Cloning of Novel Repeat-associated Small RNAs Derived from Hairpin Precursors in Oryza sativa

Chengguo YAO, Botao ZHAO, Wei LI, Yang LI, Wenming QIN, Bing HUANG, and Youxin JIN*

State Key Laboratory of Molecular Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences; Graduate School of Chinese Academy of Sciences, Shanghai 200031, China

Abstract Plant small non-coding RNAs including microRNAs (miRNAs), small interfering RNAs (siRNAs) and trans-acting siRNAs, play important roles in modulating gene expression in cells. Here we isolated 21 novel endogenous small RNA molecules, ranging from 18 to 24 nucleotides, in Oryza sativa that can be mapped to 111 hairpin precursors. Further analysis indicated that most of these hairpin sequences originated from putative miniature inverted-repeat repeat transposable elements, a major type of DNA transposon. Considering that miRNA is characteristic of hairpin-like precursor and plant endogenous siRNAs are often located at transposon regions, we hypothesized that our cloned small RNAs might represent the intermediate product in the evolutionary process between siRNAs and miRNAs. Northern blot analysis indicated that five of them were much more abundantly expressed in flower compared to other tissues, implying their potential function in inflorescence. In conclusion, our results enrich rice small RNA data and provide a meaningful perspective for small RNA annotation in plants.

Keywords small RNA; microRNA; small interfering RNA; hairpin; Oryza sativa

Received: April 4, 2007 Accepted: July 1, 2007

This work was supported by the National Natural Science Foundation of China (No. 30430210), the National Key Basic Research and Development Program (No. 2005CB124602) and Chinese Academy of Science (KSCX1-YW-R-64)

*Corresponding author: Tel, 86-21-54921222; Fax, 86-21-54921011; E-mail, yxjin@sibs.ac.cn

DOI: 10.1111/j.1745-7270.2007.00346.x

©Institute of Biochemistry and Cell Biology, SIBS, CAS
from typical organs at different development stages, and were able to obtain 3003 sequences. Subsequent analysis allowed us to pick out 21 novel small RNA sequences mapped to 111 hairpin precursors that are encoded by putative transposons. These small RNAs share characteristics with both miRNA and siRNA, and might represent the evolutionary link between both.

Materials and Methods

Cloning of small RNA sequences from rice

Whole rice plants (Oryza sativa L. ssp. japonica) were grown under natural conditions. Different plant tissues at various development stages (root, leaf, flower, and stem) were collected, washed with double distilled H2O, and frozen in liquid nitrogen. Total RNAs from O. sativa were prepared using the Trizol method (Invitrogen, Carlsbad, USA) in which the isopropanol precipitation was replaced by ethanol precipitation. In brief, small RNAs from 18 to 28 nt were size-fractionated, purified, and ligated sequentially to the 5′ DNA adapter (ACCGAATTCACTGAGACC, EcoRI site) and 3′ adapter (GCAGATCGTCA-GTTCCAG, EcoRI site) with T4 RNA ligase from New England Biolabs (Beverly, USA). The 3′ adapter was blocked by ddA with terminal transferase (NEB) at its 5′ terminus. The ligated RNA was reverse transcribed into cDNA by the Access Quick reverse transcription-polymerase chain reaction (RT-PCR) system (Promega, Madison, USA) with the 5′ adapter sequence and another primer complementary to the 3′ adapter. The RT-PCR product was amplified by PCR with the same primers. After purification, the PCR product was digested with EcoRI (NEB) and concatemerized with T4 DNA ligase (NEB) followed by purification with the kit (Bioasia, Shanghai, China). PCR was carried out with the previous primers and the DNA of approximately 800 bp was cloned into pGEM-T vector (Promega) for sequencing.

To avoid losing the cDNAs containing the EcoRI site, the adaptors with a SalI site but not an EcoRI site were also used to generate some cDNA libraries.

Sequence analysis and prediction of fold-back structures

Using BLASTN on the National Center for Biotechnology Information (NCBI) website (http://www.ncbi.nlm.nih.gov/blast), cloned sequences shorter than 16 nt and possible ribosomal RNA, messenger RNA, transfer RNA, and small nucleolar RNA fragments were discarded. This step allowed the removal of most non-siRNA and non-miRNA species and identification of cloned known miRNAs that could be confirmed in the miRNA registry (http://microrna.sanger.ac.uk/sequences/). After redundancy analysis and exclusion of previously reported rice small RNAs that were also cloned through construction of libraries, the remaining sequences were compared with the latest version of the rice genome at the Rice Genome Database (RGP; http://riceblast.dna.affrc.go.jp/). The surrounding sequences of 200 nt in both orientations were extracted and, together with the perfectly matched sequences in the RGP, were submitted for RNA folding using the Mfold program [14]. Folding results were inspected, and fold-back structures with small RNA in the stem were considered hairpin precursors.

Next, we ran BLAST search against the TIGR rice repeat database (http://tigrblast.tigr.org/euk-blast/index.cgi?project=osa1) with the hairpin precursors. Sequences with at least 80% similarity, with one hit including the mature region, were regarded as candidates for rice repeat genome data [7], and some of them were perfect matches. Genomic annotation was examined using the Rice Genome Automated Annotation System (http://ricegaas.dna.affrc.go.jp/).

Northern blot analysis of small RNAs

Approximately 10 µg total RNA was isolated from different rice samples (embryo, seedling, leaf, stem, flower, and fruit), loaded and resolved on the denaturing 15% polyacrylamide gel, and transferred to a Hybond N+ nylon membrane (Amersham, Piscataway, USA). The membrane was ultraviolet cross-linked. DNA probes complementing miRNA sequences were end-labeled with [γ-32P]ATP (3000 Ci/mM; Amersham) using T4 polynucleotide kinase (MBI, Vilnius, Lithuania). Unincorporated [γ-32P]ATP was removed using a Sephadex G-25 column (Pharmacia, Uppsala, Sweden). Methylene blue staining of membranes prior to hybridization was used to detect ribosomal RNA. The membrane was pre-hybridized and hybridized using ExpressHyb Hybridization Solution (Clontech, Palo Alto, USA) then washed according to the user manual. The membrane was briefly air-dried then exposed to a PhosphorImager (Amersham).

Results

Novel small RNA gene families identified in rice show features of both miRNA and siRNA

Construction of a small RNA library is a practical and effective way to study small RNAs. In fact, many rice
miRNAs, endogenous siRNAs and other uncharacterized tiny non-coding RNAs were identified through this starting point. We constructed six small RNA libraries ranging from embryo to flower in rice plant (*O. sativa* L. ssp. japonica). Small RNAs of 18−26 nt in length were isolated by size fractionation and ligated to 5′ and 3′ adapters, then cloned and sequenced. A total of ~460 clones comprising 3003 small cDNA insertion sequences were collected. After removal of sequences shorter than 16 nt, we blasted the remained sequences against the rice genome in the NCBI database to discard the query sequences bearing no hits, and those that could be degradation products of ribosomal RNA, transfer RNA, small nucleolar RNA, or small nuclear RNA that constituted approximately half of the overall sequences. The BLAST results in the NCBI database and miRNA registry (http://www.sanger.ac.uk/Software/Rfam/mirna/search.shtml) also showed that we had cloned many known miRNAs and endogenous siRNAs. A total of 1416 small RNA sequences were matched to the rice genome (japonica genome database at RGP). One hundred and eleven loci corresponding to 21 small RNAs were predicted to hold hairpin-like fold-backs (Table 1), characteristic of miRNAs. However, sequences blasted against the rice repeat database indicated that they all showed high sequence similarity with one or more hits. Most of them are putative MITEs, and others are putative unclassified transposons or putative retrotransposons (data not shown), which are obviously features of siRNA-generating sequences.

### Genomic annotation of new rice small RNAs

Non-coding small RNAs are usually located in the genomic segments that are distinct from protein-coding regions. Examination results of the corresponding loci with the Rice Genome Automated Annotation System (http://ricegaas.dna.affrc.go.jp/) conformed to the traditional knowledge, in that almost all of the loci were in the intergenic or intron, non-coding expressed sequence tag regions, except for a few in the exon of a hypothetical protein and cDNA region, which could be pseudo genes. The proteins specified by the intron region are nearly all

<table>
<thead>
<tr>
<th>ID</th>
<th>Sequence (5′→3′)</th>
<th>Annotation</th>
<th>Length</th>
<th>No. of hairpin precursors</th>
<th>Arm*</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV-136</td>
<td>CAGCUCCACAACUCUACUCCA</td>
<td>EST region</td>
<td>21</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>IV-165</td>
<td>UGACACCCAUCAUCGACUACC</td>
<td>Intergenic</td>
<td>21</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>IV-304</td>
<td>CUGGGAGUGGAGUUGGGUGGAG</td>
<td>EST region</td>
<td>22</td>
<td>2</td>
<td>5, 3</td>
</tr>
<tr>
<td>III-977</td>
<td>UUUGACCCGGAUGGCGAAGGGG</td>
<td>Intergenic</td>
<td>22</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>III-130</td>
<td>AACGCUUCACGGAACCUAGGG</td>
<td>Intergenic</td>
<td>24</td>
<td>24</td>
<td>3</td>
</tr>
<tr>
<td>III-196</td>
<td>AAAAUCUUCAANAGCAGCAACGU</td>
<td>Intron of putative peroxidase</td>
<td>23</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>III-200</td>
<td>GACUGACCUUGACUCGCUUU</td>
<td>EST, intergenic</td>
<td>22</td>
<td>30</td>
<td>5</td>
</tr>
<tr>
<td>II-187</td>
<td>AAUCCAUAUAGCAGCAUUGCUGG</td>
<td>EST, intergenic, exon of hypothetical protein</td>
<td>24</td>
<td>24</td>
<td>5</td>
</tr>
<tr>
<td>II-316</td>
<td>UUCACCAUAUUCUGGAAA</td>
<td>Intergenic</td>
<td>22</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>II-419</td>
<td>AAUCUGACCGUAAUGCUUGAUGU</td>
<td>EST, intergenic</td>
<td>23</td>
<td>3</td>
<td>5, 3</td>
</tr>
<tr>
<td>II-492</td>
<td>ACAAAAAAAAAACGAGCCAAACAGG</td>
<td>Intron of putative reverse transcriptase</td>
<td>22</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>IV-8</td>
<td>GGAGCAGGCUAUAGACCGCUAGAU</td>
<td>Intergenic</td>
<td>24</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>IV-25</td>
<td>GGCCCCACGUGUCACACCGCGUG</td>
<td>Intron of hypothetical protein</td>
<td>24</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>IV-123</td>
<td>UAUUGGAAACCGCUAGAUGU</td>
<td>EST</td>
<td>19</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>IV-149</td>
<td>ACACUCAAGCCGUGUGUGGUGG</td>
<td>Unknown</td>
<td>20</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>IV-235</td>
<td>ACCGCCAGCACAUCUUUUGAGCC</td>
<td>Intergenic, intron of unknown protein</td>
<td>24</td>
<td>17</td>
<td>5</td>
</tr>
<tr>
<td>IV-298</td>
<td>AAGUUGUGUGUAGAUGGUGAGA</td>
<td>Intron of unknown protein</td>
<td>24</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>IV-299</td>
<td>AAGGACUGAAGGUGGUGAUGA</td>
<td>EST, intergenic</td>
<td>18</td>
<td>7</td>
<td>5, 3</td>
</tr>
<tr>
<td>IV-351</td>
<td>GACGGAGGUCACAAAAAGCUGAGCA</td>
<td>Intergenic</td>
<td>24</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>V3-27</td>
<td>GGUAGCGGCGUGCCGCUGCGG</td>
<td>Intergenic, exon of putative protein</td>
<td>24</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>V2-16</td>
<td>AGAUGACGUGACGUUUACC</td>
<td>Intergenic</td>
<td>21</td>
<td>8</td>
<td>5</td>
</tr>
</tbody>
</table>

* Location of small RNAs in certain fold-back hairpins. 3, located in the arm close to the 3′ terminus; 5, located in the arm close to the 5′ terminus; EST, expressed sequence tag.
hypothesical or putative, indicating that they might also be pseudo genes. For those small RNAs encoded by multiple loci, we could not assign their exact origin unambiguously, and some of them might not be genuine genes, as referred to above.

Cloning of putative influorescence-associated small RNAs

To validate the cloning results, we carried out Northern blot analysis to examine the existence of these small RNAs using total RNA samples isolated from diverse organs from different development stages. Interestingly, five of them were exclusively abundant in flower compared to other tissues (Fig. 1), suggesting that they might specially function in the process of inflorescence. However, their targets and the mechanisms through which they play the role are yet to be discovered. As a control, miR159c was also examined, and showed ubiquitous expression in the samples listed in our experimental conditions. Some small RNAs were not detected by Northern blot analysis with our available rice tissues. They might be expressed at very low levels or confined to some specific tissues or cell types, therefore the amount of these miRNAs was inappreciable in the limited total RNA samples.

Discussion

As more and more detailed information about plant small RNAs, including miRNAs, endogenous siRNAs, and trans-acting siRNAs are being proffered [15–17], the differences between them seem to be much clearer. However, some aspects still remain controversial, such as their conservation, origin, and length [18], which convinced us of the view that the diversity of plant small RNAs is more complex than previously expected. Recent cloning and analysis of rice endogenous small non-coding RNAs corroborate this point [13]. To date, most reports have emphasized their distinction and classification, but ignored the correlation and possible evolutionary intermediates. Here we present data to provide a link between plant siRNAs and miRNAs. A full-length DNA-type transposable element contains an open reading frame flanking two TIRs. When it becomes a non-autonomous MITE, it is deprived of an open reading frame, and the hairpin structure would be formed by base-pair interaction of MITE TIRs. Previously, an inverted duplication model for miRNA gene evolution in plants was proposed [19], however, the role of MITEs was not discussed in that the researchers used known miRNA fold-back sequences. There is a postulation that miRNAs could have evolved from TE-encoded siRNAs in the way that the TIRs that are processed from longer RNAs to form siRNAs could be similarly processed to form miRNAs [8]. Thus we assumed that small RNAs that originated from MITEs were matured by miRNA biogenesis pathways (Fig. 2). However, additional analysis of these small RNAs in rice mutants relating to the biogenesis of miRNAs, such as DCL1, is necessary to validate this hypothesis [20,21]. Predictably, the evolutionary relationship and distinction between plant miRNA and

![Fig. 1](image-url)  Expression pattern analysis of some newly cloned rice small RNAs using Northern blot

Total RNAs from *Oryza sativa* L. ssp. japonica fruit of autumn stage (lane 1), flower of inflorescence (lane 2), leaf of tillering stage (lane 3), stem of jointing stage (lane 4), seedling (lane 5), and embryo (lane 6) were analyzed on the denaturing 15% polyacrylamide gel. Known miR159c was used as the positive control; 5S ribosomal RNA (Rrna) as the loading control. Positions are indicated with size markers. nt, nucleotide.
siRNA will become clearer as the biogenesis knowledge about these MITE-derived small RNAs accumulates.

The average copy number of rice MITEs ranges from dozens to thousands. Here we only extracted hairpin precursors perfectly matched to the cloned sequences, thus many homologs might have been missed. In view of our limited number of small RNA libraries, predicting the average copy number of the MITE-derived small RNAs is not practical. However, taking account of the variability of plant fold-back precursors and 90,000 MITEs in rice, it is a really interesting area of study, and our cloned small RNA sequences from MITEs just show the tip of the iceberg.

Although plant small RNA target prediction is convenient for the extensive complementarity between small RNA and its target according to the accepted principles [22, 23], the authentic targets of the vast majority of plant small RNAs are not characterized, and the phenotypic consequences of disrupted or altered small RNA regulation are also obscure because projects often take place over long time periods, and chance events take place, especially in the study of rice [11,24–26]. Northern blot data revealed that we have cloned five flower-exclusive small RNAs among the tissues analyzed, hinting that they might act like miRNAs to get involved in inflorescence. The molecular scenarios are waiting to be unraveled.

Fig. 2  Schematic representation of the possible biogenesis pathway of rice repeat-associated small RNAs derived from hairpin precursors

(A) Typical fold-back structure predicted with the Mfold program. Green regions indicate the mature small RNA sequences; red rectangles include the genome repeats. MITE-adh type B, type D are subtypes defined by MITE (miniature inverted-repeat transposable element) structural features. (B) Assumption model of microRNA (miRNA) biogenesis pathways for small RNAs that are originated from MITEs. Red sections indicate mature small RNA region. TIR, terminal inverted repeat.

References


17 Johnson C, Sundaresan V. Regulatory small RNAs in plants. EXS 2007, 97: 99–113


26 Jones-Rhoades MW, Bartel DP. Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Mol Cell 2004, 14: 787–799

Edited by
Blake C. MEYERS