Supplemental results and discussion

Four *TAS3* loci are expressed in developing grains

Four *TAS3* loci have been predicted in rice with experimental evidence for the production of the ta-siRNA, ta-siARF, from three of these loci (Williams et al. 2005; Axtell et al. 2006; Liu et al. 2007; Lu et al. 2008). We searched our dataset for *TAS* genes using published methods (Chen et al. 2007; Howell et al. 2007) and found ta-siRNAs produced from all four predicted *TAS3* loci (Supplemental Fig. 9A). These loci all contain dual miR390 target sites which in *Arabidopsis* set the register for production of *TAS3* ta-siRNAs (Howell et al. 2007; Supplemental Fig. 9B). A 5′ RACE analysis of *TAS3a1, TAS3b1* and *TAS3b2* transcripts showed cleavage only at the 3′ miR390 target site and not at the 5′ miR390 target site as observed in *Arabidopsis* (Howell et al. 2007; Supplemental Fig. 9C). These ta-siARFs are predicted to target one *ARF2* and four *ARF3* mRNAs (each with two ta-siARF target sites). We demonstrated cleavage of *ARF2* and three *ARF3* transcripts, consistent with activity of TAS3b1_5′D6(+) and TAS3b2_5′D5(+) (43 reads in total), or TAS3a2_5′D7(+) which was not present in the small RNA dataset (Supplemental Fig. 9D and 9E). There was no evidence for activity of the most abundant ta-siARF, TAS3a1_5′D6(+) with 110 reads, perhaps because it has two to three consecutive mismatches to its target sequences (Supplemental Fig. 9D and 9E).

When analysing cleavage at the miR390 binding sites, we found evidence of transcript cleavage directed by ta-siRNAs generated from the antisense strand. Of 16 cloned 5′ RACE products from *TAS3a1*, 11 were consistent with cleavage products of TAS3a1_5′D10(-), the most abundant ta-siRNA from this locus (Supplemental Fig. 9A and 9C). At the *TAS3b1* locus, of 47 cloned 5′ RACE products around the miR390 binding sites, a total of 24 were consistent with cleavage by three different ta-siRNAs represented in our dataset (Supplemental Fig. 9C). Similarly, at the *Arabidopsis TAS3a* locus, a cleavage site was detected 33 nt upstream of the 3′ miR390 cleavage site, the expected cleavage site of 5′D2(-) even though this ta-siRNA was not cloned at that time (Allen et al. 2005). By searching the *Arabidopsis* small RNA database (http://asrp.cgrb.oregonstate.edu/db/), 47 reads were found for 5′D2(-). These results suggest that *TAS3* transcripts might be controlled by a posttranscriptional autoregulatory loop.

References

Allen, E., Xie, Z., Gustafson, A.M., and Carrington, J.C. 2005. microRNA-directed phasing during *trans*-acting siRNA biogenesis in plants. *Cell* **121**:207-221.

Axtell, M.J., Jan, C., Rajagopalan, R., and Bartel, D.P. 2006. A two-hit trigger for siRNA biogenesis in plants. *Cell* **127**:565-577.

Chen, H.M., Li, Y.H., and Wu, S.H. 2007. Bioinformatic prediction and experimental validation of a microRNA-directed tandem *trans*-acting siRNA cascade in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA.* **104**:3318-23.

Howell, M.D., Fahlgren, N., Chapman, E.J., Cumbie, J.S., Sullivan, C.M., Givan, S.A., Kasschau, K.D., and Carrington, J.C. 2007. Genome-wide analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 pathway in *Arabidopsis* reveals dependency on miRNA- and tasiRNA-directed targeting. *Plant Cell* **19**:926-942.

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, C., Jeong, D.H., Kulkarni, K., Pillay, M., Nobuta, K., German, R., Thatcher, S.R., Maher, C., Zhang, L., Ware, D., Liu, B., Cao, X., Meyers, B.C., and Green, P.J. 2008. Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs). *Proc. Natl. Acad. Sci. USA.* **105**:4951-4956.

Notredame, C., Higgins, D.G., and Heringa, J. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol*. **302**:205-217.

Williams, L., Carles, C.C., Osmont, K.S., and Fletcher, J.C. 2005. A database analysis method identifies an endogenous *trans*-acting short-interfering RNA that targets the *Arabidopsis ARF2*, *ARF3*, and *ARF4* genes. *Proc. Natl. Acad. Sci. USA.* **102**:9703-9708.

Legends for Supplemental Figures

Supplemental Figure 1. Size distribution of small RNA sequences aligning to the rice genome. Only small RNAs, 18 nt or longer, from the Illumina dataset are shown.

Supplemental Figure 2. Alignment of small RNAs against the precursors of selected known miRNAs. The reported pre-miRNA sequences together with the predicted fold-back structures indicated by bracket notation are shown. The reported miRNA and the inferred miRNA* are shown in red and blue, respectively. Small RNAs generated from the pre-miRNAs and their antisense strands are shown above and below the pre-miRNA sequences, respectively, except for miR439a, for which the pre-miRNA and its antisense strand are shown separately and small RNAs generated are shown below the pre-miRNA or its antisense strand. Only small RNAs from the Illumina sequencing are shown.

Supplemental Figure 3. (*A*) and (*B*) Alignment of miR408 and miR408* sequences. miR408 sequences were retrieved from miRBase (http://microrna.sanger.ac.uk/sequences/) and miR408* sequences are predicted based on the secondary structures of the miR408 precursors. Alignment was generated by T-coffee (Notredame et al. 2000). (*C*) Predicted target genes of miR408, miR408*, miR529 and miR529*. Scores of the target genes are indicated after the identifier of miRNA. Ath: *Arabidopsis thaliana*; Osa: *Oryza sativa*; Ptc: *Populus trichocarpa*; Ppt: *Physcomitrella patens*; Pta: *Pinus taeda*; Smo: *Selaginella moellendorffii*; Sof: *Saccharum officinarum*; Tae: *Triticum aestivum*; Vvi: *Vitis vinifera*; Zma: *Zea mays*.

Supplemental Figure 4. (*A*) Precursors and fold-back structures of the newly-identified miRNAs. (*B*) Precursors and fold-back structures of the candidate miRNAs. The pre-miRNA sequences together with their predicted fold-back structures indicated by bracket notation are shown. The miRNA and the inferred miRNA* are shown in red and blue, respectively. Small RNAs generated from the pre-miRNAs are shown below the pre-miRNA sequences. For each pre-miRNA, the miRNA identifier is followed by the length of miRNA, genomic location, and annotation of the locus from which the miRNA derived. Only small RNAs from the Illumina sequencing are shown.

Supplemental Figure 5. Tandem miRNAs with conserved pre-miRNAs. (*A*) Alignment of pre-miRNAs of the miR1861 family, in which b and c, d and e, f and g, j and k, and l and m are tandem miRNA pairs. The miRNA and miRNA* sequences are bordered by vertical red lines. (*B*) Alignment of genomic

sequences containing tandem pre-miRNAs. The sequenced miRNAs are underlined in green. The second putative member in miR1861a, h, i and n, and both members in miR1861_5 and miR1861_7 were not annotated as miRNAs because they were not cloned. (*C*) Secondary structure of the genomic region containing miR1861j and k. (*D*) Alignment of pre-miRNAs of the miR1428 family. (*E*) Secondary structure of the genomic region containing miR1428d_3p and e_3p. (*F*) Alternative splicing of pri-miRNAs. miR1428e_3p and miR1428d_3p are tandem miRNAs encoded by a single transcript with varying transcription start sites (TSSs) and 3′ polyadenylation sites. Expression of miR1428d_3p and miR1850 seems to be controlled by alternative splicing because the region containing their precursors is spliced out in some transcripts. Pre-miRNAs and miRNAs are represented by grey and black boxes, respectively. Exons are represented by open boxes. Numbers represent the nucleotide position of TSSs, 3′ polyadenylation sites, start or end of exons, or start positions of pre-miRNAs. In each case, sequence of the longest transcript has been deposited into NCBI.

Supplemental Figure 6. Predicted target genes of the newly-identified miRNAs, candidate miRNAs and phased small RNAs from the newly identified miRNA-like long hairpins. Target gene score is indicated after the miRNA identifier.

Supplemental Figure 7. Sequences of miRNA-like long hairpins generating phased small RNAs. (*A*) The precursor of miR436 with miR436 and miR436* shown in red and blue, respectively. (*B*) A miRNA-like long hairpin located antisense to LOC_Os06g21900. The hairpin sequences are shown with each phase indicated by pink dotted lines and the loop represented by an open circle. Phases are named with the first phase closest to the loop. The phases producing exact phased small RNAs are in bold type.

Supplemental Figure 8. Candidate mirtrons identified in rice. The intron sequence together with its predicted fold-back structure indicated by bracket notation is shown for each candidate mirtron. The identifier of the sequenced small RNA is followed by the genomic location of the putative mirtron, the locus from which the mirtron derived and annotation of the locus.

Supplemental Figure 9. Rice *TAS3* loci and cleavage of *ARF* genes by ta-siARF. (*A*) Read number and distribution of phased 21 nt ta-siRNAs. ta-siRNAs from the sense (black bar) and antisense (grey bar) strands are plotted against the 21 nt phases represented by black and grey arrows, respectively, which are

set by cleavage at the 3′ miR390-binding site. miR390 and its binding sites are indicated by black and green bars, respectively, while red and blue arrows indicate ta-siARFs. (*B*) Alignment of *TAS3a1* and *TAS3a2*, *TAS3b1* and *TAS3b2*. The miR390 binding sites are underlined in red and ta-siARFs are underlined in blue and green. (*C*) Diagrammatic representation of the primary *TAS3* transcripts and validation of 3′ miR390 cleavage by 5′ RACE. The antisense 21 nt phases between the 5′ and 3′ miR390 binding sites are indicated by grey arrows. Cleavage sites are indicated by arrows with the cleavage frequency amongst sequenced clones shown above. Red arrows indicate expected or confirmed miR390-guided cleavage. Black arrows indicate cleavages guided by ta-siRNAs generated from the antisense strands of *TAS3* transcripts, other cleavage sites are indicated by light blue arrows. Solid arrows indicate cleavage sites mapped using primers downstream of the 3′ miR390 cleavage site, hatched arrows indicate the cleavage sites mapped using primers between the two miR390 binding sites. (*D*) T-coffee alignment (Notredame et al. 2000) of ta-siARFs from four *TAS3* loci. Colors indicate alignment quality in a regional context with red to green change indicating high to low quality. The cloned or predicted ta-siARF sequences are underlined with the number of sequence reads shown at right. The identifier for each phase was based on conventions used for *Arabidopsis* ta-siRNAs (Allen et al. 2005). (*E*) Cleavage of *ARF2* and *ARF3* genes by ta-siARFs**.** Positions of the ta-siARF target sites within each *ARF2* and *ARF3* coding sequence (black box) are shown in grey with the sequences expanded below. Target sequences of all ta-siARFs are shown with a bold red letter indicating the expected cleavage positions. An attempt to clone cleavage products of LOC_Os05g48870 was unsuccessful.