A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains

Qian-Hao Zhu,[1] Andrew Spriggs,[1] Louisa Matthew,[1] Longjiang Fan,[2] Gavin Kennedy,[1] Frank Gubler,[1] and Chris Helliwell[1,3]

[1]CSIRO Plant Industry, Canberra, ACT 2601, Australia; [2]Institute of Crop Science, Zhejiang University, Hangzhou 310029, China

Endogenous small RNAs, including microRNAs (miRNAs) and short-interfering RNAs (siRNAs), function as post-transcriptional or transcriptional regulators in plants. miRNA function is essential for normal plant development and therefore is likely to be important in the growth of the rice grain. To investigate the roles of miRNAs in rice grain development, we carried out deep sequencing of the small RNA populations of rice grains at two developmental stages. In a data set of ~5.5 million sequences, we found representatives of all 20 conserved plant miRNA families. We used an approach based on the presence of miRNA and miRNA* sequences to identify 39 novel, nonconserved rice miRNA families expressed in grains. Cleavage of predicted target mRNAs was confirmed for a number of the new miRNAs. We identified a putative mirtron, indicating that plants may also use spliced introns as a source of miRNAs. We also identified a miRNA-like long hairpin that generates phased 21 nt small RNAs, strongly expressed in developing grains, and show that these small RNAs act in *trans* to cleave target mRNAs. Comparison of the population of miRNAs and miRNA-like siRNAs in grains to those in other parts of the rice plant reveals that many are expressed in an organ-specific manner.

[Supplemental material is available online at www.genome.org. All small RNA sequences have been submitted to Gene Expression Omnibus in GenBank under the series identifier GSE11014. Transcripts of *TAS3a1*, *TAS3b1*, and *TAS3b2* have been submitted to GenBank under the accession nos. EU293143, EU293144, and EU293145, respectively. Primary transcripts of miR1850, miR1428e_3p, and their alternative splicing forms have been submitted to GenBank under accession nos. EU563913–EU563918. Novel miRNA sequences have been submitted to miRBase under accession nos. miR1846–miR1884.]

Two major classes of endogenous noncoding small RNAs (21–24 nt), microRNAs (miRNAs) and short-interfering RNAs (siRNAs), have been shown to be important regulators of gene expression in eukaryotes (Bartel 2004; Zamore and Haley 2005; Vazquez 2006). In plants, three types of siRNAs have been described: *trans*-acting siRNAs (ta-siRNAs), natural antisense transcript-derived siRNAs (nat-siRNAs), and repeat-associated siRNAs (ra-siRNAs) (Peragine et al. 2004; Vazquez et al. 2004; Borsani et al. 2005; Vazquez 2006). They are processed from either long bimolecular RNA duplexes generated by RNA-dependent RNA polymerases (RDRs) or single-stranded (ss) hairpin RNAs formed by inverted repeats (Peragine et al. 2004; Borsani et al. 2005; Yoshikawa et al. 2005; Zhang et al. 2007). Usually, ra-siRNAs are 24 nt long and DCL3 dependent, ta-siRNAs are 21 nt in length and DCL1 and DCL4 dependent, while nat-siRNAs are 21 or 24 nt in length and both DCL2 and DCL1 dependent (Vazquez 2006).

miRNAs are generated from primary miRNA transcripts (pri-miRNAs) that are generally transcribed by RNA polymerase II (Lee et al. 2004; Parizotto et al. 2004; Xie et al. 2005) and are capable of forming a characteristic hairpin structure by intramolecular pairing (Bartel 2004). Generally, two sequential RNase III enzyme-mediated cleavages are required to produce mature miRNAs. First, DICER-LIKE1 (DCL1, in plants) or Drosha (in animals, also known as RNASEN) cleaves near the base of the stem of the pri-miRNA to produce a miRNA precursor (pre-miRNA);

DCL1 (in plants) or Dicer (in animals, also known as DICER1) then cleaves at a second position near the loop of the pre-miRNA to generate a miRNA/miRNA* duplex (Bartel 2004; Kim and Nam 2006). Recently, an alternative to Drosha-Dicer miRNA biogenesis, the mirtron pathway, has been identified in *Drosophila melanogaster*, *Caenorhabditis elegans*, and some mammals in which spliced debranched introns with hairpin structures equivalent to pre-miRNAs enter the miRNA processing pathway to produce mature miRNAs without Drosha-mediated pri-miRNA cleavage (Berezikov et al. 2007; Okamura et al. 2007; Ruby et al. 2007). Both the miRNA and miRNA* of the duplex are methylated at their 3′ ends (by HUA ENHANCER 1 in plants) to maintain their correct size and/or to protect them from polyuridylation (Li et al. 2005; Yu et al. 2005). After being loaded into the RNA-induced silencing complex (RISC), the miRNA guides sequence-specific post-transcriptional repression of target mRNA(s) by degradation or translational repression (Bartel 2004; Jones-Rhoades et al. 2006).

The sequencing of small RNA populations in plants has established that there are 20 families of conserved miRNAs. An increasing number of miRNAs are being discovered in both plants and animals (Rajagopalan et al. 2006; Fahlgren et al. 2007; Landgraf et al. 2007; Molnar et al. 2007; Zhao et al. 2007; Lu et al. 2008; Morin et al. 2008; Sunkar et al. 2008). Most of these newly identified miRNAs are not conserved between species, suggesting a recent evolutionary origin (Fahlgren et al. 2007; Molnar et al. 2007). One route by which miRNAs are thought to evolve is through inverted duplication of target gene sequences (Allen et al. 2004). As such a locus evolves toward a miRNA locus, various

[3]Corresponding author.
E-mail chris.helliwell@csiro.au; fax 61-2-6246-5000.

intermediates, such as long hairpins, are expected to be formed. Eventually, a self-complementary transcript with a fold-back structure adapted to the miRNA biogenesis machinery will predominantly generate DCL1-dependent miRNAs (Allen et al. 2004).

The majority of nonconserved miRNAs exhibit low or tissue-specific expression. Deep sequencing technologies such as sequencing by synthesis (Margulies et al. 2005) or massively parallel signature sequencing (MPSS) (for review, see Meyers et al. 2006) have been used in rice for miRNA discovery (Johnson et al. 2007; Lu et al. 2008; Morin et al. 2008; Sunkar et al. 2008) and comparison of small RNA populations in various tissues, including inflorescence (Nobuta et al. 2007).

miRNAs have been shown to function in many plant processes including development, growth control, cell differentiation, abiotic, and biotic stress resistance (Jones-Rhoades et al. 2006; Mallory and Vaucheret 2006; Navarro et al. 2006). Many of the conserved miRNAs target mRNAs encoding transcription factors; mutations in the miRNA, overexpression of the miRNA, or disruption of the complementarity between the miRNA and its target mRNA usually results in dramatic phenotypic change (Mallory and Vaucheret 2006). The development of the rice grain involves the coordinate development of the embryo and endosperm, a complex series of events that would be expected to be regulated in part by miRNAs. To investigate the roles of known miRNAs in rice grain development and to identify potentially grain-specific miRNAs, we sequenced small RNA populations of developing rice grains using high-throughput sequencing technology. We generated ~5.5 million small RNA sequences that aligned to the rice genome, including representatives of all known conserved miRNA families. From this data set, we identified 39 new miRNA families that are predicted to target mRNAs encoding proteins with a diverse range of functions. In addition, our sequence data provides evidence that miR1429.2 could be a mirtron. We also identified a miRNA-like long hairpin locus expressed in developing grains that produces 21-nt small RNAs from nine phases and show that small RNAs from this long hairpin locus act in *trans* to cleave target mRNAs.

Results

The developing rice grain has a complex small RNA population

The rice grain undergoes a complex series of developmental events leading to the production of a mature embryo and endosperm. While rice grains take around 30 d to fully mature, the endosperm fills the embryo sac by 5 d after fertilization (DAF), and the morphology of the embryo is determined by 10 DAF (Itoh et al. 2005). Therefore, we sequenced small RNAs from a mixture of grains from 1–5 DAF or 6–10 DAF to determine the expression levels of known miRNAs and identify novel, potentially grain-specific miRNAs during these early stages of grain development. We obtained two datasets of 96,091 and 5,379,724 genome-matched small RNA reads using 454 Life Sciences (Roche) and Illumina sequencing (formerly Solexa sequencing), respectively (Table 1). Approximately 94% of these small RNAs are 20–24 nt in length with 21 and 24 nt as the major size classes (Supplemental Fig. 1), consistent with being products of cleavage by DCL enzymes. The 454 data set contained 77,193 unique sequences, of which 92% were singletons compared with 1,968,352 unique sequences and 82% singletons in the Illumina data set (Table 1). These results indicate that rice grain has a highly complex small RNA population to which repetitive sequences are the major contributors. An MPSS study of rice small RNAs also found that repetitive sequences were the major contributor to the rice small RNA population (Nobuta et al. 2007). By comparison, Rajagopalan et al. (2006) found only 65% singletons in a population of ~890,000 *Arabidopsis* small RNAs, which included ~340,000 unique sequences, suggesting that the rice small RNA population is more complex.

The conserved miRNAs are all expressed in developing rice grains

Previous computational prediction and cloning studies have identified ~300 miRNAs from 86 miRNA families in rice (Reinhart et al. 2002; Wang et al. 2004; Arazi et al. 2005; Liu et al. 2005; Sunkar et al. 2005, 2008; Luo et al. 2006; Lu et al. 2008; Morin et al. 2008). Of these 86 miRNA families, 20 are conserved between *Arabidopsis* and rice (Table 2). Of the 20 conserved miRNA families, 18 were detected in the 454 data set, while all 20 families were found in the Illumina data set (Supplemental Table 1). The miRNAs detected only in the larger sequencing data set, miR395 and miR399, are induced by low sulphate and phosphate-starvation conditions, respectively (Jones-Rhoades and Bartel 2004; Fujii et al. 2005; Chiou et al. 2006), and are not expected to be abundant in plants grown under normal nutrient conditions. We found that the frequency of sequencing varied widely between miRNA family members (Supplemental Table 2), suggesting functional divergence within the families. In contrast, of the 66 identified rice miRNA families not found in *Arabidopsis*, only 21 were represented in our data set (Table 2).

Table 1. Summary of high-throughput small RNA sequencing

Tissue	454 sequencing			Illumina sequencing		
	No. of sequences generated	No. of sequences matching the rice genome[a]	No. of unique sequences[a]	No. of sequences generated	No. of sequences matching the rice genome[a]	No. of unique sequences[a]
Shoot[b]	11,866	9,176	8,060	NA	NA	NA
Root[b]	21,697	14,691	12,249	NA	NA	NA
1–5 DAF grains	46,879	33,228	27,064	3,428,861	2,798,128	1,136,672
6–10 DAF grains	55,391	38,996	33,246	3,412,945	2,581,596	1,116,230
Total	135,833	96,091	77,193	6,841,806	5,379,724	1,968,352

[a]18 nt or longer.
[b]From 7-d-old seedlings.
NA, Not applicable.

Table 2. Detection of known rice miRNAs

| | | | | Classification of miRNA loci | | | |
| | No. of miRNA families | No. of miRNA loci | No. of miRNA families cloned | A: Consistent with DCL processing of a single-stranded precursor[a] | B: Consistent with DCL processing of a single-stranded precursor with no miRNA detected[b] | C: Consistent with DCL processing of a double-stranded RNA[c] | D: No small RNA detected from the precursor[d] |
Type of miRNA family							
Conserved in *Arabidopsis*[e]	20	138	20	107	6	2[f]	23
Nonconserved in *Arabidopsis*	66	163	21	30	21	83	29
Total	86	301	41	137	27	85	52

[a]Exact miRNA and/or miRNA* or their variants (offset by up to 2 nt) were sequenced and all small RNAs from the miRNA precursor are in the same direction.
[b]No miRNA or miRNA* or their variants were cloned, even though other small RNAs (in the same direction as the miRNA precursor) from the precursor were sequenced.
[c]Small RNAs derived from both strands or just from the antisense strand of the reported miRNA precursor were detected.
[d]No small RNAs derived from the miRNA precursor were sequenced. These loci might not be expressed, or are expressed in a tissue specific manner.
[e]Based on Jones-Rhoades et al. (2006); miR159 and miR319 are combined as a single family.
[f]Both loci belong to the miR395 family.

To determine the expression and identity of previously identified miRNA loci, we examined the sequencing frequency and distribution of small RNAs across precursor sequences, as this indicates how small RNAs are generated from a particular locus (Table 2). Processing of a pre-miRNA by DCL1 is expected to predominantly give rise to miRNA with a lower frequency of the miRNA* and possibly a low frequency of small RNAs from elsewhere in the precursor (but not antisense to the precursor). Of the 115 conserved miRNA loci from which we obtained sequences, 107 generated the expected miRNA (with or without the miRNA*), six loci gave rise to small RNAs other than the miRNA and miRNA*, but in the same direction as the expected miRNA (Supplemental Fig. 2A), and two loci of the miR395 family gave rise to small RNAs from their antisense strands (Supplemental Fig. 2B). In general, our small RNA sequencing results support the identity of the miRNA loci of the 20 conserved families. In contrast, evidence supporting the identity of some previously annotated nonconserved miRNA loci was weak, especially for those identified from rice callus (Luo et al. 2006) or by prediction methods (Wang et al. 2004). Small RNA sequences from the nonconserved and predicted miRNA loci were detected less frequently (Table 2; Supplemental Table 1). Of the 163 nonconserved miRNA loci, only 30 showed a small RNA distribution pattern consistent with that of a miRNA locus (Table 2). Of the remainder, 83 loci generated small RNAs from both strands, just from the antisense strand of the precursor, or showed an overlapping distribution pattern of small RNAs (Table 2; Supplemental Fig. 2B). Recent deep sequencing identified 24 miRNA families in rice (Lu et al. 2008; Morin et al. 2008; Sunkar et al. 2008). Ten of these miRNAs were present in our sequencing data (the precise miRNA was sequenced for only five), nine were not expressed in our grain samples, and five with multiple members were not well supported due to generation of small RNAs from both strands (Supplemental Table 2). Of the four miRNAs annotated in wheat (Yao et al. 2007) and predicted to be monocot specific, we did not find expression of three (tae-miR1122, tae-miR1126, and tae-miR1130) and found 0–4 reads of 19–24 nt variants of tae-miR1132, which mapped to 18S rRNA. Therefore, we are unable to confirm that there are rice homologs of these miRNAs from our data.

We observed a few cases where the miRNA* was sequenced more often than the miRNA (Supplemental Fig. 2C). For miR529

there were 22 reads of the miRNA and 594 reads of the miRNA*, consistent with the results obtained from the moss *Physcomitrella patens* (Axtell et al. 2007), where it is suggested that the annotated miR529* is likely the true miRNA. For miR408 we obtained 72 reads of the miRNA and 631 reads of the miRNA*; however, the strong conservation of miR408 across plant species (Supplemental Fig. 3A) and confirmed target mRNAs for the annotated miR408 in *Populus trichocarpa* (Lu et al. 2005) suggest that the annotation should not be changed. In some cases, the annotated miRNA was not the most abundant small RNA generated from the pre-miRNA. For instance, the reported miR820 (21 nt) was only sequenced once in our data set, whereas its 22-, 23-, and 24-nt variants had 386, 1397, and 2208 reads, respectively (Supplemental Fig. 2D).

miRNA expression changes during seed development

It has been shown that high-throughput sequencing has good reproducibility and can be used as a tool for miRNA expression profiling (Fahlgren et al. 2007). In the 454 data set, most known miRNAs were detected less than 10 times in each sample, making quantitative comparisons unreliable. In contrast, 18 of the 20 conserved miRNA families (except miR395 and miR398) and eight other previously annotated rice miRNA families had at least 10 reads in the Illumina data set (Supplemental Table 1), enabling comparison of the expression levels of these miRNAs between the two developmental stages (Fig. 1A). Most miRNAs were expressed approximately equally in the two libraries or more highly in 6–10 DAF grains than in 1–5 DAF grains, with the highest expression ratio of 18.4:1 observed for miR408 (Fig. 1A). miR528 was highly expressed in developing rice grains with 7.5-fold more expression in 6–10 DAF grains. miR159/miR319, miR393, miR529, and miR1318/1432 had lower expression in the 6–10 DAF grains than the 1–5 DAF grains. miR168a was the most abundant miRNA in our two sequencing datasets (Supplemental Table 1), accounting for 5.1% of all sequence reads matched with the rice genome and ~60% of the sequence reads from previously annotated miRNAs. miR168a was also the most abundant miRNA in previous high-throughput sequencing studies in rice (Johnson et al. 2007; Sunkar et al. 2008). We note that miR168a is not the most abundant miRNA in *Arabidopsis* (http://asrp.cgrb.oregonstate.edu/db/microRNA.html?fid=11/), where it targets
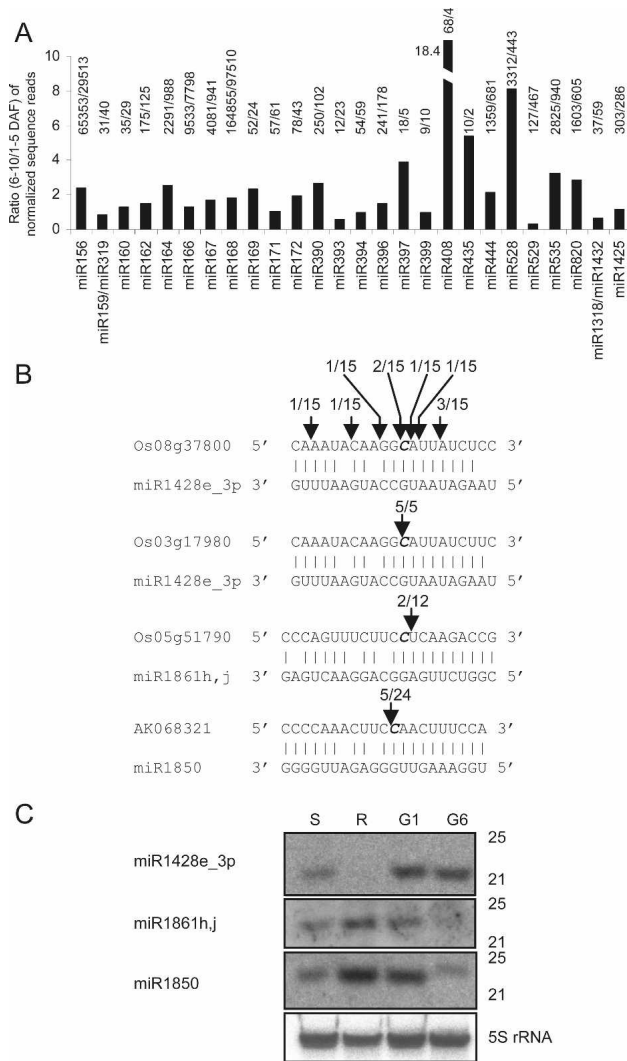
Figure 1. (*A*) Ratio (6–10 DAF/1–5 DAF) of the normalized miRNA sequence reads of the known miRNA families with at least 10 reads in total in the Illumina data set. The raw sequence reads of each miRNA family is shown on *top* of the bars with numerator and denominator representing number of reads in 6–10 DAF and 1–5 DAF grains, respectively. (*B*) Mapping of miRNA-guided cleavage sites in predicted target genes. Predicted cleavage sites are indicated by a bold italic nucleotide at position 10 relative to the 5′ end of the miRNA. Mapped cleavage positions are indicated by arrows with the frequency amongst 5′ RACE clones sequenced. (*C*) Northern blot detection of expression of miRNAs in different tissues. Uniform RNA loading was demonstrated by reprobing blots with 5S rRNA (representative blot shown). (S) Shoots; (R) roots; (G1) 1–5 DAF grains; (G6) 6–10 DAF grains.

ARGONAUTE 1 (*AGO1*), a key component of RISC of the miRNA pathway.

Novel miRNAs are present in the rice grain small RNA data set

We developed a bioinformatics pipeline to identify novel miRNAs from our sequence data set (see Methods for details). Briefly, genomic sequences flanked by two small RNA sequences on the same strand were extracted from the rice genome. The small RNA sequences had to be within 400 nt of each other and have less than 10 matches to the rice genome. Genomic regions

with a smeared distribution of small RNAs across the locus, or with many small RNAs generated from the antisense strand, were excluded. The genomic sequences were then tested for their ability to form a hairpin structure. One of the key criteria for miRNA identification by this procedure was the presence of the miRNA*. While our sequencing results support the identity of 137 known miRNA loci (Table 2), both miRNA and miRNA* were sequenced for only 56 of these loci (Supplemental Table 2); these 56 were identified by the prediction pipeline. Our method is conservative, as novel miRNAs sequenced without the corresponding miRNA* will not be identified; however, the predicted miRNAs should be products of DCL-processed hairpins and are thus good miRNA candidates. In addition to the known miRNAs, 324 potential miRNA/miRNA* pairs derived from intergenic regions or introns were identified. They were further checked manually to ensure that they were from good quality single-stranded hairpins and that a miRNA/miRNA* pair had 0–2 nt 3′ overhangs. Thirty-nine were designated miRNA families, and a further 15 were considered as candidate miRNAs because their miRNA and miRNA* were detected with less than five reads (Supplemental Table 3). For the candidate miRNA loci, both miRNA and miRNA* were annotated to represent the miRNA locus. None of these new miRNAs have been found in *Arabidopsis*. Predicted pre-miRNA sequences for these newly identified miRNAs and candidate miRNAs, their hairpin structures, and distribution of small RNAs are presented in Supplemental Figure 4.

For seven of the 39 novel miRNAs, additional family members were identified from the rice genome sequence. The miR1861 family has 14 members with identical or slightly different miRNA sequences. Ten members of the miR1861 family are found in five pairs of tandem miRNAs and the other four members have putative tandem miRNAs, which were not confirmed by our sequencing data (Supplemental Table 3; Supplemental Fig. 5B). In each pair, the two pre-miRNAs are separated by ~100 nt (Supplemental Fig. 5B,C). In all cases, the sequences of the first and second miRNA precursors in each pair are highly conserved, except that the second precursors have a 22-nt deletion immediately downstream of the mature miRNA (Supplemental Fig. 5A). As this deletion exists in all of the second precursors, we suggest that the original precursor was generated by a gene duplication event followed by a 22-nt deletion in the second precursor or a 22-nt insertion in the first precursor. This gene duplication was then followed by segmental duplications to generate other family members, which are always in tandem configuration. For miR1861i, its putative second precursor has diverged significantly and become almost unrecognizable (Supplemental Fig. 5B). In addition, there were two more tandem pairs on chromosome 5 (miR1861_5) and 7 (miR1861_7), which were not sequenced in our data set and have diverged in sequence from the other family members (Supplemental Fig. 5B). The conservation of the precursor sequences suggests that miR1861 is a relatively young miRNA family.

A tandem genomic configuration of conserved miRNA precursors was also identified for the miR1428 family (Supplemental Fig. 5D,E). Transcript cloning confirmed that miR1428d_3p and miR1428e_3p are expressed in the same transcript, which shows variable transcription start sites and 3′ polyadenylation sites (Supplemental Fig. 5F), a feature previously observed for some *Arabidopsis* miRNA loci (Xie et al. 2005; Nikovics et al. 2006). Some transcripts were splice variants that lacked the second precursor, pre-miR1428d_3p. Consistent with this, more sequence reads were found for miR1428e_3p (27 reads) than for

miR1428d_3p (one read). Alternative pri-miRNA splicing was also observed in miR1850 (Supplemental Fig. 5F). Other examples of tandem miRNAs were also identified, including pairs or triplets of unrelated miRNAs, such as miR1423b_5p and miR1868, and miR1862d, miR1876, and miR1884b (Supplemental Table 3).

We confirmed the size and abundance of a subset of the newly identified miRNAs in different tissues using Northern blots. miR1428e_3p has very low or no expression in roots but is highly expressed in grains, consistent with the small RNA sequencing results; miR1850 and miR1861 are expressed in grains, but are also abundant in shoots and roots (Fig. 1C).

Identification of the target mRNAs of newly identified miRNAs assists in understanding their biological roles. Of the 39 miRNA families we identified, 18 have at least one predicted target gene with a score of up to 3.5 (a maximum of three mismatches and a G-U pair; Supplemental Table 3). The majority of the 21- and 22-nt miRNAs have predicted targets, whereas more than half of the 24-nt miRNAs do not (Supplemental Table 3). These miRNAs may genuinely have no target, their target mRNAs may not be annotated in the rice genome, or their targets have greater mismatch to the miRNA sequence. Altogether, 95 target mRNAs were predicted; 10 encode transcription factors, 24 encode proteins without annotated functions, and 61 encode other proteins, including several protein kinases and leucine-rich repeat family proteins (Supplemental Fig. 6A,B). This is in contrast

with the targets of conserved miRNAs, ~2/3 of which encode transcription factors. Cleavage of selected target genes by the newly identified miRNAs has been confirmed using 5′ RACE (rapid amplification of 5′ complementary DNA ends). miR1428e_3p targets Os08g37800 and Os03g17980 mRNAs, both encoding kinases of the sucrose nonfermenting-related kinase 1 (SnRK1) family. Several members of the miR1861 family target Os05g51790, encoding a protein of the STRUCTURAL MAINTENANCE OF CHROMOSOMES family, in this case the cleavage site was 1 nt downstream from the expected site (Fig. 1B).

A miRNA-like long hairpin locus produces *trans*-acting phased small RNAs

A search for phased siRNAs generated from ta-siRNA loci identified an additional *TAS3* locus (see Supplemental material) and also identified two loci-encoding transcripts that can form long hairpin structures that are the source of large numbers of phased 21-nt sequences in our grain data set (Fig. 2A,B; Supplemental Fig. 7A,B). The first locus generates phased 21-nt siRNAs that have been found in other small RNA datasets (Johnson et al. 2007; Nobuta et al. 2007) and shown to be DCL4 dependent (Liu et al. 2007); one of these 21-nt siRNAs was originally annotated as miR436 (Sunkar et al. 2005) but is no longer considered a miRNA. We obtained 15,580 sequence reads from the miR436 hairpin, our analysis shows they are produced from 12 phases of
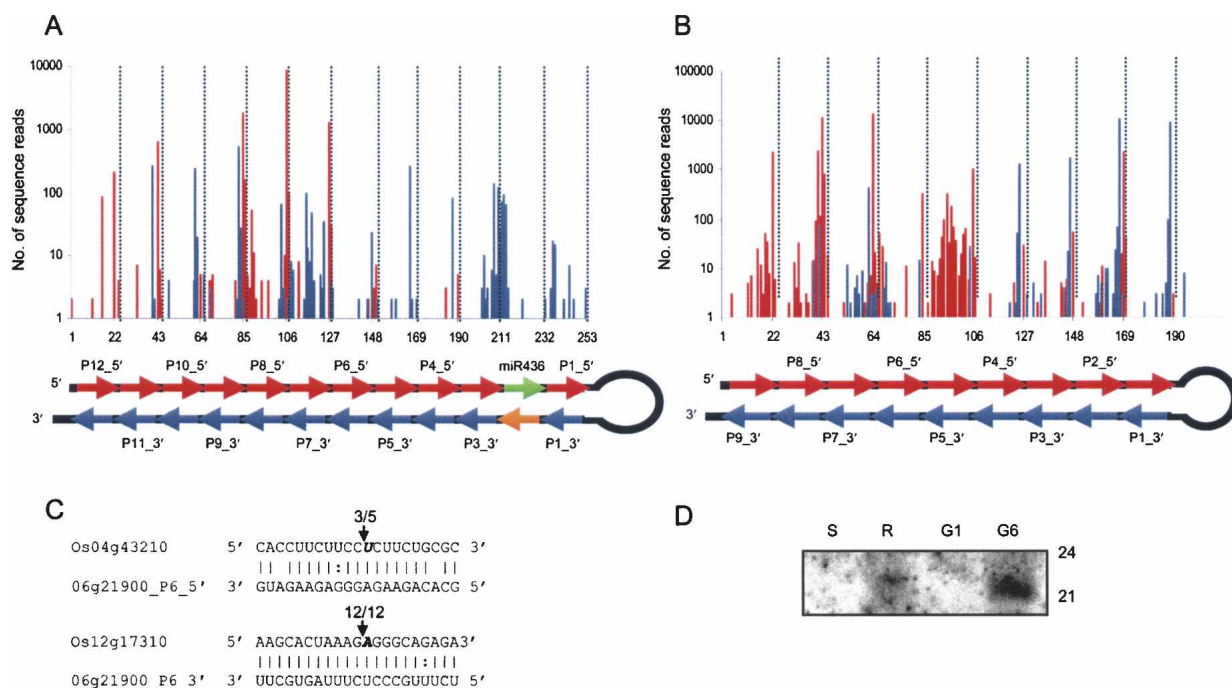
Figure 2. miRNA-like long hairpins producing phased 21-nt small RNAs. (*A*) Hairpin of the pre-miR436. In total, 13,141 and 2439 sequence reads were generated from the 5′ and the 3′ arms of the hairpin, respectively, and only 19 reads were mapped to the antisense strand. Of these sequences, 91.2% were 21 nt in length and 95.9% of these 21-nt small RNAs fell in phases of 21-nt spacing. The number of reads for all small RNAs with a common 5′ terminus (regardless of size) is plotted, with red and blue bars representing sequences from the 5′ and the 3′ arms, respectively. The 21-nt phases corresponding to the 5′ arm are indicated by the vertical dotted lines. Phase identifiers are shown above and below the hairpin. Nucleotide positions shown on the *X*-axis are based on the 5′ arm of the hairpin starting from the 5′ end of phase P12_5′. miR436 and miR436* are shown in green and orange, respectively. (*B*) A miRNA-like long hairpin generated from the antisense strand of LOC_Os06g21900. In total, 59,150 sequence reads were generated (~60% from the 5′ arm and ~40% from the 3′ arm; ~97% of the reads were from 6–10 DAF grains), including 97 reads mapped to the antisense strand; 54,857 (92.7%) reads were 21 nt in length and 93.9% of these 21-nt small RNAs were in exact 21-nt phases. The frequency and distribution of small RNAs is plotted as described in *A*. (*C*) 5′ RACE validation of cleavage of predicted target mRNAs by P6_5′ and P6_3′. (*D*) Expression of small RNAs from the newly identified long hairpin in different tissues detected by a probe antisense to the 3′ arm of the hairpin (218 nt in length). The blot used was the same as that for miR1861 shown in Figure 1C. (S) Shoots; (R) roots; (G1) 1–5 DAF grains; (G6) 6–10 DAF grains.

21 nt, with miR436 and miR436* represented by only three and one read, respectively. The second locus is complementary to Os06g21900 (Fig. 2B; Supplemental Fig. 7B) and produces siRNAs from nine 21-nt phases. As this locus has similar properties to the miR436 hairpin, the phased 21-nt siRNAs are likely to be produced by DCL4. This hairpin was not detected by our new miRNA discovery pipeline, because there is a ~1.3-kb spacer between the two hairpin arms. The siRNAs from this hairpin were more abundant than those from the miR436 hairpin in our data set and are predominantly expressed in 6–10 DAF grains (Fig. 2D). Consistent with this, small RNAs from this locus have only been found previously in a mixed sample, including inflorescence (Johnson et al. 2007), where 44 reads of 11 small RNAs were detected. Target genes could be predicted for eight of the 18 phased small RNAs from both hairpin arms (Supplemental Fig. 6C). We were unable to predict targets for the three most abundant phased small RNAs (P3_3′, P7_5′, and P8_5′, each of which had more than 10,000 reads). Cleavage of Os04g43210 by P6_5′ and Os12g17310 by P6_3′ was confirmed by 5′ RACE (Fig. 2C), indicating that these small RNAs act in *trans*. Os04g43210 encodes a putative myoinositol transporter, while Os12g17310 encodes a putative myosin heavy-chain protein.

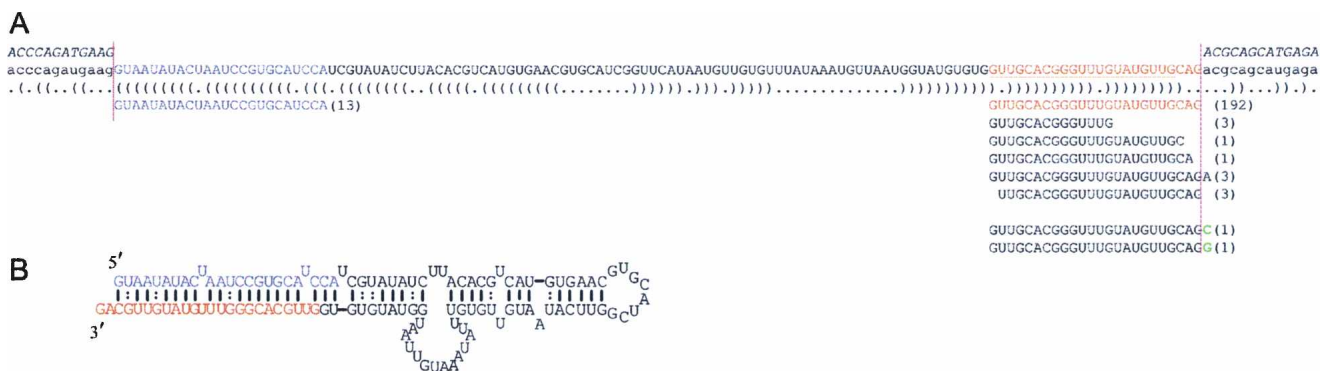### miR1429.2 is a putative mirtron

Most miRNA loci are encoded by independent transcription units, which are transcribed by RNA polymerase II (Lee et al. 2004); however, a number of miRNAs are located in the introns of mRNA-encoding host genes. All miRNAs identified so far in plants are thought to be derived from pri-miRNAs by sequential DCL1 cleavages. We sequenced 24-nt variants of miR1429 and miR1429* (miR1429.2 and miR1429.2*), but did not find miR1429 or miR1429* in our data. We observed that the 5′ terminus of miR1429.2* and the 3′ terminus of miR1429.2 coincide precisely with the 5′ and 3′ borders of an annotated intron of a putative homeobox gene; these intron boundaries are supported by cDNA data (Fig. 3A). This distribution of small RNAs is consistent with that of recently identified mirtrons in *Drosophila melanogaster* and *Caenorhabditis elegans* (Ruby et al. 2007) and suggests that spliced, debranched introns can be DCL substrates in plants. A comparison of small RNA sequences with annotated rice introns did not reveal other mirtrons, although six potential

candidates were identified (Supplemental Fig. 8). By comparison, in *D. melanogaster*, *C. elegans*, and mammals, 14, 4, and 19 mirtrons have been identified, respectively (Berezikov et al. 2007; Ruby et al. 2007). This may be due to the imperfect annotation of the rice genome, or perhaps because the average rice intron, at 412 nt, (http://www.tigr.org/tdb/e2k1/osa1/riceInfo/info.shtml#Genes) is an unsuitable mimic of pre-miRNAs (~150 nt in rice). It is also possible that mirtrons are tissue or developmental time dependent because their biogenesis requires expression of their host genes and intron splicing. In addition, although miR1429.2 is located at the 3′ hairpin arm, as are those of invertebrate mirtrons (Ruby et al. 2007), unlike all other known 3′ mirtron products from flies, worms, or mammals (Berezikov et al. 2007; Ruby et al. 2007), it begins with a G, suggesting that divergent characteristics may be tolerated in plant mirtrons.

Discussion

Our small RNA sequencing results revealed a diverse and complex small RNA population and directly demonstrated the expression of all conserved plant miRNA families in developing rice grains. We re-evaluated previously annotated rice miRNAs using our data, based on the production of small RNAs from their precursors, and found that precursors of most conserved miRNAs had a small RNA distribution pattern consistent with processing of ssRNAs by DCL proteins. However, in precursors of approximately half of the previously annotated, nonconserved miRNAs, the distribution pattern of small RNA reads deviated significantly from that expected. In these cases, small RNAs were generated from the antisense strand or from the entire precursor, making their classification as miRNA loci questionable. With this in mind, extreme caution was taken when designating 39 novel miRNAs, and we required detection of a miRNA* to confidently annotate nonconserved miRNAs. All of our newly identified miRNAs do not appear to be conserved in *Arabidopsis* and have predicted target genes with more diverse functions than those of conserved miRNAs. In particular, the nonconserved grain miRNAs have a number of potential target mRNAs that encode proteins likely to be involved in primary metabolic pathways.

We have identified additional routes by which siRNAs with miRNA-like properties can arise in plants. We have evidence that

Figure 3. miR1429.2 is a putative mirtron. (*A*) Pre-miR1429.2 (shown in capital letters; adjacent exon sequences are shown in lowercase letters; miR1429 is underlined) with aligned small RNAs. The mature miRNA (miR1429.2 in red) and its miRNA* (in blue), and the termini of the intron (vertical pink lines), are indicated with the predicted secondary structure shown in bracket notation. Positions of the annotated splicing donor and acceptor were supported by cDNA sequence AB101650 (sequences flanking the intron are shown in italic at the first row). The 3′ terminal nucleotide in the last two small RNAs, which do not match the genomic sequence, might be generated by untemplated addition. (*B*) Predicted secondary structure of pre-miR1429.2 showing 2-nt 3′ overhangs.

spliced, debranched introns can be processed by a DCL activity to yield a mature miRNA-like molecule. This appears to be analogous to the mirtrons that have been found in *Drosophila*, *C. elegans*, and mammals. We also found that phased 21-nt siRNAs from long hairpin RNAs with structures similar to pre-miRNAs can cleave target mRNAs in *trans*. These siRNAs are likely to be produced via the DCL4 pathway (Liu et al. 2007) and so their double-stranded RNA (dsRNA) precursors may feed into the ta-siRNA pathway. Together, these results suggest that many sources of dsRNA can be adopted to produce miRNA-like molecules to regulate mRNA abundance.

Comparison of our rice grain small RNA sequences to deep sequencing results generated from other parts of the rice plant suggest that the nonconserved miRNAs are expressed in a tissue-specific manner. Of the 24 new rice miRNAs identified by deep sequencing in samples derived mainly from vegetative tissues (Lu et al. 2008; Morin et al. 2008; Sunkar et al. 2008), we found evidence for only 10 in developing grains, while we identified 39 miRNAs in grains that were not previously detected in other organs. This contrasts with the conserved miRNA families that are all present in deep sequencing data from different parts of the plant, although some family members showed tissue or organ specificity. For example, the expression level of miR156l was similar to that of miR156a-j in seedlings (Sunkar et al. 2008) but was much lower than that of miR156a-j in grains (Supplemental Table 2). All four members of the miR172 family were equally highly expressed in seedlings (Sunkar et al. 2008), whereas miR172c was not detected in grains (Supplemental Table 2). The phased 21-nt small RNAs from the Os06g21900 locus also show grain-specific expression. The 24-nt putative mirtron product, miR1429.2, has only been identified in grains with the 21-nt miR1429 (which does not correspond to intron boundaries) found elsewhere. Thus, the same locus appears to be generating miRNAs by differing pathways in different parts of the plant.

The presence of a large set of miRNAs and miRNA-like molecules in the developing grains suggests that many processes are under miRNA regulation during grain development. The conserved miRNAs generally target the mRNAs of transcription factors and other regulatory proteins; given the roles in development that have been demonstrated for many of the conserved miRNAs in plants, it is not surprising to find them in developing rice grains. The sequence data suggests that miRNA expression increases over the first 10 d of grain development; this is coincident with development of both the endosperm and embryo. The largest change in expression between our two grain samples was for miR408 with 18-fold more expression in the 6–10 DAF grains. The conserved target mRNA of this miRNA encodes a plastocyanin-like protein (Lu et al. 2005); it is not clear what role this protein would play in grains; however, there are other predicted targets of miR408 (Supplemental Fig. 3) encoding proteins with other functions. The predicted target genes for the nonconserved miRNAs found in grains (Supplemental Table 3) do not show the same bias toward targeting the mRNAs of annotated transcription factors as the conserved miRNAs. Our data provides some evidence that miRNAs are acting on metabolic pathways in the grain. We confirmed that miR1428e_3p cleaves two SnRK1b kinases, one of which (Os08g37800) has been suggested to have a role in regulating starch accumulation based on its expression in the endosperm and aleurone (Kanegae et al. 2005). Expression of miR1428e_3p is high in both grain samples and coincides with the onset of deposition of starch in these grains. The Os06g21900 siRNAs are predominantly expressed in 6–10 DAF grain, so could

be involved in regulating a transitional step in the developmental program of the grain. The confirmed mRNA target of an Os06g21900 siRNA, Os04g43210, encodes a putative plasma membrane myoinositol transporter (closely related to *Arabidopsis* INOSITOL TRANSPORTER 2) (Schneider et al. 2007). This protein could be involved in transport of myoinositol or its derivatives, which could act in either metabolic or signaling roles.

The miRNAs and other small RNAs from rice grains and their likely target mRNAs identified in this study indicate that small RNA regulation is likely to have many roles in grains, and this data set should provide a rich resource for future studies of small RNA functions in rice grains.

Methods

All rice (*Oryza sativa* spp. *japonica*) tissue samples were collected from rice cultivar Nipponbare plants grown in a controlled glasshouse at $25 \pm 3°C$ with 16 h of light.

Small RNA isolation, small RNA library construction, and sequencing

Total RNA was extracted from ~1.5 g of shoots and roots of 7-d-old seedlings, and 1–5 and 6–10 DAF dehusked rice grains. Tissues were ground into a fine powder in liquid nitrogen and processed using RNA extraction buffer (0.1 M NaCl, 2% SDS, 50 mM Tris-HCl, 10 mM EDTA, and 20 mM β-mercaptoethanol) and an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1). The aqueous fraction was subsequently extracted two times using equal volumes of phenol/chloroform/isoamyl alcohol (25:24:1) and chloroform. Total RNA was then precipitated using LiCl (2 M final concentration) overnight at $-20°C$. Low-molecular-weight (LMW) RNA was precipitated from the supernatant with an equal volume of isopropanol in the presence of 0.1% SDS at $-20°C$ for 2 d. The precipitated LMW RNA was resuspended in DEPC-treated H_2O. Approximately 150 μg of LMW RNA was then mixed with a trace amount of 5′-^{32}P-labeled RNA markers (18 and 25 nt) and size fractionated on a 15% polyacrylamide, 8 M urea gel. The gel fragment spanning the 18–25-nt RNA markers was excised, crushed into fine pieces, and the RNA was eluted overnight in 0.3 M NaCl at 4°C with gentle shaking. After removing excess acrylamide through a 0.4-μm filter, the eluate was precipitated using an equal volume of isopropanol and 5 μg/mL of glycogen on ice for at least 2 h. The isolated small RNAs were then ligated to a 3′ adaptor (pCTGTAGGCACCATCAAT-idT; (p) phosphate; (id) O-C3 linker that blocks the 3′ terminus, preventing it from self-ligating) using T4 RNA ligase (Amersham Biosciences) overnight at 4°C according to the manufacturer's instructions. The ligated product was size fractionated using ligated forms of the RNA markers as a size guide and gel eluted as above. Another T4 RNA ligase-mediated ligation was used to add a 5′ adaptor (Dharmacon Research; ATCGTAG**GCAC**CUGAUA, ATC GTA**GCCA**CCUGAUA, ATCGTA**GCUG**CUGAUA, and ATCG TA**GCGU**CUGAUA, for 7-d-old shoots and roots, and 1–5 and 6–10 DAF grains, respectively; the bold nucleotides are bar-code sequences for distinguishing sequence reads from each library after sequencing the mixed small RNA library). The ligated product was fractionated using ligated forms of the RNA markers as a size guide and gel eluted using the above method. The purified ligation was then reverse transcribed (SuperScript II, Invitrogen) with the RT primer (ATTGATGGTGCCTACAG). The cDNA was further amplified using a pair of 454 sequencing primers (Supplemental Table 4). The final PCR products were purified using Qiagen's PCR purification kit, mixed, and sequenced using the 454

technology (Margulies et al. 2005). Small RNA populations of 1–5 and 6-10 DAF grains were further sequenced using the Ilumina technology.

Bioinformatic analysis of small RNAs and miRNA identification

The 454 adaptor sequences were identified and masked using Cross_Match (http://www.phrap.org/phredphrapconsed.html#block_phrap) and then trimmed with a custom script. All Illumina sequences were 33 nt in length, with the length of adaptor sequence present at the 3′ end, depending upon the length of the sequence it tailed. A custom script was fashioned to find and trim the maximum possible adaptor sequence from each.

The remaining analysis was repeated for both sequencing sets. Duplicate sequences were tallied, then excised to form a nonredundant set of sequences. Sequences of <18 nt were removed. To find the origins of the small RNAs, BLAST was used to search the *Oryza sativa* genome sequence (TIGR Rice Annotation Release 5.0, http://www.tigr.org/tdb/e2k1/osa1/overview.shtml). The BLAST search used a small word size of seven, query sequence filtering disabled, and a large expectation value cut-off of 1000 to maximize the likelihood of finding all matches for the small sequences. The results were parsed to extract the chromosomal coordinates of all perfect, full-length matches. BLAST was similarly used to test which sequences had matches in TIGR's Oryza Repeats Database (version 3.3) and which had matches to relevant members of the Rfam RNA family database (http://www.sanger.ac.uk/Software/Rfam). Sequences matching either of these datasets were excluded from further analysis. Coordinates from the genomic BLAST results were compared with coordinates from the miRBase microRNA database (version 9.0; http://microrna.sanger.ac.uk/sequences/index.shtml) to determine which genomic hits corresponded to coordinates of known miRNAs. This was in addition to annotating which sequences directly matched known mature miRNA sequences according to miRBase.

The presence of both a miRNA and a miRNA* was the basis of the method to predict novel miRNAs. Pairs of sequences that mapped to the same chromosome, on the same strand, within 400 nt, with a similar length (<5 nt difference in length), without too many other sequences mapped between them (disallowed if the other sequences overlap to form a smear of sequences between the pair), without too many sequences mapped to the other strand between the pair (one sequence allowed on opposing strand per five sequences on the same strand), and with neither having too many hits throughout the genome (no more than 10 hits), were selected as candidates for testing as potential miRNA/miRNA* pairs. Surrounding genomic sequence for selected pairs was extracted and folded with Vienna RNAFold (http://www.tbi.univie.ac.at/~ivo/RNA/). A custom script was used to test the quality of the fold and, in particular, that the potential miRNA and miRNA* was paired in a hairpin. The pairing was tested based on rules described previously (Jones-Rhoades et al. 2006), i.e., no more than seven unpaired nucleotides between the miRNA and miRNA*, with no more than three consecutive unpaired nucleotides, and no more than two unpaired nucleotides that do not have a corresponding unpaired nucleotide in the other sequence. This test was also run allowing only four unpaired nucleotides to give a reduced result list.

For predicted novel miRNA sequences, conservation in other plant species was examined by searching for matches to the sequence, with up to three mismatches allowed in plant sequence datasets using PatScan (Dsouza et al. 1997). The *Arabidopsis* chromosome sequences (http://www.arabidopsis.org/),

Populus trichocarpa, assembled scaffold sequences (http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html), and *Zea mays* contig sequences (http://www.plantgdb.org/prj/GSSAssembly/zeamays/index.php) were searched. For any matches, sequences surrounding the match, for 200 nt to either side, were extracted and run through RNAFold. The results were analyzed using custom scripts to see whether the matched sequence could form a suitable hairpin for it to be a conserved miRNA. Analysis of folds in this instance was as above, but in this case, all parts of the sequence had to be tested as a potential miRNA* matching to the potential conserved miRNA.

Identification of mirtron

The presence of mirtrons was examined by identifying small RNA sequences that mapped to the 5′ or 3′ end of an intron. The intron sequences involved were folded with RNAFold and any resulting hairpin was analyzed with custom scripts to determine whether the mapped small RNA might be a valid mirtron and the intron a valid precursor. The folds were tested as above for miRNA/miRNA* pairs, except for assuming the miRNA* to be a sequence of equal length to the small RNA being tested at the other end of the intron sequence.

Target gene prediction

Target sequences of predicted novel miRNAs were searched for using PatScan to search for matching sequences, with up to four differences, in the *Oryza Sativa* cDNA, CDS, and UTR sequence sets from TIGR (TIGR Rice Annotation Release 5.0, http://www.tigr.org/tdb/e2k1/osa1/data_download.shtml), the *Oryza Sativa* UniGene sequence sets from NCBI (http://www.ncbi.nlm.nih.gov/sites/entrez?db=unigene), and *Oryza Sativa* EST sequences available from the NCBI sequence database. Matches were given a score ranking, with mismatches assigned one point, G–U bonds between the miRNA and target assigned 0.5 points, and mismatch at position one assigned two points. No mismatch was allowed at position 10 or 11. Matches were dismissed if they contained consecutive mismatches, more than one mismatch in the first 10 bases, or more than three mismatches in the remaining bases.

Northern blot hybridization

Approximately 14 µg of LMW RNA (prepared as described above) was separated on 18% polyacrylamide denaturing gels, including labeled RNA oligonucleotides as size markers. RNAs were transferred to Hybond-N$^+$ membrane (Amersham Bioscience) and hybridized with locked nucleic acid DNA oligonucleotides complementary to predicted miRNA sequences, which had been T4 kinase labeled with [γ-^{32}P]ATP. Blots were prehybridized and hybridized at 37°C in 125 mM Na$_2$HPO$_4$ (pH 7.2), 250 mM NaCl$_2$, 7% SDS, and 50% formamide, and washed at 37°C twice with 2× SSC, 0.2% SDS, followed by a higher stringency wash of 1× SSC, 0.1% SDS at 37°C if required. Blots were imaged using an FLA-5000 PhosphorImager (Fujifilm). Blots were reprobed with an RNA probe complementary to 5S rRNA to confirm uniform loading.

5′ and 3′ RACE

5′ RACE was used to determine the 5′ ends of pri-miRNAs and to map the cleavage sites on target mRNAs of miRNAs or ta-siARFs. 3′ RACE was used to determine the 3′ ends of pri-miRNAs. RACE was carried out on poly(A)$^+$ RNA, isolated using the PolyATract mRNA isolation system (Promega) using the GeneRacer kit (Invitrogen) according to the manufacturer's instructions. For cloning pri-miRNAs, poly(A)$^+$ RNA was treated with calf intestinal

alkaline phosphatase (CIP) to remove truncated mRNA and non-mRNA, and de-capped using tobacco acid pyrophosphatase (TAP) before ligation of the 5′ adaptor, whereas poly(A)+ RNA was directly ligated to the 5′ adaptor when mapping the cleavage sites of the miRNA target genes. Two gene-specific primers were used for each RACE (Supplemental Table 4). The PCR products from a positive 5′ or 3′ RACE reaction were gel purified and cloned (pGEM-T Easy, Promega). A minimum of four clones were sequenced for each PCR product.

Acknowledgments

References

Allen, E., Xie, Z., Gustafson, A.M., Sung, G.H., Spatafora, J.W., and Carrington, J.C. 2004. Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. *Nat. Genet.* **36:** 1282–1290.

Arazi, T., Talmor-Neiman, M., Stav, R., Riese, M., Huijser, P., and Baulcombe, D.C. 2005. Cloning and characterization of microRNAs from moss. *Plant J.* **43:** 837–848.

Axtell, M.J., Snyder, J.A., and Bartel, D.P. 2007. Common functions for diverse small RNAs of land plants. *Plant Cell* **19:** 1750–1769.

Bartel, D.P. 2004. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* **116:** 281–297.

Berezikov, E., Chung, W.-J., Willis, J., Cuppen, E., and Lai, E.C. 2007. Mammalian mirtron genes. *Mol. Cell* **28:** 328–336.

Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R., and Zhu, J.K. 2005. Endogenous siRNAs derived from a pair of natural *cis*-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell* **123:** 1279–1291.

Chiou, T.J., Aung, K., Lin, S.I., Wu, C.C., Chiang, S.F., and Su, C.L. 2006. Regulation of phosphate homeostasis by microRNA in *Arabidopsis*. *Plant Cell* **18:** 412–421.

Dsouza, M., Larsen, N., and Overbeek, R. 1997. Searching for patterns in genomic data. *Trends Genet.* **13:** 497–498.

Fahlgren, N., Howell, M.D., Kasschau, K.D., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., Law, T.F., Grant, S.R., Dangl, J.L., et al. 2007. High-throughput sequencing of *Arabidopsis* microRNAs: Evidence for frequent birth and death of *MIRNA* genes. *PLoS ONE* **2:** e219. doi: 10.1371/journal.pone.0000219.

Fujii, H., Chiou, T.J., Lin, S.I., Aung, K., and Zhu, J.K. 2005. A miRNA involved in phosphate-starvation response in *Arabidopsis*. *Curr. Biol.* **15:** 2038–2043.

Itoh, J., Nonomura, K., Ikeda, K., Yamaki, S., Inukai, Y., Yamagishi, H., Kitano, H., and Nagato, Y. 2005. Rice plant development: From zygote to spikelet. *Plant Cell Physiol.* **46:** 23–47.

Johnson, C., Bowman, L., Adai, A.T., Vance, V., and Sundaresan, V. 2007. CSRDB: A small RNA integrated database and browser resource for cereals. *Nucleic Acids Res.* **35:** D829–D833.

Jones-Rhoades, M.W. and Bartel, D.P. 2004. Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol. Cell* **14:** 787–799.

Jones-Rhoades, M.W., Bartel, D.P., and Bartel, B. 2006. MicroRNAs and their regulatory roles in plants. *Annu. Rev. Plant Biol.* **57:** 19–53.

Kanegae, H., Miyoshi, K., Hirose, T., Tsuchimoto, S., Mori, M., Nagato, Y., and Takano, M. 2005. Expressions of rice *sucrose non-fermenting-1 related protein kinase 1* genes are differently regulated during the caryopsis development. *Plant Physiol. Biochem.* **43:** 669–679.

Kim, V.N. and Nam, J.W. 2006. Genomics of microRNA. *Trends Genet.* **22:** 165–173.

Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., Pfeffer, S., Rice, A., Kamphorst, A.O., Landthaler, M., et al. 2007. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129:** 1401–1414.

Lee, Y., Kim, M., Han, J., Yeom, K.H., Lee, S., Baek, S.H., and Kim, V.N. 2004. MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* **23:** 4051–4060.

Li, J., Yang, Z., Yu, B., Liu, J., and Chen, X. 2005. Methylation protects miRNAs and siRNAs from a 3′-end uridylation activity in *Arabidopsis*. *Curr. Biol.* **15:** 1501–1507.

Liu, B., Li, P., Li, X., Liu, C., Cao, S., Chu, C., and Cao, X. 2005. Loss of function of *OsDCL1* affects microRNA accumulation and causes developmental defects in rice. *Plant Physiol.* **139:** 296–305.

Liu, B., Chen, Z., Song, X., Liu, C., Cui, X., Zhao, X., Fang, J., Xu, W., Zhang, H., Wang, X., et al. 2007. Oryza sativa *Dicer-like4* reveals a key role for small interfering RNA silencing in plant development. *Plant Cell* **19:** 2705–2718.

Lu, S., Sun, Y.H., Shi, R., Clark, C., Li, L., and Chiang, V.L. 2005. Novel and mechanical stress-responsive microRNAs in *Populus trichocarpa* that are absent from *Arabidopsis*. *Plant Cell* **17:** 2186–2203.

Lu, C., Jeong, D.H., Kulkarni, K., Pillay, M., Nobuta, K., German, R., Thatcher, S.R., Maher, C., Zhang, L., Ware, D., et al. 2008. Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs). *Proc. Natl. Acad. Sci.* **105:** 4951–4956.

Luo, Y.C., Zhou, H., Li, Y., Chen, J.Y., Yang, J.H., Chen, Y.Q., and Qu, L.H. 2006. Rice embryogenic calli express a unique set of microRNAs, suggesting regulatory roles of microRNAs in plant post-embryogenic development. *FEBS Lett.* **580:** 5111–5116.

Mallory, A.C. and Vaucheret, H. 2006. Functions of microRNAs and related small RNAs in plants. *Nat. Genet.* Suppl: S31–S36.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437:** 376–380.

Meyers, B.C., Souret, F.F., Lu, C., and Green, P.J. 2006. Sweating the small stuff: MicroRNA discovery in plants. *Curr. Opin. Biotechnol.* **17:** 139–146.

Molnar, A., Schwach, F., Studholme, D.J., Thuenemann, E.C., and Baulcombe, D.C. 2007. miRNAs control gene expression in the single-cell alga *Chlamydomonas reinhardtii*. *Nature* **447:** 1126–1129.

Morin, R.D., Aksay, G., Dolgosheina, E., Ebhardt, H.A., Magrini, V., Mardis, E.R., Sahinalp, S.C., and Unrau, P.J. 2008. Comparative analysis of the small RNA transcriptomes of *Pinus contorta* and *Oryza sativa*. *Genome Res.* **18:** 571–584.

Navarro, L., Dunoyer, P., Jay, F., Arnold, B., Dharmasiri, N., Estelle, M., Voinnet, O., and Jones, J.D. 2006. A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. *Science* **312:** 436–439.

Nikovics, K., Blein, T., Peaucelle, A., Ishida, T., Morin, H., Aida, M., and Laufs, P. 2006. The balance between the *MIR164A* and *CUC2* genes controls leaf margin serration in *Arabidopsis*. *Plant Cell* **18:** 2929–2945.

Nobuta, K., Venu, R.C., Lu, C., Beló, A., Vemaraju, K., Kulkarni, K., Wang, W., Pillay, M., Green, P.J., Wang, G.L., et al. 2007. An expression atlas of rice mRNAs and small RNAs. *Nat. Biotechnol.* **25:** 473–477.

Okamura, K., Hagen, J.W., Duan, H., Tyler, D.M., and Lai, E.C. 2007. The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell* **130:** 89–100.

Parizotto, E.A., Dunoyer, P., Rahm, N., Himber, C., and Voinnet, O. 2004. In vivo investigation of the transcription, processing, endonucleolytic activity, and functional relevance of the spatial distribution of a plant miRNA. *Genes & Dev.* **18:** 2237–2242.

Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H.L., and Poethig, R.S. 2004. SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of *trans*-acting siRNAs in *Arabidopsis*. *Genes & Dev.* **18:** 2368–2379.

Rajagopalan, R., Vaucheret, H., Trejo, J., and Bartel, D.P. 2006. A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes & Dev.* **20:** 3407–3425.

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., and Bartel, D.P. 2002. MicroRNAs in plants. *Genes & Dev.* **16:** 1616–1626.

Ruby, J.G., Jan, C.H., and Bartel, D.P. 2007. Intronic microRNA precursors that bypass Drosha processing. *Nature* **448:** 83–86.

Schneider, S., Schneidereit, A., Udvardi, P., Hammes, U., Gramann, M., Dietrich, P., and Sauer, N. 2007. *Arabidopsis* INOSITOL TRANSPORTER2 mediates H+ symport of different inositol epimers and derivatives across the plasma membrane. *Plant Physiol.* **145:** 1395–1407.

Sunkar, R., Girke, T., Jain, P.K., and Zhu, J.K. 2005. Cloning and characterization of microRNAs from rice. *Plant Cell* **17:** 1397–1411.

Sunkar, R., Zhou, X., Zheng, Y., Zhang, W., and Zhu, J.K. 2008. Identification of novel and candidate miRNAs in rice by high throughput sequencing. *BMC Plant Biol.* **8:** 25. doi: 10.1186/1471-2229-8-25.

Vazquez, F. 2006. *Arabidopsis* endogenous small RNAs: Highways and byways. *Trends Plant Sci.* **11:** 460–468.

Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gasciolli, V., Mallory, A.C., Hilbert, J.L., Bartel, D.P., and Crété, P. 2004. Endogenous *trans*-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs. *Mol. Cell* **16:** 69–79.

Wang, X.J., Reyes, J.L., Chua, N.H., and Gaasterland, T. 2004. Prediction

and identification of *Arabidopsis thaliana* microRNAs and their mRNA targets. *Genome Biol.* **5:** R65. doi: 10.1186/gb-2004-5-r65.

Xie, Z., Allen, E., Fahlgren, N., Calamar, A., Givan, S.A., and Carrington, J.C. 2005. Expression of *Arabidopsis MIRNA* genes. *Plant Physiol.* **138:** 2145–2154.

Yao, Y., Guo, G., Ni, Z., Sunkar, R., Du, J., Zhu, J.K., and Sun, Q. 2007. Cloning and characterization of microRNAs from wheat (*Triticum aestivum* L.). *Genome Biol.* **8:** R96. doi: 10.1186/gb-2007-8-6-r96.

Yoshikawa, M., Peragine, A., Park, M.Y., and Poethig, R.S. 2005. A pathway for the biogenesis of *trans*-acting siRNAs in *Arabidopsis*. *Genes & Dev.* **19:** 2164–2175.

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R., and Chen, X. 2005. Methylation as a crucial step in plant microRNA biogenesis. *Science* **307:** 932–935.

Zamore, P.D. and Haley, B. 2005. Ribo-genome: The big world of small RNAs. *Science* **309:** 1519–1524.

Zhang, X., Henderson, I.R., Lu, C., Green, P.J., and Jacobsen, S.E. 2007. Role of RNA polymerase IV in plant small RNA metabolism. *Proc. Natl. Acad. Sci.* **104:** 4536–4541.

Zhao, T., Li, G., Mi, S., Li, S., Hannon, G.J., Wang, X.J., and Qi, Y. 2007. A complex system of small RNAs in the unicellular green alga *Chlamydomonas reinhardtii*. *Genes & Dev.* **21:** 1190–1203.