Characterization of expression patterns of small RNAs among various organs in *Arabidopsis* and rice based on 454 platform-generated high-throughput sequencing data

Chaogang Shao¹, Xiaoxia Ma², Ming Chen³*, Yijun Meng²*

¹College of Life Sciences, Huzhou Teachers College, Huzhou 313000, P. R. China
²College of Life and Environmental Sciences, Hangzhou Normal University, Hangzhou 310036, P. R. China
³Department of Bioinformatics, College of Life Sciences, Zhejiang University, Hangzhou 310058, P. R. China

* Corresponding authors: mengyijun@zju.edu.cn²; mchen@zju.edu.cn³

Abstract

The advent of high-throughput sequencing (HTS) strengthened our capacity for small RNA (sRNA) discovery. Here, we did a transcriptome-wide survey of sRNAs, mainly focusing on the microRNAs, with organ-specific expression patterns in both *Arabidopsis* (*Arabidopsis thaliana*) and rice (*Oryza sativa*). By using sRNA HTS data generated by 454 sequencing technology, four organs in *Arabidopsis* (i.e. six-day-old seedling, rosette leaf, flower and silique), and four in rice (shoot apex, leaf, root apex and inflorescence) were investigated. Chromosome-wide distribution patterns of the organ-specific sRNAs were obtained. We found that, in rice, the 21-nt (nucleotide) sRNAs occupy a large portion of the sRNA population highly expressed in inflorescences. In contrast, the sRNAs not expressed in rice floral organ are predominantly 24 nt in length. Through literature mining, ath-miR156d, ath-miR400, ath-miR822 and ath-miR824 were suggested to be involved in seedling development in *Arabidopsis*, and osa-miR169 induced by drought and high salinity was indicated to regulate leaf growth in rice.

Keywords: High-throughput sequencing, MicroRNA, Organ-specific, Plant, Small RNA.

Abbreviations: GEO - Gene Expression Omnibus; HTS - high-throughput sequencing; kb - kilobases; miRNA - microRNA; NF-Y - nuclear factor Y; nt - nucleotide; PPR - pentatricopeptide repeat; pre-miRNA - precursor microRNA; RISC - RNA-induced silencing complex; RNA Pol II - RNA polymerase II; RPM - reads per million; sRNA - small RNA; TIGR - The Institute for Genome Research.

Introduction

Since the discovery of ~21-nucleotide (nt) microRNAs (miRNAs) in *Caenorhabditis elegans*, i.e. lin-4 and let-7 (Lee et al., 1993; Reinhart et al., 2000), many miRNA genes, either highly conserved or newly evolved, have been cloned in organisms (Carthew & Sontheimer, 2009; Voinnet, 2009). Recently, the powerful technology, called high-throughput sequencing (HTS), has presented us with an unprecedentedly huge and complex sRNA world. In plants, several sRNA species have been characterized, such as miRNAs (Voinnet, 2009), nat-siRNAs (natural antisense small interfering RNAs) (Borsani et al., 2005), and ta-siRNAs (trans-acting small interfering RNAs) (Allen et al., 2005; Williams et al., 2005). However, much more research efforts are needed to depict a clear view of sRNA-mediated gene regulatory networks.

As one of the most sophisticatedly characterized sRNA species, the biological functions of the miRNAs have been well recognized in plants (Chen, 2009). Dozens of miRNA targets have been verified by using modified 5’ RACE (rapid amplification of cDNA ends) (Jones-Rhoades et al., 2006) or degradome sequencing method (Addo-Quaye et al., 2008; German et al., 2008; Li et al., 2010). Further functional studies could be carried out by focusing on these identified targets through forward or reverse genetics approach. However, one notion has been widely recognized that coincident spatio-temporal expression patterns between the miRNA regulators and their target genes are the requisite factor for successful regulation (Wu et al., 2006; Allen et al., 2007; Palatnik et al., 2007; Sieber et al., 2007; Raman et al., 2008; Nogueira et al., 2009; Rubio-Somoza et al., 2009). In other word, the expression pattern of a specific miRNA gene predetermines its effective range. For example, miR319 shares high sequence similarity with miR159 in *Arabidopsis*. Interestingly, *MYB* transcripts are efficiently targeted by miR159 but not by miR319. Obviously, it can not be simply explained by the sequence complementarity-based target recognition rule for miRNA-mediated regulation in plants. Instead, the organ/tissue-specific expression patterns of both miR319 and its *MYB* targets restrict the regulatory effect of this miRNA on the *MYB* tRNAs (Palatnik et al., 2007).

In this study, we investigated the expression patterns of the sRNAs in various organs of *Arabidopsis* and rice based on the publicly available sRNA HTS data. Several miRNAs were demonstrated to be organ-specifically expressed, while some were constitutively expressed among the organs investigated. Literature mining showed that ath-miR156d, ath-miR400, ath-miR822 and ath-miR824 were likely to be implicated in young seedling development in *Arabidopsis*, and the osa-miR169 family in rice had a potential role in the regulation of leaf growth, both of which were consistent with the results of our expression-based analysis. Taken together, dozens of organ-specifically expressed sRNAs were identified by employing HTS data-based bioinformatics approach. These sRNAs showing great potential in controlling organ growth and development provide a basis for further functional studies on plant sRNAs.
Results and discussion

Expression patterns

The raw sRNA HTS data was parsed, and the normalized expression values (in RPM, reads per million; see Materials and methods for detail of normalization) were calculated for each signature. Four organs (six-day-old seedling, rosette leaf, flower and silique) in Arabidopsis and four (shoot apex, leaf, root apex and inflorescence) in rice were investigated separately. At first glance, more miRNAs were constitutively expressed in the four organs in Arabidopsis than in rice (57 versus 10). This discrepancy is more significant when considering the factor that less miRNAs of Arabidopsis (a total of 199 miRNAs) have been registered in miRBase (release 15) (Griffiths-Jones et al., 2008) than those of rice (447 miRNAs).

Moreover, a great difference also exists in the constitutively expressed sRNAs (including miRNAs) between Arabidopsis and rice (2315 versus 28). One possible reason for this discrepancy turned out to be that distinct organs were investigated for the two plants. And another potential reason could be the different sequencing depth of the two data series of Arabidopsis and rice, although both were generated by using the 454 sequencing technology. Our results also showed that, in both plants, the organ-specific miRNAs tend to be poorly conserved while the constitutive ones were more conserved within the plant kingdom (Table 1). For example, in Arabidopsis, the poorly conserved miRNAs, ath-miR780 and ath-miR832, were exclusively cloned from flowers and siliques, respectively. The highly conserved families, such as ath-miR159, ath-miR164, ath-miR169, ath-miR171, and ath-miR172, were constitutively expressed. Similarly, the less conserved miRNAs, osa-miR1317 and osa-miR2118 were only detected in rice leaves and inflorescences respectively, while the highly conserved ones, osa-miR160 and osa-miR169 were expressed in four organs. The 24-nt sRNAs were reported to occupy a dominant portion of plant endogenous sRNAs (Chen et al., 2010; Kasschau et al., 2007). This notion is confirmed by this study (Fig. S1). However, we found some exceptions in rice that the 21-nt sRNAs occupied a dominant portion of the sRNAs specifically expressed in the inflorescences, expressed in leaves, shoot apexes and inflorescences, and the sRNAs constitutively expressed in the four organs, respectively (Fig. S2). It is notable that all the three expression patterns are associated with the rice reproductive organ, inflorescence. Previous study also discovered that the 21-nt sRNAs were preferentially expressed in the rice developing inflorescences, which were suggested to function in early reproductive development (Johnson et al., 2009). However, the enrichment of the 21-nt sRNAs in rice reproductive organ was not detected in Arabidopsis, further supporting the previous hypothesis that the reproduction-associated 21-nt sRNAs might be specifically conserved in the monocots (Johnson et al., 2009). From this point of view, the biological relevance of these 21-nt sRNAs needs to be elucidated, especially for rice reproduction.

Chromosome-wide distribution patterns

Distribution patterns of miRNAs

The miRNAs detected by 454 sequencing were mapped to the chromosomes, based on their genomic positions retrieved from miRBase (release 15) (Griffiths-Jones et al., 2008) (Fig. S3). In-depth investigation showed that the miRNA coordinates generated from the same miRNA precursor shared a common expression pattern. For instance, ath-miR832-3p and ath-miR832-5p were specifically expressed in siliques, ath-miR780.1 and ath-miR780.2 were specifically expressed in flowers, and ath-miR161.1 and ath-miR161.2 were constitutively expressed in the four organs of Arabidopsis. In rice, osa-miR1317-5p and osa-miR1317-3p were leaf-specifically expressed. Hence, it is interesting to characterize the organ-specific promoters involved in transcriptional regulation of these miRNA genes. To partially address this issue, we retrieved the promoters of the organ-specific miRNA genes from PMRD (plant microRNA database; http://bioinformatics.cau.edu.cn/PMRD/) (Zhang et al., 2010) for cis-regulatory element analysis. The 1-kb (kilobase) sequence upstream of each pre-miRNA (precursor microRNA) was subjected to this analysis by using the online tool PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Lescot et al., 2002). Consistent with the well-established notion that most microRNA genes are transcribed by RNA polymerase II (Pol II) and possess the canonical Pol II-dependent cis-elements TATA-box and CAAT-box (Lee et al., 2004; Meng et al., 2009; Xie et al., 2005), nearly all the organ-specific miRNAs investigated possess the two core promoter elements (Table 2). Some interesting cis-elements were found to be possessed by certain organ-specific miRNA groups. For example, MBS that involved in drought inducibility was identified in all the miRNA genes expressed in the 6-day-old whole seedlings in Arabidopsis, and LTR involved in low-temperature responsiveness was only detected in the upstream regions of the miRNA genes specifically expressed in the rice shoot apexes. Notably, Box-W1, a fungal elicitor responsive element, exists in all the miRNA genes specifically expressed in the root apexes of rice. Considering the widespread interactions between plant root systems and the microorganisms in the soil, this kind of cis-elements may play a critical role in defining root-specific expression patterns of certain genes. However, some precursors generated two miRNAs with distinct expression patterns. ath-miR2112-3p was expressed in the siliques whereas ath-miR2112-5p was detected in the young seedlings and the flowers. osa-miR159a.2 was expressed in the leaves, but osa-miR159a.1 was constitutively expressed. Some neighboring miRNA genes may share a common promoter based on the similar expression patterns, such as the cluster formed by ath-miR157b, ath-miR163 and ath-miR839, the cluster formed by ath-miR171c and ath-miR399b, the cluster formed by ath-miR398a and ath-miR840, and the cluster formed by osa-miR2118d, osa-miR2118c and osa-miR2118f. For miRNAs within a specific family, their expression patterns might be quite distinct from each other. For example, osa-miR169a was expressed in the leaves, osa-miR169c was expressed in the leaves and the shoot apexes, osa-miR169e was expressed in the inflorescences, and osa-miR169n, osa-miR169g and osa-miR169m were constitutively expressed.

Distribution patterns of sRNAs

All the short reads were mapped to the corresponding genomes. After BLAST, the numbers of signatures were reduced from 340,114 to 331,037 containing 228,265 uniquely mapped ones in Arabidopsis, and from 2,722 to 2,720 including 955 unique ones in rice. Further statistical results presented a quite distinct vision of the endogenous sRNAs compared to the miRNAs. In both plants, the organ-specific sRNA populations are much larger than the constitutive ones (Fig. S4 A and B). In contrast, the constitutively expressed miRNAs constitute a large population, especially in Arabidopsis (Table 1). The silique-specific and the flower-specific sRNAs in Arabidopsis, and the inflorescence-specific sRNAs in rice are the abundant sRNA portions (Fig. S4 A, B and C), suggesting an active role of plant sRNAs in reproduction. Previous reports showed that plant endogenous sRNAs were highly enriched in the centromeric and pericentromeric regions (Chen et al., 2010; Kasschau et al., 2007). We found that the inflorescence-specific sRNAs in rice were not enriched in these regions (Fig. S4 C).
### Table 1. List of organ-specifically and constitutively expressed microRNAs.

<table>
<thead>
<tr>
<th>Arabidopsis</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Whole seedling (6-day-old)</strong></td>
<td>ath-miR773, ath-miR842, ath-miR870, and ath-miR1886.1</td>
</tr>
<tr>
<td><strong>Rosette leaf</strong></td>
<td></td>
</tr>
<tr>
<td>Flower</td>
<td>ath-miR778, ath-miR780.1, ath-miR780.2, ath-miR835-5p, ath-miR867, and ath-miR2111b*</td>
</tr>
<tr>
<td>Silique</td>
<td>ath-miR832-3p, ath-miR832-5p, ath-miR836, ath-miR859, ath-miR864-5p, ath-miR866-3p, ath-miR2111a*, and ath-miR2112-3p</td>
</tr>
<tr>
<td>Four organs</td>
<td>ath-miR156d, ath-miR157b, ath-miR157d, ath-miR158a, ath-miR159a, ath-miR159b, ath-miR160c, ath-miR161.1, ath-miR161.2, ath-miR162a, ath-miR163, ath-miR164b, ath-miR164c, ath-miR165a, ath-miR166d, ath-miR167b, ath-miR167d, ath-miR168a, ath-miR169a, ath-miR169c, ath-miR169d, ath-miR169m, ath-miR170, ath-miR171a, ath-miR171c, ath-miR172b, ath-miR172b*, ath-miR172c, ath-miR172e, ath-miR173, ath-miR319b, ath-miR390a, ath-miR391, ath-miR394b, ath-miR396a, ath-miR396b, ath-miR397a, ath-miR397b, ath-miR398a, ath-miR398c, ath-miR399a, ath-miR399b, ath-miR399f, ath-miR400, ath-miR403, ath-miR408, ath-miR447b, ath-miR775, ath-miR822, ath-miR823, ath-miR824, ath-miR825, ath-miR839, ath-miR840, ath-miR845a, ath-miR847, and ath-miR848</td>
</tr>
</tbody>
</table>

| Rice |  |
| Shoot apex | osa-miR1428e-3p, and osa-miR1859 |
| Leaf | osa-miR440, osa-miR1850.1, and osa-miR1874-3p |
| Root apex | osa-miR1469e, osa-miR1423b, osa-miR2118d, osa-miR2118e, osa-miR2118f, and osa-miR2118q |
| Inflorescence | osa-miR159a.1, osa-miR160d, osa-miR160e, osa-miR166b, osa-miR168a, osa-miR169g, osa-miR169m, osa-miR169n, osa-miR444c.2, and osa-miR1862d |
| Four organs | The microRNAs belonging to the same families were indicated by gray shadow. |
However, we should clarify that only the sRNAs with unique genomic loci were retained for this distribution analysis. The previously observed sRNAs highly enriched in centromeric and pericentromeric regions were largely associated with repeat sequences. Consistent with this notion, we found that most of the sRNAs resided within the centromeric/pericentromeric regions could be mapped to multiple genomic loci. Since we were not able to tell from which loci these sRNAs were exactly originated, these sRNAs were removed from our distribution analysis.

**MicroRNAs differentially expressed in various organs**

Although 57 and 10 miRNAs were constitutively expressed in *Arabidopsis* and rice respectively, further insights into the expression patterns of these miRNAs provided us with a quite intriguing view. In *Arabidopsis*, ath-miR156d, ath-miR400, ath-miR822 and ath-miR824 were highly expressed in the young seedlings (Fig. 1 A). The ath-miR172 family members, including ath-miR172b and its partner ath-miR172b*, ath-miR172c, and ath-miR172e, were dominantly expressed in the reproductive organs, i.e. the flowers and the siliques (Fig. 1 B). It is consistent with the previous result that the regulatory pathway ath-miR172—APETALA2 is involved in flower development (Chen, 2004). It was also evident that osa-miR172a was specifically detected in the rice leaves, and ath-miR172b was highly expressed in the rosette leaves of *Arabidopsis* (Table 1). In addition to the involvement of miR172 in floral organ development as mentioned above, miR172 was also demonstrated to be involved in vegetative leaf emergence through the miR172—SCHNARCHZAPFEN pathway (Nonogaki, 2010). It is a *bona fide* support for our detected expression patterns of the miR172 family members in both plants. From another point of view, it indicates that our cross-library (i.e. cross-organ) comparison is highly reliable. Moreover, ath-miR167b and ath-miR168a were highly expressed in the rosette leaves (Fig. S5). In rice, osa-miR160d and osa-miR160e were highly expressed in the shoot apexes (Fig. 1 C). The osa-miR169 family (including osa-miR169g, osa-miR169m and osa-miR169n), osa-miR159a.1, and osa-miR168a were dominantly expressed in the rice leaves (Fig. 1 D and Fig. S5). ath-miR160 was demonstrated to be critical for root cap formation (Wang et al., 2005). Here, the high expression level of osa-miR160 in the shoot apexes indicates its potential role in both underground and aboveground apex development. This notion is supported by the previous finding that the miR160—auxin response factor 17 pathway played an important role in leaf development in *Arabidopsis* (Mallory et al., 2005). In this regard, the highly conserved miR160 family may play an essential role in apex meristem formation and maintenance in the plants. Moreover, we noticed that no miRNA specifically expressed in the rosette leaves was identified in *Arabidopsis* (Table 1). On the other hand, numerous sRNAs specifically expressed in the *Arabidopsis* leaves were identified by using the same HTS data sets. However, we still could not exclude the possibility that some miRNAs with relatively low expression levels were specifically expressed in *Arabidopsis* leaves, which were not cloned by the 454 technology due to its insufficient sequencing depth.

**MicroRNAs involved in seedling development in *Arabidopsis***

The aforementioned expression data of ath-miR156d, ath-miR400, ath-miR822, and ath-miR824 suggest their potential roles in young seedling development (Fig. 1 A). We obtained additional evidences to support this presumption (Fig. 1 E). Wu et al.’s study (2009) demonstrated that ath-miR156 was necessary during the juvenile phase, and regulated the
juvenile-to-adult transition in Arabidopsis, both of which were accomplished by repressing the SPL transcription factors. Target prediction by using miRU (Zhang, 2005) showed that the genes encoding embryo defective 1025 and embryo defective 2745 (AT4G20090 and AT5G39710), pentatricopeptide repeat (PPR) proteins, and seed storage cupin family protein (AT2G28680), were targeted by ath-miR400. The former three genes are essential for embryogenesis (Cushing et al., 2005; Lurin et al., 2004). In this regard, we proposed that ath-miR400 was involved in embryogenesis, seed maturation, and possibly embryo-to-seedling transition. The ath-miR824—AGL16 regulatory pathway is involved in stomatal development (Kutter et al., 2007), and AGL16 is specifically expressed in leaf guard cells and trichomes (Alvarez-Buylla et al., 2000). Considering the active stomatal development during leaf growth, the ath-miR824—AGL16 pathway may be indispensable for young seedling survival. The genes, AT1G66450, AT2G02620, AT2G13900 and AT3G26250 encoding DC1 domain-containing proteins, were predicted to be targeted by ath-miR822. Recent study shows that the functions of light-responsive genes in the shoot apaxes and cotyledons of Arabidopsis young seedlings are highly enriched in zinc ion binding activities (Lopez-Juez et al., 2008). Notably, the abovementioned DC1 domain-containing proteins possess zinc binding activity based on the TAIR annotations (Huala et al., 2001). Certain targets of ath-miR156d (AT2G21840) and ath-miR824 (AT4G04885) also have zinc binding activities.

Hence, ath-miR156d, ath-miR824, and ath-miR822 may be involved in light response pathways during seedling development.

### MicroRNAs involved in leaf development

Previous studies showed that certain osa-miR169 family members were induced by either drought or high salinity treatment. However, there were both overlapping and specific responses of osa-miR169 members to the drought and salt stress (Zhao et al., 2009; Zhao et al., 2007). Our target prediction results suggest that a dominant portion of osa-miR169 targets belong to the nuclear factor Y (NF-Y)
family. One predicted target (LOC_Os03g29760) was validated by a recent study (Zhao et al., 2009). Interestingly, the expression of several NF-Y family members appeared to be responsive to drought stress in the leaves of wheat (Triticum aestivum) seedlings, suggesting their involvement in drought adaptation (Stephenson et al., 2007). Considering the close evolutionary relationship between wheat and rice, and the high expression levels of osa-miR169g, osa-miR169m, and osa-miR169n in leaves, potential role of osa-miR169 in leaf development was proposed here (Fig. 1 F). Moreover, both ath-miR168a and osa-miR168a were constitutively expressed, but with higher levels in the leaves. Previous study demonstrated that ath-miR168 was implicated in leaf development (Vaucheret et al., 2006). Hence, the mechanism of miR168-mediated regulation of leaf development may be conserved in plants.

Materials and methods

Data collection

Small RNA high-throughput sequencing data sets of Arabidopsis (GEO series number: GSE5228) and rice (GSE16350), which were generated by the 454 sequencing technology, were retrieved from Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) (Barrett et al., 2009) (see Table S1 for details). Only the HTS data sets belonging to one GEO series (normally, one series was contributed by one research project) were utilized for each plant to ensure the uniformity of the experimental procedure, thus increasing the reliability of cross-library (i.e. cross-organ) comparison. There are four sRNA libraries prepared from four different organs of Arabidopsis including six-day-old seedling, rosette leaf, flower and siliques, and 12 libraries containing three biological replications prepared from four different organs of rice including shoot apex, leaf, root apex and inflorescence. The whole genome sequences of Arabidopsis and rice were obtained from the Arabidopsis information resource (TAIR release 9, ftp://ftp.arabidopsis.org/home/tair/Sequences/) (Huala et al., 2001), and the rice genome annotation project established by the institute for genome research (currently named as the J. Craig Venter institute) (TIGR rice genome release 6.1, ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryota_c_Projects/o_sativa/annotation_dbs/) (Yuan et al., 2003). The mature miRNA sequences were retrieved from miRBase (release 15, http://www.mirbase.org/index.shtml) (Griffiths-Jones et al., 2008).

Sequencing data parsing

For each sRNA HTS data set, the short read sequences with “0” raw read counts (i.e. the ones without detectable expression levels by using the 454 sequencing technology), and those containing ambiguous nucleotides “N” were removed. Then, expression level normalization was performed for each data set to enable cross-library (i.e. cross-organ) comparison. For each data set, the normalized read count of a short sequence was calculated by dividing the raw read count of this short sequence by the total read counts of all the sequences within the data set, and then multiplied by 106. The expression levels of all the short reads were measured by RPM (reads per million) after normalization. There are three biological replications for each rice organ used for sRNA HTS library preparation. In this case, the normalization was performed for each replication, and then the average RPM values of the three replications were calculated to serve as the final expression levels for subsequent analyses. To increase the reliability of this expression-based analysis, only the intersection among the three replications (i.e. the sRNAs could be detected by all the replications) were utilized to search for the organ-specific sRNAs. The normalized HTS data sets were then subjected to expression pattern and chromosome-wide distribution analyses.

Conclusion

By utilizing the publicly available sRNA HTS data, a systemic search for the organ-specific sRNAs in Arabidopsis and rice was performed. Both the sequence length and the chromosome-wide distribution patterns were analyzed for the extracted organ-specific sRNAs. Based on the organ-specific expression patterns and literature mining, potential regulatory roles of several organ-specific miRNAs in plant organ development were uncovered. However, the biological implications of numerous organ-specific sRNAs identified in this study need further investigation. Taken together, our study provided a basic list of organ-specifically expressed sRNAs for the following functional studies on plant sRNAs.

Acknowledgements

We would like to thank all the publicly available data sets and the scientists behind them. This work was funded by the National Natural Sciences Foundation of China [31100937] and [30971743], the Starting Grant funded by Hangzhou Normal University to Yijun Meng [2011QDL60], and Zhejiang Provincial Natural Science Foundation of China [Y5110136].

Supplementary materials are available online.

References


