.5	putative transcription factor	Solanum tuberosum
can-miR319g	2.5	putative transcription factor	Solanum tuberosum
can-miR390a-b	3.5	internal-motor kinesin	Nicotiana tabacum
can-miR390a-b	2.5	serine/threonine-protein kinase bri1, putative	Ricinus communis
can-miR390c	2	putative leucine rich repeat-type serine/threonine receptor-like kinase	Daucus carota

can-miR390c	4	eucine-rich repeat transmembrane protein kinase, putative	Arabidopsis thaliana
can-miR390c	4	erine-threonine protein kinase, plant-type, putative	Ricinus communis
can-miR390c	3	internal-motor kinesin	Nicotiana tabacum
can-miR390c	4	leucine-rich repeat transmembrane protein kinase, putative	Arabidopsis thaliana
can-miR390c	4	leucine rich repeat receptor kinase, putative	Ricinus communis
can-miR390c	4	protein phosphatase, putative	Ricinus communis
can-miR390c	4	serine-threonine protein kinase, plant-type, putative	Ricinus communis
can-miR393a-b	2.5	TIR1-like protein	Solanum lycopersicum
can-miR394a	1	F-box family protein	Citrus trifoliata
can-miR394b	1	F-box family protein	Citrus trifoliata
can-miR394b	4	ATP binding / microtubule motor	Arabidopsis thaliana
can-miR395a-j	4	DRB2 (DSRNA-BINDING PROTEIN 2)	Arabidopsis thaliana
can-miR395a-j	3.5	serine-threonine protein kinase, plant-type, putative	Ricinus communis
can-miR395a-j	4	carotenoid isomerase	Chrysanthemum x morifolium
can-miR395a-j	2	sulfate/bicarbonate/oxalate exchanger and transporter sat-1	Populus trichocarpa
can-miR395a-j	2.5	sulfate/bicarbonate/oxalate exchanger and transporter sat-1	Populus trichocarpa
can-miR395a-j	3	sulfate/bicarbonate/oxalate exchanger and transporter sat-1	Populus trichocarpa
can-miR395a-j	3.5	sulfate adenylyltransferase	Solanum tuberosum
can-miR396a	4	auxin response factor 8	Solanum lycopersicum
can-miR396a	3	AtGRF4 (GROWTH-REGULATING FACTOR 4)	Arabidopsis thaliana
can-miR396a	3.5	putative DNA cytosine 5-methyltransferase	Solanum lycopersicum
can-miR396a	3.5	putative DNA cytosine 5-methyltransferase	Solanum lycopersicum
can-miR396a	3.5	putative DNA cytosine 5-methyltransferase	Solanum lycopersicum
can-miR396a	4	zinc ion binding	Arabidopsis thaliana
can-miR396a	3	AtGRF3 (GROWTH-REGULATING FACTOR 3)	Arabidopsis thaliana
can-miR396a	4	ceramidase family protein	Arabidopsis thaliana
can-miR396a	3.5	UPA17	Capsicum annuum
can-miR396a	4	zinc finger protein, putative	Ricinus communis
can-miR396a	3	GRF domain class transcription factor	Malus x domestica
can-miR396a	4	heat shock protein binding protein, putative	Ricinus communis
can-miR396b	4	AtGRF4 (GROWTH-REGULATING FACTOR 4)	Arabidopsis thaliana
can-miR396b	2.5	putative DNA cytosine 5-methyltransferase	Solanum lycopersicum
can-miR396b	2.5	putative DNA cytosine 5-methyltransferase	Solanum lycopersicum
can-miR396b	2.5	putative DNA cytosine 5-methyltransferase	Solanum lycopersicum
can-miR396b	4	glutamate decarboxylase isoform2	Solanum lycopersicum

can-miR396b	4	putative WD-repeat protein	Arabidopsis thaliana
can-miR396b	4	AtGRF3 (GROWTH-REGULATING FACTOR 3); transcription activator	Arabidopsis thaliana
can-miR396b	4	polyadenylate-binding protein, putative	Ricinus communis
can-miR396b	3.5	AtALMT9 (aluminum-activated malate transporter 9); anion channel	Arabidopsis thaliana
can-miR396b	3.5	heat shock protein, putative	Ricinus communis
can-miR396b	4	UPA17	Capsicum annuum
can-miR396b	3.5	potyviral helper component protease-interacting protein 1	Solanum tuberosum subsp. andigenum
can-miR396b	4	GRF domain class transcription factor	Malus x domestica
can-miR396b	4	ring finger protein, putative	Ricinus communis
can-miR396b	4	disease resistance protein BS2	Capsicum chacoense
can-miR396c	2.5	ARGONAUTE 1	Nicotiana tabacum
can-miR396c	4	GRF domain class transcription factor	Malus x domestica
can-miR396c	4	AtGRF4 (GROWTH-REGULATING FACTOR 4)	Arabidopsis thaliana
can-miR396c	3.5	putative DNA cytosine 5-methyltransferase	Solanum lycopersicum
can-miR396c	3.5	putative DNA cytosine 5-methyltransferase	Solanum lycopersicum
can-miR396c	3.5	putative DNA cytosine 5-methyltransferase	Solanum lycopersicum
can-miR396c	3.5	FRS5 (FAR1-related sequence 5); zinc ion binding	Arabidopsis thaliana
can-miR396c	4	AtGRF3 (GROWTH-REGULATING FACTOR 3)	Arabidopsis thaliana
can-miR396c	3	UPA17	Capsicum annuum
can-miR396c	4	GRF domain class transcription factor	Malus x domestica
can-miR397a	4	3'-N-debenzoyl-2'-deoxytaxol N-benzoyltransferase	Ricinus communis
can-miR397a	1.5	diphenol oxidase	Nicotiana tabacum
can-miR397a	2.5	diphenol oxidase	Nicotiana tabacum
can-miR397a	2	diphenol oxidase	Nicotiana tabacum
can-miR397a	2.5	laccase	Solanum lycopersicum
can-miR397a	3	accase 110c	Populus trichocarpa
can-miR397a	2.5	laccase 1a	Populus trichocarpa
can-miR397a	2.5	laccase, putative	Ricinus communis
can-miR397a	2.5	laccase, putative	Ricinus communis
can-miR397a	3	LAC3 (laccase 3); laccase	Arabidopsis thaliana
can-miR397a	3	laccase 110b	Populus trichocarpa
can-miR397b	4	subtilase family protein	Arabidopsis thaliana
can-miR397b	2	diphenol oxidase	Nicotiana tabacum

can-miR397b	3	diphenol oxidase	Nicotiana tabacum
can-miR397b	2	diphenol oxidase	Nicotiana tabacum
can-miR397b	2	laccase	Solanum lycopersicum
can-miR397b	4	laccase 1a	Populus trichocarpa
can-miR397b	4	laccase, putative	Ricinus communis
can-miR397b	4	laccase, putative	Ricinus communis
can-miR397b	3.5	ankyrin repeat family protein	Arabidopsis thaliana
can-miR397b	4	RNA-binding region RNP-1 (RNA recognition motif)	Medicago truncatula
can-miR397b	1.5	laccase	Arabidopsis thaliana
can-miR397b	4	laccase 110b	Populus trichocarpa
can-miR398c	3.5	short-chain dehydrogenase/reductase	Cucumis melo subsp. melo
can-miR398c	3.5	myb family transcription factor	Arabidopsis thaliana
can-miR399a-e	4	two component sensor kinase	Agrobacterium tumefaciens str. C58
can-miR399a-e	4	Nucleotide pyrophosphatase/phosphodiesterase	Ricinus communis
can-miR399a-e	3.5	protein kinase family protein	Arabidopsis thaliana
can-miR399f-g	2.5	inorganic phosphate transporter	Solanum lycopersicum
can-miR399f-g	3	inorganic phosphate transporter	Solanum lycopersicum
can-miR408a	4	phosphoglucomutase, putative	Ricinus communis
can-miR408b	4	phosphoglucomutase, putative	Ricinus communis
can-miR408b	3	laccase 90c	Populus trichocarpa
can-miR408b	3	laccase 90a	Populus trichocarpa
can-miR408b	4	CYP72A56	Nicotiana tabacum
can-miR482b	3	nematode resistance-like protein	Solanum tuberosum
can-miR482c	3	resistance protein PSH-RGH6	Solanum tuberosum
can-miR482d	4	resistance protein PSH-RGH6	Solanum tuberosum
can-miR482d	4	resistance protein PSH-RGH6	Solanum tuberosum
can-miR482d	3.5	nematode resistance-like protein	Solanum tuberosum
can-miR482d	3	bacterial spot disease resistance protein 4	Solanum lycopersicum
can-miR482d	4	disease resistance protein Gpa2	Solanum tuberosum
can-miR482e	3.5	peptidase M3 family protein / thimet oligopeptidase family protein	Arabidopsis thaliana
can-miR482e	3.5	bacterial spot disease resistance protein 4	Solanum lycopersicum
can-miR482e	4	Carotenoid isomerase, chloroplast precursor, putative	Ricinus communis
can-miR482f	4	PR-protein	Capsicum annuum
can-miR482f	4	RRP6-like protein 2	Arabidopsis thaliana
can-miR482f	4	20S proteasome alpha 6 subunit	Nicotiana benthamiana

can-miR482f	2.5	blight resistance protein RGA3	Solanum bulbocastanum
can-miR482f	3	blight resistance protein RGA3	Solanum bulbocastanum
can-miR482f	4	nematode resistance-like protein	Solanum tuberosum
can-miR482f	4	nematode resistance-like protein	Solanum tuberosum
can-miR530	3.5	SlTCP3	Solanum lycopersicum
can-miR530	4	80 kD MCM3-associated protein, putative	Ricinus communis
can-miR530	1	TZP; DNA binding / nucleic acid binding / zinc ion binding	Arabidopsis thaliana
can-miR530	3.5	ukaryotic translation initiation factor 3 subunit, putative	Ricinus communis
can-miR530	4	PFK5 (PHOSPHOFRUCTOKINASE 5); 6-phosphofructokinase	Arabidopsis thaliana
can-miR530	3.5	Galactose oxidase precursor, putative	Ricinus communis
can-miR827	4	egulator of chromosome condensation (RCC1) family protein	Arabidopsis thaliana
can-miR827	4	chaperone protein dnaJ-related	Arabidopsis thaliana
can-miR827	4	S-locus lectin protein kinase family protein	Glycine max
can-miR1446a-b	3	GRAS family transcription factor	Populus trichocarpa
can-miR4376	4	casein kinase, putative	Ricinus communis
can-miR4414a	4	heat shock protein binding protein, putative	Ricinus communis
can-miR4414a	4	ABC transporter ATP-binding protein	Listeria monocytogenes Clip81459
can-miR4414a	3.5	TOC159 (TRANSLOCON AT THE OUTER ENVELOPE MEMBRANE OF CHLOROPLASTS 159); transmembrane receptor	Arabidopsis thaliana
can-miR4414b	2.5	regulator of chromosome condensation (RCC1) family protein	Arabidopsis thaliana
can-miR4414b	3.5	phosphoribulokinase/uridine kinase family protein	Arabidopsis thaliana
can-miR4414b	0	peroxin-3 family protein	Arabidopsis thaliana
can-miR4414b	4	putative auxin-regulated protein	Oryza sativa Japonica Group

S means total penalty score. This was carried out in collaboration with Jae Yun Lim.

Table 5. Predicted targets of novel miRNAs

Novel miRNA	S	Predicted targets	Reference species
can-miR-n001	4	3-hydroxy-3-methylglutaryl coenzyme A reductase	Withania somnifera
can-miR-n001	3.5	dehydroquinate dehydratase	Solanum tuberosum
can-miR-n002a-c	0	F-box family protein-1	Populus trichocarpa
can-miR-n002a-c	1	F-box family protein-2	Populus trichocarpa
can-miR-n002a-c	3	F-box family protein-3	Populus trichocarpa
can-miR-n002a-c	3	F-box family protein-4	Populus trichocarpa
can-miR-n002a-c	4	F-box family protein-5	Populus trichocarpa
can-miR-n002a-c	4	F-box family protein-6	Populus trichocarpa
can-miR-n003a	4	HGWP repeat containing protein-like	Oryza sativa Japonica Group
can-miR-n003b-d	3	S-adenosylmethionine-dependent methyltransferase	Ricinus communis
can-miR-n003b-d	3.5	pentatricopeptide repeat-containing protein-1	Ricinus communis
can-miR-n003b-d	3.5	pentatricopeptide repeat-containing protein-2	Ricinus communis
can-miR-n004	3.5	SPla/RYanodine receptor (SPRY) domain-containing protein	Arabidopsis thaliana
can-miR-n005	0	F-box family protein-7	Populus trichocarpa
can-miR-n005	4	F-box family protein-8	Populus trichocarpa
can-miR-n006	4	ATL2; protein binding / zinc ion binding	Arabidopsis thaliana
can-miR-n006	4	ATFH8 (formin 8)	Arabidopsis thaliana
can-miR-n006	4	WRKY27; transcription factor	Arabidopsis thaliana
can-miR-n006	4	cytochrome b5 isoform Cb5-D	Vernicia fordii
can-miR-n007	4	B-block binding subunit of TFIIIC	Arabidopsis thaliana
can-miR-n007	3	IBS1 (IMPAIRED IN BABA-INDUCED STERILITY 1)	Arabidopsis thaliana
can-miR-n007	4	PIP5K9	Arabidopsis thaliana
can-miR-n007	4	serine-threonine protein kinase, plant-type	Ricinus communis
can-miR-n007	4	EMB2730 (EMBRYO DEFECTIVE 2730); 3'-5'-exoribonuclease	Arabidopsis thaliana
can-miR-n009a,b-5p	3	pumilio	Ricinus communis
can-miR-n009a,b-3p	4	salt-inducible protein-like	Arabidopsis thaliana
can-miR-n009a,b-3p	3	leucine-rich repeat receptor-like protein kinase	Arabidopsis thaliana
can-miR-n009a,b-3p	4	protein serine/threonine kinase	Arabidopsis thaliana
can-miR-n009a,b-3p	4	Hcr9-OR2C	Solanum pimpinellifolium
can-miR-n009a,b-3p	3	verticillium wilt disease resistance protein Ve2	Solanum lycopersicum
can-miR-n009a,b-3p	2.5	peru 1	Solanum peruvianum
can-miR-n010	4	leucine-rich repeat transmembrane protein kinase, putative	Arabidopsis thaliana

can-miR-n013	2.5	putative disease resistance protein	Solanum tuberosum
can-miR-n013	4	NBS-LRR type disease resistance protein	Ipomoea batatas
can-miR-n013	3.5	nbs-lrr resistance protein	Populus trichocarpa
can-miR-n013	3.5	subtilisin-like protein	Arabidopsis thaliana
can-miR-n013	4	leucine-rich repeat-containing protein, putative	Ricinus communis
can-miR-n013	3	NRC1	Solanum lycopersicum
can-miR-n014	3.5	serine/threonine protein kinase	Ricinus communis
can-miR-n014	4	Putative late blight resistance protein	Solanum demissum
can-miR-n017	4	phytochrome A	Solanum tuberosum
can-miR-n019a-d	4	F-box family protein-9	Populus trichocarpa
can-miR-n024	2	Methyltransferase FkbM	Limnobacter sp. MED105
can-miR-n024	4	mitochondrial processing peptidase-like	Solanum tuberosum
can-miR-n026	4	MAPKKK19	Arabidopsis thaliana
can-miR-n026	4	serine/threonine-protein kinase bri1	Ricinus communis
can-miR-n026	4	esterase	Zea mays
can-miR-n026	2	GDSL lipase-like chlorogenate-dependent caffeoyltransferase precursor-1	Solanum lycopersicum
can-miR-n026	2	GDSL lipase-like chlorogenate-dependent caffeoyltransferase precursor-2	Solanum lycopersicum
can-miR-n026	4	GDSL lipase-like chlorogenate-dependent caffeoyltransferase precursor-3	Solanum lycopersicum
can-miR-n027	4	short-chain dehydrogenase/reductase (SDR) family protein	Arabidopsis thaliana
can-miR-n031	4	calmodulin binding protein-like protein	Arabidopsis thaliana
can-miR-n031	4	neutral leucine aminopeptidase protein	Solanum lycopersicum
can-miR-n032	4	phytoene desaturase	Nicotiana benthamiana
can-miR-n032	4	small multi-drug export protein	Zea mays
can-miR-n033	4	nbs-lrr resistance protein	Populus trichocarpa
can-miR-n033	3.5	Putative late blight resistance protein	Solanum demissum

S means total penalty score. This was carried out in collaboration with Jae Yun Lim.

Table 6. Probe sequences used for northern blot analysis

miRNA	probe sequence(5'→3')
can-miR157a-c	GTGCTCTCTATCTTCTGTCAA
can-miR159a-c	TAGAGCTCCCTTCAATCCAAA
can-miR160	TGGCATACAGGGAGCCAGGCA
can-miR162a-e	CCGGATGCAGAGGTTTATCGA
can-miR164a-b	TGCACGTGCCCTGCTTCTCCA
can-miR166a-h	GGGGAATGAAGCCTGGTCCGA
can-miR167a-c	TAGATCATGCTGGCAGCTTCA
can-miR168c	TTCCCGACCTGCACCAAGCGA
can-miR171a-e	GATATTGGCACGGCTCAATCA
can-miR319a	GGGAGCTCCCTTCAGTCCAA
can-miR390a-b	GGCGCTATCCCTCCTGAGCTT
can-miR394a	GGAGGTGGACAGAATGCCAA
can-miR395a-j	GAGTTCCCCCAAACACTTCAG
can-miR396b	AAGTTCAAGAAAGCTGTGGAA
can-miR398a-b	CAGGGGCGACCTGAGAACACA
can-miR403	CGAGTTTGTGCGTGAATCTAA
can-miR408a	AGCCAGGGAAGAGGCAGTGCA
can-miR477	CAGAAGCCTTTGAGGGAGAGT
can-miR530	CAGGTGCAGGTGCAAATGCA
can-miR-n001	AGGCCTACTATCAAAACAGAAA
can-miR-n002a-c	ACGATTCAGGTGTGTTTGCAA
can-miR-n003a	AATAGAATGGCTACAACTACA
can-miR-n004	AACAACTTGTGCTCGATCTTA
can-miR-n005	GCGATTCAGGTGCGTATCGAA
can-miR-n006	AAAGCCCAAAGGATGATTGTTG
can-miR-n007	GCAACTCACCTGAAGGTTCAA
can-miR-n008	TTCTCAATATTTTCTCTCATGCCG
can-miR-n009a,b-3p	TCCCTGAACGGAACAATACCA
can-miR-n010	CGTGTTGAACTTATCTCATAA
can-miR-n013	GGGAAAACAACTCTTGCTAAAA
can-miR-n014	GACAACAGTAACGAACTTCAG
can-miR-n015	AGTTTTTCGTTTCTCAGGCCTGAA
can-miR-n016a-b	GCTGTGGACGATGTATTGGAT
can-miR-n017	ACTCGACATAGCGTTGCTTAG
can-miR-n026	TGGGATGGAGTTCATTGTGACA
can-miR-n027	GTTGGCATCGACAACTTTCAT
can-miR-n030	TATGCTACTGACGGTCCTCGAA
can-miR-n032	AGTGGCTTATCTACAGGAACA
can-miR-n033	ACCTGAAGCCATGCTTTCTCTA

Table 7. Gene sepcific primers used for 5' RACE

Predicted conservd miRNA targets	Touchdown PCR reverse primer	Nested PCR reverse primer
squamosa promoter-binding protein-1	GGACCTGAGTAGTGATTGACATGAACAGCA	GAACAGCATGGGAACCTGATGGTTGA
squamosa promoter-binding protein-2	ATTGGTAGGTGCAACATGGGAACCTG	CCTGAAGGTTGAATGACGTGTGTCTCA
GAMYB-like1	CTGTTTCCAGCTGGAGGGGATTGAATCAGA	GGATTGAATCAGAATATCATCCGACTCTA
GAMYB-like2	CAAATTGCATGATTCAGTATGCTCAGTTGG	GTATGCTCAGTTGGAGGGGACTGAATCA
auxine response factor-1	TTCAAGGGTCTACGGGGTGCTGC	GCTGCTGCAGCCTGATCAAGTGG
no apical meristem	AAGCGAACTTGCAGCGAAAACATCC	TCCATTGGATTTGACACAACAGGCAA
ago1 and ago2	CTGACAACTCAGGAGGCTGAGATGAAGA	TGAAGAACCCACTTCCATAGATGTGTCAGT
auxine response factor-2	TCTTACTGCCAAGGACCGTCACCG	CGTCACCGAGGAGGAGGACATCATT
TCP transcription factor	CTTCCCCTTGAATTCTTGGTGCACTCGAAGT	CACTCGAAGTTACTCCTGAGAAACCATCA
sulfate transporter	ATAGCAGGCCTGGTGGTTCCGG	TCCGGCACATTTAGTACCCAATCGC
F-box family protein	AGCAAACCTTTGTTACCCAAGCAAGTCA	GCAAGTCATTCGAACGTGCTTTCCATA
caDRM1c	CCAACCCACACCAAATTCCACTTC	TCCACTTCCGACATTGATCAAGAACATAC
caDRM1b	CTTCATCAGGCTCCAGAGGAGCAACT	GCAACTTTGTTTCTCCCAACCCAGA
caDRM1a	CCAACCCACACCAAATTCCACTTC	TCCACTTCCGACATTGATCAAGAACATAC
Predicted novel miRNA targets	**Touchdown PCR reverse primer**	**Nested PCR reverse primer**
F-box family protein	ACTCGGCATGACGAGTGTGTATAAGAATG	TATAAGAATGGAGGAACCACACTGCTGC
F-box family protein	TTGATGGCGGCTGCCTATAATAAGGGACG	ATAAGGGACGAACCGGAACGCTGCAAAGA
F-box family protein	GGCGGTAATGGGTGACGATGAAATC	CGATGAAATCGGGCTTGTAAATGAG
F-box family protein	GTTTTGTAGTGGGCTTCAAGAATAAGGTTC	AATAAGGTTCAGATTTGCATTTGGGTTAGG
GDSL lipase-like chlorogenate-dependent caffeoyltransferase precursor	GGGAAACATATCATCAACTAGCCATCTTG	TCAACTAGCCATCTTGCTAACCAACTG

Table 8. 5'RACE results of novel miRNA targets

MicroRNA	S	Predicted targets	Target cleavage
can-miR-n002a-c	**0**	f-box family protein	**Yes**
can-miR-n005	**0**	f-box family protein	**Yes**
can-miR-n002a-c	**1**	f-box family protein	**Yes**
can-miR-n026	2	GDSL lipase-like chlorogenate-dependent caffeoyltransferase precursor	No
can-miR-n026	**2**	GDSL lipase-like chlorogenate-dependent caffeoyltransferase precursor	**Yes**
can-miR-n002a-c	**3**	f-box family protein	**Yes**
can-miR-n002a-c	3	f-box family protein	No
can-miR-n009a,b-5p	3	pumilio, putative	No
can-miR-n009a,b-3p	3	leucine-rich repeat receptor-like protein kinase	No
can-miR-n003b-c	3.5	pentatricopeptide repeat-containing protein, putative	No
can-miR-n003b-c	3.5	pentatricopeptide repeat-containing protein, putative	No
can-miR-n004	3.5	SPla/RYanodine receptor (SPRY) domain-containing protein	No
can-miR-n013	3.5	subtilisin-like protein	No
can-miR-n014	3.5	serine/threonine protein kinase, putative	No
can-miR-n002a-c	4	f-box family protein	No
can-miR-n005	4	f-box family protein	No
can-miR-n006	4	ATFH8 (formin 8); actin binding / actin filament binding / profilin binding	No
can-miR-n006	4	WRKY27; transcription factor	No
can-miR-n006	4	cytochrome b5 isoform Cb5-D	No
can-miR-n007	4	PIP5K9 (PHOSPHATIDYL INOSITOL MONOPHOSPHATE 5 KINASE)	No
can-miR-n007	4	serine-threonine protein kinase, plant-type, putative	No
can-miR-n007	4	EMB2730 (EMBRYO DEFECTIVE 2730)	No
can-miR-n010	4	leucine-rich repeat transmembrane protein kinase, putative	No
can-miR-n026	4	MAPKKK19	No
can-miR-n026	4	serine/threonine-protein kinase bri1, putative	No
can-miR-n026	4	GDSL lipase-like chlorogenate-dependent caffeoyltransferase precursor	No

FIGURES AND LEGENDS

Figure 1. Length distribution and 5' end analysis of small RNAs in pepper

A small RNA sequencing libraries were constructed from pepper tissue samples, which includes leaf, stem, root, flower, Fruit-1 (6 DAP), Fruit-2 (16 DAP), Fruit-MG (36 DAP), Fruit-B (38 DAP), Fruit-B5 (43 DAP), Fruit-B10 (48 DAP), DAP for days after pollination, MG for mature green. The normalizing factors were calculated using the DESeq library in the R statistical software package. This analysis was performed by Jae Yun Lim.

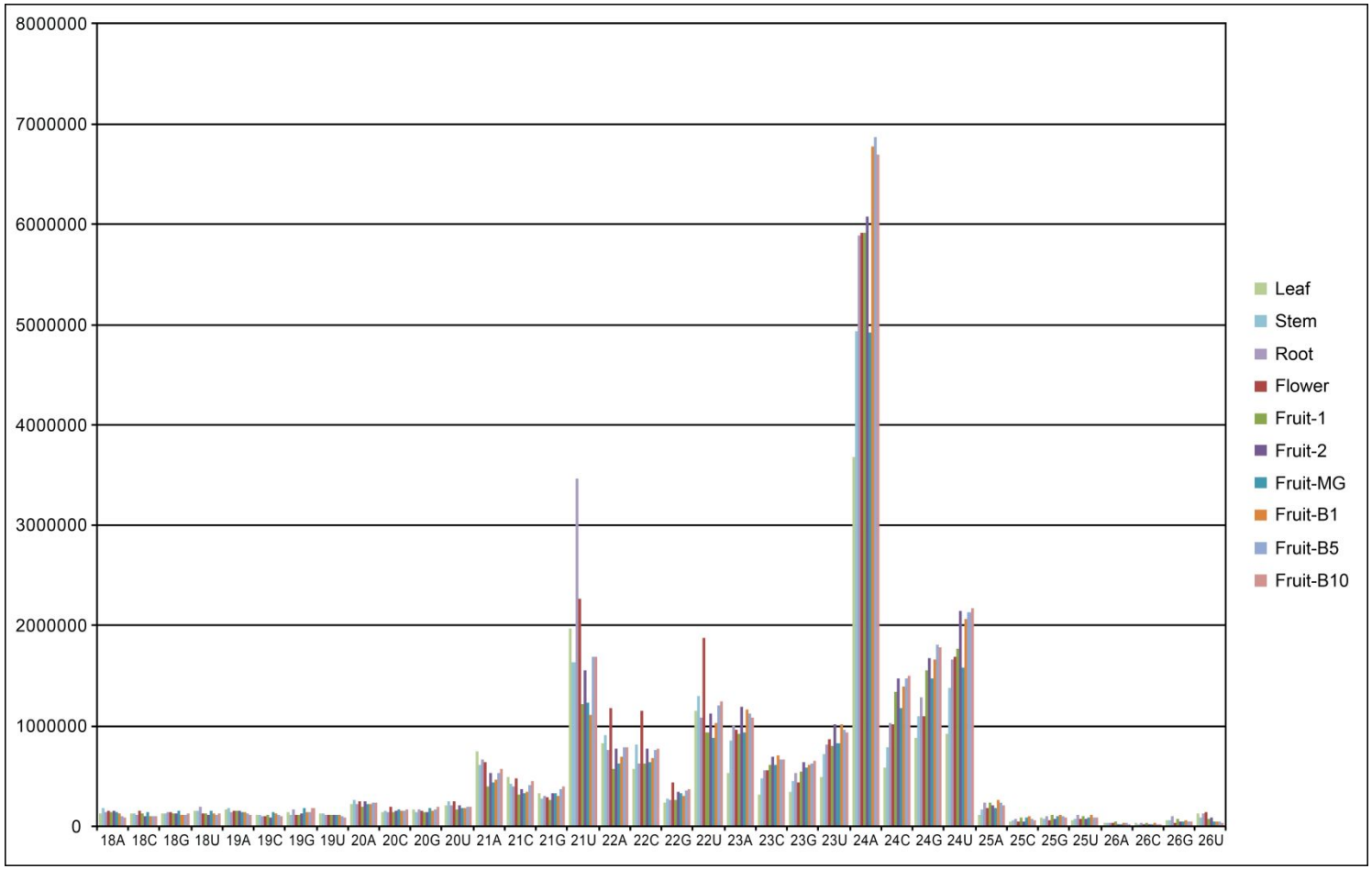

8000000
7000000
6000000
5000000
4000000
3000000
2000000
1000000
0
18A 18C 18G 18U 19A 19C 19G 19U 20A 20C 20G 20U 21A 21C 21G 21U 22A 22C 22G 22U 23A 23C 23G 23U 24A 24C 24G 24U 25A 25C 25G 25U 26A 26C 26G 26U
Leaf
Stem
Root
Flower
Fruit-1
Fruit-2
Fruit-MG
Fruit-B1
Fruit-B5
Fruit-B10

Figure 2. Representative hairpin precursors of conserved miRNA families

Among 29 conserved miRNA families, hairpin structures for 29 miRNAs were identified. Some of the examples were shown here. The mature miRNAs are colored in red. This figure was drawn by Prof. Chansoek Shin.

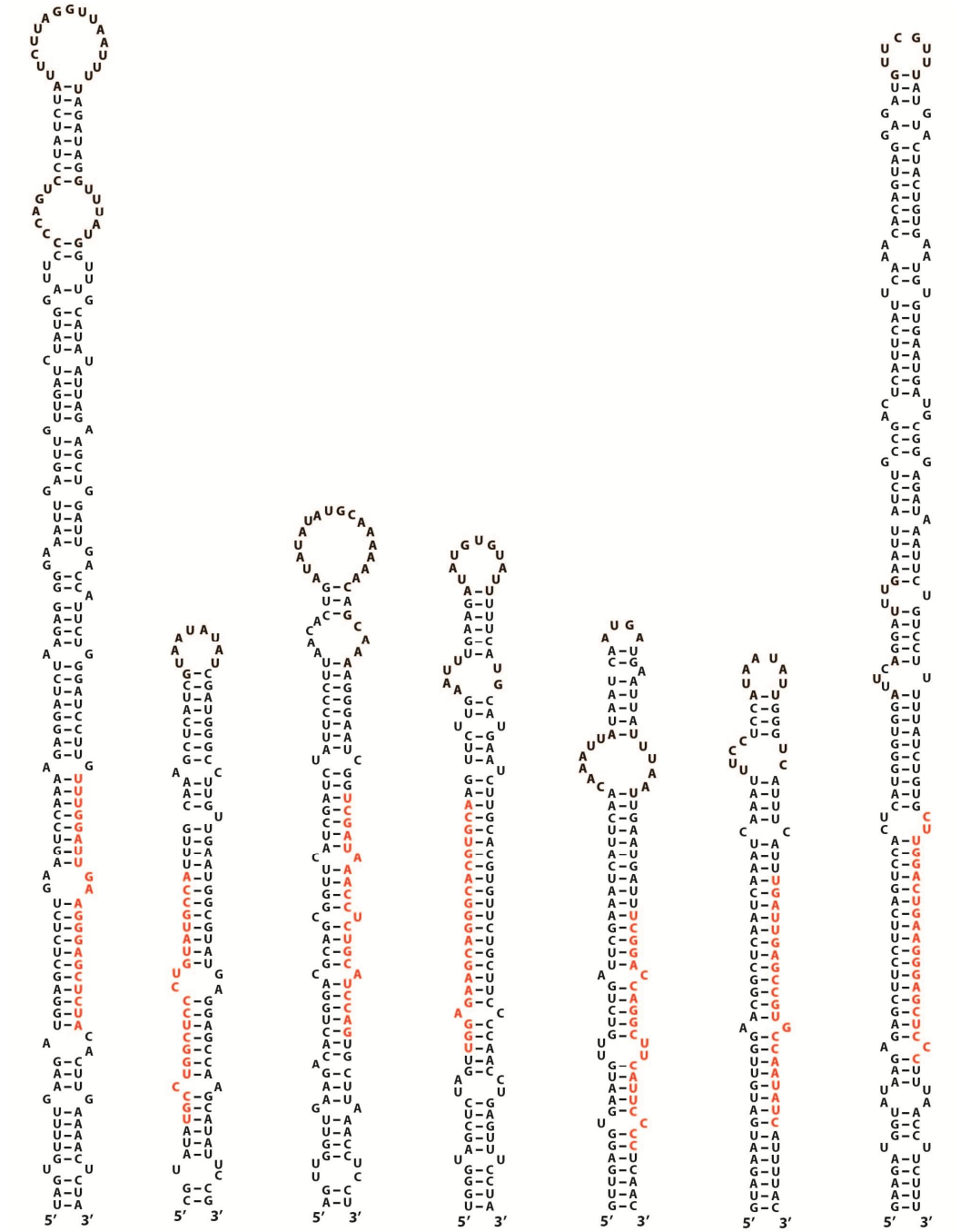

can-miR159a-c
can-miR160
can-miR162a-e
can-miR164a-b
can-miR166a-h
can-miR171a-e
can-miR319a

Figure 3. Representative hairpin precursors of novel miRNA families

Among 35 novel miRNA families, hairpin structures for 35 miRNAs were identified. Some of the examples were shown here. The mature miRNAs are colored in red. This figure was drawn by Soyoung Kim.

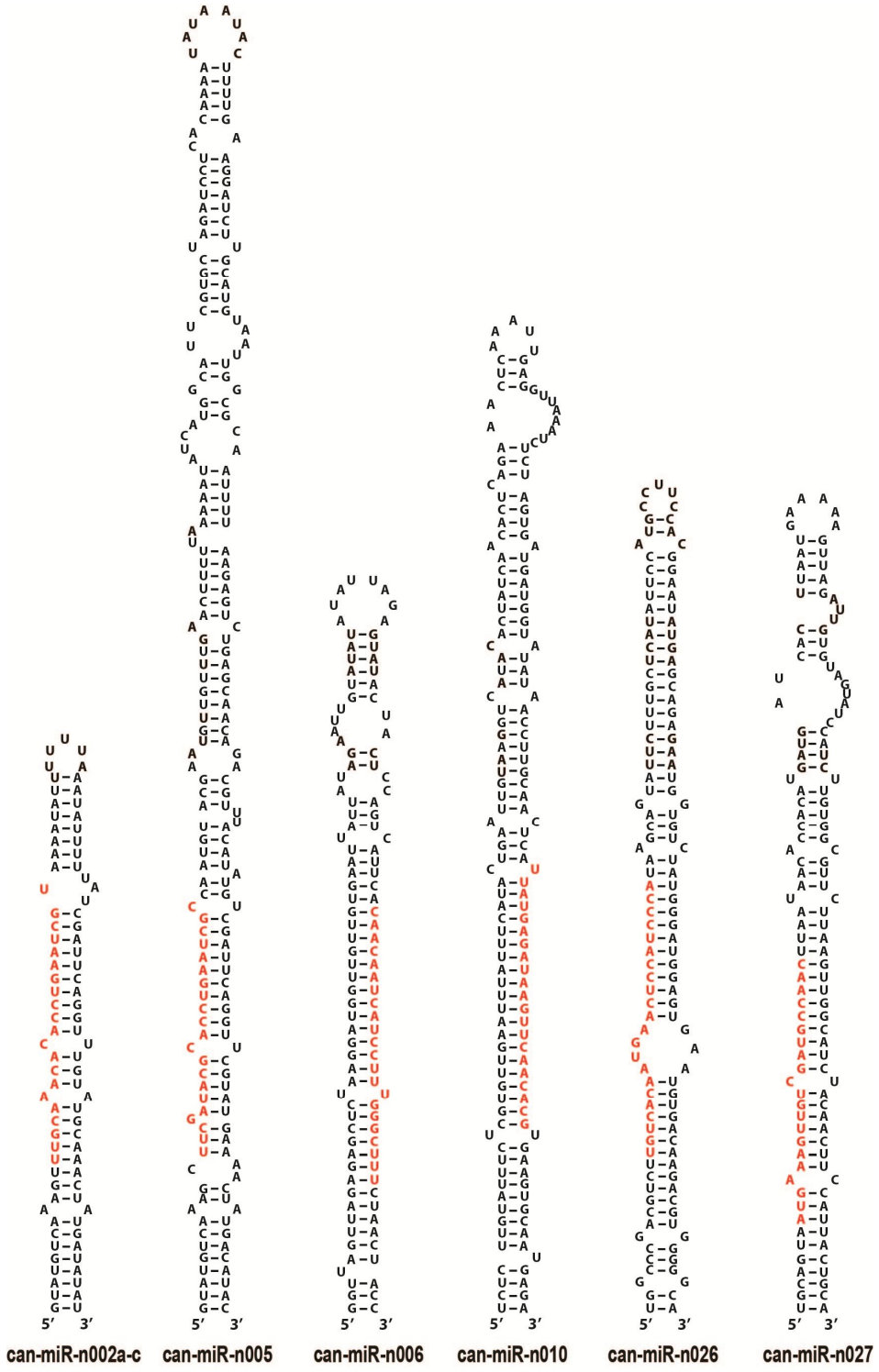

can-miR-n002a-c
can-miR-n005
can-miR-n006
can-miR-n010
can-miR-n026
can-miR-n027

Figure 4. Expression patterns of conserved miRNAs in pepper

Total RNAs were extracted from different tissues including leaf (L), stem (S), root (R), inflorescence (I), mixed stage of fruits (F) and seedling (Se). One of the blots probed for U6 snRNA as a loading control was representatively shown here.

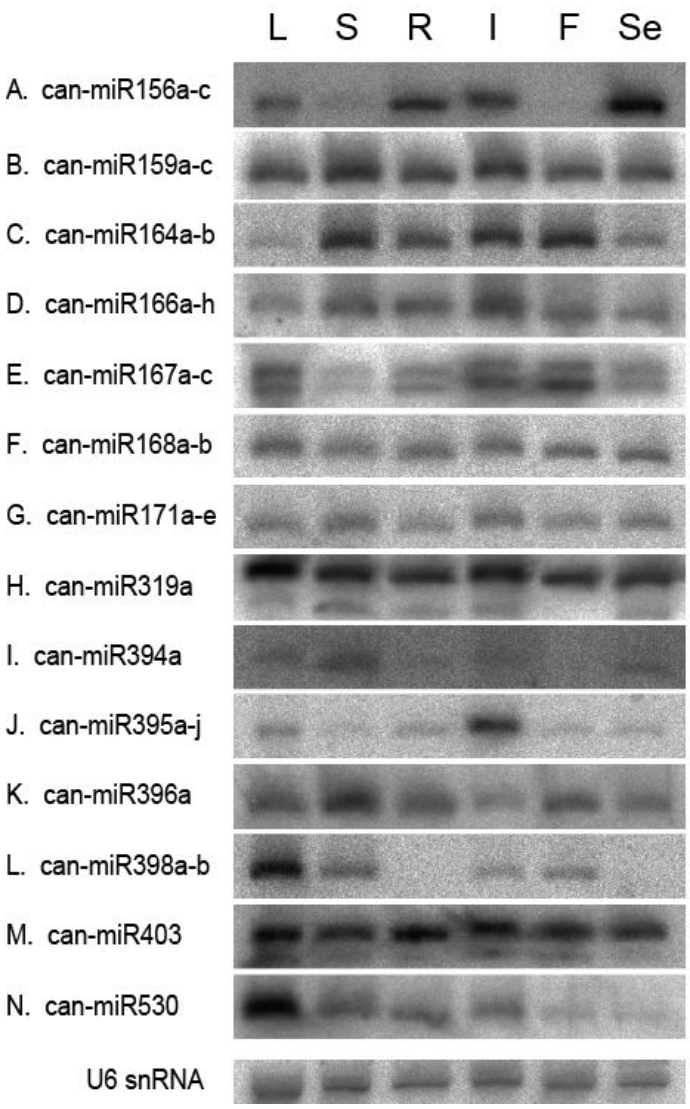

L S R I F Se
A. can-miR156a-c
B. can-miR159a-c
C. can-miR164a-b
D. can-miR166a-h
E. can-miR167a-c
F. can-miR168a-b
G. can-miR171a-e
H. can-miR319a
I. can-miR394a
J. can-miR395a-j
K. can-miR396a
L. can-miR398a-b
M. can-miR403
N. can-miR530
U6 snRNA

Figure 5. Expression patterns of novel miRNAs in pepper

Total RNAs were extracted from different tissues including leaf (L), stem (S), root (R), inflorescence (I), mixed stage of fruits (F) and seedling (Se). U6 snRNAs were shown as a loading control.

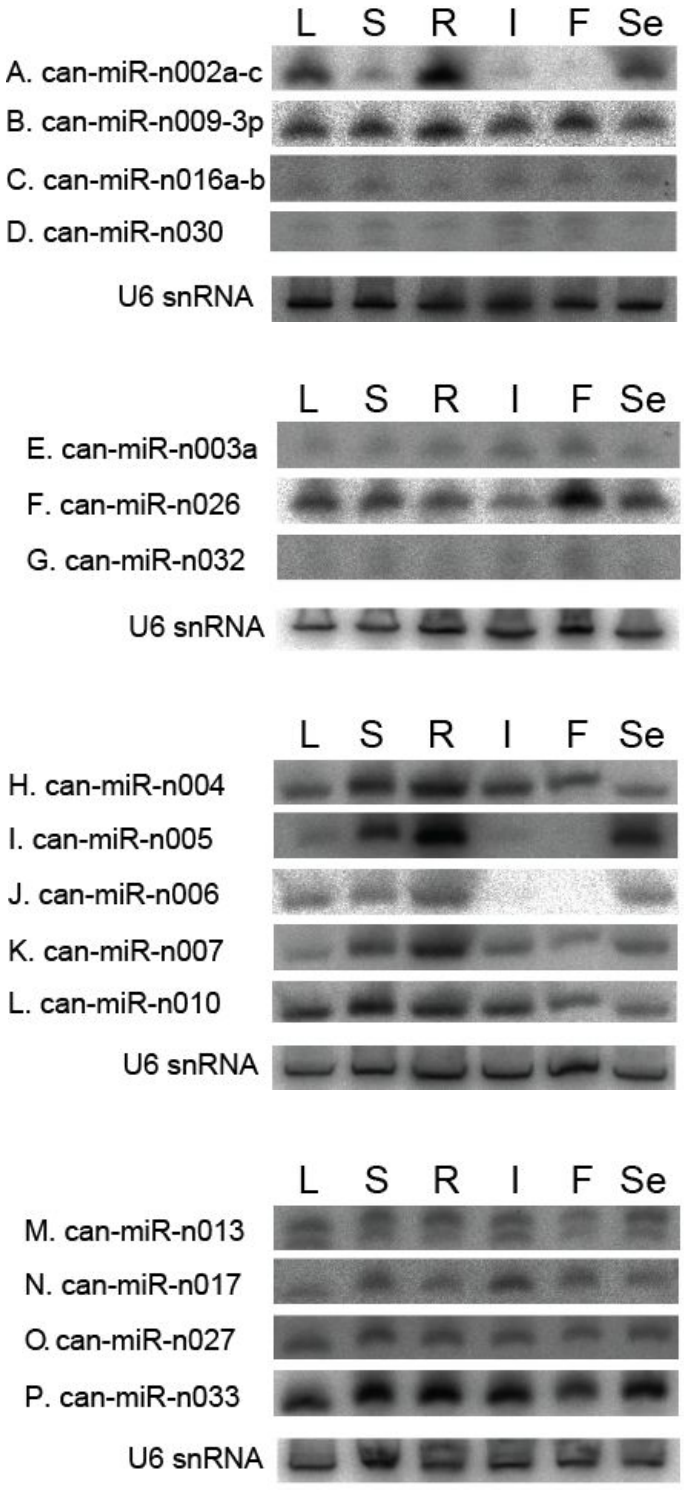

L S R I F Se
A. can-miR-n002a-c
B. can-miR-n009-3p
C. can-miR-n016a-b
D. can-miR-n030
U6 snRNA
L S R I F Se
E. can-miR-n003a
F. can-miR-n026
G. can-miR-n032
U6 snRNA
L S R I F Se
H. can-miR-n004
I. can-miR-n005
J. can-miR-n006
K. can-miR-n007
L. can-miR-n010
U6 snRNA
L S R I F Se
M. can-miR-n013
N. can-miR-n017
O. can-miR-n027
P. can-miR-n033
U6 snRNA

Figure 6. Validation of conserved miRNA-directed target cleavage in pepper

5' RACE-PCR products terminating at a given position indicated above the each miRNA-target duplex. Fractions refer to the number of independently cloned 5' RACE products whose 5′ end terminated at the indicated position (numerator) over the total number of sequenced clones (denominator). Note that different cleavage sites were mapped to the same target gene, depending on sequence specificity of can-miR156 (Figure 5B). This experiment was performed by June Hyun Park, Donghyun Kim, and Yourim Choi.

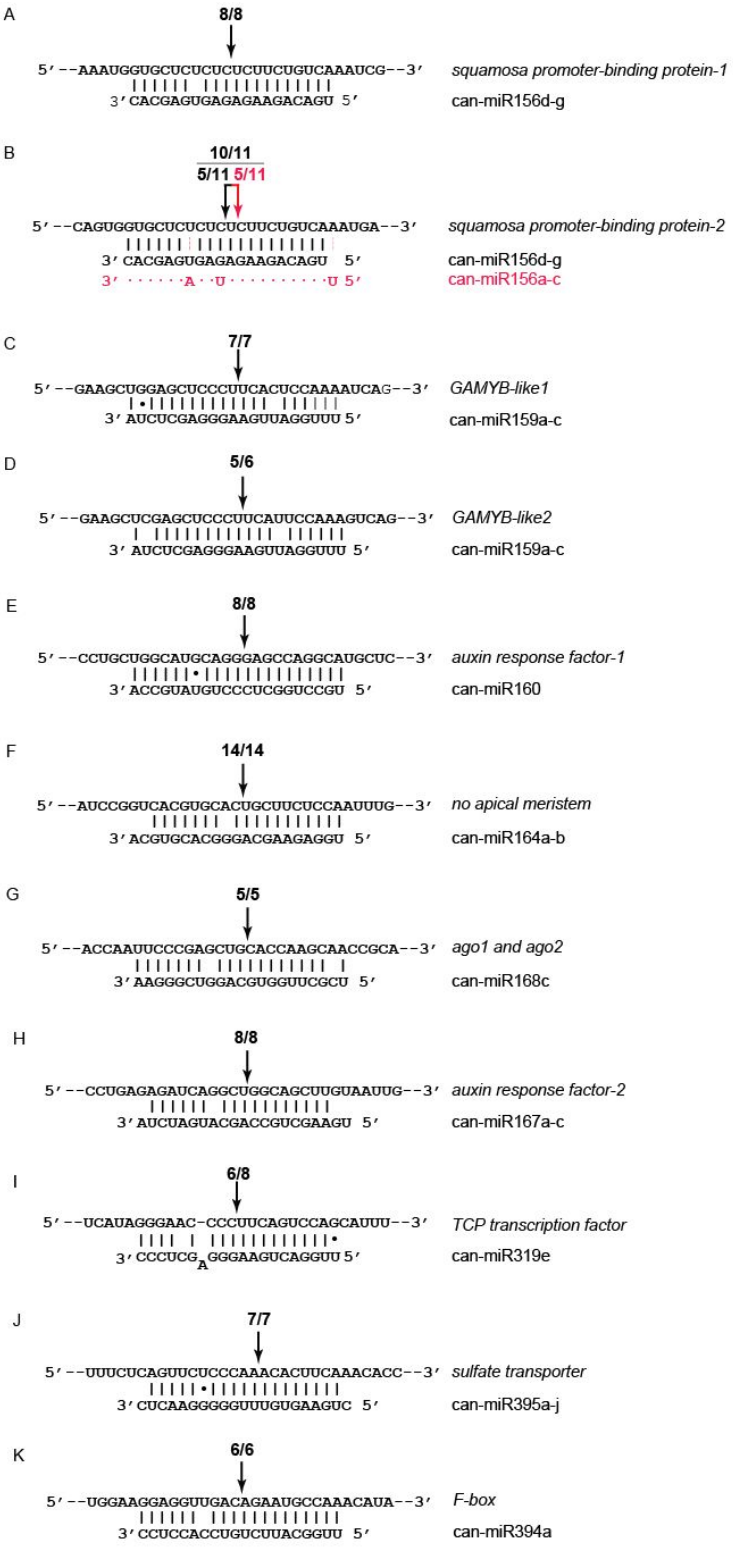
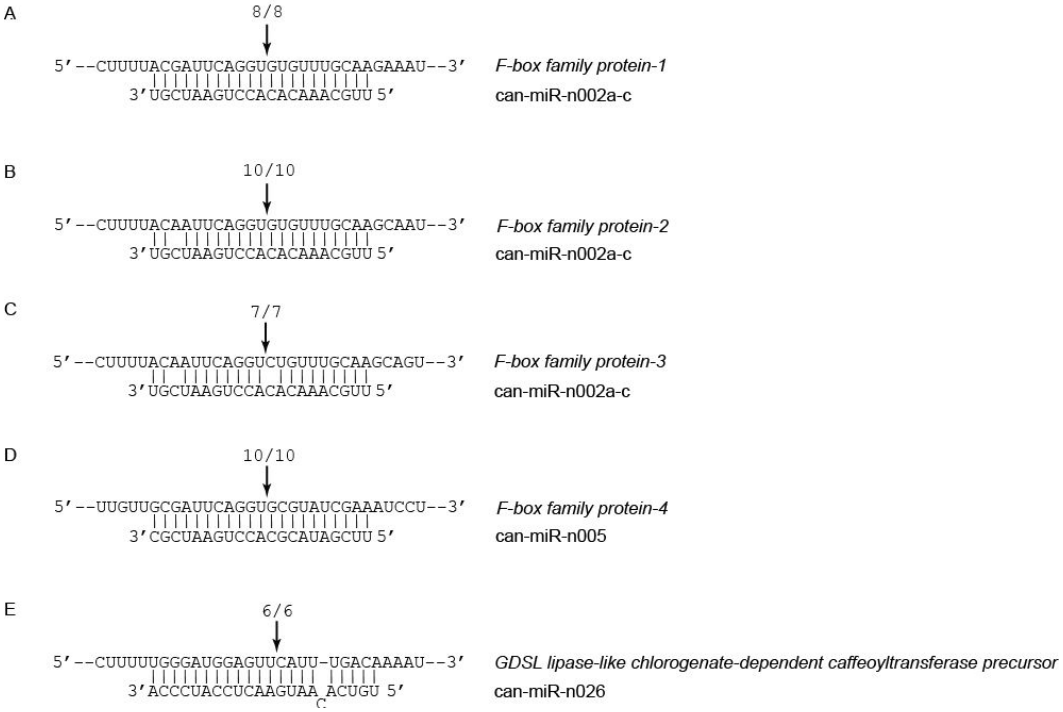

A
8/8
5′--AAAUGGUGCUCUCUCUCUUCUGUCAAAUCG--3′
3′CACGAGUGAGAGAAGACAGU 5′
squamosa promoter-binding protein-1
can-miR156d-g
B
10/11
5/11 5/11
5′--CAGUGGUGCUCUCUCUCUUCUGUCAAAUGA--3′
3′CACGAGUGAGAGAAGACAGU 5′
3′ ·····A··U·········U 5′
squamosa promoter-binding protein-2
can-miR156d-g
can-miR156a-c
C
7/7
5′--GAAGCUGGAGCUCCCUUCACUCCAAAAUCAG--3′
3′AUCUCGAGGGAAGUUAGGUUU 5′
GAMYB-like1
can-miR159a-c
D
5/6
5′--GAAGCUCGAGCUCCCUUCAUUCCAAAGUCAG--3′
3′AUCUCGAGGGAAGUUAGGUUU 5′
GAMYB-like2
can-miR159a-c
E
8/8
5′--CCUGCUGGCAUGCAGGGAGCCAGGCAUGCUC--3′
3′ACCGUAUGUCCCUCGGUCCGU 5′
auxin response factor-1
can-miR160
F
14/14
5′--AUCCGGUCACGUGCACUGCUUCUCCAAUUUG--3′
3′ACGUGCACGGGACGAAGAGGU 5′
no apical meristem
can-miR164a-b
G
5/5
5′--ACCAAUUCCCGAGCUGCACCAAGCAACCGCA--3′
3′AAGGGCUGGACGUGGUUCGCU 5′
ago1 and ago2
can-miR168c
H
8/8
5′--CCUGAGAGAUCAGGCUGGCAGCUUGUAAUUG--3′
3′AUCUAGUACGACCGUCGAAGU 5′
auxin response factor-2
can-miR167a-c
I
6/8
5′--UCAUAGGGAAC-CCCUUCAGUCCAGCAUUU--3′
3′CCCUCG A GGGAAGUCAGGUU 5′
TCP transcription factor
can-miR319e
J
7/7
5′--UUUCUCAGUUCUCCCAAACACUUCAAACACC--3′
3′CUCAAGGGGGUUUGUGAAGUC 5′
sulfate transporter
can-miR395a-j
K
6/6
5′--UGGAAGGAGGUUGACAGAAUGCCAAACAUA--3′
3′CCUCCACCUGUCUUACGGUU 5′
F-box
can-miR394a

Figure 7. Validation of novel miRNA-directed target cleavage in pepper

5' RACE-PCR products terminating at a given position indicated above the each miRNA-target duplex with the frequency of clones. Fractions refer to the number of independently cloned 5’ RACE products whose 5′ end terminated at the indicated position (numerator) over the total number of sequenced clones (denominator).


A
8/8
5′--CUUUUACGAUUCAGGUGUGUUUGCAAGAAAU--3′
3′UGCUAAGUCCACACAAACGUU 5′
F-box family protein-1
can-miR-n002a-c
B
10/10
5′--CUUUUACAAUUCAGGUGUGUUUGCAAGCAAU--3′
3′UGCUAAGUCCACACAAACGUU 5′
F-box family protein-2
can-miR-n002a-c
C
7/7
5′--CUUUUACAAUUCAGGUCUGUUUGCAAGCAGU--3′
3′UGCUAAGUCCACACAAACGUU 5′
F-box family protein-3
can-miR-n002a-c
D
10/10
5′--UUGUUGCGAUUCAGGUGCGUAUCGAAAUCCU--3′
3′CGCUAAGUCCACGCAUAGCUU 5′
F-box family protein-4
can-miR-n005
E
6/6
5′--CUUUUUGGGAUGGAGUUCAUU-UGACAAAAU--3′
3′ACCCUACCUCAAGUAA ACUGU 5′
C
GDSL lipase-like chlorogenate-dependent caffeoyltransferase precursor
can-miR-n026

REFERENCE

1. Voinnet O (2009) Origin, biogenesis, and activity of plant microRNAs. Cell 136: 669-687.
2. Zhang B, Pan X, Cobb GP, Anderson TA (2006) Plant microRNA: a small regulatory molecule with big impact. Dev Biol 289: 3-16.
3. Jones-Rhoades MW, Bartel DP, Bartel B (2006) MicroRNAs and their regulatory roles in plants. Annual Review of Plant Biology 57: 19-53.
4. Lanet E, Delannoy E, Sormani R, Floris M, Brodersen P, et al. (2009) Biochemical evidence for translational repression by Arabidopsis microRNAs. Plant Cell 21: 1762-1768.
5. Brodersen P, Sakvarelidze-Achard L, Bruun-Rasmussen M, Dunoyer P, Yamamoto YY, et al. (2008) Widespread translational inhibition by plant miRNAs and siRNAs. Science 320: 1185-1190.
6. Yang Z, Ebright YW, Yu B, Chen X (2006) HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. Nucleic Acids Res 34: 667-675.
7. Bollman KM, Aukerman MJ, Park MY, Hunter C, Berardini TZ, et al. (2003) HASTY, the Arabidopsis ortholog of exportin 5/MSN5, regulates phase change and morphogenesis. Development 130: 1493-1504.

8. Zhang JG, Zeng R, Chen JS, Liu X, Liao QS (2008) Identification of conserved microRNAs and their targets from Solanum lycopersicum Mill. Gene 423: 1-7.

9. Moxon S, Jing RC, Szittya G, Schwach F, Pilcher RLR, et al. (2008) Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening. Genome Research 18: 1602-1609.

10. Itaya A, Bundschuh R, Archual AJ, Joung JG, Fei ZJ, et al. (2008) Small RNAs in tomato fruit and leaf development. Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms 1779: 99-107.

11. Knapp S (2002) Tobacco to tomatoes: a phylogenetic perspective on fruit diversity in the Solanaceae. J Exp Bot 53: 2001-2022.

12. Oyama K, Hernandez-Verdugo S, Sanchez C, Gonzalez-Rodriguez A, Sanchez-Pena P, et al. (2006) Genetic structure of wild and domesticated populations of Capsicum annuum (Solanaceae) from northwestern Mexico analyzed by RAPDs. Genetic Resources and Crop Evolution 53: 553-562.

13. Kim HJ, Baek KH, Lee BW, Choi D, Hur CG (2011) In silico identification and characterization of microRNAs and their putative target genes in Solanaceae plants. Genome 54: 91-98.

14. Park M, Park J, Kim S, Kwon JK, Park HM, et al. (2012) Evolution of the

large genome in *Capsicum annuum* occurred through accumulation of single-type long terminal repeat retrotransposons and their derivatives. Plant J 69: 1018-1029.

15. Llave C, Xie Z, Kasschau KD, Carrington JC (2002) Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. Science 297: 2053-2056.
16. Blankenberg D, Gordon A, Von Kuster G, Coraor N, Taylor J, et al. (2010) Manipulation of FASTQ data with Galaxy. Bioinformatics 26: 1783-1785.
17. Emde AK, Grunert M, Weese D, Reinert K, Sperling SR (2010) MicroRazerS: rapid alignment of small RNA reads. Bioinformatics 26: 123-124.
18. Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, et al. (1994) Fast Folding and Comparison of Rna Secondary Structures. Monatshefte Fur Chemie 125: 167-188.
19. Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, et al. (2008) Criteria for annotation of plant MicroRNAs. Plant Cell 20: 3186-3190.
20. Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11: R106.
21. Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, et al. (2007) High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. PLoS One 2:

e219.

22. Fahlgren N, Carrington JC (2010) miRNA Target Prediction in Plants. Methods Mol Biol 592: 51-57.

23. Allen E, Xie Z, Gustafson AM, Carrington JC (2005) microRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell 121: 207-221.

24. Pall GS, Hamilton AJ (2008) Improved northern blot method for enhanced detection of small RNA. Nat Protoc 3: 1077-1084.

25. Sunkar R, Zhu JK (2004) Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis. Plant Cell 16: 2001-2019.

26. Czech B, Hannon GJ (2011) Small RNA sorting: matchmaking for Argonautes. Nat Rev Genet 12: 19-31.

27. Rajagopalan R, Vaucheret H, Trejo J, Bartel DP (2006) A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. Genes Dev 20: 3407-3425.

28. Song C, Fang J, Li X, Liu H, Thomas Chao C (2009) Identification and characterization of 27 conserved microRNAs in citrus. Planta 230: 671-685.

29. Pantaleo V, Szittya G, Moxon S, Miozzi L, Moulton V, et al. (2010) Identification of grapevine microRNAs and their targets using high-

throughput sequencing and degradome analysis. Plant Journal 62: 960-976.

30. Song CN, Fang JG, Li XY, Liu H, Chao CT (2009) Identification and characterization of 27 conserved microRNAs in citrus. Planta 230: 671-685.

31. Mallory AC, Bouche N (2008) MicroRNA-directed regulation: to cleave or not to cleave. Trends Plant Sci 13: 359-367.

32. Schwartz S, Oren R, Ast G (2011) Detection and removal of biases in the analysis of next-generation sequencing reads. PLoS One 6: e16685.

33. Chen C, Ridzon DA, Broomer AJ, Zhou Z, Lee DH, et al. (2005) Real-time quantification of microRNAs by stem-loop RT-PCR. Nucleic Acids Res 33: e179.

34. Allen E, Xie Z, Gustafson AM, Sung GH, Spatafora JW, et al. (2004) Evolution of microRNA genes by inverted duplication of target gene sequences in Arabidopsis thaliana. Nat Genet 36: 1282-1290.

35. Yao Y, Guo G, Ni Z, Sunkar R, Du J, et al. (2007) Cloning and characterization of microRNAs from wheat (Triticum aestivum L.). Genome Biol 8: R96.

36. Ma Z, Coruh C, Axtell MJ (2010) Arabidopsis lyrata small RNAs: transient MIRNA and small interfering RNA loci within the Arabidopsis genus.

Plant Cell 22: 1090-1103.

37. Axtell MJ (2008) Evolution of microRNAs and their targets: Are all microRNAs biologically relevant? Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms 1779: 725-734.

38. Rhoades MW, Reinhart BJ, Lim LP, Burge CB, Bartel B, et al. (2002) Prediction of plant microRNA targets. Cell 110: 513-520.

39. Park W, Li J, Song R, Messing J, Chen X (2002) CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana. Curr Biol 12: 1484-1495.

40. Yanhui C, Xiaoyuan Y, Kun H, Meihua L, Jigang L, et al. (2006) The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. Plant Mol Biol 60: 107-124.

41. Martin C, Paz-Ares J (1997) MYB transcription factors in plants. Trends Genet 13: 67-73.

42. Tiwari SB, Hagen G, Guilfoyle T (2003) The roles of auxin response factor domains in auxin-responsive transcription. Plant Cell 15: 533-543.

43. Shivaprasad PV, Chen HM, Patel K, Bond DM, Santos BA, et al. (2012) A microRNA superfamily regulates nucleotide binding site-leucine-rich repeats and other mRNAs. Plant Cell 24: 859-874.

44. Li F, Pignatta D, Bendix C, Brunkard JO, Cohn MM, et al. (2012) MicroRNA regulation of plant innate immune receptors. Proc Natl Acad Sci U S A 109: 1790-1795.

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

46. Jones JD, Dangl JL (2006) The plant immune system. Nature 444: 323-329.

47. Dodds PN, Rathjen JP (2010) Plant immunity: towards an integrated view of plant-pathogen interactions. Nat Rev Genet 11: 539-548.

48. Chepyshko H, Lai CP, Huang LM, Liu JH, Shaw JF (2012) Multifunctionality and diversity of GDSL esterase/lipase gene family in rice (Oryza sativa L. japonica) genome: new insights from bioinformatics analysis. BMC Genomics 13: 309.

49. Schulman BA, Carrano AC, Jeffrey PD, Bowen Z, Kinnucan ERE, et al. (2000) Insights into SCF ubiquitin ligases from the structure of the Skp1-Skp2 complex. Nature 408: 381-386.

50. Craig KL, Tyers M (1999) The F-box: a new motif for ubiquitin dependent proteolysis in cell cycle regulation and signal transduction. Prog Biophys Mol Biol 72: 299-328.

51. Chi XY, Yang QL, Chen XP, Wang JY, Pan LJ, et al. (2011) Identification and Characterization of microRNAs from Peanut (Arachis hypogaea L.) by High-Throughput Sequencing. PLoS One 6.

52. Jin W, Li N, Zhang B, Wu F, Li W, et al. (2008) Identification and verification of microRNA in wheat (Triticum aestivum). J Plant Res 121: 351-355.

53. Liang CW, Zhang XW, Zou J, Xu D, Su F, et al. (2010) Identification of miRNA from Porphyra yezoensis by High-Throughput Sequencing and Bioinformatics Analysis. PLoS One 5.

54. Zhao CZ, Xia H, Frazier TP, Yao YY, Bi YP, et al. (2010) Deep sequencing identifies novel and conserved microRNAs in peanuts (Arachis hypogaea L.). BMC Plant Biol 10: 3.

55. Carrington JC, Ambros V (2003) Role of microRNAs in plant and animal development. Science 301: 336-338.

56. Chen J, Li WX, Xie DX, Peng JR, Ding SW (2004) Viral virulence protein suppresses RNA silencing-mediated defense but upregulates the role of MicroRNA in host gene expression. Plant Cell 16: 1302-1313.

57. Kasschau KD, Xie Z, Allen E, Llave C, Chapman EJ, et al. (2003) P1/HC-Pro, a viral suppressor of RNA silencing, interferes with Arabidopsis development and miRNA unction. Dev Cell 4: 205-217.

58. Vazquez F, Vaucheret H, Rajagopalan R, Lepers C, Gasciolli V, et al. (2004) Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. Mol Cell 16: 69-79.

59. Axtell MJ, Bartel DP (2005) Antiquity of microRNAs and their targets in land plants. Plant Cell 17: 1658-1673.

60. Fahlgren N, Montgomery TA, Howell MD, Allen E, Dvorak SK, et al. (2006) Regulation of AUXIN RESPONSE FACTOR3 by TAS3 ta-siRNA affects developmental timing and patterning in Arabidopsis. Current Biology 16: 939-944.

61. Xie KB, Wu CQ, Xiong LZ (2006) Genomic organization, differential expression, and interaction of SQUAMOSA promoter-binding-like transcription factors and microRNA156 in rice. Plant Physiology 142: 280-293.

62. Wang JW, Czech B, Weigel D (2009) miR156-Regulated SPL Transcription Factors Define an Endogenous Flowering Pathway in Arabidopsis thaliana. Cell 138: 738-749.

63. Wu G, Poethig RS (2006) Temporal regulation of shoot development in Arabidopsis thaliana by miR156 and its target SPL3. Development 133: 3539-3547.

64. Xie K, Shen J, Hou X, Yao J, Li X, et al. (2012) Gradual Increase of miR156

Regulates Temporal Expression Changes of Numerous Genes during Leaf Development in Rice. Plant Physiol 158: 1382-1394.

65. Huijser P, Schmid M (2011) The control of developmental phase transitions in plants. Development 138: 4117-4129.

66. Schwab R, Palatnik JF, Riester M, Schommer C, Schmid M, et al. (2005) Specific effects of MicroRNAs on the plant transcriptome. Developmental Cell 8: 517-527.

67. Chuck G, O'Connor D (2010) Small RNAs going the distance during plant development. Current Opinion in Plant Biology 13: 40-45.

68. Kim JH, Woo HR, Kim J, Lim PO, Lee IC, et al. (2009) Trifurcate feed-forward regulation of age-dependent cell death involving miR164 in Arabidopsis. Science 323: 1053-1057.

69. Zhang BH, Pan XP, Wang QL, Cobb GP, Anderson TA (2005) Identification and characterization of new plant microRNAs using EST analysis. Cell Research 15: 336-360.

70. Kim J, Jung JH, Reyes JL, Kim YS, Kim SY, et al. (2005) microRNA-directed cleavage of ATHB15 mRNA regulates vascular development in Arabidopsis inflorescence stems. Plant Journal 42: 84-94.

71. Williams L, Grigg SP, Xie MT, Christensen S, Fletcher JC (2005) Regulation of Arabidopsis shoot apical meristem and lateral organ formation by

microRNA miR166g and its AtHD-ZIP target genes. Development 132: 3657-3668.

72. Palatnik JF, Allen E, Wu XL, Schommer C, Schwab R, et al. (2003) Control of leaf morphogenesis by microRNAs. Nature 425: 257-263.

73. Rhoades MW, Reinhart BJ, Lim LP, Burge CB, Bartel B, et al. (2002) Prediction of plant microRNA targets. Cell 110: 513-520.

74. Palatnik JF, Wollmann H, Schommer C, Schwab R, Boisbouvier J, et al. (2007) Sequence and expression differences underlie functional specialization of Arabidopsis microRNAs miR159 and miR319. Dev Cell 13: 115-125.

75. Jones-Rhoades MW, Bartel DP (2004) Computational identification of plant MicroRNAs and their targets, including a stress-induced miRNA. Molecular Cell 14: 787-799.

76. Kawashima CG, Yoshimoto N, Maruyama-Nakashita A, Tsuchiya YN, Saito K, et al. (2009) Sulphur starvation induces the expression of microRNA-395 and one of its target genes but in different cell types. Plant J 57: 313-321.

77. Liang G, Yang F, Yu D (2010) MicroRNA395 mediates regulation of sulfate accumulation and allocation in Arabidopsis thaliana. Plant J 62: 1046-1057.

78. Jagadeeswaran G, Zheng Y, Li YF, Shukla LI, Matts J, et al. (2009) Cloning and characterization of small RNAs from Medicago truncatula reveals four novel legume-specific microRNA families. New Phytol 184: 85-98.

79. Karginov FV, Cheloufi S, Chong MM, Stark A, Smith AD, et al. (2010) Diverse endonucleolytic cleavage sites in the mammalian transcriptome depend upon microRNAs, Drosha, and additional nucleases. Mol Cell 38: 781-788.

80. Cuperus JT, Fahlgren N, Carrington JC (2011) Evolution and functional diversification of MIRNA genes. Plant Cell 23: 431-442.

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

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

83. Mi SJ, Cai T, Hu YG, Chen Y, Hodges E, et al. (2008) Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5 ' terminal nucleotide. Cell 133: 116-127.

84. Montgomery TA, Howell MD, Cuperus JT, Li D, Hansen JE, et al. (2008) Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. Cell 133: 128-141.

85. Pantaleo V, Szittya G, Moxon S, Miozzi L, Moulton V, et al. (2010) Identification of grapevine microRNAs and their targets using high-throughput sequencing and degradome analysis. Plant J 62: 960-976.

ABSTRACT IN KOREAN

고추 microRNA와 표적유전자의 대량 발굴

이번 연구에서는 고추의 서로 다른 10개의 small RNA library를 이용하여 high-throughput sequencing을 진행하였고, 그 결과 29개의 family에 해당하는 128개의 conserved microRNA (miRNA)와 35개의 family에 해당하는 50개의 novel miRNA 를 대량으로 발굴하였다. 이들에 대한 northern blot 분석실험을 진행하여, 그들의 in vivo 발현을 다시 증명하고 조직 특이적인 또는 발달단계 특이적인 발현양상을 분석하였다. 또한, 생물정보학을 기반으로 331개의 conserved miRNA 표적유전자와 57개의 novel miRNA의 표적유전자들을 예측하였고, 5' RNA liagse mediated-rapid amplification of cDNA ends (5' RLM-RACE) 실험으로 이들 중 많은 표적유전자를 증명하였다. 이번 연구는 고추 miRNA들을 대규모로 발굴한 최초의 연구이며, 고추를 비롯한 가지과 식물 전반에 걸쳐 miRNA의 기능을 연구하고 이해하는데 중요한 기반이 될 것이다.

주요어: MicroRNA, *C. annuum*, High-throughput sequencing, Northern blot anlaysis, 5′-RACE, miRNA-directed target cleavage

학번: 2011-21329

ACKNOWLEDGEMENT

본 석사학위논문은 석사과정동안 연구한 내용을 *PLoS ONE* 저널에 제출한 논문에서 본인이 한 부분을 중심으로 추린 것이지만, 공동으로 작업된 부분이 많기 때문에 이에 대해 솔직히 밝히고 도움에 대한 감사의 말씀을 전하고자 합니다.

먼저 최도일 교수님 실험실에서 고추 샘플을 제공해 주셨습니다. 생물정보학 작업 부분에서 임재윤 선생님이 대량의 자료를 정리하는데 많은 도움을 주셨습니다. 실험적인 부분에서 conserved miRNA의 표적유전자에 대한 5′RLM-RACE 실험은 우리 실험실의 준현이 형, 동현이, 유림이가 전담해주었습니다. 또한, 저널에 제출된 논문의 글쓰기는 준현이 형과 제가 함께 작업했지만 대부분 준현이 형의 노력으로 이루어졌고, 특히 이 글의 discussion writing을 전담해주셨습니다. 이렇게 논문이 나올 수 있게끔 도와주신 분들께 진심으로 감사드립니다.

논문을 완성하기 위해서는 양질의 실험자료와 논리정연한 글쓰기 솜씨만이 전부는 아닙니다. 전반적인 연구진행 상황을 파악하고 방향을 제시하는 일 또는 공동연구자와 의사소통하고 의견을 조율하는 작업 등이 잘 이루어져야 합니다. 이러한 부분들을 저 혼자 했다면 굉장히 힘들었을 텐데 준현이 형께서 2년 내내 관심을 가지고 같이 이끌어 주셨습니다. 이렇게 이끌어주셔서 정말 깊이 감사드립니다. 신찬석 교수님께서도 논문이 나오기까지 같이 관리해주시고 조언해주신 점에 대해 감사의 말씀 드리겠습니다. 마지막으로 석사과정동안 실험과 연구에 전념할 수 있도록 전적으로 지원해준 우리가족에게 진심으로 감사합니다. 특히, 제 논문을 염원하며 응원해주신 아버지께 감사드리며 더 노력하겠다는 다짐을 전합니다.