Characterization and expression profiles of miRNAs in rice seeds

Liang-Jiao Xue, Jing-Jing Zhang and Hong-Wei Xue*

National Key Laboratory of Plant Molecular Genetics, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 300, Fenglin Road, 200032 Shanghai, China

Received July 21, 2008; Revised November 25, 2008; Accepted November 27, 2008

ABSTRACT
Small RNAs (sRNAs) are common and effective modulators of gene expression in eukaryotic organisms. To characterize the sRNAs expressed during rice seed development, massively parallel signature sequencing (MPSS) was performed, resulting in the obtainment of 797 399 22-nt sequence signatures, of which 111 161 are distinct ones. Analysis on the distributions of sRNAs on chromosomes showed that most sRNAs originate from interspersed repeats that mainly consist of transposable elements, suggesting the major function of sRNAs in rice seeds is transposon silencing. Through integrative analysis, 26 novel miRNAs and 12 miRNA candidates were identified. Further analysis on the expression profiles of the known and novel miRNAs through hybridizing the generated chips revealed that most miRNAs were expressed preferentially in one or two rice tissues. Detailed comparison of the expression patterns of miRNAs and corresponding target genes revealed the negative correlation between them, while few of them are positively correlated. In addition, differential accumulations of miRNAs and corresponding miRNA’s suggest the functions of miRNA’s other than being passenger strands of mature miRNAs, and in regulating the miRNA functions.

INTRODUCTION
Small RNA (sRNA) molecules are widely recognized as common and effective modulators of gene expression in many eukaryotic organisms (1–3). According to the present knowledge, sRNAs are generally divided into several categories, including microRNAs (miRNAs), short-interfering RNAs (siRNAs), trans-acting siRNAs (ta-siRNAs; 4), natural antisense transcript siRNAs (nat-siRNAs; 5) and Piwi-interacting RNAs (piRNAs) in metazoans (6). In plants, ta-siRNAs and nat-siRNAs guide the cleavage of target mRNAs, whereas heterochromatic siRNAs are involved in DNA and histone modification, leading to the silencing of gene transcription (7). In addition to nat-siRNAs, rice natural antisense miRNAs (nat-miRNAs) were recently identified (8).

miRNAs (~21 nt) are produced from noncoding imperfectly complementary, stem–loop RNA precursors. The miRNA genes can be transcribed by RNA polymerase II (9) to form the primary miRNAs (pri-miRNAs). In higher plants, pri-miRNAs are processed by dicer-like 1 (DCL1; 10) or DCL4 (11) to form miRNA duplexes through multiple cleavage steps, and then miRNA duplexes are methylated by Hua Enhancer1 (HEN1) at the 3′-end (12) and loaded onto Argonaute 1 (AGO1) where the mature miRNAs guide the cleavage of target mRNAs. In the meanwhile, the passenger strands (referred as miRNA’s) are gradually degraded (13).

Although the first miRNA (lin-4) was discovered by forward genetic screening in Caenorhabditis elegans (14), the majority of currently known plant miRNAs were identified by size-selected cloning and sequencing, especially those in Arabidopsis (15–18) and rice (19–21). Recently developed high-throughput sequencing strategies have expanded the depth of sRNA cloning coverage. In Arabidopsis, analysis through massively parallel signature sequencing (MPSS) identified 67 528 and 27 833 unique 17-nt sequences from inflorescences and seedlings, respectively (22). More than 340 000 unique sRNAs from Arabidopsis seedlings, rosette leaves, flowers and siliques were sequenced using pyrophosphate-based high-throughput sequencing technique (11), and 48 new miRNAs were identified with similar strategy (23). In rice, ~20 miRNAs were identified by large-scale sequencing of sRNAs in panicles, seedlings and stems (8,24–26).

Several guidelines have been proposed for miRNAs annotation (27). The miRNA precursors should contain stable and conserved stem–loop structures that can be predicted by Mfold (28), and mature miRNAs should be detected by northern blotting or sequencing. In addition, as miRNA genes are transcribed by RNA polymerase II,
capped and polyadenylated as normal mRNAs (9), EST analysis is a powerful approach to identify the new miRNAs (29). Identification of a miRNA* sequence, a product of Dicer cleavage corresponding to miRNA (11), also strongly indicates that the corresponding sRNA molecule was indeed processed by Dicer-like RNase III enzyme (11,23,30).

Rice is an important food resource for human daily life and serves as model species of monocotyledon plants. Development and maturation of rice seed, a highly specialized organ of nutrient storage and reproductive development, involve meticulous and fine gene regulations at transcriptional and post-transcriptional levels (31). To further study the complicated regulatory network of rice seed development, and to elucidate the functions of sRNAs during this process, MPSS and integrated bioinformatics analysis were performed, resulting in the identification of novel and candidate miRNAs. Further, expression profiles of miRNAs were analyzed through miRNA microarray hybridization, which have been widely used to study the miRNA expression levels in several species (32-35). Comparison of expression patterns revealed the positive or negative correlations between miRNAs and the corresponding target genes, which greatly expand the understanding of how miRNAs were involved in rice seed development.

MATERIALS AND METHODS

cDNA library construction and MPSS analysis

Rice (Oryza sativa, japonica cultivar Zhonghua 11) immature seeds at 3, 6, 9 and 12 days after anthesis (DAA) were collected and total RNAs were extracted separately using Trizol Reagent (Invitrogen, Nottingham, UK), which were then mixed with equal amounts from individual samples. After separation on denatured polyacrylamide gel electrophoresis (PAGE), the fraction of 18-26-nt small RNAs (sRNAs) was recovered by sRNAs gel extraction Kit (TaKaRa Bio, Otsu, Japan).

The cDNA library construction and MPSS analysis were carried out by TaKaRa Bio Inc. (Otsu, Japan) as described previously (36). In brief, 20 ng of sRNAs fraction was dephosphorylated and ligated with RNA-DNA chimeric 3’-adapter. After purification by denatured PAGE, the RNA fraction with attached adapters was phosphorylated and ligated with RNA-DNA chimeric 5’-adapter. Resultant products were reversely transcribed and amplified by PCR with primers (with a SfaNl site). Then the amplified PCR products were digested by SfaNl and cloned into the Tag vector pMBS I (Solexa). The resultant library, which has more than 1.3 million total complexity, was amplified and loaded onto microbeads. Finally, more than 1 million microbeads were loaded into each flowcell and the signature sequences of 22 bases were determined by a series of enzymatic reactions.

The adaptor sequences of signatures were masked. After removing signatures corresponding to ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs), the total abundances of signatures were around 250,000.

sRNAs were normalized to transcripts per quarter million (TPQ) as previously described (22,36). The signatures and abundances of sRNAs in rice seed library can be found in National Center for Biotechnology Information, Gene Expression Omnibus (NCBI, GEO) under series number GSE11974.

Data sets

Genome sequences and annotations of rice (version 5.0) and assembled Zea mays sequence (AZMS) were obtained from TIGR (the Institute for Genomic Research). Sequences of rRNAs, tRNAs, snRNAs and snoRNAs were downloaded from databases including the European ribosomal RNA database (http://www.psb.ugent.be/rRNA/, for rRNA), the Genomic tRNA database (http://lowelab.ucsc.edu/GtRNAdb/, for tRNA) and NONCODE (http://www.bioinfo.org.cn/NONCODE/, for snRNAs and snoRNAs). Mature miRNAs and annotated stem–loop sequences were obtained from miRBase (versions 10.0 and 11.0, http://microrna.sanger.ac.uk/; 37). sRNAs sequences of rice, Arabidopsis and C. elegans were downloaded from rice MPSS database (http://mpss.udel.edu/rice/), Arabidopsis Small RNA Project (ASRP, http://asrp.cgrb.oregonstate.edu/) and GenBank data libraries (GEO accession number GSE5990, sample GSM139137), respectively.

Identification of sRNA clusters and hotspots

The sRNAs were grouped into clusters dependent on their locations on the genome as described previously, i.e. sRNAs within 500 bp of each other were fallen under a cluster (22). To identify the sRNA hotspots, abundance of each signature was firstly normalized by hitting times of signature on the genome, and then the sums of abundances of all signatures in no overlapping 500-bp windows were calculated. The top-ranking windows were used as seeds for extension in both directions until a window hits no signatures (11).

Predictions of miRNAs and corresponding mRNA targets

All the known rice miRNAs, whose precursors contain no repetitive sequences, matched genome for <30 times. Our analysis on the obtained signatures that matched genome for more than 30 times indicated that 72.9% of them (3470 out of 4760) originated from repetitive sequences (TIGR Oryza Repeat Database v3.3). Signatures matched genome for more than 30 times were thus filtered out during miRNA prediction and those corresponding to rRNAs, tRNA, snRNAs and snoRNAs were eliminated. The 400-bp genome sequences on each side of these signatures were extracted and a sliding window of 450 bp with an increment of five bases was scanned along the extracted sequences. All the fragment sequences containing each signature were folded with the help of MFOLD3.2 (28). Structures of the sequences with minimum energy were further analyzed using a Perl script to check whether these structures satisfy the criteria for majority of known plant miRNAs (38). The sequences of candidate precursors were analyzed using RepeatMasker (http://www.repeatmasker.org) to...
eliminate the repetitive sequences. Finally, signatures whose predicted secondary structures passed the in silico check were grouped by genomic positions. The candidates with miRNA* and/or EST supporting evidences were identified as miRNAs.

To predict the target sites of miRNAs, FASTA Program (39) was used to search the rice cDNA database (TIGR, version 5.0) with +15/−10 scoring matrix. Obtained reverse complement results (maximal 100 000) of each miRNA were used to calculate the mispair scores and Minimum Free Energy (MFE) ratios. In brief, the MFE (ΔGMFE) of a hypothetical duplex containing miRNA target paired with a perfectly complementary sequence and free energy of each actual miRNA-predicted target duplex (ΔGtarget) were determined respectively. MFE ratio was calculated as ΔGtarget/ΔGMFE (4). The mispair score and MFE ratio tolerance limits to ≤4 and ≥0.73 were assigned, and the match up of all miRNAs and predicted targets were checked manually. All the Perl scripts used in miRNA prediction and target prediction were available upon request.

miRNA chip generation and hybridization

The miRNAs in miRBase version 8.0, newly identified ones and other candidates in this study, were collected to design the probes. Independent sequences (polyT18 or polyT10) were attached to the full-length miRNA probes at 3'-end to acquire the final 40-nt probes so that they have appropriate extension space at the chip surface. The probes were then attached to the activated slide surface via a C6 5'-amino-modifier (35). Probes for tRNA and U6 RNA were designed and used as internal positive controls. In addition, eight DNA sequences having no homology with any sRNAs were designed and used as external positive controls. All probes were synthesized at MWG Company (Ebersberg, Germany), and dissolved into DNA spotting solution (40 µmol/l), which was spotted onto aldehyde-decorated slides for three repeated dots.

Labeling of sRNAs and chip hybridization were performed as described previously (35). In detail, 50–100 µg of total RNAs from different samples (embryo, endosperm, leaf and root of 7-day-old seedlings, 10-day-old seedling) were size-fractionated using Ambion’s miRNA Isolation Kit (Cat#:1560) and labeled with T4 polymerase (32). RNA was dissolved in hybridization buffer containing 15% formamide, 0.2% SDS, 3x SSC and 50 µg/ml Denhardt’s solution. After overnight hybridization at 42°C, slides were washed in 0.2% SDS, 2x SSC for 4 min and 0.2x SSC for another 4 min. Hybridized slides were scanned with CapitalBio® LuxScan™ 10K/A (Beijing, China) and analyzed with the GenePix Pro 4.0 (Axon Instruments, Union City, CA). The maximal signal level of background probes was 200, and 800 was set as detection signal threshold of miRNAs.

Hybridization signals were normalized using medium center method. For each tissue (except embryo) two hybridizations were performed, thus six hybridization signals were finally obtained for each miRNA from one tissue (three signals for embryo). As the immature embryo tissues were difficult to collect and considering that there were three repeat dots of probes on the chip, only one hybridization was performed. At last, mean of hybridization signals was calculated and visualized using TIGR Mev (version 3.1, http://www.tm4.org/mev.html).

Gene expression analysis through microarray hybridization

GeneChip® rice genome array (Affymetrix, Santa Clara, CA) was used for DNA microarray analysis. The RNA samples were same as those for miRNA microarray hybridization (total RNAs were fractioned, the fraction containing large or small RNA molecules were used for DNA and miRNA microarray analysis, respectively). Each sample has two biological repeats, and 10-µg cRNA was used in each hybridization. Washing, staining and scanning of chips were performed according to supplier’s protocol. Hybridization signals were normalized using the Affymetrix Microarray Suit program (version 5.0) and visualized using TIGR Mev. The microarray data of Affymetrix arrays are under series number GSE1966 in the database of NCBI, GEO. To check the replicate quality of microarray, least-square linear regression for each tissue was carried out (40) (Table S1).

Expression analysis of mature miRNAs through quantitative real-time RT-PCR (qPCR)

Total RNAs were isolated from various tissues (same as those used in miRNA microarray hybridizations), treated with RNase-free DNaseI (MBI Fermentas, Lithuania), and then polyadenylated (2 µg) by poly(A) polymerase (NEB, USA) at 37°C for 1 h in a 20-µl reaction mixture. After phenol–chloroform extraction and ethanol precipitation, the RNAs were dissolved and reversely transcribed using poly(T) adapter (41). qPCR was performed using SYBR® Green Realtime PCR Master Mix (Toyobo, Japan) and all the primers used were listed in Table S2. The values of threshold cycle (Ct, the fractional cycle number at which the fluorescence passes the fixed threshold), were calculated by Rotor-Gene 6 software (Corbett Robotics, Australia). Ct values were converted into relative copy numbers using a standard curve (42). As reference gene, 5.8S rRNA was used in qPCR detection of miRNAs in Arabidopsis previously (41). To check whether rice 5.8S rRNA is suitable, we firstly surveyed the expression level of 5.8S rRNA and some other suggested reference genes (43) in GeneChip® rice genome array, and the results showed that the expression levels of these genes were relative stable (Supplementary Figure 1A). Further, expression level of 5.8S rRNA was characterized by qPCR using rice Actin gene (Os03g50890) as reference gene (43), and results confirmed the stable expression of 5.8S rRNA in the samples used in the study (Supplementary Figure 1B). These indicated that it is suitable using rice 5.8S rRNA to normalize the qPCR data.

In addition, reaction efficiencies of qPCR assays for individual miRNAs were determined using Rotor-Gene 6 software. A 10-fold dilution series of purified product of the tested miRNA (or guide genes) were used as the template for qPCR to generate a plot of Log copy numbers of the tested miRNA (or guide genes) versus the
corresponding $C_T$ as described (41). The slope of the linear plot is defined as $-1/\log (E + 1)$, where $E$ is the reaction efficiency. In Rotor-Gene 6 software, reaction efficiency, whose optimal value is 1, is defined as amplification value $-1$. The results showed that the reaction efficiency of tested miRNAs approach the optimal value (Supplementary Figure 2).

**Sequence alignment and phylogenetic analysis**

Precursor sequences of miR169 and miR396 families were aligned with ClustalW (http://www.ebi.ac.uk/Tools/clustalw2; 44), and the conserved fragments were visualized with the help of GeneDoc program (http://www.nrbsc.org/gfx/genedoc). To construct the phylogenetic tree of AGO proteins, the protein sequences were aligned using ClustalW (44), and MEGA4 (45) was employed to generate the phylogenetic tree using neighbor-joining method (46) and Poisson correction method (47). The bootstrap test of 1000 replications was performed.

**RESULTS**

**sRNAs from rice seeds matched classes of genomic features**

sRNAs extracted from rice seeds at different developmental stages were used to isolate short RNAs, which were then used to construct the library. Subsequent massively parallel signature sequencing (MPSS) analysis resulted in the obtainment of 797,399 22-nt sequences (or 'signatures') after removing low-quality ones, and 111,161 of them were distinct (Figure 1A). Analysis showed that 70.5% of distinct signatures matched the rice genome (the Institute for Genomic Research, TIGR, version 5.0), representing 93.8% of the total signatures (Table S3). There was only a small part of signatures (2.3%) that matched the chloroplast or mitochondria genome (Table S3). Around 9.7% of the distinct signatures matched noncoding RNAs (ncRNAs) including ribosomal rRNAs, tRNAs, snoRNAs and snRNAs, which account for 61.5% of the total signatures (Table S3), similar to the previous report in rice (21). All the genome-matching signatures excluding ncRNAs were then regarded as sRNAs and used for further analysis.

In rice seeds, signatures matching miRNA precursors account for $<1.5\%$ of the total abundance of genome-matching signatures, which is much lower than that in rice inoforescences, seedlings and stems (24), and Arabidopsis flowers and seedlings (22). This may because rice genome contains many interspersed and inverted repeat sequences, which produce abundant siRNAs (24). Additionally, 15.6% of the distinct sRNAs match protein-coding genes in sense and/or antisense orientation (Table 1), and the corresponding encoded proteins involve in various molecular functions, biological processes and cellular components, suggesting the complexity of sRNAs’ origin (Table S4).

In addition, 70.7% of the genome-matching sRNAs locate on genome only once (the location numbers account for 4.5% of the total locations), while $\sim3\%$ sRNAs match the genome for more than 100 times

---

**Figure 1.** Analysis approaches and distributions of small RNAs in rice seeds. (A) A flow diagram showing the analysis process of sRNAs in rice seeds. Obtained signatures through MPSS were compared to RNAs from various RNA databases and those do not match the known RNAs were aligned on the genome to extract the flanking sequences for miRNA prediction. (B) Distributions of sRNAs on rice chromosome 1. The $x$-axis indicates the position on chromosome 1 (Mb). The left $y$-axis indicates the length of expressed genes (blue line) and repetitive sequence (pink line) per 200 kb of the chromosome. The right $y$-axis indicates the abundance (per 200 kb) of total sRNAs (black line) on the chromosome. Abundance of each signature was normalized by location numbers on the genome. Triangle indicates the position of centromere.

**Table 1.** Summary of signatures that match various RNAs

<table>
<thead>
<tr>
<th>Locus class</th>
<th>Distinct signatures</th>
<th>Total signatures</th>
<th>Mean frequency$^a$</th>
<th>Mean hits$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>miRNA precursors</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>snoRNA</td>
<td>20 (0.02%)</td>
<td>56 (0.01%)</td>
<td>2.8</td>
<td>2.1</td>
</tr>
<tr>
<td>snRNA</td>
<td>293 (0.26%)</td>
<td>5413 (0.68%)</td>
<td>18.47</td>
<td>10.08</td>
</tr>
<tr>
<td>rRNA</td>
<td>1133 (1.02%)</td>
<td>38404 (4.82%)</td>
<td>33.9</td>
<td>8.56</td>
</tr>
<tr>
<td>rRNA</td>
<td>9332 (8.40%)</td>
<td>446596 (56.01%)</td>
<td>47.86</td>
<td>5.52</td>
</tr>
<tr>
<td>Small RNAs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small RNAs matching</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>protein-coding genes</td>
<td>Sense</td>
<td>10429 (9.38%)</td>
<td>62041 (7.78%)</td>
<td>5.95</td>
</tr>
<tr>
<td></td>
<td>Antisense</td>
<td>5896 (5.30%)</td>
<td>20805 (2.61%)</td>
<td>3.53</td>
</tr>
<tr>
<td></td>
<td>Sense and</td>
<td>1032 (0.93%)</td>
<td>4885 (0.61%)</td>
<td>4.73</td>
</tr>
<tr>
<td></td>
<td>antisense</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>miRNA</td>
<td>Known</td>
<td>228 (0.21%)</td>
<td>4029 (0.50%)</td>
<td>17.67</td>
</tr>
<tr>
<td></td>
<td>Novel</td>
<td>328 (0.30%)</td>
<td>7359 (0.92%)</td>
<td>22.44</td>
</tr>
<tr>
<td>Other small RNAs</td>
<td>82470 (74.2%)</td>
<td>207811 (26.1%)</td>
<td>2.52</td>
<td>–</td>
</tr>
<tr>
<td>Total</td>
<td>111161 (100%)</td>
<td>797399 (100%)</td>
<td>7.17</td>
<td>–</td>
</tr>
</tbody>
</table>

$^a$The average frequency of sequenced signatures matching different type RNAs.

$^b$The average hit numbers on the chromosomes of signatures.

The mean hits of ‘other small RNAs’ and total small RNAs are not indicated as ‘-’.

Numbers in brackets indicate the percentage of the distinct signatures or total signatures.
(Table S5). Characterization of the distribution of sRNAs, protein-encoding genes and repetitive sequences on chromosome shows that some loci have high propensity to produce sRNAs, and these sRNA hotspots originated preferentially from the regions other than centromeres (Figure 1B, Supplementary Figure 3).

**Most sRNAs in rice seeds originated from transposable elements**

A total of 81,530 sRNA clusters were identified depending on the distance between every two signatures, and 72,173 of them contain interspersed repeats (annotated by RepeatMasker, http://www.repeatmasker.org; Table S6), 2967 were marked as tandem repeats (by Tandem repeats finder; 48), indicating that repeat-associated sRNAs (rasiRNAs) take up a large portion of sRNAs. The presence of highly abundant rasiRNAs in seeds is consistent with their roles in regulating transposon activity during seed development. Further identification of the top 50 hotspots, according to the normalized abundances of sRNAs, showed that tandem repeat loci inclined to generate sRNAs at high abundance. Nineteen of the top 50 hotspots contain tandem repeats, while only 3.6% of the total sRNA clusters were tandem repeats. This supported the generation mechanism of sRNAs from tandem repeats, proposed by Martienssen (49). In addition, some miRNAs including osa-MIR168a, osa-MIR172b, and osa-MIR159a were also found with high abundance in sRNAs hotspots (Table S7).

**Validation of registered miRNAs using sRNA signatures**

One of the most critical criteria for miRNA annotation is that the miRNA should originate from a single-strand RNA transcript (27). However, some miRNA precursor sequences of registered miRNAs matched the short signatures in both sense and antisense orientations when the sRNAs sequences were mapped onto the precursor sequences (version 11.0), suggesting that they might not be bona fide miRNAs. In addition, most loci of these miRNAs appear to be repetitive sequences. Among the 269 miRNA loci, 91 precursors were annotated as transposable element repetitive sequences, and most of them were miniature inverted-repeat transposable elements (MITEs, Table S8; 50).

sRNAs that match miRNA precursors can be grouped into several classes (Table S9). For bona fide miRNAs, most of the sRNAs should originate from the positions of mature miRNAs or miRNA’s at sense orientation. Our analysis on the sRNAs sequences from rice seeds (present study), flowers, seedlings and stems (24) showed that the repetitive sequences-containing miRNA loci were different from those without repetitive sequences in at least two aspects. First, the antisense sRNAs generated from the miRNA loci containing repetitive sequences were comparable to sense sRNAs; second, many sRNAs from these loci were generated from the sites other than mature miRNAs or miRNA’s (Table S9). This indicated that the annotated repetitive sequences-containing miRNA loci might be either sRNA clusters producing sRNAs randomly from two strands, or borderline miRNA candidates (8). It is worthy to notice that three miRNAs (osa-MIR531, osa-MIR1426 and osa-MIR1431), which generated sRNAs only from the mature miRNA positions, were also annotated as repetitive sequences (Table S9). Therefore, during the miRNA annotation in the present study, all the sequences containing repetitive sequences or putative precursors matching sRNAs in two orientations at similar abundance were excluded.

Two annotated miRNA loci, osa-MIR169p and osa-MIR169q, were significantly different from others (Table S9), as they generated many sRNAs in antisense orientation from the sites of mature miRNA and miRNA’. Among the generated sRNAs, some are uniquely mapped on the genome, indicating these two miRNAs may be transcribed from the antisense orientation of the annotated miRNA loci. Alignment on the precursor sequences of all osa-miR169 family members and the reversely complemented sequences of osa-MIR169p and osa-MIR169q showed that the reversely complemented sequences of osa-MIR169p and osa-MIR169q were conserved with osa-MIR169a-o (Figure 2A). This indicated that directions of osa-MIR169p and osa-MIR169q might be mis-annotated in miRBase (version 11.0).

Many conserved miRNA families have multiple paralogous loci, and some of them generate same mature miRNAs, making it difficult to identify from which locus the mature miRNAs were generated. In addition, detection of miRNA’s and rare side products of miRNA loci will help to confirm the expression of unique locus (11). Through mapping the signatures obtained from seeds and other tissues (8,24) onto the miRNA precursors, 96 loci from 47 miRNA families were confirmed using signatures uniquely hitting on the genome (Table S10). In addition, our analysis showed that the unique signatures mapping osa-miR156j and osa-miR166j were identical to those mapping osa-miR156h and osa-miR166i, respectively. Considering that osa-miR156j and osa-miR166j were not successfully mapped on the genome (TIGR 4.0), the loci of them might be identical to osa-miR156h and osa-miR166i.

Among the unique signatures matching miRNA precursors, several ones matched the precursors in antisense orientation. The unique origin of each signature on the genome suggested that it might be produced exclusively from this site (Table S11). Similarly, analysis on the precursor sequences from Arabidopsis (from Arabidopsis Small RNA Project, ASRP, http://asrp.cgrb.oregonstate.edu/) and C. elegans (GenBank data libraries, GEO accession number GSE5909, sample GSM139137) showed that this phenomenon is common in various species, and many sRNAs can match mature miRNAs or miRNA’s in antisense orientation (Figure 2B). Recent report showed that a single Hox locus in Drosophila produces functional miRNAs from opposite DNA strands (51), indicating this phenomenon is conserved throughout the animal and plant kingdoms.

**Identification of 26 novel and 12 candidate miRNAs**

To identify the novel miRNA loci, the consensus surrounding genomic sequence of each sRNA from this
A study was extracted, and secondary structure of each sequence was predicted with the help of Mfold program (28). A Perl script was used to evaluate the paired regions of predicted RNA secondary structures according to the criteria by MIRcheck (38). Sequences of putative miRNA loci were then analyzed by RepeatMasker to exclude the repetitive sequences. Those contained no repetitive sequences were regarded as candidate precursors and used for further analysis.

As previously reported, accumulations of miRNA/C3s were strong supporting evidences for the cleavage by DCLs during the miRNA biogenesis (11), thus the miRNA identification in this study is mainly dependent on the detection of miRNA’s. Eleven novel miRNA loci were finally identified, excluding three miRNAs (osa-miR810b.1, osa-miR1429 and osa-miR1432; references 8 and 26, which were published during the manuscript revision). Interestingly, the abundance of osa-miR1429/C3 (126 TPQ) were higher than osa-miR1429 (15 TPQ) in seed library, suggesting that the accumulation of miRNA/C3s may differ in different tissues. Further, sRNAs sequences from other six libraries reported by Lu et al. (8) were used to check the presence of miRNA/C3s, resulting in the identification of eight other miRNA loci (Table 2)
Two candidate precursors producing sRNAs at very high abundance (>180 TPQ) were identified as miRNAs (osa-miR827b and osa-miR1859), of which osa-miR827b was the ortholog of Arabidopsis ath-miR827. Comparison of the candidate precursors with rice ESTs showed that five previously annotated noncoding RNAs were indeed miRNAs. Six precursors, which had matched ESTs and were previously annotated as 'conserved hypothetical protein', were identified as miRNA candidates (osa-miRc1-c6). Additionally, six candidate miRNAs with abundance >15 TPQ were identified. As a whole, 26 novel (Table 2) and 12 candidate miRNAs (Table S12) were identified from this study (predicted structures were shown in Supplementary Dataset 1).

Among the novel miRNAs, osa-miR396f was a homolog of osa-miR396 family, which was one of the two novel miRNAs having orthologs in other species. As shown in Figure 2C, miR396 genes can be grouped into two subgroups according to sequences of mature miRNAs. The two annotated novel miRNAs in maize, zma-miR396c from AZM5_13 014 and zma-miR396d from AZM5_11 578, were more conserved with osa-miR396d-e than osa-miR396a-c.

The osa-miR827b was conserved in both rice and Arabidopsis. In miRBase (V11.0), osa-miR827 was recently identified through similarity analysis (52), however, no sRNA from the mature miRNA or miRNA* sites of osa-miR827 precursor was obtained in seed library or other public libraries (8). In contrast, the mature miRNA from osa-miR827b loci was abundant in seed library (150 TPQ), suggesting that osa-miR827b (rather than osa-miR827) has existent physiological functions.

Targets of these identified miRNAs were further predicted using the methodology described by Allen et al. (4), and targets of 18 novel miRNAs were captured with assignment of mispair score and MFE ratio tolerance limits to ≤4 and >0.73, respectively. Analysis and annotation of the predicted target genes using GO terms.
showed that they were with diverse functions (Tables 3 and S13), and involved in multiple cellular processes including response to stress, signal transduction and metabolic processes (Table S14).

Preferential expression of miRNAs in specific tissues

Further, miRNAs chips containing the previously and newly identified miRNAs were generated to study the expression profiles of miRNAs. Several detected miRNA’s by this study were also included on the chips. To analyze the miRNA expression in seeds in detail, RNAs from embryo and endosperm were extracted individually and used for hybridization. In addition, expression profiles of miRNAs in three vegetative tissues (root, leaf and seedlings) were analyzed and compared with those in seed tissues (Table S15). Analysis showed that most miRNAs were expressed preferentially in one or two tissues, and only some of them were highly expressed in most tested tissues (Figure 3A).

To corroborate the expression profiles obtained from chip hybridization, quantitative real-time RT-PCR (qPCR) was performed to study the transcripts of several mature miRNAs. In previous report, 5.8S rRNA, which can be polyadenylated by poly(A) polymerase at the 3'-end as mature miRNAs, was used as reference gene in qPCR analysis of miRNAs expression in Arabidopsis (41). Detection on the expression levels of rice 5.8S rRNA from GeneChip® genome array or by qPCR analysis showed that it was relative stable in the tested tissues (Supplementary Figure 1). We thus select rice 5.8S rRNA as reference gene in the analysis. Results showed that the tested miRNAs were indeed preferentially expressed in specific tissues (Figure 3B, Supplementary Figure 4A), being consistent with the data from miRNA chip hybridization. Comparison of the miRNAs expression patterns revealed by chip hybridization with the data of a recent publication (GSE11014; 53), in which sRNAs of four tissues were sequenced using 454 technology, revealed the consistence of the chip data with the differentially expressed miRNAs from 454 dataset, after filtering out the low abundance sequences (Supplementary Figure 4B). These indicated that the expression profiles of miRNAs by microarray hybridization were accurate and reliable.

Expression patterns of miRNAs were closely related to their functions. osa-miR160, osa-miR171g/h and osa-miR393 were preferentially expressed in root, and all of

---

**Table 3. Predicted target genes of novel miRNAs**

<table>
<thead>
<tr>
<th>miRNA</th>
<th>Target</th>
<th>MFE</th>
<th>Score</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>osa-miR1866-3p</td>
<td>Os09g21110</td>
<td>0.775</td>
<td>4</td>
<td>Leucyl-tRNA synthetase, cytoplasmic</td>
</tr>
<tr>
<td>osa-miR1846c</td>
<td>Os02g04100</td>
<td>0.858</td>
<td>3</td>
<td>Proteasome subunit alpha type 1</td>
</tr>
<tr>
<td>osa-miR1846-5p</td>
<td>Os10g12320</td>
<td>0.977</td>
<td>0.5</td>
<td>Hypothetical protein</td>
</tr>
<tr>
<td>osa-miR1846-5p</td>
<td>Os08g10350</td>
<td>0.841</td>
<td>2.5</td>
<td>Expressed protein</td>
</tr>
<tr>
<td>osa-miR2091-5p</td>
<td>Os12g17660</td>
<td>0.881</td>
<td>2</td>
<td>Hypothetical protein</td>
</tr>
<tr>
<td>osa-miR2092-5p</td>
<td>Os02g45054</td>
<td>0.786</td>
<td>3</td>
<td>INDETERMINATE-related protein 9</td>
</tr>
<tr>
<td>osa-miR2093-5p</td>
<td>Os08g14440</td>
<td>0.794</td>
<td>3</td>
<td>Expressed protein</td>
</tr>
<tr>
<td>osa-miR2094-5p</td>
<td>Os09g37834</td>
<td>0.745</td>
<td>4</td>
<td>Serine/threonine-protein kinase receptor precursor</td>
</tr>
<tr>
<td>osa-miR1850-3p-2</td>
<td>Os06g05440</td>
<td>0.735</td>
<td>4</td>
<td>Expressed protein</td>
</tr>
<tr>
<td>osa-miR1850-5p-2</td>
<td>Os07g07330</td>
<td>0.812</td>
<td>4</td>
<td>Expressed protein</td>
</tr>
<tr>
<td>osa-miR2097-5p</td>
<td>Os02g02680</td>
<td>0.846</td>
<td>3</td>
<td>NBS-LRR disease resistance protein</td>
</tr>
<tr>
<td>osa-miR2098-3p</td>
<td>Os05g08950</td>
<td>0.873</td>
<td>3</td>
<td>Bifunctional protein tilS/hrpT</td>
</tr>
<tr>
<td>osa-miR2098-3p</td>
<td>Os12g38560</td>
<td>0.790</td>
<td>3</td>
<td>Expressed protein</td>
</tr>
<tr>
<td>osa-miR2098-3p-2</td>
<td>Os11g05930</td>
<td>0.971</td>
<td>3.5</td>
<td>CCT motif family protein</td>
</tr>
<tr>
<td>osa-miR2098-5p</td>
<td>Os08g37940</td>
<td>0.815</td>
<td>2.5</td>
<td>HAD-superfamily hydrolase subfamily 1A</td>
</tr>
<tr>
<td>osa-miR2098-5p</td>
<td>Os02g48880</td>
<td>0.841</td>
<td>3</td>
<td>Chloride channel protein CLC-f</td>
</tr>
<tr>
<td>osa-miR2099-5p</td>
<td>Os01g28810</td>
<td>0.913</td>
<td>2</td>
<td>Expressed protein</td>
</tr>
<tr>
<td>osa-miR2099-5p</td>
<td>Os03g31320</td>
<td>0.768</td>
<td>3</td>
<td>RING zinc finger protein-like</td>
</tr>
<tr>
<td>osa-miR2100-5p</td>
<td>Os08g43160</td>
<td>0.884</td>
<td>3</td>
<td>TCP family transcription factor containing protein</td>
</tr>
<tr>
<td>osa-miR2100-5p</td>
<td>Os04g59470</td>
<td>0.932</td>
<td>2.5</td>
<td>NAC domain-containing protein 76</td>
</tr>
<tr>
<td>osa-miR2100-5p</td>
<td>Os04g31290</td>
<td>0.919</td>
<td>2</td>
<td>Helix-loop-helix DNA-binding domain containing protein</td>
</tr>
<tr>
<td>osa-miR2102-5p</td>
<td>Os05g25390</td>
<td>0.866</td>
<td>3</td>
<td>Protein kinase</td>
</tr>
<tr>
<td>osa-miR2102-5p</td>
<td>Os01g12280</td>
<td>0.852</td>
<td>3</td>
<td>Protein dimerization</td>
</tr>
<tr>
<td>osa-miR2103</td>
<td>Os09g09200</td>
<td>0.905</td>
<td>2</td>
<td>Hypothetical protein</td>
</tr>
<tr>
<td>osa-miR2103</td>
<td>Os01g50190</td>
<td>0.905</td>
<td>1</td>
<td>Indole-3-acetic acid-amido synthetase GH3.2</td>
</tr>
<tr>
<td>osa-miR2103</td>
<td>Os05g37960</td>
<td>0.901</td>
<td>1</td>
<td>BTB/POZ domain containing protein</td>
</tr>
<tr>
<td>osa-miR2103</td>
<td>Os08g41230</td>
<td>0.867</td>
<td>3</td>
<td>Hypothetical protein</td>
</tr>
<tr>
<td>osa-miR2103</td>
<td>Os03g47920</td>
<td>0.783</td>
<td>3.5</td>
<td>Hypothetical protein</td>
</tr>
<tr>
<td>osa-miR2104</td>
<td>Os06g11120</td>
<td>0.899</td>
<td>3</td>
<td>Conserved hypothetical protein</td>
</tr>
<tr>
<td>osa-miR2104</td>
<td>Os02g27380</td>
<td>0.826</td>
<td>2.5</td>
<td>Calcium-transporting ATPase 2</td>
</tr>
<tr>
<td>osa-miR2104</td>
<td>Os03g10640</td>
<td>0.805</td>
<td>3.5</td>
<td>Conserved hypothetical protein</td>
</tr>
<tr>
<td>osa-miR2105</td>
<td>Os03g39520</td>
<td>0.798</td>
<td>3.5</td>
<td>Conserved hypothetical protein</td>
</tr>
<tr>
<td>osa-miR2105</td>
<td>Os12g13170</td>
<td>0.763</td>
<td>3</td>
<td>Transcription factor HBP-1a</td>
</tr>
<tr>
<td>osa-miR2105</td>
<td>Os03g11140</td>
<td>0.848</td>
<td>3.5</td>
<td>RHO GTPase activator</td>
</tr>
<tr>
<td>osa-miR2106</td>
<td>Os07g41694</td>
<td>0.791</td>
<td>4</td>
<td>Acidic leucine-rich nuclear phosphoprotein 32 family</td>
</tr>
<tr>
<td>osa-miR2106</td>
<td>Os04g48390</td>
<td>0.851</td>
<td>3</td>
<td>beta-lactamase, class A</td>
</tr>
</tbody>
</table>

‘MFE’ indicates the minimum free energy ratio of miRNA:target duplex and target:complementary sequence duplex.
negative correlation of most miRNAs with the corresponding targets

Expression profiles of miRNAs were compared with those of the corresponding targets (all the targets of known and novel miRNAs used in the analysis were listed in Table S16), which were analyzed through hybridizing the Affymetrix Rice genome DNA chip. The results showed that the expression of most miRNAs was negatively correlated with their targets (Figure 4A), being consistent with their functions in guiding the cleavage of target mRNAs in plants (1, 59, 60). In addition, some miRNAs (osa-miR160, osa-miR164 and osa-miR172) were positively correlated with their targets (Figure 4B), which
may due to the consequence of feedback regulation (61). As reported in Arabidopsis, both MIR164A and CUC2 are transcribed at the margins of young leaf primordia, and the balance between them determines the extent of serration (62). The positive correlations of miRNAs and their targets also suggested that these miRNAs may function by translational repression (60), and indicated that the regulation mechanisms of conserved miRNAs, for example miR164 and miR172, were conserved throughout monocot and dicot plants.

Differential regulations of miRNAs and miRNA’s

Besides being sequenced, some miRNA’s were detected with high-level expressions in several tissues by chip hybridization. As shown in Figure 5A, some mature miRNAs and corresponding miRNA’s accumulated highly in same tissue(s) (osa-miR408, osa-miR1428e-5p and osa-miR397b). This was reasonable as both mature miRNAs and corresponding miRNA’s were produced from same precursors. However, the mature miRNAs and miRNA’s of some miRNAs (osa-miR168a, osa-miR1866-3p and osa-miR396a) accumulated differently in these tissues, suggesting the differential regulation of mature miRNAs and miRNA’s. In addition, for some miRNAs, expression level of miRNA* were even higher than the corresponding miRNAs (Figure 5A and B).

To confirm our observation, sRNAs from seeds and other six tissues (8) were surveyed to compare the expression profiles of miRNAs and miRNA’s. All the expression abundances of sRNAs were normalized to transcripts per quarter million (TPQ) after filtering out signatures corresponding to rRNAs, tRNAs, snRNAs and snoRNAs, to ensure the normalized values were comparable across the tissues (22,63). Being consistent with the results from chip hybridization, the abundance relations of mature miRNAs and miRNA’s were diverse in these tissues. For osa-miR408 and osa-miR390, mature miRNAs and miRNA’s were closely related, while for osa-miR168a and osa-miR396a, mature miRNAs and miRNA’s accumulated differently in these tissues (Figure 5C).

The differential accumulation of miRNAs and miRNA’s in specific tissue may due to the different activities of proteins directly involved in sRNAs biogenesis. Recent report showed that different argonaute (AGO) proteins in Arabidopsis harbor sRNAs with different 5’ terminal nucleotide, and the miRNAs and miRNA’s are directed in different AGO complexes based on their 5’ terminal nucleotide (64). Several homolog AGO proteins were identified in rice (Figure6A; 65) and our analysis showed the different expression levels of them in the tested tissues (Figure 6B). Especially, two AGO-coding genes (Os03g58600, Os07g09020) homologous to AtAGO5 were predominantly expressed in rice embryo, and an AGO-coding gene (Os06g51310) homologous to AtAGO1 was low expressed in rice endosperm and embryo. The variety of AGO gene abundances in these tissues may determine the abundance of miRNA and miRNA’s in RISCs.

DISCUSSIONS

MPSS analysis using rice seeds at different developmental stages was performed and turned out to be effective in identifying a large number of sRNAs including novel miRNAs. Several recent studies have reported the large-scale sequencing of sRNAs in rice (8,24–26); however, most of them were obtained from the tissues other than seeds. In the study by Lu et al. (8), only sRNAs with the abundance more than 100 TPQ were used to predict the putative precursors, which may miss some uncharacterized miRNAs. In present study, all the sRNAs
osa-miR1866-3p was reported the scale of the relative expression levels of miRNAs and was presented by Pearson’s correlation coefficient (PCC). The bar represented the scale of the relative expression levels of miRNAs among them, which was set as 1.0. (A) Phylogenetic analysis of the Arabidopsis and rice AGO proteins. The protein sequences of AtAGO1 to AtAGO10 were obtained from the Arabidopsis Information Resource (http://www.arabidopsis.org). Rice protein sequences were from the Institute for Genomic Research. ClustalW (www.ebi.ac.uk/Tools/clustalw2) was used to align the sequences, and MEGA4 was used to generate the phylogenetic tree using the neighbor-joining and the Poisson correction methods. Bootstrap values were calculated with 1000 permutations. The scale bar represented 0.1-amino-acid substitutions per residue. (B) Expression profiles of rice AGO genes. The most homologous Arabidopsis genes were labeled. The bar represented the scale of relative expression levels of AGO genes and the expression signals were Log2 transformed.

(excluding rRNAs, tRNAs, snRNAs and snoRNAs) were analyzed (Figure 1A), and as a result, 26 novel miRNAs were identified mostly dependent on the accumulations of miRNA*s. The targets of 18 novel miRNAs were predicted, and hybridization with the generated miRNA chip provided a detailed expression profile of known and novel miRNAs in vegetative (root, seedling, young leaf) and reproductive (embryo and endosperm) tissues. These
will significantly contribute to the discovery and studies of miRNAs in rice, especially rice seeds.

**Most rice sRNAs are derived from transposable elements**

Of the 81,530 sRNA clusters 72,173 contained interspersed repetitive sequences (Table S6), and of all the identified signatures from rice seeds, miRNAs account for <1.5%, which is much lower than in other rice tissues (8.8%, 36.6% and 12.6% from the inflorescences, seedlings and stems, respectively, 24) or Arabidopsis inflorescences (40%) and seedlings (62%, 22). The low ratios of miRNAs in rice tissues may because that the complex rice genome contains more interspersed and inverted repetitive sequences, and produces many repeat-associated sRNAs (24). The even lower ratio of miRNAs (higher ratio of rasiRNAs) in rice seeds suggested the important roles of rasiRNAs in seed development. As indicated by Girard and Hannon (66), the robust defense of transposition activity coincides with the transposon challenge in the germline, where new insertions of transposons can be transmitted to next generation. High abundance of rasiRNAs in seeds may serve as a suppressor of transposon to prevent its propagation in rice genome. Our results also suggested that miRNAs were low expressed in pre-differentiation tissues including embryo. Indeed, a recent study showed that genes targeted by miRNAs are expressed in a higher level in embryo than in mature tissues in mouse and Drosophila, suggesting that miRNAs are low expressed in embryo (67). Furthermore, analysis on the top 10 highly expressed individual miRNAs revealed that seven of them were with the highest expression in tissues other than embryo or endosperm (Table S15). Although several miRNAs were highly expressed in embryo or endosperm (Figure 3A), the numbers of these miRNAs were low, and the expression levels of them were much lower than those top highly expressed miRNAs in other tissues (Figure 3A, Table S9). As whole, the total abundance of miRNAs was low in rice seeds.

nat-siRNA may be formed when natural antisense transcripts are expressed in same cell (5). After comparing the 687 bidirectional transcript pairs (68), we failed to identify any signatures matching both pairs, which supports the hypothesis that the bidirectional transcripts pairs were mainly to induce the alternative splicing or polyadenylation, other than serving as origins of sRNAs by forming RNA duplex (69,70).

**Some registered miRNAs appear to be repetitive sequences**

In the guidelines proposed by Ambros et al. (27), miRNAs can be distinguished from siRNAs in several aspects, including genomic origin characters and biogenesis mechanisms. Our study showed that some registered miRNA precursors contain transposon or retrotransponson-like repetitive sequences, even the whole structure of them were annotated as retrotransposon (such as Stowaway) (Table S8). Furthermore, our analysis showed that most of these repeat-associated precursors could generate sRNAs from two strands randomly (Table S9), thus these miRNA precursors might not be bona fide miRNAs (8). Several recently reported miRNAs (26) also contained transposable elements (Table S8), among which three (osa-miR1440, osa-miR1441 and osa-miR1442) can produce sRNAs from two strands randomly (Table S9). During their analysis, repeat-associated borderline precursors were used as positive controls, thus some candidate miRNA precursors similar to repeat elements were also identified as miRNAs (26). This indicates that it is necessary to distinguish the bona fide miRNAs from borderline ones for annotating the novel miRNAs. According to the definition of a miRNA, one of the most critical criteria is that the miRNA should originate from a single-strand RNA transcript (27). Thus, if a locus produces sRNAs from both two strands, this locus should be siRNA locus; if a candidate precursor contains repeat elements, it should be very careful to annotate it as an miRNA, as most of the precursors of this type can produce sRNAs from both strands and many of these sRNAs are not from the site of mature miRNAs or miRNA*’s (Table S9).

**Diverse regulatory effects of miRNAs in rice seeds**

Auxin plays crucial roles during seed development, including pattern formation, cell division and cell expansion (71). The mutation of AUXIN RESPONSE FACTOR 2 (ARF2) leads to dramatic increase of seed size and weight in Arabidopsis (71), osa-miR167, the target genes of which encode transcription factors ARF6 and ARF8 (56), was preferentially expressed in rice seeds (Figure 3B). In addition, osa-miR167 was induced by auxin, resulting in the decrease of ARF8 mRNA abundance. ARF8 positively regulated the expression of OsGh3.2, a rice indole-3-acetic acid (IAA)-conjugating enzyme, which in turn participated in the control of cellular-free IAA levels (56). In rice seeds, the IAA level is around 40-fold higher than in other tissues (72), which is consistent with the high abundance of osa-miR167 in rice seeds and suggests the regulatory roles of auxin-miR167-ARF8-OsGh3.2 pathway in auxin amount and hence the rice seed development.

Many target genes other than transcription factor coding genes are also regulated by miRNAs preferentially expressed in seeds, including those encode copper-binding proteins and carbohydrate metabolism enzymes. Such kind of functional relation is similar to miRNAs isolated from developing xylem of stems, which are predicted to target genes involved in biosynthesis of cell wall metabolites in Populus trichocarpa (73). Among the miRNAs that highly expressed in embryo, four miRNA families (osa-miR397, osa-miR398, osa-miR408 and osa-miR528) targeted genes coding copper-binding proteins (17,20,38; Table S16). Interestingly, all these target genes were suppressed in embryo (Figure 4A), suggesting the roles of these miRNAs in maintaining the copper homeostasis (57,58) and/or the ROS levels (74) in rice embryo.

osa-miR397 and osa-miR528 are predicted to target the miRNAs of L-ascorbate oxidases (AO), which catalyze the oxidation of ascorbic acid (AA) to yield dehydroascorbate (DHA) to regulate the apoplastic redox state and modulate plant growth and defense responses by regulating...
signal transduction cascades (75). It has been reported that the accumulation of DHA oxidized by AO could trigger the arrest of cell division (76). The high amounts of these two miRNAs will result in the low level of AO genes, which may in turn maintain the low level of apoplastic DHA in seeds. This is consistent with the higher cell-division ability of seeds at early development stage.

**Differential accumulations of miRNAs and miRNA’s**

It is believed that miRNA*s are gradually degraded while mature miRNAs guide the silencing of target miRNAs in miRNA-induced silencing complex (miRISC) (11). Our study indicates that some miRNA*s were also accumulated at certain tissues (Figure 5). Abundance of osa-miR390* was even higher than miR390 in rice embryo (Table S15). The relations between the accumulation of p21 and the Beet yellows virus duplex. It has been reported that the accumulation of DHA oxidized by AO could trigger the arrest of cell division (76). The high amounts of these two miRNAs will result in the low level of AO genes, which may in turn maintain the low level of apoplastic DHA in seeds. This is consistent with the higher cell-division ability of seeds at early development stage.

**SUPPLEMENTARY DATA**

Supplementary data are available at NAR Online.

**FUNDING**

State Key Project of Basic Research (2005CB120803); Chinese Academy of Sciences (KSCX2-YW-N-058); Nature Science Foundation (No. 30721061); and High-Tech program (2006AA10A102). Funding for open access charge: Nature Science Foundation.

Conflict of interest statement. None declared.

**REFERENCES**


