Identification of the highly accumulated microRNA*s in Arabidopsis (Arabidopsis thaliana) and rice (Oryza sativa)

Chaogang Shao a,b,*, Xiaoxia Ma a, Xiufang Xu a, Yijun Meng b,***

a College of Life Sciences, Huzhou Teachers College, Huzhou 313000, PR China
b College of Life and Environmental Sciences, Hangzhou Normal University, Hangzhou 310036, PR China

ABSTRACT

Plant microRNAs (miRNAs) are crucial for the regulation of gene expression, which is involved in almost all the important biological processes. In the cytoplasm, the miRNA strand is selectively incorporated into a specific Argonaute (AGO)-associated gene silencing complex, while the miRNA* is degraded rapidly. Thus, most miRNA*s were thought to be biologically meaningless. Interestingly, several recent reports in both plants and animals have shaken this notion. Many miRNA*s were demonstrated to possess regulatory roles in gene expression. However, the low accumulation levels of most miRNA*s raise the question whether the activities of this small RNA (sRNA) species are widespread in plants. Here, by using publicly available sRNA high-throughput sequencing data, we found that the accumulation levels of several miRNA*s could be much higher than those of their miRNA partners in certain organs, mutants and/or AGO-associated silencing complexes of both Arabidopsis (Arabidopsis thaliana) and rice (Oryza sativa). Based on target prediction and degradome sequencing data-based validation, some of these highly accumulated miRNA*s were indicated to possess cleavage-based potential regulatory role on certain targets. Besides, some interesting biological interpretations were obtained based on the accumulation patterns of the miRNA*s, the annotations of the target genes, and literature mining. Taken together, the expanded list of the highly accumulated miRNA*s along with their potential target genes discovered in this study further strengthened the current notion that certain members of the miRNA* species are biologically relevant, which needs further inspection.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

MicroRNA (miRNA)-mediated gene regulation is widespread in the organisms, and was demonstrated to be essential or even indispensable for numerous biological processes during plant/animal growth and development (Carthew and Sontheimer, 2009; Chen, 2009; Jones-Rhoades et al., 2006). In plants, the primary transcript of a miRNA gene, i.e. pri-miRNA (primary microRNA), is processed into stem–loop structured (also known as the hairpin structure) pre-miRNA (precursor microRNA), and then into the miRNA/miRNA* duplex with 2′-nt 3′ overhangs at both ends (Jones-Rhoades et al., 2006; Voinnet, 2009). This procedure relies on the RNase III activity of the Dicer-like 1 (DCL1) protein. The miRNA/miRNA* duplex is then exported to the cytoplasm, and the miRNA strand is selectively loaded into a specific Argonaute (AGO)-associated (AGO1 in most cases) gene silencing complex, called miRISC (microRNA-induced silencing complex). Thus, the short miRNA strand is stabilized in the miRISC, and could guide the silencing complex to the target transcript containing a highly complementary binding site. As a result, the target gene could be post-transcriptionally regulated through miRNA-mediated target cleavages. On the other hand, the miRNA*, also defined as the passenger strand, is subjected to rapid degradation, which leads to a much lower accumulation level compared to that of the miRNA (Jones-Rhoades et al., 2006; Voinnet, 2009). In this regard, the miRNA* species were thought to be biologically meaningless, and their activities have not been recognized for a long time.

Until recently, several studies in both plants and animals unraveled the fact that some miRNA*s could modulate the expression of the target genes post-transcriptionally by loading into specific AGO complexes (Devers et al., 2011; Ghildiyal et al., 2010; Meng et al., 2011b; Okamura et al., 2008; Packer et al., 2008; Yang et al., 2011; Zhang et al., 2011). In plants, several examples of miRNA*-mediated target cleavages have been provided (Devers et al., 2011; Meng et al., 2011b; Zhang et al., 2011). However, we should acknowledge that most...
miRNA*s are not stable enough to reach high accumulation levels as the miRNAs, since they are not preferentially recruited by the AGO-associated silencing complexes in vivo. This raises the questions how widespread the regulatory activities of the miRNA*s are in plants, and whether the miRNA*s are accumulated and activated in certain organs or under certain circumstances.

The recently developed high-throughput sequencing (HTS) technology is a powerful tool to investigate the accumulation levels of the small RNAs (sRNAs) (Ozsolak and Milos, 2010). In this study, the valuable sRNA HTS data sets retrieved from public databases were utilized to search for the highly accumulated miRNA*s in Arabidopsis (Arabidopsis thaliana) and rice (Oryza sativa). As a result, several miRNA*s were found to be accumulated at much higher levels than the corresponding miRNAs. However, the high accumulation was discovered in certain organs, plant mutants and/or AGO-associated silencing complexes. For example, ath-miR172d*, ath-miR172e*, ath-miR393b* and ath-miR396b* were highly accumulated in the ago1 mutant of Arabidopsis, indicating the novel role of AGO1 in destabilization of these miRNA*s. Certain miRNA*s were observed to be highly accumulated in certain biological samples, such as osa-miR169g* and osa-miR169p* in the rice seedlings, osa-miR396a* in the rice roots, and osa-miR1433* in the panicles. Some miRNA*s, such as ath-miR393b* and osa-miR810a*, were highly accumulated in various organs/tissues. For the miRNA*s enriched in the AGO4 protein complexes, a large portion of their predicted target genes were annotated to encode transposable elements (TEs), which well correlates with the biological role of the AGO4-associated sRNAs in TE silencing through chromosome modifications. Besides, a large-scale prediction was performed to extract the potential target transcripts of the highly accumulated miRNA*s. Degradome sequencing data-based validation enabled us to obtain a list of miRNA*—target pairs in Arabidopsis and rice, which is useful for further functional studies. Taken together, our results demonstrated that the accumulation levels of certain plant miRNA*s could be elevated in certain organs, mutants and/or AGO-associated silencing complexes, and might be highly correlated with their biological functions which need further experimental investigations.

2. Results and discussion

2.1. Search for the highly accumulated miRNA* species in Arabidopsis and rice

A number of sRNA HTS data sets of Arabidopsis and rice were retrieved from the public databases including Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) (Barrett et al., 2009), Next-Gen Sequence Databases (http://mpss.udel.edu/) (Nakano et al., 2006), and Cereal Small RNAs Database (CSRDDB; http://sundarlab.ucdavis.edu/smrnas/) (Johnson et al., 2007) (see detailed information of the HTS data sets in Tables S1 and S2). In order to allow cross-library comparison, the normalized read count (in RPM, reads per million) of a short sequence from a specific sRNA library was calculated by dividing the raw count of this sequence by the total counts of the library, and then multiplied by 106. Then, by using the BLAST algorithm (Altschul et al., 1990), all the sRNA sequences were mapped to the pre-miRNA*s retrieved from miRbase (http://www.mirbase.org/; release 18) (Griffith-Jones et al., 2008), and only the perfectly matched sRNAs were retained for further analysis.

To extract the highly accumulated microRNA* candidates, the following criteria were set for filtering the perfectly matched sRNAs obtained above: (1) Comparing to the levels of the corresponding mature miRNAs on the pre-miRNAs, the accumulation levels of the miRNA*s should be ten times or more, and should be 10 RPM or higher. (2) For the cases that the mature miRNAs were not detected by HTS, i.e. the accumulation levels of the miRNAs are 0 RPM based on the sRNA HTS data, the levels of the corresponding miRNA*s should be 10 RPM or higher. As a result, 28 miRNA*s in Arabidopsis and 18 in rice were observed to accumulate much higher than their miRNA partners (Tables S3 and S4), which indicates that the high accumulation of the miRNA*s is not an occasional phenomenon in plants. To further validate that the extracted highly accumulated sRNAs correctly located within the miRNA*—coding regions on the pre-miRNAs, secondary structure-based examination was performed. The structures of the pre-miRNAs were predicted by RNAshapes under the mode “Shape folding” with default parameters (Steffen et al., 2006). The miRBase-registered mature miRNAs and the highly accumulated sRNAs obtained in this study were marked on the stem-loop-structured pre-miRNAs. Manual examination showed that all of the extracted highly accumulated sRNAs resided within the stem regions of the hairpin structures and highly overlapped with the mature miRNAs, and most miRNA/miRNA* duplexes possessed 2-3’ overhangs at both ends (see examples in Figs. 1A to K). This result supports that the highly accumulated sRNAs are the true miRNA*s.

2.2. Certain miRNA*s were highly accumulated in certain organs, mutants and/or AGO-associated silencing complexes

Based on the descriptions of the sRNA HTS libraries, the high accumulation of the miRNA*s was specifically discovered in certain organs, mutants and/or AGO-associated silencing complexes. For example, in the ago1 mutant of Arabidopsis, the accumulation levels of some miRNA*s, including ath-miR158a*, ath-miR171c*, ath-miR172d*, ath-miR172e*, ath-miR396b*, ath-miR396e*, ath-miR402* and ath-miR434*, are much higher than those of the mature miRNAs (see tables of Fig. 1, and Table S3). This observation raises the possibility that the AGO1 protein is not only essential for the stability and the activity of the mature miRNAs, but also plays an important role in the degradation of certain miRNA*s. In rice, osa-miR169g* and osa-miR169p* were highly accumulated in the seedlings (Fig. 1). Along with the accumulation patterns of ath-miR172d* and ath-miR172e* described above, we deduced that the miRNA*s belonging to the same miRNA family might share a common accumulation pattern. Furthermore, osa-miR396a*, osa-miR396b* and osa-miR2121a* were highly accumulated in the rice roots, and osa-miR5159* was accumulated in the shoot apexes (Table S3). osa-miR1428f* was highly accumulated in the inflorescences, and osa-miR1433* and osa-miR5533* were accumulated in the panicles (Table S3). Based on the literatures, some evidences were obtained to explain the observed accumulation patterns of certain miRNA*s. In the previous study on the expression of the miRNAs in superior and inferior spikelets in rice, the high accumulation of osa-miR1433* was also observed, indicating its potential role in rice flowering and reproduction (Peng et al., 2011). More interestingly, miR396 was demonstrated to be involved in the reprogramming of root cells during cyst nematode infection in Arabidopsis (Hewezi et al., 2012). The exclusive high-accumulation pattern of miR396* in the rice roots indicates its potential role in rice root development. However, the biological relevance of the organ-specific accumulation patterns of the miRNA*s still needs further interpretations. The specific accumulation patterns do not apply to all the miRNA*s uncovered in this study. For instance, ath-miR393b*, osa-miR810a* and osa-miR1433* tend to be highly expressed in various biological samples investigated in this study (Table S3), indicating their widespread activities in the plants.

2.3. Biological insights gained from identification of the miRNA* targets

To further investigate the biological activities of the highly accumulated miRNA*s, a large-scale target prediction was performed by employing the miRU algorithm (Dai and Zhao, 2011; Zhang, 2005). Degradome sequencing combining modified 5’ rapid amplification of cDNA ends (RACE) with HTS is quite efficient for target identification and slicing site mapping (Addo-Quaye et al., 2008; German et al., 2008, 2009; Li et al., 2010; Pantaleo et al., 2010). By utilizing the publicly available degradome sequencing data, the predicted target...
transcripts were subjected to strict validation through target plot (t-plot)-based approach as described previously (see detailed procedure in the Materials and methods section) (German et al., 2008, 2009; Meng et al., 2011a). As a result, one miRNA*—target pair in Arabidopsis involving ath-miR173* and 18 pairs in rice involving osa-miR169g*, osa-miR1433* and osa-miR2878* were discovered (Figs. 2 and S1). These results indicate that these miRNA*s might regulate specific target transcripts through cleavages, which need further experimental validations. Since limited degradome data sets were available for this survey, the cleavage-based activities of the remaining portion of the highly accumulated miRNA*s discovered in this study could not be excluded.
We also observed that certain miRNA*s, including ath-miR168a*, ath-miR169a*, ath-miR173*, ath-miR393b*, ath-miR396a* and ath-miR396b* in Arabidopsis, and osa-miR1423* and osa-miR1882e* in rice, were highly enriched in AGO4-associated silencing complexes (see accumulation levels in Tables S3 and S4). The recent report shows that the miRNAs associated with AGO4 clade proteins could mediate DNA methylation in rice (Wu et al., 2010). In this regard, we examined the predicted targets of these AGO4-enriched miRNA*s. Based on the gene annotations retrieved from the Arabidopsis information resource (TAIR, release 10) (Huala et al., 2001) and the rice genome annotation project established by the institute for genome research (TIGR rice, release 6.1) (Yuan et al., 2003), a large portion of the target genes encode transposable elements in both plants (Data S1 and S2). Considering the notion that the activities of the TEVs are under strict control by plant endogenous sRNAs through chromosome modification (Bender, 2004; Kasschau et al., 2007; Li et al., 2008; Lister et al., 2008; Paszkowski and Whitham, 2001; Piriyapongsa and Jordan, 2008; Zhang et al., 2006; Zilberman et al., 2007), we proposed that the AGO4-enriched miRNA*s might be implicated in the repression of the TEs in the plants.

Taken together, our study on the highly accumulated miRNA*s demonstrates that except for some miRNA*s such as ath-miR393b*, osa-miR810a* and osa-miR1433*, most miRNA*s investigated accumulated at higher levels than the mature miRNAs in certain biological samples. The biological relevance of the specific accumulation patterns (e.g. organ-specific) of these miRNA*s needs further interrogation. Target prediction and degradome sequencing data-based validation indicate that the highly accumulated miRNA*s possess potential regulatory roles in gene expression either through target cleavages or through chromosome modifications.

3. Materials and methods

3.1. Data resources used in this study

The sequence information of the mature miRNAs and the pre-miRNAs of Arabidopsis and rice were retrieved from miBase (http://www.mirbase.org/; release 18) (Griffiths-Jones et al., 2008). The sRNA high-throughput sequencing data sets were downloaded from Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) (Barrett et al., 2009), Next-Gen Sequence Databases (http://mpss.udel.edu/) (Nakano et al., 2006), and Cereal Small RNAs Database (CSRDB; http://sundarlab.ucdavis.edu/smrnas/) (Johnson et al., 2007). Please see Tables S1 and S2 for detailed information. The Arabidopsis degradome sequencing data sets GSM278332, GSM278334, GSM278335 and GSM278370 were retrieved from GEO, and AxIDT, AxIRP, AxSRP, Col,

![Image](46x357 to 540x741)
3.2. Target prediction and degradome sequencing data-based validation

Target prediction was performed by using miRUV algorithm (Dai and Zhao, 2011; Zhang, 2005) with default parameters. The degradome sequencing data were utilized to validate the predicted miRNA*–target pairs. First, in order to allow cross-library comparison, the normalized read count (in RPM, reads per million) of a short sequence from a specific degradome library was calculated by dividing the raw count of this sequence by the total counts of the library, and then multiplied by 10^6. Then, two-step filtering was performed to extract the most likely miRNA*–target interactions. During the first step, the predicted miRNA* binding sites along with the 50-nt surrounding sequences at both ends were collected in order to reduce the BLAST (Altschul et al., 1990) time. For the BLAST, all the collected degradome data sets were utilized at the same time to do a comprehensive search. It was based on the scenario that a miRNA*–target pair was considered to be the candidate once the cleavage signal(s) existed in any data set(s). The target transcripts met the following criteria that were retained: (1) there must be perfectly matched degradome signatures with five′ ends resided within 8–14 nt region away from the 5′ ends of the target binding sites; and (2) for a specific position within the 8–14 nt region, which could be regarded as the potential cleavage site, there must be two or more distinct degradome signatures with accordant 5′ ends to support this position. These retained transcripts were subjected to a second BLAST, and the degradome signals along each cDNA were obtained to provide a global view of the signal noise when compared to the signal intensity within a specific target binding site. Referring to our previous study (Meng et al., 2011a), the target plots (t-plots) (German et al., 2008, 2009) were drawn for the subsequent manual filtering. Only the targets with cleavage signals easy to be recognized were extracted as the potential miRNA*–target pairs.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ijpe.2012.11.015.

Acknowledgments

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 31271380], the Starting Grant funded by Hangzhou Normal University to Yijun Meng [2011QDL60], Zhejiang Provincial Natural Science Foundation of China [YS110136], and Scientific Research Fund of Zhejiang Provincial Education Department [Y201222905].

References


