Large-Scale Identification of Mirtrons in *Arabidopsis* and Rice

Yijun Meng[1]*, **Chaogang Shao**[2]*

1 College of Life and Environmental Sciences, Hangzhou Normal University, Hangzhou, People's Republic of China, 2 College of Life Sciences, Huzhou University, Huzhou, People's Republic of China

Abstract

A new catalog of microRNA (miRNA) species called mirtrons has been discovered in animals recently, which originate from spliced introns of the gene transcripts. However, only one putative mirtron, osa-MIR1429, has been identified in rice (*Oryza sativa*). We employed a high-throughput sequencing (HTS) data- and structure-based approach to do a genome-wide search for the mirtron candidate in both *Arabidopsis* (*Arabidopsis thaliana*) and rice. Five and eighteen candidates were discovered in the two plants respectively. To investigate their biological roles, the targets of these mirtrons were predicted and validated based on degradome sequencing data. The result indicates that the mirtrons could guide target cleavages to exert their regulatory roles post-transcriptionally, which needs further experimental validation.

Introduction

MicroRNAs, the well-known small RNA (sRNA) species of ~21-nucleotide (nt) in length, play essential regulatory roles in gene expression in a vast range of organisms [1–3]. Different from the miRNA genes in animals [1,4], which were mainly discovered within the introns or the exons of the coding or non-coding genes, most plant miRNAs were found to reside in the intergenic regions [5]. Generally, the primary transcripts of the miRNA genes should be subjected to two sequential cleavages by Drosha and Dicer in animals [1,4], or Dicer-like 1 (DCL1) in plants [5], in order to generate functional mature miRNA molecules for post-transcriptional gene silencing (PTGS). However, some non-canonical biogenesis pathways have been discovered. Recently, a novel class of miRNAs named as mirtrons was widely recognized in animals [4,6–14]. These miRNAs are generated from the spliced and subsequently debranched introns of the gene transcripts. Making them quite distinguishable from the other miRNA genes resided with the intronic regions, the mirtrons are processed through a Drosha-independent pathway [4,8,11,13]. After Dicing, the miRNAs are separated from the miRNA/miRNA* duplexes, and incorporated into the Argonaute (AGO)-associated miRNA-induced silencing complexes (miRISCs) to mediate PTGS.

To date tens of mirtron genes have been uncovered in several animal species, such as *Drosophila melanogaster*, *Caenorhabditis elegans*, *Gallus gallus* and mammals [6,7,9,10,12,13]. However, only one putative mirtron, osa-MIR1429, was identified in rice [15]. Thus, the question whether the mirtrons are also widespread in plants as in animals remain to be addressed. In this study, we interrogated this issue in both *Arabidopsis* and rice. Taking advantage of sRNA HTS data and secondary structure prediction, five and eighteen mirtron candidates were discovered in the two plants, respectively. The biological relevance of these mirtrons was primarily illustrated

through degradome data-based target identification. Taken together, we did a first systemic search for the mirtron genes in the two model plants, and hoped that these findings could inspire further research efforts on this interesting topic.

Results and Discussion

Genome-wide Identification of Mirtron Candidates

The two model plants, *Arabidopsis* and rice, with well-annotated genomes were selected for this study. Considering the fact that the sequences of the currently registered plant miRNA precursors [according to miRBase (Release 17) [16]] are generally less than 300 nt in length (Figure S1), the introns with this length range were selected for secondary structure prediction by using RNAshapes [17]. Since almost all the canonical miRNA precursors could form simple stem-loop structures, the simplest structure (with a single hairpin structure in most cases) among the prediction results of an intron was selected for manual check. The ones capable of forming stable hairpin structures were retained for further filtering.

Then all the short reads from the retrieved sRNA HTS data sets were mapped onto these intron sequences, and all the perfectly matched ones were retained. The findings in animals demonstrated that the processing of the mirtron precursors could bypass the Drosha cleavages [13], which were required for canonical miRNA maturation [4]. Additionally, another two kinds of mirtrons, i.e. 5′ and 3′ tailed introns, were identified in mouse [13,18]. Different from the canonical mirtrons as mentioned above, the processing of these mirtron precursors requires 5′-to-3′ and 3′-to-5′ trimming before they are subjected to Dicing respectively [14]. In this regard, the mirtron-generating introns were retained, and were classified into three catalogues (i.e. "match both ends", "match 5′ end", and "match 3′ end"; see "Materials and Methods" for details) based on HTS data mapping results.

Although the canonical miRNA/miRNA* duplex possesses 2-nt 3′ overhangs at both ends which result from Drosha/Dicer- (in animals) or DCL1- (in plants) mediated cropping [1,4,5], recent studies on mirtrons in animals showed a wide-spread scene of unusual configuration of the duplex overhangs [6]. Thus, the mature mirtron duplex candidates were selected from the perfectly matched short reads (see "Materials and Methods" for details), and mapped onto the secondary structures of the corresponding introns for manual check. Only the introns that could generate short mirtron duplexes with 0 to 3-nt 3′ overhangs at both ends were finally considered to be the mirtron candidates. As a result, five and eighteen mirtron candidates were identified in *Arabidopsis* and rice, respectively (Figure 1 and Figure S2). The previously reported putative mirtron in rice, osa-MIR1429 [15], was also uncovered in this study, indicating the reliability of our HTS- and structure-based filtering criteria. Notably, one out of the five mirtrons in *Arabidopsis*, and 12 out of 18 in rice reside within the introns of the transposable element (TE) genes (Table S3 and S4). Thus, whether the introns embedded within the TE genes are the hotspots for the birth mirtrons especially in rice needs to be investigated. Moreover, different from the canonical miRNA/ miRNA* duplexes, a dominant portion of the mirtrons (17 out of 23) generate mature mirtron duplexes with indistinguishable expression levels. Thus, it is hard to tell the mature and the star species separately. Except for the putative mirtron osa-MIR1429, nearly all the mirtrons are expressed at considerably low levels with normalized read counts less than 10 RPM (reads per million) (Figure S2). Many newly evolved miRNA genes with low conservation were also observed to be expressed weakly [19]. From this point of view, whether the mirtron candidates identified here have evolved recently, and whether they could serve as a novel source for new canonical miRNA genes through a TE gene-mediated pathway need to be studied.

Degradome Sequencing Data-based Identification of Mirtron Targets

Since the expression levels of most mature mirtrons generated from 5′ and 3′ arms of the precursors are indistinguishable, all the mature mirtron candidates identified on the stem-loop-structured precursors (all were marked in Figure S2; see sequence lists in Table S3 and S4) were recruited for functional analysis.

First, the sequence characteristics of all the mature mirtrons were analyzed. Different from the ~21-nt miRNAs predominantly starting with 5′ U (uridine), a dominant portion of mature mirtrons are 24 nt in length, and begin with 5′ G (guanine) and 5′ A (adenosine) in both *Arabidopsis* and rice (Figure 2).

A transcriptome-wide target prediction was performed by using miRU algorithm [20,21]. Then, degradome sequencing data-based validation of these predicted targets was carried out by employing t-plot (target plot)-based approach [22,23] (see details in "Materials and Methods"). As a result, dozens of mirtron—target pairs were identified, most of which were supported by compelling cleavage signals in the middle of the target recognition sites of the mature mirtrons (Figure 3, Figure S3 and S4, and Table S5 and S6), indicating their target cleavage-based role in PTGS. Intriguingly, some targets in rice, such as LOC_Os03g40600.1 (Figure 3D), LOC_Os02g48390.1, and LOC_Os04g45665.1 (Figure S4), were found to be cleaved at two different sites of the transcripts by distinct mirtrons, which could be defined as mirtron-mediated co-regulation. It is likely that the co-regulation mechanism could serve to enhance the efficiency of the mirtron-involved PTGS, although it still needs experimental validation.

According to the gene annotations provided by TAIR and TIGR rice, only a few targets of rice mirtron encode transcription factors (TFs) (Table S6), which is different from the plant miRNAs that target numerous TF genes [3]. Instead, several mirtron targets were annotated to be involved in RNA metabolism in rice (e.g. LOC_Os03g62080.1, LOC_Os06g31210.1, and LOC_Os07g46600.1), and photosynthesis in both plants (e.g. AT2G30570.1 in Arabidopsis, and LOC_Os02g10390.1/2/3 in rice) (see details in Table S5 and S6). More interestingly, three target genes in *Arabidopsis* (AT1G36360.1, AT4G10460.1, and AT5G33386.1) and four in rice (LOC_Os01g05030.1, LOC_Os01g22770.1, LOC_Os04g16010.1, and LOC_Os09g09820.1) encode TEs (see details in Table S5 and S6). Considering the origination of many mirtrons from the introns of TE genes (e.g. Arab_mirtron_candidate_10/11 from AT4G05280.1; Rice_mirtron_candidate_1/2 from LOC_Os03g36170.1, Rice_mirtron_candidate_3/4/5 from LOC_Os04g09380.1, Rice_mirtron_candidate_16/17/18/19/20 from LOC_Os07g32220.1, etc. See details in Table S3 and S4), this observation raised a feedback regulatory circuit that TEs might be hotspots for mirtron generation, and in turn, the corresponding mirtrons could strictly modulate the expression of these TE genes at the transcriptional or the post-transcriptional level.

Taken together, the mirtrons and their targets identified based on our bioinformatics approach provide a basis for further experimental studies on the origin and the functions of the mirtrons in plants.

Materials and Methods

Data Sets Used in this Study

The sRNA HTS data sets of *Arabidopsis* and rice were retrieved from GEO (Gene Expression Omnibus; http://www.ncbi.nlm.nih.gov/geo/) [24] and CSRDB (Cereal Small RNAs Database; http://sundarlab.ucdavis.edu/smrnas/) [25]. See Table S1 for the accession numbers. The degradome sequencing data sets were retrieved from GEO and NGSDBs (Next-Gen Sequence Databases; http://mpss.udel.edu/) [26]. See Table S2 for the accessions. The gene annotation and sequence information of *Arabidopsis* and rice were retrieved from the FTP sites of The *Arabidopsis* Information Resource (TAIR, Release 7 and 10; ftp://ftp.arabidopsis.org/home/tair/Sequences/blast_datasets/) [27] and the rice genome annotation project established by The Institute for Genome Research (currently named the J. Craig Venter institute) (TIGR rice, Release 5 and 6.1; ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/) [28], respectively. The miRNA precursor sequences of the 15 plant species were downloaded from miRBase (Release 17; http://www.mirbase.org/) [16].

Secondary Structure- and HTS Data-based Identification of Mirtron Candidates

Considering the average length of the pre-miRNAs (precursor microRNAs) of the currently annotated miRNA genes [according to miRBase (Release 17; http://www.mirbase.org/) [16]], the intron sequences retrieved from TAIR and TIGR rice that less than 300 nt in length were subjected to secondary structure prediction by using RNAshapes [17] in "Shape folding" mode with default parameters. The simplest structure (with single stem-loop region in most cases) among all the predicted results of an intron sequence was selected for manual check. The introns capable of forming stable hairpin-like structures were retained. Then, the sRNA HTS data were mapped onto these selected introns by BLAST algorithm [29], and all the perfectly matched ones were retained. In order to allow cross-library
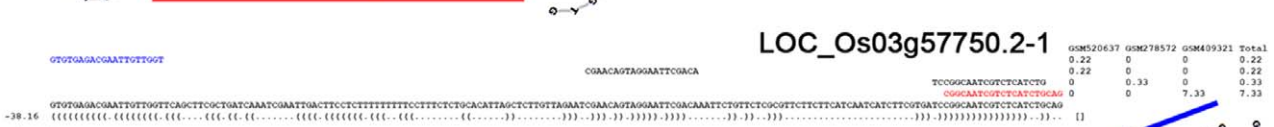
A — AT3G60950.1-15

	GSM343002	GSM366867	GSM518432	GSM342999	GSM343004	GSM442934	GSM343005	GSM304283	GSM442935	GSM642338	GSM642337	GSM366866	GSM442933	GSM642335	GSM366865	Total
GTTCATTGTTTAAACATGTGGGAC	0.28	1.14	0	0	0	0	0	0	0	0	0	0	0	0	0	1.42
TTCATTGTTTAAACATGTGGGAC	0	0	0.24	0.38	0	0	0	0	0	0	0	0	0	0	0	0.62
TTCATTGTTTAAACATGTGGGACA	0	0	0	0	0.91	0	0	0	0	0	0	0	0	0	0	0.91
ATTGTTTAAACATGTGGGACATTT	0	0	0	0	0	0.53	0	0	0	0	0	0	0	0	0	0.53
ATTGTTTAAACATGTGGGAC	0	0	0	0.38	0	0	0	0	0	0	0	0	0	0	0	0.74
ATTGTTTAAACATGTGGGAC	0	0	0	0	0	0.36	0	0.68	0	0	0	0	0	0	0	0.68
AAACATGTGGGACATTTTTGT	0	0	0	0	0	0	0.27	0	0	0	0	0	0	0	0	0.27
ACATGTGGGACATTTTTGTATTT	0	0	0	0	0	0	0	0	0.15	0	0	0	0	0	0	0.15
ATGTGGGACATTTTGTATTTGT	0	0	0	0	0	0	0	0	0	0.13	0	0	0	0	0	0.13
TTGTGTTACAATTGTTTTGGTTA	0.28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.28
GTGTTACAATTGTTTTGGTTA	0	0	0	0	0	0	0	0	0	0	0.43	0	0	0	0	0.43
GTTTGGTTATAGAAAAGTGG	0	0	0	0	0	0	0	0	0	0	0	0.29	0	0	0	0.29
ATAGAAAAGTGGGCTCCCTGTGA	0	0	0	0	0	0	0	0	0	0	0	0	0	0.76	0	0.76
GTTCATTGTTTAAACATGTGGGACATTTTGTATTTGTGTTACAATTGTTTTGGTTATAGAAAAGTGGGCTCCCTGTGAAAAG	0	0	0	0	0	0.68	0	0	0	0	0	0	0	0	0.58	1.26

-11.00 ((((...(((((((....(((((....(((((....)))))..)))))).-))).-)))).....))))).......... []

B — AT1G76680.1-2

	GSM366866	GSM304283	GSM366870	GSM304284	Total
GTAAGCACATTTTGGATTTGGA	0.86	0	0	0	0.86
GCACATTTTGGATTTGGA	0	0.68	0	0	0.68
TGTGTCCTTGTTGTTGCAG	0	0	1.74	0	1.74
TGTGTCCTTGTTGTTGCA	0	0	3.67	0	3.67
GTAAGCACATTTTGGATTTGGAATTCAAGAAGCTGAAACCATTCAAGATTCTTATATGTGTCCTTGTTGTTGCAG	0	0	0	0	

-13.10 ....(((((.....(((((..((..(((((.(((((....)))))...))))))..))..)).)))))...)))).. []

C — AT4G27410.1-1

	GSM642338	GSM442934	Total
GTAAATAAAGACTGAAACTTTCC	0.2	0	0.2
GTAAATAAAGACTGAAACTTTCCCAGATC	0.05	0	0.05
ATAAAGTTTGCTTCTTTGTTACAG	0	0.26	0.26
AAAGTTTGCTTCTTTGTTACAG	0.6	0	0.6
GTAAATAAAGACTGAAACTTTCCCAGATCCAAATTCGTTAAAATTGAAATCAAAATCTGATAAAGTTTGCTTCTTTGTTACAG			

-17.10 ((.(((((((((....(((((((..(((((.....(((.......)))....)))))....)))))..))))))))...)))))))))). []

D — LOC_Os04g09380.1-5

	GSM693280	GSM693279	GSM571078	GSM278572	GSM278571	GSM520635	GSM520637	GSM520634	GSM520638	GSM686039	GSM520636	Total
GTGGAGCACGTGGATGATCAGGAA	0.75	1.1	0.49	0.33	0.92	0	0	0	0	0	0	3.59
GTGGAGCACGTGGATGATCAGGA	0.11	0.1	0	0	0	0	0	0	0	0	0	0.21
TGGAGCACGTGGATGATCAGGAA	0.11	0.1	0	0	0	0	0	0	0	0	0	0.21
TGGAGCACGTGGATGATCAGGA	0.11	0	0	0	0	0	0	0	0	0	0	0.11
TGGAGCACGTGGATGATCAGGA	0.22	0	0	0.33	0	0	0	0	0	0	0	0.55
TGGAGCACGTGGATGATCA	0.11	0	0	0	0	0	0	0	0	0	0	0.11
TGGAGCACGTGGATGATCAGG	0	0	0	0	0.31	0	0	0	0	0	0	0.31
TGGAGCACGTGGATGATCAGGAAA	0.32	0	0	0.33	0	0	0	0	0	0	0	0.65
GGAGCACGTGGATGATCAGGAAAC	0.11	0	0	0	0	0	0	0	0	0	0	0.11
GAGCACGTGGATGATCAGGAAACT	0	0	0	0	0.31	0	0	0	0	0	0	0.31
AGCACGTGGATGATCAGGAAACT	0	0	0	0	0.31	0	0	0	0	0	0	0.31
AGCACGTGGATGATCAGGAAAC	0	0.1	0	0	0	0	0	0	0	0	0	0.1
GGATGATCAGGAAACTAAAATGT	0	0	0	0	0	0.16	0	0	0	0	0	0.16
AATAATGTTCATGATACTGTTCAG	0	0	0	0	0	0	0.22	0	0	0	0	0.22
ATGATACTGTTCAGCACTCACCTC	0	0	0	0	0	0	0	0.37	0.2	0	0	0.57
GATACTGTTCAGCACTCACCTCCT	0	0	0	0	0	0.16	0	0	0	0	0	0.16
TACTGTTCAGCACTCACCTCCTGT	0	0	0	0	0	0	0	0	0	0.08	0	0.08
ACTGTTCAGCACTCACCTCCTGTT	0	0	0	0	0	0.16	0	1.47	0	0.94	0	2.57
GTGGAGCACGTGGATGATCAGGAAACTAAAATGTTGTTAGTAATAAATGTTCATGATACTGTTCAGCACTCACCTCCTGTTTGTAG												

-13.61 ..(((((...(((...(((.((...(((..........(((((..))))))........))).-))).-))).....)))) []

E — LOC_Os09g04260.1-13

	GSM686040	GSM693279	GSM329296	Total
CAAGACCAAGTTGATTATGCCG	0.08	0	0	0.08
ATGCCTTGATCAACCAATCT	0	0.1	0	0.1
CCTTGATCAACCAATCTGGAG	0	0	8.15	8.15
GTGCACAAATTTCCAAGACCAAGTTGATTATGCCGTACAACATGCCTTGATCAACCAATCTGGAG				

-12.20 ..........(((.-(((.....((((((((((..((............))..)))))))))...)))))).. []

LOC_Os03g57750.2-1

	GSM520637	GSM278572	GSM409321	Total
GTGTGAGACGAATTGTTGGT	0.22	0	0	0.22
CGAACAGTAGGAAATTCGACA	0.22	0	0	0.22
TCCGGCAATCGTCTCATCTG	0	0.33	0	0.33
CGGCAATCGTCTCATCTGCAG	0	0	7.33	7.33

GTGTGAGACGAATTGTTGGTTCAGCTTCGCTGATCAAATCGAATTGACTTCCTCTTTTTTTTTCCTTTCTCTGCACATTAGCTCTTGTTGAATGCAACAGTAGGAAATTCGACAAATTCGTTCTCGCGTTCTTCTTCATCAATCATCTTCGTGATCCGGCAATCGTCTCATCTGCAG

-38.16 ((((((((((.(((((((...(((...((.((...(((((((.(((.-((((.....((...))..))).-)))).)))))))......)).-))..)))...))))))).....)))))))))))..)).-.. []
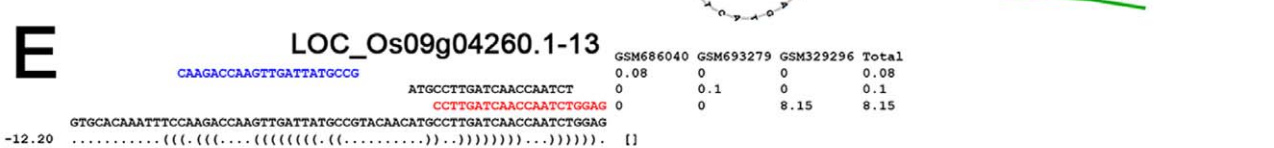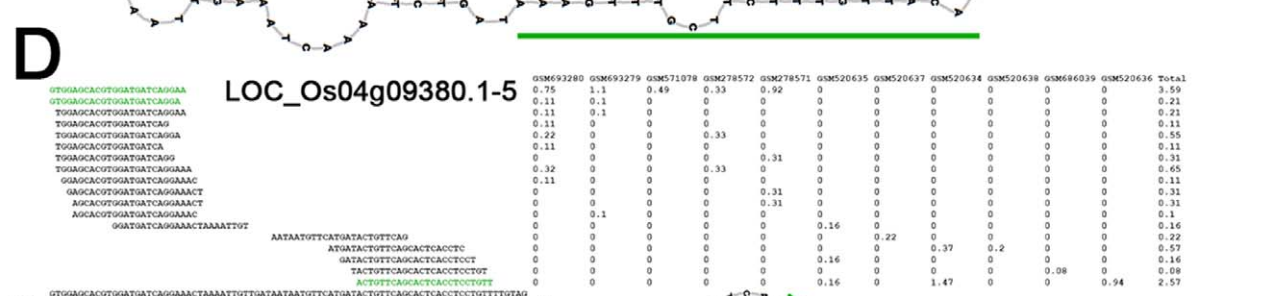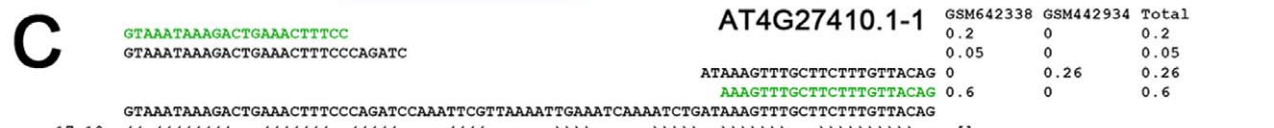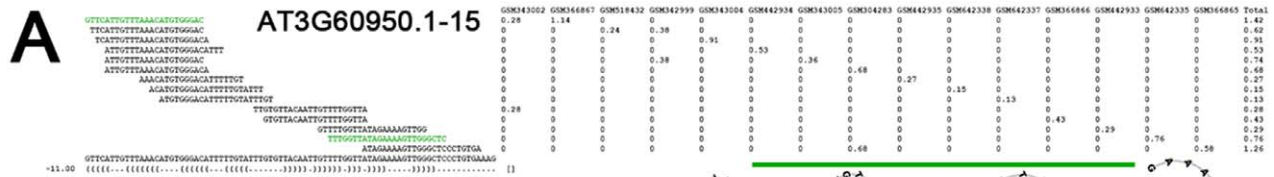
Figure 1. High-throughput sequencing (HTS) data- and structure-based identification of mirtrons in *Arabidopsis* and rice. (A) A "match 5' end" mirtron resided within the 15th intron of AT3G60950.1. (B) A "match 3' end" mirtron resided within the second intron of AT1G76680.1. (C) A "match both ends" mirtron resided within the first intron of AT4G27410.1. (D) A "match 5' end" mirtron resided within the 5th intron of LOC_Os04g09380.1. (E) A "match 3' end" mirtron resided within the 13th intron of LOC_Os09g04260.1. (F) A "match both ends" mirtron resided within the first intron of LOC_Os03g57750.2. For all the panels, the short reads perfectly mapped to the mirtron precursors along with their normalized read counts in RPM (reads per million) are shown (see Table S1 for the small RNA HTS data sources and see "Materials and Methods" for read count normalization). The mature mirtrons with significantly higher expression levels compared to the coordinates on the other arms were highlighted in red color, and the coordinates were in blue. For the mirtron precursors generating mirtrons with indistinguishable expression levels on both arms, their mature mirtrons were highlighted in green color. The mature mirtrons and their coordinates were also indicated in the stem-loop structures of their precursors. The parenthesis-dot formed secondary structure expression along with the free energy, and the stem-loop structures were all predicted and generated by RNAshapes [17].
doi:10.1371/journal.pone.0031163.g001

comparison, the normalized read count (in RPM, reads per million) of a short read from a specific library was calculated by dividing the raw count of this read by the total counts of the library, and then multiplied by 10^6. According to the mapping results, the stem-loop-structure introns were classified into three categories: (1) "Match both ends": the introns with HTS short reads perfectly mapped onto both the 5' and the 3' ends; (2) "Match 5' end": with reads mapped onto the 5' ends of the introns; (3) "Match 3' end": with reads mapped onto the 3' ends of the introns. Finally, the mature mirtron candidates along with their partners on the other ends were selected from the short reads clusters on the corresponding introns, and mapped onto the stem-loop structures for manual check. For the "match both ends" introns, the short reads mapped to the two ends of the introns were selected as mature mirtron candidates. For the "match 5' end" and the "match 3' end" introns, the reads mapped to the 5' ends and the 3' ends of the introns were considered as one of the mature mirtron candidates of the

corresponding introns, respectively. And, the remaining candidates on the other arms of the stem-loop-structured introns were selected manually. For all three categories of the introns, the ones possessing pairs of mature mirtron candidates with 0 to 3-nt 3' overhangs at both ends were finally considered to be the mirtron candidates.

Prediction and Validation of the Targets of the Mirtrons

Target prediction was performed by using miRU algorithm [20,21] with default parameters. The degradome sequencing data were utilized to validate the predicted mirtron—target pairs. First, the read counts of all the degradome reads from each library were normalized as described in the above section. Then, two-step filtering was performed to extract the most likely mirtron—target pairs. During the first step, the predicted mirtron binding sites along with the 50-nt surrounding sequences at both ends were collected in order to reduce the BLAST time. For the BLAST, all the collected degradome data sets (eleven of *Arabidopsis* and four of
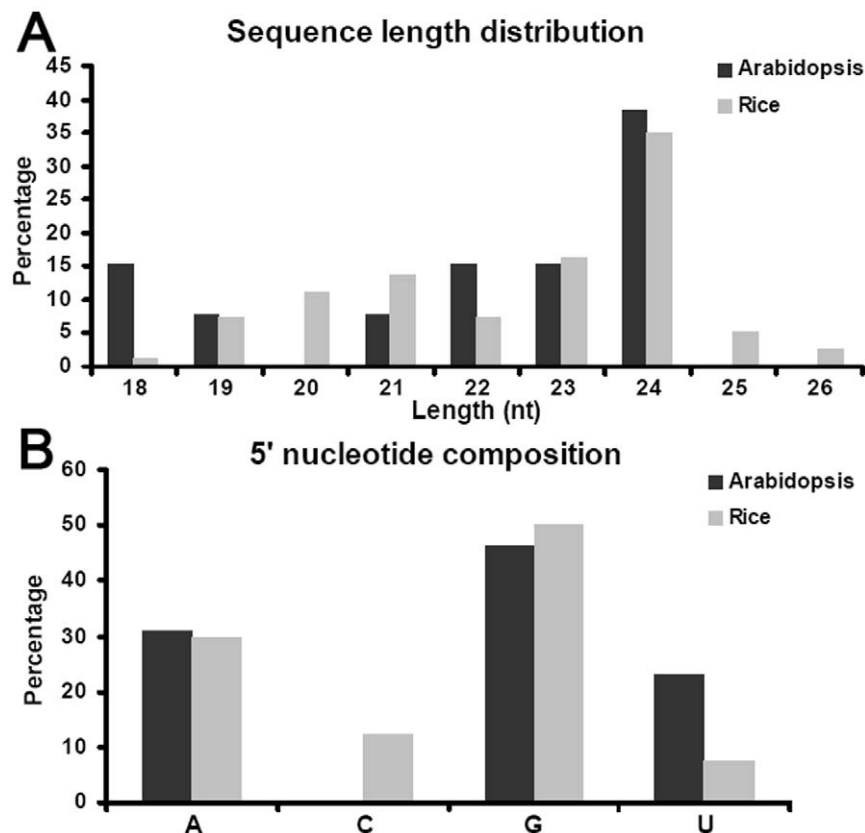

Figure 2. Sequence characteristics of the mature mirtrons identified in *Arabidopsis* and rice. (A) Sequence length distribution patterns. (B) 5' terminal nucleotide compositions. See Table S3 and S4 for sequence information.
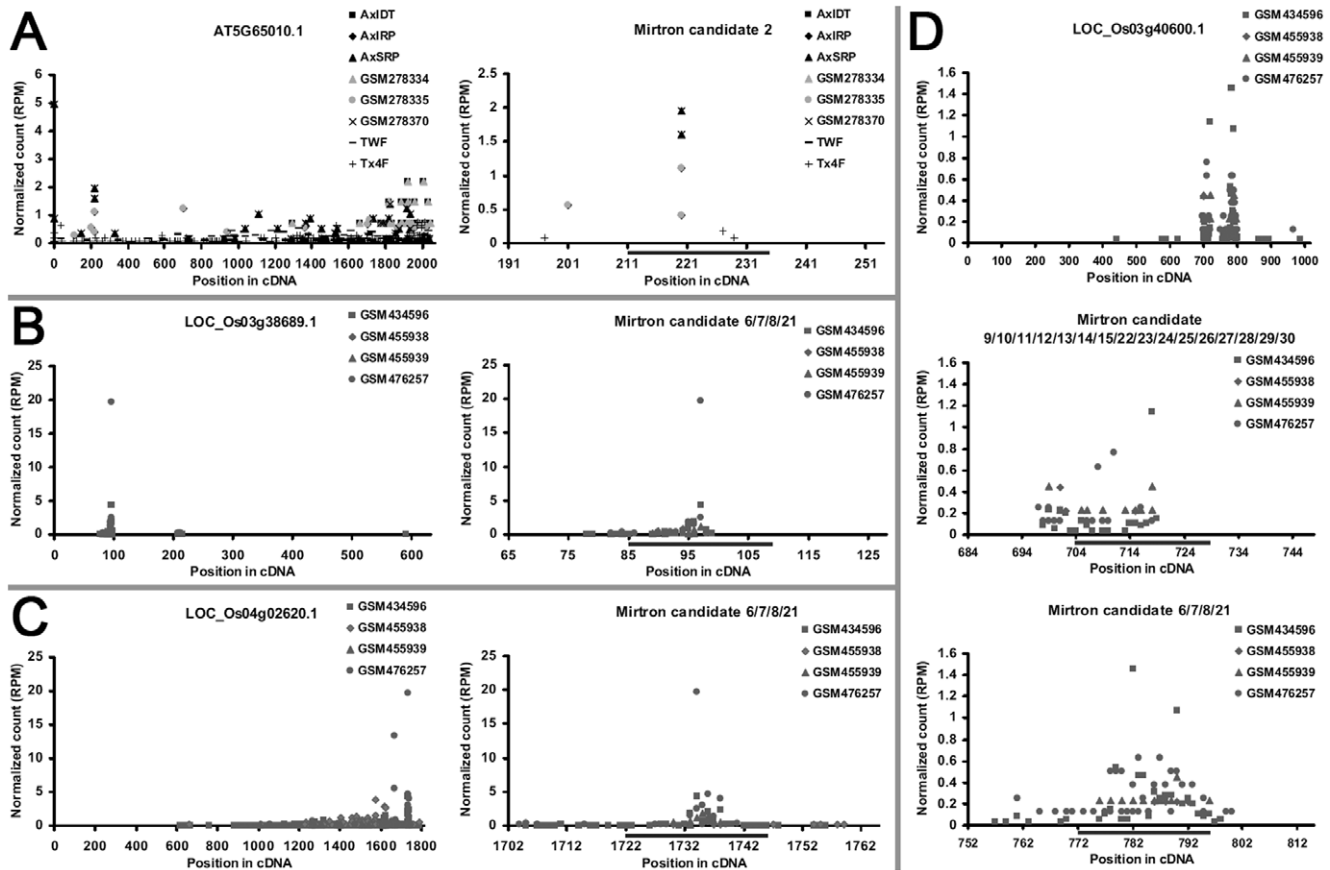doi:10.1371/journal.pone.0031163.g002

Figure 3. Degradome sequencing data-based identification of the targets of the mature mirtrons in *Arabidopsis* and rice. For all the sub-figures (A to D), the first panels depict the degradome signals all along the target transcripts, and the other panels provide detailed views of the cleavage signals within the regions surrounding the target recognition sites (denoted by gray horizontal lines). The transcript IDs are shown in the first panels, and the mirtron IDs are listed in the other panels (see Table S3 and S4 for the sequence information corresponding to the mirtron IDs). The *x* axes measure the positions of the signals along the transcripts, and the *y* axes measure the signal intensities based on normalized counts (in RPM, reads per million), allowing cross-library comparison. See Table S2 for the degradome data sets used in this analysis.
doi:10.1371/journal.pone.0031163.g003

rice; see Table S2) were utilized at the same time to do a comprehensive search. It was based on the scenario that a mirtron—target pair was considered to be the candidate once the cleavage signal(s) existed in any data set(s). Two types of predicted targets were retained for further filtering: (1) there must be perfectly matched degradome reads with their 5′ ends resided within 8–14 nt region away from the 5′ ends of the target binding sites; or (2) the target transcripts should possess degradome reads at least partially located within the target binding sites, and their normalized counts should be significantly higher than the surrounding signals. These transcripts were subjected to a second BLAST, and the degradome signals along each transcript 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 [30], both the global and the local t-plots were drawn. Exhaustive manual filtering was performed, and only the transcripts with cleavage signals easy to be recognized were extracted as the potential mirtron—target pairs.

Supporting Information

Figure S1 Sequence length distribution of the miRBase-registered microRNA precursors belonging to 15 plant species. The *x* axis marks the sequence length range, and the *y* axis measures the percentage of the microRNA precursors resided within a specific length range in a plant.
(PDF)

Figure S2 All the mirtrons in *Arabidopsis* and rice. The short reads perfectly mapped to the mirtron precursors along with their normalized read counts in RPM (reads per million) are shown (see Table S1 for the small RNA HTS data sources and see METHODS in the text for read count normalization). The mature mirtrons with significantly higher expression levels compared to the coordinates on the other arms were highlighted in red color, and the coordinates were in blue. For the mirtron precursors generating mirtrons with indistinguishable expression levels on both arms, their mature mirtrons were highlighted in green color. The mature mirtrons and their coordinates were also indicated in the stem-loop structures of their precursors. The parenthesis-dot formed secondary structure expression along with the free energy, and the stem-loop structures were all predicted and generated by RNAshapes (Steffen *et al.*, 2006).
(PDF)

Figure S3 Degradome sequencing data-based identification of the targets of the mature mirtrons in *Arabidopsis*. For all the sub-figures, the first panels depict the degradome signals all along the target transcripts, and the other

panels provide detailed views of the cleavage signals within the regions surrounding the target recognition sites (denoted by blue horizontal lines). The transcript IDs are shown in the first panels, and the mirtron IDs are listed in the other panels (see Table S3 and S4 for the sequence information corresponding to the mirtron IDs). The x axes measure the positions of the signals along the transcripts, and the y axes measure the signal intensities based on normalized counts (in RPM, reads per million), allowing cross-library comparison. See Table S2 for the degradome data sets used in this analysis.
(PDF)

Figure S4 Degradome sequencing data-based identification of the targets of the mature mirtrons in rice. For all the sub-figures, the first panels depict the degradome signals all along the target transcripts, and the other panels provide detailed views of the cleavage signals within the regions surrounding the target recognition sites (denoted by blue horizontal lines). The transcript IDs are shown in the first panels, and the mirtron IDs are listed in the other panels (see Table S3 and S4 for the sequence information corresponding to the mirtron IDs). The x axes measure the positions of the signals along the transcripts, and the y axes measure the signal intensities based on normalized counts (in RPM, reads per million), allowing cross-library comparison. See Table S2 for the degradome data sets used in this analysis.
(PDF)

Table S1 Plant small RNA high-throughput sequencing data sets used in this study.
(PDF)

Table S2 Plant degradome sequencing data sets used in this study.
(PDF)

Table S3 List of the mirtrons identified in *Arabidopsis*. The sequences of the mature mirtrons, the IDs of the mirtron-generating introns (for example, AT3G60950.1-15 represents the 15th intron resided within the transcript AT3G60950.1), and the corresponding annotations (TAIR 10) of the mirtron-containing genes are provided.
(XLS)

Table S4 List of the mirtrons identified in rice. The sequences of the mature mirtrons, the IDs of the mirtron-generating introns (for example, LOC_Os03g36170.1-6 represents the 6th intron resided within the transcript LOC_Os03g36170.1), and the corresponding annotations (TIGR rice 6.1) of the mirtron-containing genes are provided.
(XLS)

Table S5 List of the target genes of the mature mirtrons in *Arabidopsis*. The mirtron IDs, the transcript IDs of the target genes, and the target gene annotations (TAIR 10) are provided.
(XLS)

Table S6 List of the target genes of the mature mirtrons in rice. The mirtron IDs, the transcript IDs of the target genes, and the target gene annotations (TIGR rice 6.1) are provided.
(XLS)

References

1. Carthew RW, Sontheimer EJ (2009) Origins and mechanisms of miRNAs and siRNAs. Cell 136: 642–655.
2. Chen X (2009) Small RNAs and their roles in plant development. Annu Rev Cell Dev Biol 25: 21–44.
3. Jones-Rhoades MW, Bartel DP, Bartel B (2006) MicroRNAs and their regulatory roles in plants. Annu Rev Plant Biol 57: 19–53.
4. Kim VN, Han J, Siomi MC (2009) Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol 10: 126–139.
5. Voinnet O (2009) Origin, biogenesis, and activity of plant microRNAs. Cell 136: 669–687.
6. Berezikov E, Chung WJ, Willis J, Cuppen E, Lai EC (2007) Mammalian mirtron genes. Mol Cell 28: 328–336.
7. Berezikov E, Liu N, Flynt AS, Hodges E, Rooks M, et al. (2010) Evolutionary flux of canonical microRNAs and mirtrons in Drosophila. Nat Genet 42: 6–9.
8. Brown JW, Marshall DF, Echeverria M (2008) Intronic noncoding RNAs and splicing. Trends Plant Sci 13: 335–342.
9. Chung WJ, Agius P, Westholm JO, Chen M, Okamura K, et al. (2011) Computational and experimental identification of mirtrons in Drosophila melanogaster and Caenorhabditis elegans. Genome Res 21: 286–300.
10. Glazov EA, Cottee PA, Barris WC, Moore RJ, Dalrymple BP, et al. (2008) A microRNA catalog of the developing chicken embryo identified by a deep sequencing approach. Genome Res 18: 957–964.
11. Miyoshi K, Miyoshi T, Siomi H (2010) Many ways to generate microRNA-like small RNAs: non-canonical pathways for microRNA production. Mol Genet Genomics 284: 95–103.
12. Okamura K, Hagen JW, Duan H, Tyler DM, Lai EC (2007) The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. Cell 130: 89–100.
13. Ruby JG, Jan CH, Bartel DP (2007) Intronic microRNA precursors that bypass Drosha processing. Nature 448: 83–86.
14. Westholm JO, Lai EC (2011) Mirtrons: microRNA biogenesis via splicing. Biochimie. doi:10.1016/j.biochi.2011.06.017.
15. Zhu QH, Spriggs A, Matthew L, Fan L, Kennedy G, et al. (2008) A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains. Genome Res 18: 1456–1465.
16. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ (2008) miRBase: tools for microRNA genomics. Nucleic Acids Res 36: D154–158.
17. Steffen P, Voss B, Rehmsmeier M, Reeder J, Giegerich R (2006) RNAshapes: an integrated RNA analysis package based on abstract shapes. Bioinformatics 22: 500–503.
18. Babiarz JE, Ruby JG, Wang Y, Bartel DP, Blelloch R (2008) Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. Genes Dev 22: 2773–2785.
19. Axtell MJ, Bowman JL (2008) Evolution of plant microRNAs and their targets. Trends Plant Sci 13: 343–349.
20. Dai X, Zhao PX (2011) psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res 39: W155–159.
21. Zhang Y (2005) miRU: an automated plant miRNA target prediction server. Nucleic Acids Res 33: W701–704.
22. German MA, Luo S, Schroth G, Meyers BC, Green PJ (2009) Construction of Parallel Analysis of RNA Ends (PARE) libraries for the study of cleaved miRNA targets and the RNA degradome. Nat Protoc 4: 356–362.
23. German MA, Pillay M, Jeong DH, Hetawal A, Luo S, et al. (2008) Global identification of microRNA—target RNA pairs by parallel analysis of RNA ends. Nat Biotechnol 26: 941–946.
24. Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, et al. (2009) NCBI GEO: archive for high-throughput functional genomic data. Nucleic Acids Res 37: D885–890.
25. Johnson C, Bowman L, Adai AT, Vance V, Sundaresan V (2007) CSRDB: a small RNA integrated database and browser resource for cereals. Nucleic Acids Res 35: D829–833.
26. Nakano M, Nobuta K, Vemaraju K, Tej SS, Skogen JW, et al. (2006) Plant MPSS databases: signature-based transcriptional resources for analyses of mRNA and small RNA. Nucleic Acids Res 34: D731–735.
27. Huala E, Dickerman AW, Garcia-Hernandez M, Weems D, Reiser L, et al. (2001) The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. Nucleic Acids Res 29: 102–105.
28. Yuan Q, Ouyang S, Liu J, Suh B, Cheung F, et al. (2003) The TIGR rice genome annotation resource: annotating the rice genome and creating resources for plant biologists. Nucleic Acids Res 31: 229–233.
29. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410.
30. Meng Y, Shao C, Chen M (2011) Toward microRNA-mediated gene regulatory networks in plants. Brief Bioinform;doi: 10.1093/bib/bbq091.