*Journal of Heredity* 2012:103(2):268–277
doi:10.1093/jhered/esr127
Advance Access publication January 27, 2012



# Computational Identification of MicroRNAs in Strawberry Expressed Sequence Tags and Validation of Their Precise Sequences by miR-RACE

Qing-Hua Dong*, Jian Han*, Hua-Ping Yu, Chen Wang, Mi-Zhen Zhao, Hong Liu, An-Jing Ge, and Jing-Gui Fang

From the Beijing Key Laboratory of New Technology in Agricultural Application, Plant Science and Technology College, Beijing University of Agriculture, Beijing, P. R. China (Dong and Ge); College of Horticulture, Nanjing Agricultural University, Nanjing 210095, P. R. China (Han, Yu, Wang, Liu, and Fang); and Institute of Horticulture, Jiangsu Academy of Agricultural Sciences, Nanjing, P. R. China (Zhao).

*These authors contributed equally to this work.

Address correspondence to Jing-Gui Fang at the address above, or e-mail: fanggg@njau.edu.cn.

## Abstract

MicroRNAs (miRNAs) are small, endogenously expressed, nonprotein-coding RNAs that regulate gene expression at the post-transcriptional level in both animals and plants through repressing translation or inducing mRNA degradation. A comprehensive strategy to identify new miRNA homologs by mining the repository of available strawberry expressed sequence tags (ESTs) was developed. By adopting a range of filtering criteria, we identified 11 potential miRNAs belonging to 5 miRNA families from 47 890 *Fragaria vesca* EST sequences. Using 2 specific 5′ and 3′ miRNA RACE PCR reactions and a sequence-directed cloning method, we accurately determined both end sequences of 5 candidate miRNAs. Meanwhile, qRT-PCR was used to detect the expression of these 5 miRNAs in different strawberry organs and tissues at several growing stages. These newly identified *F. vesca* miRNAs (fve-miRNAs) and their expression information can improve our understanding of possible roles of fve-miRNAs in regulating the growth and development of *F. vesca*.

**Key words:** Fragaria vesca, *microRNA, miR-RACE, qRT-PCR*

MicroRNAs (miRNAs) are a newly discovered class of nonprotein-coding small RNAs with 21–24 nucleotides in length. miRNAs regulate the expression of functional genes during plant development and in other physiological processes (Bartel 2004). The maturation of miRNAs in plants involves several steps that require key enzymes such as Dicer-like 1 enzyme (DCL1) (Bartel 2004) and HASTY (Papp et al. 2003; Park et al. 2005). Mature miRNAs are incorporated into RNA-induced silencing complex that is induced by miRNAs to cause the cleavage of target genes' mRNA (Bartel 2004). Increasing evidence demonstrates that plant miRNAs negatively regulate their target genes, which in turn function in a range of developmental processes, including meristem cell identity (Laufs et al. 2004; Mallory, Dugas, et al. 2004; Mallory, Reinhart, et al. 2004; Guo et al. 2005), leaf organ morphogenesis and polarity (Emery et al. 2003; Juarez et al. 2004), floral differentiation and development (Aukerman and Sakai 2003; Chen 2004), auxin signaling, boundary formation/organ separation, and stress responses (Kasschau et al. 2003; Sunkar and Zhu 2004; Sunkar et al. 2006), as well as fruit growth and development (Moxon et al. 2008; Carra et al. 2009).

Cloning and characterization of miRNA is still a popular research area since the number of identified miRNAs is still far from saturation, and their roles in organisms have elicited great interest among scientists. Several reports about miRNAs in fruit crops, such as citrus, apple, and grape, have been published and more studies are ongoing to date (Gleave et al. 2008; Lu et al. 2008; Song et al. 2009; Pantaleo et al. 2010; Varkonyi-Gasic et al. 2010; Yu et al. 2011). Even though there are some reports on identification of miRNAs from *Fragaria vesca* (fve-miRNAs) (Li et al. 2009), the total number of these identified fve-miRNAs and their families are still low. This reflects a disparity between the economic

importance of this plant species and available information on its molecular and genetic studies, including small RNA-mediated gene regulation in *F. vesca*. Sequence conservation of miRNAs has laid a sound foundation for predicting and studying miRNAs in nonmodel plants whose whole genome sequences are unknown. Many known miRNAs are conserved in plant species, for example, a majority of 42 miRNA families comprising 117 miRNAs in *Arabidopsis* have been found in other species such as rice, maize, and sorghum (Griffiths-Jones 2004). This makes it possible to perform computational searches for the homologs or orthologs of miRNAs based on the highly conserved sequences of mature miRNAs as well as their hairpin structures in precursors (Wang, Reyes, et al. 2004). It is reported that the computational or bioinformatics-based approach is useful in the prediction of conserved miRNAs that usually cannot be detected easily by direct cloning due to their low expression level and/or spatiotemporal expression. As of now, miRNAs have been found in tomato, soybean, maize, *Brassica napus*, cotton, apple, grape, citrus, *Euphorbiaceous*, and other plants by bioinformatics prediction (Zhang, Pan, and Anderson 2006; Qiu et al. 2007; Xie et al. 2007; Zhang et al. 2007; Gleave et al. 2008; Yin et al. 2008; Zhang, Pan, and Stellwag 2008; Zhang, Zeng, et al. 2008; Carra et al. 2009; Song et al. 2009; Song, Jia, et al. 2010; Zeng et al. 2010).

Despite the apparent importance of bioinformatics in novel miRNA prediction, the prediction algorithms cannot always guarantee accurate prediction of the location of a mature miRNA in its precursor with nucleotide-level precision. The determination of the precise sequence of the potential miRNA, especially the termini nucleotides, is therefore necessary since precise sequences of miRNAs are essential for some important downstream researches, such as miRNA target prediction, miRNA evolution, miRNA regulatory roles, and biogenesis mechanisms.

In previous studies, a combination of computational prediction and experimental verification were employed to identify miRNAs, in which RNA blotting and/or RT-PCR were the 2 main experimental techniques used. However, these 2 techniques can only confirm the existence and size but not the full precise sequence of a miRNA predicted computationally, especially its termini nucleotides. The employment of new strategies that can validate the precise sequences of miRNAs predicted computationally is of great significance. As reported, miR-RACE is a powerful method that has been utilized successfully in the validation of citrus and apple miRNA precise sequences (Song, Fang, et al. 2010; Yu et al. 2011). This strategy, which combines miRNA-enriched library preparation, 5′ RACE and 3′ RACE reactions, and sequence-directed cloning, can amplify and determine most sequences including those of nonabundant miRNAs that are typically difficult to clone directly.

Despite the importance of strawberry as one of the most widely consumed fruits worldwide and extensively studies on plant miRNAs in the past few years, not much work has been done on experimental or computational identification of miRNAs in *F. vesca* (fve-miRNAs) (Li et al. 2009). The available reports on fve-miRNAs do not reveal the exact sequence of the potential fve-miRNAs. In this study, we not only predicted more fve-miRNAs through bioinformatics approaches but also validated the precise sequences of fve-miRNAs belonging to 5 families using miR-RACE. The ubiquitous expression of these 5 fve-miRNAs was also detected over a range of strawberry stages and tissues by qRT-PCR. This is the first report on application of miR-RACE in the validation of the precise sequences of fve-miRNA predicted bioinformatically, and the results can facilitate the ease of undertaking further studies on fve-miRNA evolution, regulatory roles, and biogenesis mechanisms.

## Materials and Methods

### Plant Material

Roots, young and old leaves, small sepals (collected from small fruits), big sepals (collected from big fruits), flower buds, flowers, and small fruits (1.5 cm in diameter), medium sized fruits (2 cm in diameter), and big fruits (3 cm in diameter) were collected from strawberry ("Sweet Charlie," *Fragaria* × *ananassa*) grown at the fruit science experimental station of Beijing University of Agriculture, China, in 2009. After collection, all the samples were immediately frozen in liquid Nitrogen and stored at −80 °C until use.

### Plant miRNAs Released Publically and *F. vesca* ESTs

In this study, we used mRNA and GenBank nucleotide databases from the National Center for Biotechnology Information (NCBI) (http://ftp.ncbi.nlm.nih.gov). All the known plant miRNAs were obtained from miRBase (Release 11.0 [Griffiths-Jones]) (Griffiths-Jones et al. 2008) and clustered by CD-HIT-EST (Li and Godzik 2006) with $c = 1$, $n = 8$, $d = 250$, and $g = 1$. Only 1 mature sequence was selected from each cluster in order to eliminate closely related sequences. All the nonredundant mature miRNAs searched were in turn used as referential miRNAs for the prediction of potential fve-miRNAs. All *F. vesca* expressed sequence tags (ESTs) were downloaded from the NCBI dbEST database.

### Software Employed

Comparative software (BLAST-2.2.14) was downloaded from NCBI GenBank. The prediction of secondary structures of miRNA precursors were processed by miRCat (http://srna-tools.cmp.uea.ac.uk/mircat/; Moxon et al. 2008).

### Prediction of Potential Fve-miRNAs

The steps of our prediction procedure were similar to those in our earlier work on citrus (Song, Jia, et al. 2010). The prediction was performed using the EST databases, which are publicly available from NCBI (http://www.ncbi.nlm.nih.gov/). Sequences of miRNAs in *Arabidopsis*, rice, and other plant species are publicly available at http://www.sanger.ac.uk/. Initially, one known miRNA from the

**Table 1** Primers used for miR-5′ RACE, miR-3′ RACE, and qRT-PCR

| miRNA | GSP1 (5′→3′) | GSP2 (5′→3′) | GSP3 (5′→3′) |
|---|---|---|---|
| fve-miR156 | TTTTTTTTTTGTGCTCTCTATCTTCTG | GGAGTAGAAATGACAGAAGATAGAGAG | TGACAGAAGATAGAGAGCAC |
| fve-miR159 | TTTTTTTTTTAGAGCTCCCTTCAATC | GGAGTAGAAATTTGGATTGAAGGGAGC | TTTGGATTGAAGGGAGCTCTA |
| fve-miR398 | TTTTTTTTTCAGGGGCGACCTGAGAA | GGAGTAGAAACGTGTTCTCAGGTCGCC | CGTGTTCTCAGGTCGCCCCTG |
| fve-miR408 | TTTTTTTTTAGCCAGGGAAGAGGCAG | GGAGTAGAAATGCACTGCCTCTTCCCT | TGCACTGCCTCTTCCCTGGCT |
| fve-miR858 | TTTTTTTTTCAGGTCGAACAGACAAC | GGAGTAGAAATTTCGTTGTCTGTTCGA | TTTCGTTGTCTGTTCGACCTG |

GSP1 is the specific primer for miR-5′ RACE, and the underlined region base pairs with the 3′ poly (A)n;
GSP2 is the specific primer used for miR-3′ RACE, and the underlined region base pairs with the 5′ adapter;
GSP3 is the specific primer used for miRNA qRT-PCR.

referential set was selected and its seed region (positions 2–8) aligned to the ESTs on both strands. At each hit position, a "raw miRNA" of same length as the known miRNA was extracted. Secondly, the known miRNA was taken as a pattern and PatScan (Dsouz et al. 1997) used to filter all the raw miRNAs, with an allowance of a maximum of 3 mismatches between the known miRNA and the raw miRNAs. For the highly conserved mature sequences, particularly in the seed region (Lewis et al. 2005), we stipulated that all the mismatches could only occur in the nonseed region. The secondary structures of these sequences were predicted using miRCat. In this prediction, the following series of criteria was set to filter the sequences: 1) filtering with an RNA secondary folding energy threshold (−18 kcal/mol); 2) at least 18 paired nucleotides in its stem region and 1 terminal loop; 3) the position of mature miRNA on the hairpin (max 1 nt overlap of the 5′ end and hairpin loop); 4) a mature miRNA should be located on the same arm of the precursor as its known homologs; 5) no loop or break in the star sequences; 6) with a minimum of 12 base pairings between the mature and the star sequence; and 7) removal of the protein coding sequence at *E*-value $\leq 1 \times 10^{-6}$ (using the entire ESTs for BLASTX analysis).

### Oligonucleotide Synthesis and Preparation

All the oligonucleotides used were synthesized by Invitrogen Technologies and purified by desalting. The set of primers used in this study are as listed in Table 1.

### Low Molecular Weight RNA Extraction

Total RNA was isolated using TRIZOL (Invitrogen, Life Technologies, Carlsbad, CA) in accordance with the manufacturer's instructions. Low molecular weight RNA was separated from the total RNA solution using 4M LiCl (Adai et al. 2005; Song et al. 2009). The concentration of the RNA was measured by a UV-1800 spectrophotometer (Shimadzu, Japan) and visually ascertained in a 2.0% agarose gel.

### Construction and Screening of cDNA Libraries of Small RNAs

We utilized the procedure reported by Fu et al. (2005) to generate the miRNA-enriched library. This procedure has been popularly used to clone miRNAs and analyze their expression via RT-PCR, in which 5′- and 3′-end adapters are linked to the miRNA molecules (Lau et al. 2001; Aravin and Tuschl 2005).

Small RNAs were polyadenylated at 37 °C for 60 min in a 50-μl reaction mixture with 1.5 μg of small RNA, 1 mM ATP, 2.5 mM $MgCl_2$, and 4U poly(A) polymerase (Ambion, Austin, TX). Poly(A)-tailed small RNA was recovered by phenol/chloroform extraction and ethanol precipitation. 5′adapter (5′-CGACUGGAGCACGAGGACACUGA-CAUGGACUGAAGGAG UAGAAA-3′) was ligated to the poly (A)-tailed RNA using T4 RNA ligase (Invitrogen, Carlsbad, CA), and the ligation products recovered by phenol/chloroform extraction followed by ethanol

precipitation. Reverse transcription was performed using 1.5 μg of small RNA and 1 μg of $(dT)_{30}$ RT primer (ATTCTAGAGGCCGAGGCGGCCGACATG-$d(T)_{30}$ [A, G, or C] [A, G, C, or T]) with 200 U of SuperScript III reverse transcriptase (Invitrogen). Poly(A)-tailed small RNA (10-μl total volume) was incubated with 1 μl of $(dT)_{30}$ RT primer and 1 μl dNTP mix (10 mM each) at 65 °C for 5 min to remove any RNA secondary structures. The reactions were chilled on ice for at least 2 min, then the remaining reagents (5 × buffers, dithiothreitol [DTT], RNaseout, SuperScript III) were added as specified in the SuperScript III manual, and the reaction left to proceed for 60 min at 50 °C. Finally, the reverse transcriptase was inactivated by incubation for 15 min at 70 °C. After preparation of the miRNA libraries from various strawberry organs and tissues, we pooled similar quantities of these library samples for further PCR amplification reactions as done by Song, Jia, et al. (2010).

### Verification of Fve-miRNA Precise Sequence by MiR-RACE

To validate the precise sequences of the computationally predicted fve-miRNAs using miR-RACE (Song, Fang, et al. 2010), a cDNA library of the small RNA pool from the RNA samples isolated from various organs and tissues was amplified with mirRacer 5′ primer (5′-GGACACTGA-CATGGACTGAAGGAGTA-3′) and the mirRacer 3′ primer (5′-ATTCTAGAGGCCGAGGCGGCCGACATG-3′) to generate a pool of nongene-specific products. These miRacer primers are complementary to the 5′ and 3′ adapters, respectively. The conditions used for this amplification were carried out for 25 cycles at a final annealing temperature of 60 °C. 5′ end reactions were performed with the mirRacer 5′ primer and miRNA gene–specific forward primers (GSP1), whereas 3′ end reactions were carried out with the mirRacer 3′ primer and miRNA gene–specific reverse primers (GSP2). GSP1 and GSP2 were complementary to 17 nucleotide length sequences of the potential fve-miRNAs and a part of Poly (T) and 5′adapter (Table 1). In each case, a unique gene-specific DNA fragment was amplified. After the amplification, the 5′ and 3′ PCR products were separated in a 2.5% agarose gel with ethidium bromide (EtBr) staining. The gel slices containing DNA with a size of about 56 bp (5′ end product) and 87 bp (3′ end product) were excised, and the DNA fragments were purified using an agarose gel DNA purification kit (Takara, Japan), according to the manufacturer's instructions. The DNA fragment was directly subcloned with the TOPO TA cloning Kit (Invitrogen), and Colony PCR was performed using the PCR-specific primer pairs just as done above. The 5′end and 3′ end clones with PCR products of about 56 and 87 bp in length, respectively, were sequenced. To ensure the accuracy of the sequencing, more than 3 products were selected for sequencing. The results from correct sequencing were manually spliced so that we could determine exact sequences of the mdo-miRNAs.

### Quantitative RT-PCR of Fve-miRNAs

The templates used for fve-miRNA qRT-PCR were the fve-miRNA-enriched cDNA libraries generated from developing organs and tissues. To amplify the miRNA from the reverse transcribed cDNAs, we used the miRNA sequence validated as the forward primer and the mirRacer 3′Primer as the reverse primer. qRT-PCR was conducted with the fluorescence quantitative polymerase real-time PCR (Bio-rad) and the Rotor-Gene software version 6.1 (Wang, Zhou, et al. 2004). For each reaction, 1 μl of diluted cDNA (equivalent to about 100 pg of total RNA) was mixed with 10 μL of 2× SYBR green reaction mix (SYBR Green qRT-PCR Master Mix; Toyobo, Osaka, Japan), and 5 pmol each of the forward and the reverse primers were added in a final volume of 20 μl. The conditions for the PCR amplification were as follows: polymerase activation at 95 °C for 1 min, followed by 40 cycles of 95 °C for 15 s, 60 °C for 20 s, and 72 °C for 20 s. The fluorescence signal was measured once every 1 °C. Negative PCR controls (no cDNA template) were prepared to detect possible contamination. The specificity of the primer amplicons was checked by a melting curve analysis. Each reaction was repeated 3 times, and the template amount of fve-mirRNA was corrected by 5.8 s rRNAs. The CT values were converted into relative copy numbers using a standard curve (Chen et al. 2005). The 5.8 S rRNA was previously used as a reference gene in the qPCR detection of miRNAs in *Arabidopsis* (Shi and Chiang 2005). Data were analyzed with an $R^2$ above 0.998 using the LinRegPCR program (Ramakers et al. 2003). The primers are as shown in Table 1.

## Results

### Computational Prediction of Potential Fve-miRNAs

Many mature miRNAs are conserved from species to species in the plant kingdom, thus providing the basis for miRNA prediction in various plant species. We initiated our BLAST to search potential miRNA sequences using an *e*-value of 0.01.We then utilized all previously identified plant miRNAs against the *F. vesca* EST database in the first round of searches, whereas the second round of screening included cutting off of repeat and protein-coding sequences. After these 2 rounds of screening, the number of potential fve-miRNAs were reduced to 11 (Table 2) that could be classified into 5 miRNA families. Fve-miRNA156, fve-miRNA408, fve-miRNA159 families had 4, 3, and 2 members, respectively, whereas fve-miRNA398 and fve-miRNA408 families had only 1 member each. During screening, these fve-miRNAs were also evaluated for their A + U content, with results showing that the A + U contents range from 42.7% to 62.7% in the fve-miRNA precursors (Table 2), which is consistent with some previous studies (Zhang, Pan, and Anderson 2006; Qiu et al. 2007). The identified potential fve-miRNAs also have highly negative minimal fold energies between 46.2 and 85.7 (Table 2). These results are an indication of the strictness in the screening criteria used to obtain the predicted miRNAs. Based on the estimation that approximately 10 000 ESTs in plants contain one miRNA (Zhang, Pan, and Anderson 2006), the 47 890 ESTs in *F. vesca* should contain about 4–5

**Table 2** List of computer predicted and verified fve-miRNAs from the *Fragaria vesca* EST database

| miRNA | Predicted miRNA sequence | Verified miRNA sequence | Gene ID | Gene family | A + U (%) | MFES | Location |
|---|---|---|---|---|---|---|---|
| fve-miR156a.1 | CAGAAGATAGAGAGCACAGA | | | | | | |
| fve-miR156a.2 | TGACAGAAGATAGAGAGCAC | TGACAGAAGATAGAGAGCAC | 89547189 | MIR156 | 57.5 | 71.4 | 5′ |
| fve-miR156a.3 | TGACAGAAGATAGAGAGCACA | | | | | | |
| fve-miR156a.4 | TTGACAGAAGATAGAGAGCAC | | | | | | |
| fve-miR159a.1 | TTGGATTGAAGGGAGCTCTA | | | | | | |
| fve-miR159a.2 | TTTGGATTGAAGGGAGCTCTA | TTTGGATTGAAGGGAGCTCTA | 89542937 | MIR159 | 56.1 | 85.7 | 5′ |
| fve-miR398 | CGTGTTCTCAGGTCGCCCCTG | CGTGTTCTCAGGTCGCCCCTG | 89541205 | MIR398 | 42.7 | 61.4 | 3′ |
| fve-miR858 | TTTCGTTGTCTGTTCGACCTG | TTTCGTTGTCTGTTCGACCTG | 89545399 | MIR858 | 62.7 | 46.2 | 5′ |
| fve-miR408a.1 | ATGCACTGCCTCTTCCCTGGC | | | | | | |
| fve-miR408a.2 | TGCACTGCCTCTTCCCTGGCTT | | | | | | |
| fve-miR408a.3 | TGCACTGCCTCTTCCCTGGCT | TGCACTGCCTCTTCCCTGGCT | 89555570 | MIR408 | 46.1 | 79.3 | 3′ |

MFEs, minimal folding free energies.

miRNAs. However, in this study, 11 miRNAs were detected, a much higher frequency than the previous prediction in other plants. *Fragaria vesca* miRNA precursors showed similar diversity in structure and size (Table 2, Figure 1) as those of other plants. The length of miRNA precursors in *F. vesca* varied from 69 to 189 nt. The different sizes of the precursors of the fve-miRNAs within the 5 families suggest that they could have unique functions in the regulation of miRNA biogenesis or gene expression (Zhang, Pan, and Anderson 2006). All the potential fve-miRNA precursors could be folded into the typical secondary structure of miRNAs (Figure 1) and they are also postulated as important validation parameters for the MIR genes predicted. The diversity of the identified miRNAs can be also found in the location of mature miRNA sequences. The sequences of fve-miR156, fve-miR159, fve-miR408, and fve-miR858 are located at the 5′ end of the miRNA precursors, whereas the sequences of the others were found at the 3′ ends of their precursors.

### Validation of Fve-miRNAs' Precise Sequences

To date, the exact sequences of quite a number of miRNAs have been identified using miR-RACE in some fruit crops, such as citrus and apple (Song et al. 2009; Yu et al. 2010). However, there is no report about the bioinformatics prediction together with verification of the exact sequence of fve-miRNAs. In this study, we computationally identified 11 potential fve-miRNAs belonging to 5 families. However, their precise sequences needed to be validated to be true-to-type so as to allow them to be used for further studies. To validate the above predicted fve-miRNAs, miR-RACE was employed (refer to Materials and Methods), similar to the rapid amplification of cDNA ends (cDNA-RACE), with its PCR reaction template being a cDNA library for small RNAs ($\leq$200 nt) constructed from different and developing organs of *F. vesca* following the method previously reported (Song, Fang, et al. 2010). The difference between our method and traditional RACE lies in the gene-specific primers used (Table 1). The 2 gene-specific primers (GSP1 and GSP2) were designed factoring in 2 additional parameters. Reliable sequences of the PCR products were cloned, sequenced, and spliced to generate whole mature miRNA sequences that could be reconfirmed from their location and the nucleotides in their corresponding precursors.

The sequencing results demonstrate that the fve-miRNAs were relatively conserved with their orthologs of *Arabidopsis* (Table 2), whereas the sequence validation results demonstrate that 3 conserved miRNAs (fve-mir156, 159, and 398) were identical both in length and nucleotide sequence to their orthologs in *Arabidopsis*; however, there exist some variations of nucleotides that occur more frequently at the terminals rather than at the interiors of the miRNAs, an observation, which is in agreement with the findings of Seitz et al. (2008). This finding of the variant nucleotides between the orthologous miRNAs of strawberry and *Arabidopsis* buttresses the need for experimental validation of precise sequences of miRNAs predicted computationally. The possible explanation for the sequence variation across these small active elements is that transcription and miRNA processing might introduce differences in various plants. This can also be considered to be the results of evolution of miRNAs.

In our prediction, the families of fve-mir156, fve-mir408, and fve-mir159 have 4, 3, and 2 potential members each, respectively. However, only 1 potential member of each of the 3 families could be validated to be true to type.

### Expression Analysis of Fve-miRNAs by qRT-PCR

qRT-PCR is one of the most important and popular tools employed in analyzing the expression of miRNAs in organisms, and the spatiotemporal expression of miRNAs not only provides clues about their physiological functions but also gives fundamental evidence supporting the existence of the miRNAs in the organisms. In order to analyze the strawberry miRNA functions, we examined the expression of fve-miRNAs in different organs and tissues (Figure 2) by qRT-PCR analysis on the cDNA libraries of LMW-RNA samples from various organs of strawberry plants, for which the primers utilized, are as listed in Table 1. The positive validation of RT-PCR was carried out by

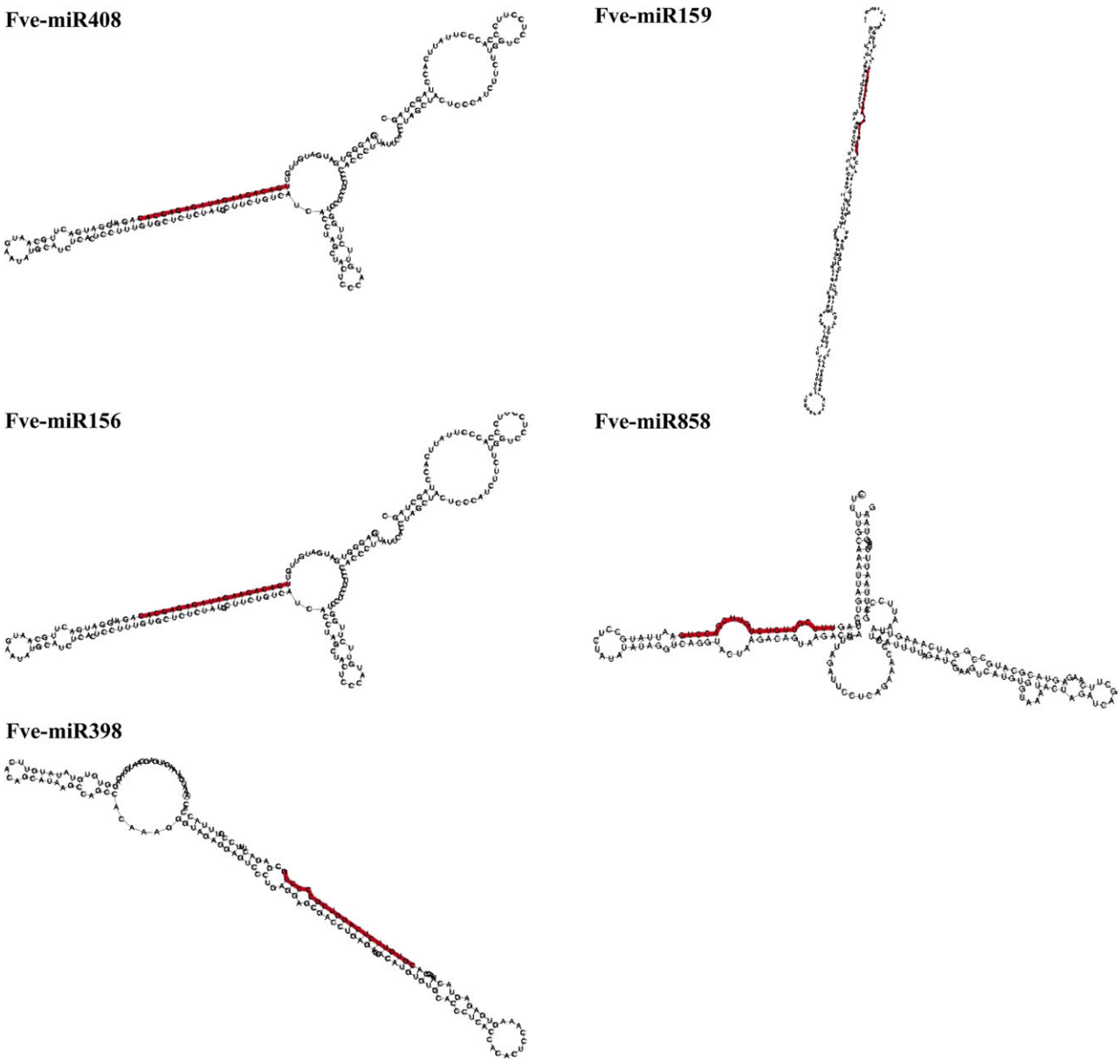


**Figure 1.** Predicted fold-back structures of identified fve-miRNAs. Mature miRNA sequences are shaded. miRNA precursors may be slightly longer than the sequences shown in this figure.

cloning and sequencing of the RT-PCR products. From the RT-PCR work, all the 5 conserved fve-miRNAs were present in the RNA fractions with different expression levels in different organs and tissues. The expression patterns of these 5 identified fve-miRNAs are shown in Figure 2. Basically, all the fve-miRNAs were expressed in most of the tissues, with some being expressed in specific patterns depending on tissue and/or growth stage of the strawberry (Figure 2). All the expression patterns of these fve-miRNAs can be categorized into a number of situations. Some miRNAs show ubiquitous expression (e.g., fve-miR408), whereas others show more specific expression patterns (e.g., fve-miR156 and fve-miR159). Fve-miR159 was expressed in most tissues examined, and its expression level in old leaves was higher than that in flowers, roots, and sepals, suggesting tissue-specific expression characteristics. Similarly, fve-miR156 appeared to be highly expressed in old leaves and moderately expressed in big flowers and small fruits, whereas fve-miR398 was moderately expressed in medium-sized sepals and old leaves. Fve-miR408 and fve-miR858 were highly expressed in the big flowers, with fve-miR408 showing faint expression in roots, whereas the expression of fve-miR858 being lowest in fruits.

In summary, qRT-PCR analysis confirmed the existence and sizes of the 5 identified miRNAs in strawberry (Figure 2). Their expression patterns indicate the aspects of tissue-,

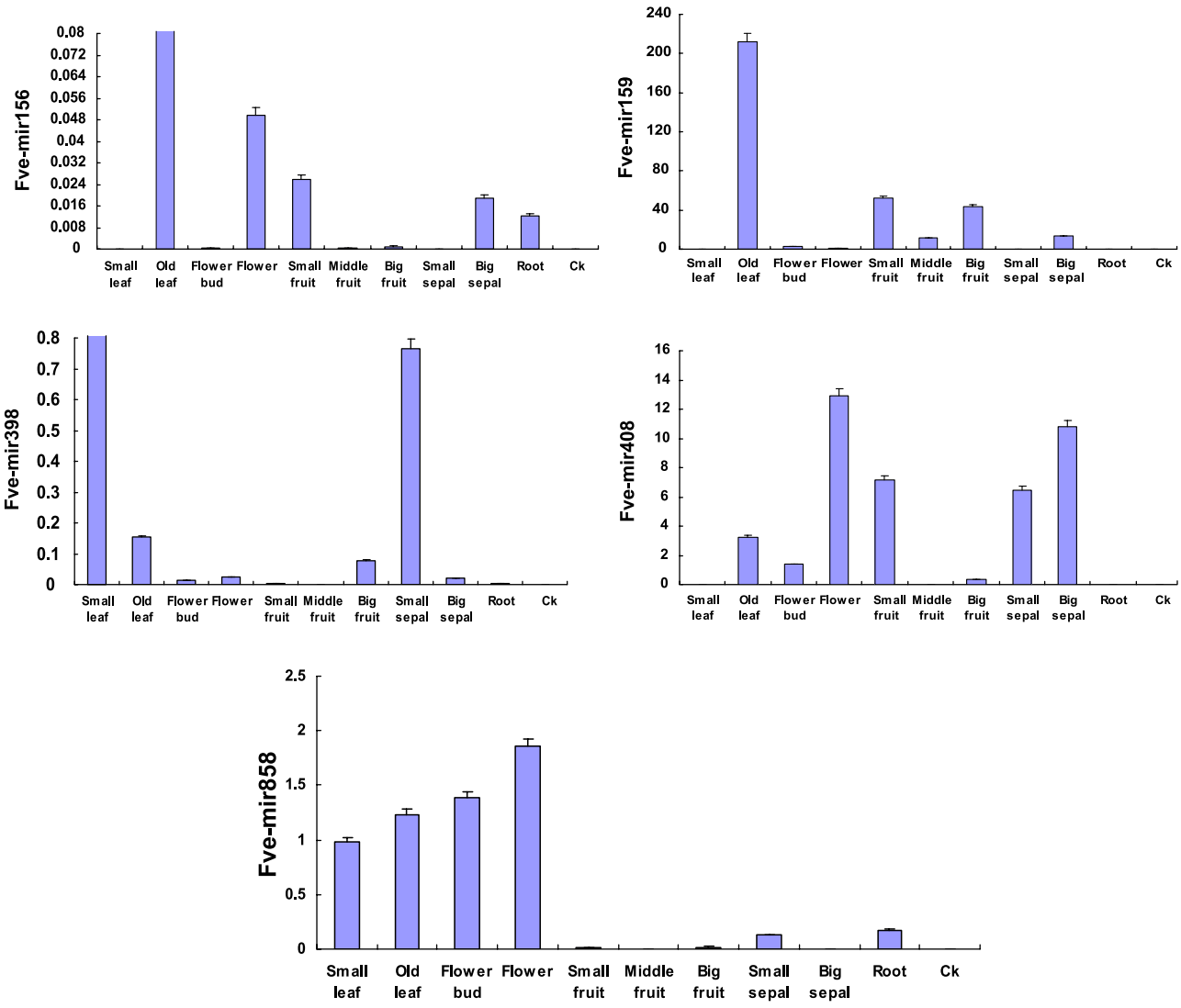


**Figure 2.** Relative expression levels of *Fragaria vesca* miRNAs in young leaves, old leaves, medium-sized sepals, big sepals, flower buds, flowers, and developing fruits.

species-, and/or growth-stage specificity as reflected by different expression levels. This can be an important reference for further comprehensive analysis of the functions of these fve-miRNAs.

## Discussion

The high level of sequence conservation of many mature miRNA sequences between distantly related species within the plant kingdom (Zhang, Pan, Cannon, et al. 2006) provides some homology search routes for identifying conserved orthologous miRNAs from various plant species. EST databases provide a valuable resource for identification of conserved miRNAs from plant species where the genome sequence is not available. In this study, computational prediction of fve-miRNAs from 47 890 *F. vesca* EST sequences was first carried out, and the verification of their precise sequences was done using the small RNA library generated from strawberry cv. Sweet Charlie. We identified miRNAs based on *F. vesca* miRNAs (diploid) and made the assumption that this identification could be relevant to Sweet Charlie (octoploid). Computational prediction methods for miRNAs have obvious advantages, including the quick prediction of a large number of miRNAs, low costs, and the prediction of novel and nonabundant miRNAs that are usually difficult to clone directly and have well been used in the identification of miRNAs in various plants, which are well dependent on the availability of ESTs and genomic sequences. In order to study Sweet Charlie miRNAs, we identified miRNAs from 47 890 *F. vesca* EST sequences, which is based on an assumption that *F. vesca* miRNAs (diploid) could be relevant to Sweet Charlie (octoploid). Our unpublished results on grape miRNAs alignment analysis (data not shown) from sequencing of small RNA libraries

illustrate that about 31.4% (38/121) of homologous Vv-miRNAs from various cultivars had divergent sequences, with the nucleotide variation happening only at either end of the sequences, signifying that validation of the precise sequences, especially the termini nucleotides, of miRNAs in some cultivars can also work as a new strategy for identifying their homologs in another cultivar. The sequence validation results demonstrated that the 2 conserved miRNAs (fve-miR156 and fve-miR159) were identical both in length and nucleotide sequence with their orthologs in *Arabidopsis* (Supplementary Table S1), but the other 3 miRNAs, which were nonconserved, varied in sequence at both internal and terminal nucleotides. The computational prediction of fve-miRNAs in this study is the first on record. Since precise sequences of miRNAs are essential for some important downstream researches, such as miRNA target prediction, miRNA evolution, miRNA regulatory roles, and biogenesis mechanisms, the determination of the precise sequence of the potential miRNA, especially the termini nucleotides, is therefore necessary. Combining bioinformatics comparison and experimental verification using the powerful miR-RACE procedure identified 11 fve-miRNAs belonging to 5 miRNA families from strawberry ESTs. Of the 5 miRNA families, 2 are highly conserved across the plant kingdom (miR156 and miR159) and 2 are moderately conserved families (miR398 and miR408) (Zhang, Pan, Cannon, et al. 2006). In addition, the sequences of the mature fve-miRNAs were also located on the same arm of the fold-back structure as the corresponding miRNA from the other plant species. The fold-back structures of the precursor sequences of fve-miRNAs also have free energy values similar to those of experimentally validated miRNAs of other plant species (Zhang, Pan, and Anderson 2006). Compared with the estimation that approximately 10 000 ESTs in plants could contain about one miRNA (Weber 2005), the frequency of 11 miRNAs identified in more than 47 890 *F. vesca* EST sequences (approximately 2.3 miRNAs per 10 000 ESTs) is slightly higher. This might be due to the different redundancy levels in various plant EST database and EST sequence quality. The current results confirm that this bioinformatics identification approach of miRNAs from ESTs is quite efficient (Zhang, Pan, and Anderson 2006).

The criteria used for acceptance of candidate fve-miRNA precursors were followed exactly in our prediction work, where 4 potential miRNAs met the requirements. The one candidate strawberry miRNA (fve-miR858) that did not meet the definitions of fve-miRNA precursors was validated experimentally to be among true-to-type miRNAs in strawberry and was therefore listed accordingly in this paper. This phenomenon could be explained by the fact that the set of criteria used for miRNA prediction were relatively powerful even before we could characterize miRNAs in plant kingdom very well. Every prediction could discover even more but not all the miRNAs in an organism due to various intervening variables.

If identified miRNAs are not subjected to any experimental work we cannot validate whether these were indeed true miRNAs. In our study, all of the 5 miRNAs belonging to 5 families could be validated by miRNA-RACE. This is first report on accurate sequences of fve-miRNAs. Their existence was also confirmed by RT-PCR, which also showed the expression patterns of these miRNAs in different tissues of strawberry ranging from young leaves, old leaves, small sepals, big sepals, flower buds, flowers, and developing fruits. To elucidate the strawberry miRNA functions, we examined their expression in different organs (Figure 2). Preferential expression of a miRNA in specific tissues might provide clues about its physiological function. These data can also be powerful evidence supporting the existence of these miRNAs in the strawberry. The majority of the fve-miRNAs exhibited expression patterns (Figure 2) similar to their orthologs in *Arabidopsis*. Although Wu et al. (2009) showed that miR156 is both necessary and sufficient for the expression of the juvenile phase and that it functions as a master regulator of this phase, the expression patterns of fve-miR156 in this study do not fit those reported for *Arabidopsis*. All 5 miRNAs families exhibited some organ and/or growing stage specificity. The expression patterns of these 5 fve-miRNAs were similar to those reported in *Arabidopsis*, where the majority was expressed ubiquitously in all tissues, whereas some exhibited tissue-, species-, and/or growth-stage-specific characteristics.

With more EST sequencing data for strawberry released publically, undertaking fve-miRNA identification by searches against the new EST database will be more fruitful. This work is important as it provides a more comprehensive understanding of the roles, biogenesis, and evolution of fve-miRNAs. Recently, Shulaev et al. (2011) identified 76 miRNAs from *F. vesca* genome sequences by bioinformatics analysis, but strawberry miRNAs still remains largely unknown and the identified strawberry miRNAs have not been confirmed by experiments. To elucidate the functions and regulation pathways of these 76 potential fve-miRNAs, it is important to validate their precise sequences and know the localization of them.

## Supplementary Material

Supplementary material can be found at http://www.jhered.oxfordjournals.org/.

## Funding

Opening Foundation of Beijing Key Laboratory of New Technology in Agricultural Application (NYXJ-2010-07); NCET Program of China (NCET-08-0796).

## References

Adai A, Johnson C, Mlotshwa S, Archer-Evans S, Manocha V, Vance V, Sundaresan V. 2005. Computational prediction of miRNAs in *Arabidopsis thaliana*. Genome Res. 15:78–91.

Aravin A, Tuschl T. 2005. Identification and characterization of small RNAs involved in RNA silencing. FEBS Lett. 579:5830–5840.

Aukerman MJ, Sakai H. 2003. Regulation of flowering time and floral organ identity by a microRNA and its *APETALA2*-like target genes. Plant Cell. 15:2730–2741.

Bartel DP. 2004. MicroRNAs, genomics, biogenesis, mechanism, and function. Cell. 116:281–297.

Carra A, Mica E, Gambino G, Pindo M, Moser C, Enrico MP, Schubert A. 2009. Cloning and characterization of small non-coding RNAs from grape. Plant J. 59:750–763.

Chen C, Ridzon DA, Broomer AJ, Zhou Z, Lee DH, Nguyen JT, Barbisin M, Xu NL, Mahuvakar VR, Andersen MR, et al. 2005. Real-time quantification of microRNAs by stem-loop RT-PCR. Nucleic Acids Res. 33:e179.

Chen X. 2004. A microRNA as a translational repressor of *APETALA2* in *Arabidopsis* flower development. Science. 303:2022–2025.

Dsouz M, Larsen N, Overbeek R. 1997. Searching for patterns in genomic data. Trends Genet. 13:497–498.

Emery JF, Floyd SK, Alvarez J, Eshed Y, Hawker NP, Izhaki A. 2003. Radial patterning of *Arabidopsis* shoots by *Class III HD-ZIP* and *KANADI* genes. Curr Biol. 13:1768–1774.

Fu H, Tie Y, Xu C, Zhang Z, Zhu J, Shi Y, Jiang H, Sun Z, Zheng X. 2005. Identification of human fetal liver miRNAs by a novel method. FEBS Lett. 579:3849–3854.

Gleave AP, Ampomah-Dwamena C, Berthold S, Dejnoprat S, Karunairetnam S, Nain B, Wang YY, Crowhurst RN, MacDiarmid RM. 2008. Identification and characterisation of primary microRNAs from apple (*Malus domestica* cv. Royal Gala) expressed sequence tags. Tree Genet Genomes. 4:343–358.

Griffiths-Jones S. 2004. The microRNA registry. Nucleic Acids Res. 32:D109–D111.

Griffiths-Jones S, Saini HK, Dongen SV, Enrigh AJ. 2008. MiRBase, tools for microRNA genomics. Nucleic Acids Res. 36 (Data issue):D154–D158.

Guo HS, Xie Q, Fei JF, Chua NH. 2005. MicroRNA directs mRNA cleavage of the transcription factor NAC1 to downregulate auxin signals for Arabidopsis lateral root development. Plant Cell. 17:1376–1386.

Juarez MT, Kui JS, Thomas J, Heller BA, Timmermans MC. 2004. MicroRNA-mediated repression of rolled leaf specifies maize leaf polarity. Nature. 428:84–88.

Kasschau KD, Xie Z, Allen E, Llave C, Chapman EJ, Krizan KA, Carrington JC. 2003. P1/HC-Pro, a viral suppressor of RNA silencing, interferes with *Arabidopsis* development and miRNA function. Dev Cell. 4:205–217.

Lau NC, Lim LP, Weinstein EG, Bartel DP. 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. Science. 294:858–862.

Laufs P, Peaucelle A, Morin H, Traas J. 2004. MicroRNA regulation of the CUC genes is required for boundary size control in *Arabidopsis* meristems. Development. 131:4311–4322.

Lewis BP, Burge CB, Bartel DP. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. 120:15–20.

Li H, Zhang ZH, Huang FF, Chang LL, Ma Y. 2009. MicroRNA expression profiles in conventional and micropropagated strawberry (*Fragaria* x *ananassa* Duch.) plants. Plant Cell Rep. 28:891–902.

Li W, Godzik A. 2006. Cd-hit, a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 22:1658–1659.

Lu YD, Gan QH, Chi XY, Qin S. 2008. Identification and characterization of microRNAs and their targets in grapevine (*Vitis vinifera*). Scientia Agricultura Sinica. 7:929–943.

Mallory AC, Dugas DV, Bartel DP, Bartel B. 2004. MicroRNA regulation of NAC-domain targets is required for proper formation and separation of adjacent embryonic, vegetative, and floral organs. Curr Biol. 14:1035–1046.

Mallory AC, Reinhart BJ, Jones-Rhoades MW, Tang G, Zamore PD, Barton MK, Bartel DP. 2004. MicroRNA control of PHABULOSA in leaf development, importance of pairing to the micro RNA 5′ region. EMBO J. 23:3356–3364.

Moxon S, Jing R, Szittya G, Schwach F, Rusholme Pilcher RL, Moulton V, Dalmay T. 2008. Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening. Genome Res. 18:1602–1609.

Pantaleo V, Szittya G, Moxon S, Miozzi L, Moulton V, Dalmay T, Burgyan J. 2010. Identification of grapevine microRNAs and their targets using high-throughput sequencing and degradome analysis. Plant J. 62:960–976.

Papp I, Mette MF, Aufsatz W. 2003. Evidence for nuclear processing plant microRNA and short interfering precursors. Plant Physiol. 132:1382–1390.

Park MY, Wu G, Gonzalez-Sulser A. 2005. Nuclear processing and export of microRNAs in *Arabidopsis*. Proc Natl Acad Sci U S A. 102:3691–3696.

Qiu CX, Xie FL, Zhu YY, Guo K, Huang SQ, Nie L, Yang ZM. 2007. Computational identification of microRNAs and their targets in *Gossypium hirsutum* expressed sequence tags. Gene. 395:49–61.

Ramakers C, Ruijter JM, Deprez RH, Moorman AF. 2003. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neurosci Lett. 339:62–66.

Seitz H, Ghildiyal M, Zamore PD. 2008. Argonaute loading improves the 5′ precision of both microRNAs and their miRNA strands in flies. Curr Biol. 18:147–151.

Shi R, Chiang VL. 2005. Facile means for quantifying microRNA expression by real-time PCR. Biotechniques. 39:519–525.

Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Shrinivasrao P, et al. 2011. The genome of woodland strawberry (*Fragaria vesca*). Nat Genet. 43:109–116.

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.

Song C, Jia Q, Fang J, Li F, Wang C, Zhang Z. 2010. Computational identification of citrus microRNAs and target analysis in citrus expressed sequence tags. Plant Biol. 12:927–934.

Song C, Fang J, Wang C, Guo L, Nicholas KK, Ma Z. 2010. miR-RACE, a new efficient approach to determine the precise sequences of computationally identified trifoliate orange (*Poncirus trifoliata*) microRNAs. PLos One. 5:e10861.

Sunkar R, Kapoor A, Zhu JK. 2006. Posttranscriptional induction of two Cu/Zn superoxide dismutase genes in *Arabidopsis* is mediated by downregulation of miR398 and important for oxidative stress tolerance. Plant Cell. 18:2051–2065.

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

Varkonyi-Gasic E, Gould N, Sandanayaka M, Sutherland P, MacDiarmid R M. 2010. Characterisation of microRNAs from apple (*Malus domestica* 'Royal Gala') vascular tissue and phloem sap. BMC Plant Biol. 10:159–171.

Wang JF, Zhou H, Chen YQ, Luo QJ, Qu LH. 2004. Identification of 20 microRNAs from Oryza sativa. Nucleic Acids Res. 32:1688–1695.

Wang XJ, Reyes JL, Chua NH, Gaasterland T. 2004. Prediction and identification of *Arabidopsis thaliana* microRNAs and their mRNA targets. Genome Biol. 5:R65.

Weber MJ. 2005. New human and mouse microRNA genes found by homology search. FEBS J. 272:59–73.

Wu G, Park MY, Conway SR, Wang JW, Weigel D, Poethig RS. 2009. The sequential action of miR156 and miR172 regulates developmental timing in *Arabidopsis*. Cell. 138(4):750–759.

Xie FL, Huang SQ, Guo K, Xiang AL, Zhu YY, Nie L, Yang ZM. 2007. Computational identification of novel microRNAs and targets in Brassica napus. FEBS Lett. 581:1464–1474.

Yin Z, Li C, Han X, Shen F. 2008. Identification of conserved microRNAs and their target genes in tomato (*Lycopersicon esculentum*). Gene. 414:60–66.

Yu HP, Song CN, Jia QD, Wang C, Li F, Nicholas KK, Zhang XY, Fang JG. 2011. Computational identification of microRNAs in apple expressed sequence tags and validation of their precise sequences by miR-RACE. Physiologia Plantarum. 141:56–70.

Zeng C, Wang W, Zheng Y, Chen X, Bo W, Song S, Zhang W, Peng M. 2010. Conservation and divergence of microRNAs and their functions in *Euphorbiaceous* plants. Nucleic Acids Res. 38:981–995.

Zhang BH, Pan X, Anderson TA. 2006. Identification of 188 conserved maize microRNAs and their targets. FEBS Lett. 580:3753–3762.

Zhang BH, Pan X, Cannon CH, Cobb GP, Anderson TA. 2006. Conservation and divergence of plant microRNA genes. Plant J. 46:243–259.

Zhang BH, Pan XP, Stellwag EJ. 2008. Identification of soybean microRNAs and their targets. Planta. 229:161–182.

Zhang BH, Wang QL, Wang KB, Pan XP, Liu F, Guo TL, Cobb GP, Anderson TA. 2007. Identification of cotton microRNAs and their targets. Gene. 397:26–37.

Zhang J, Zeng R, Chen J, Liu X, Liao Q. 2008. Identification of conserved microRNAs and their targets from *Solanum lycopersicum* Mill. Gene. 423:1–7.

Received June 18, 2011; Revised September 30, 2011; Accepted October 18, 2011

Corresponding Editor: John Stommel