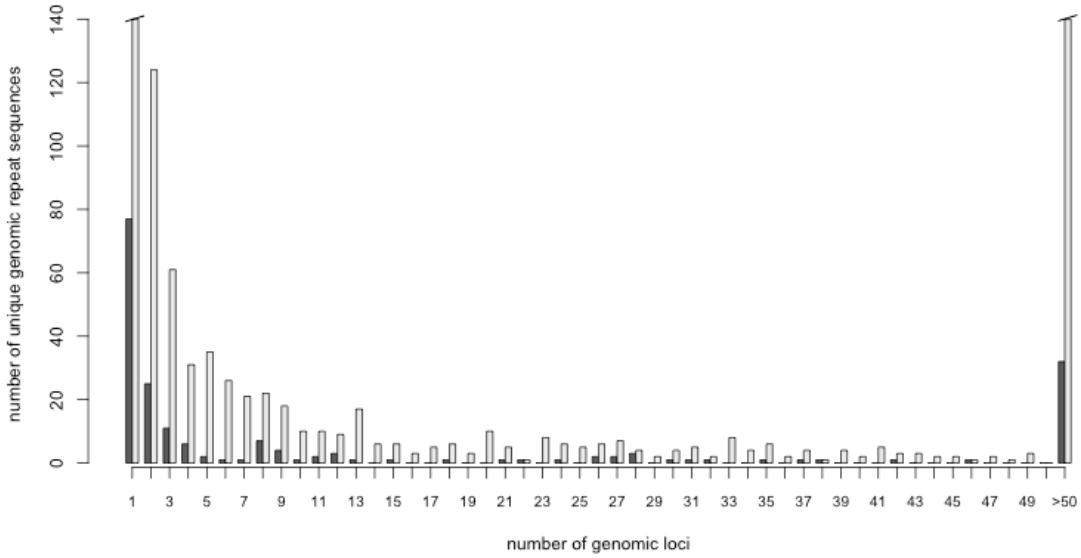
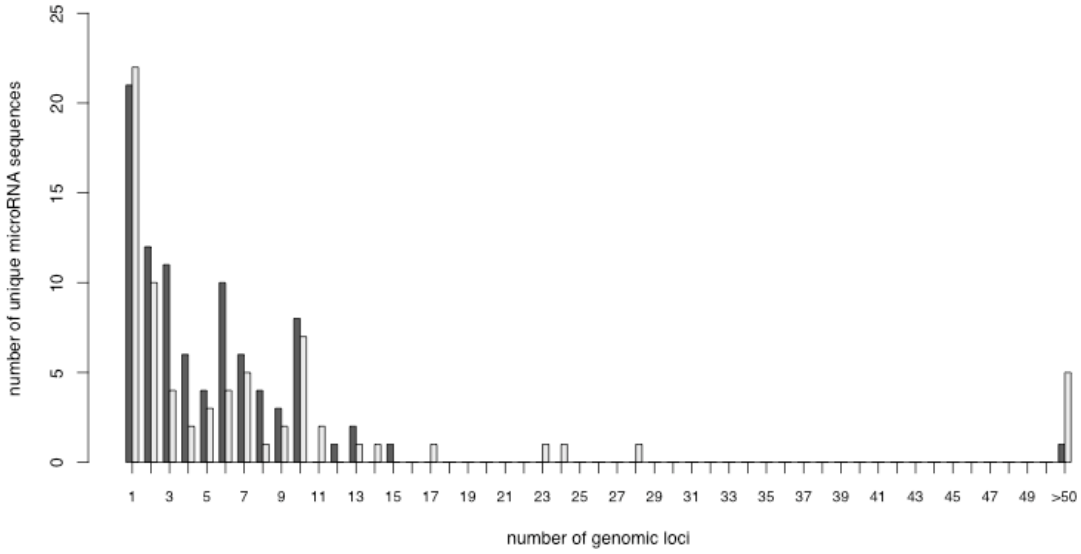
Supplementary Figures

Supplementary Figure 1

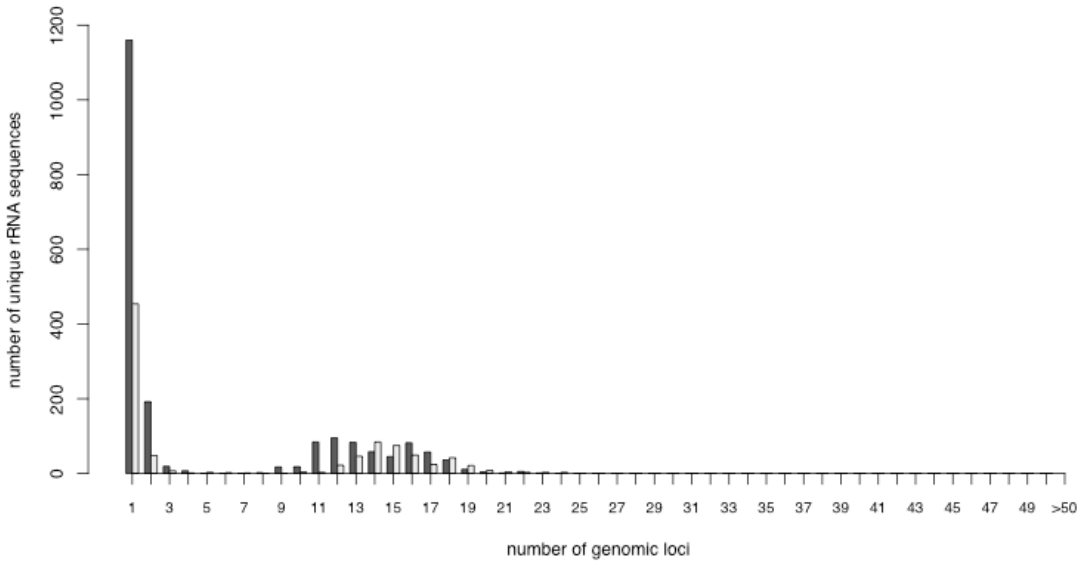
Supplementary Figure 1a



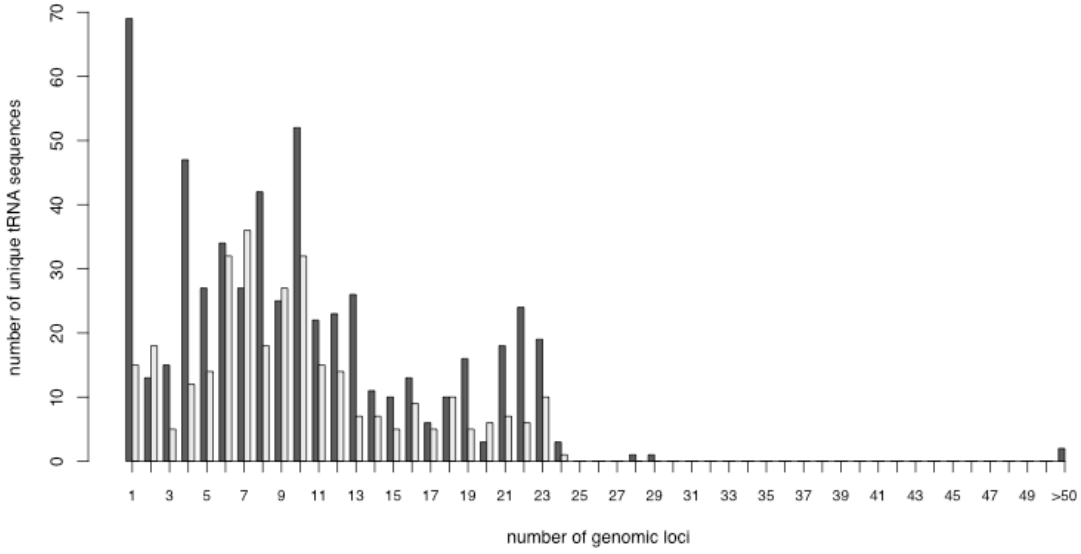
Supplementary Figure 1b

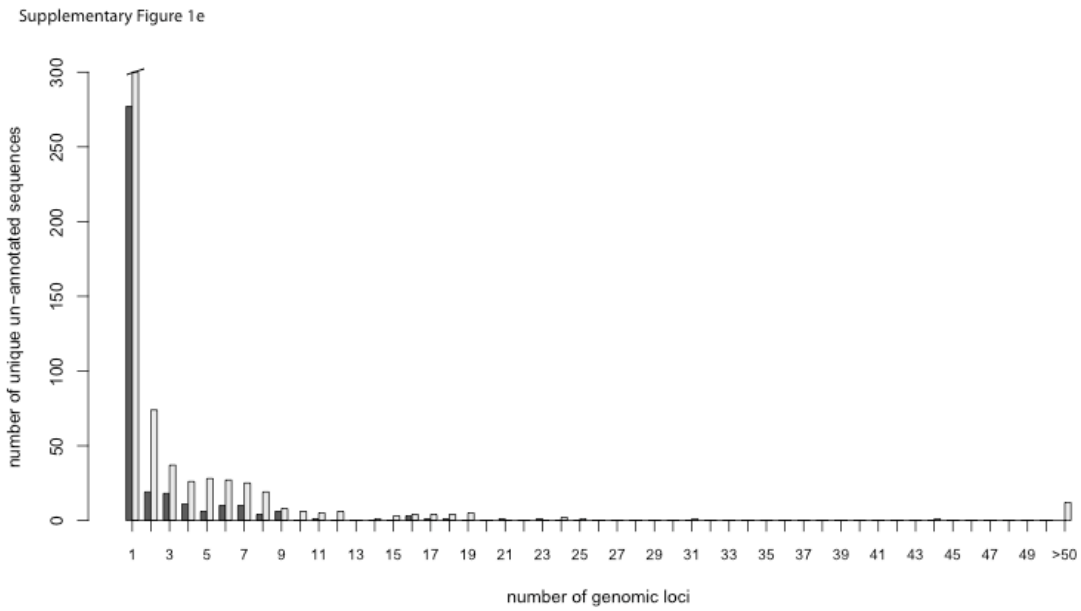


Supplementary Figure 1c



Supplementary Figure 1d





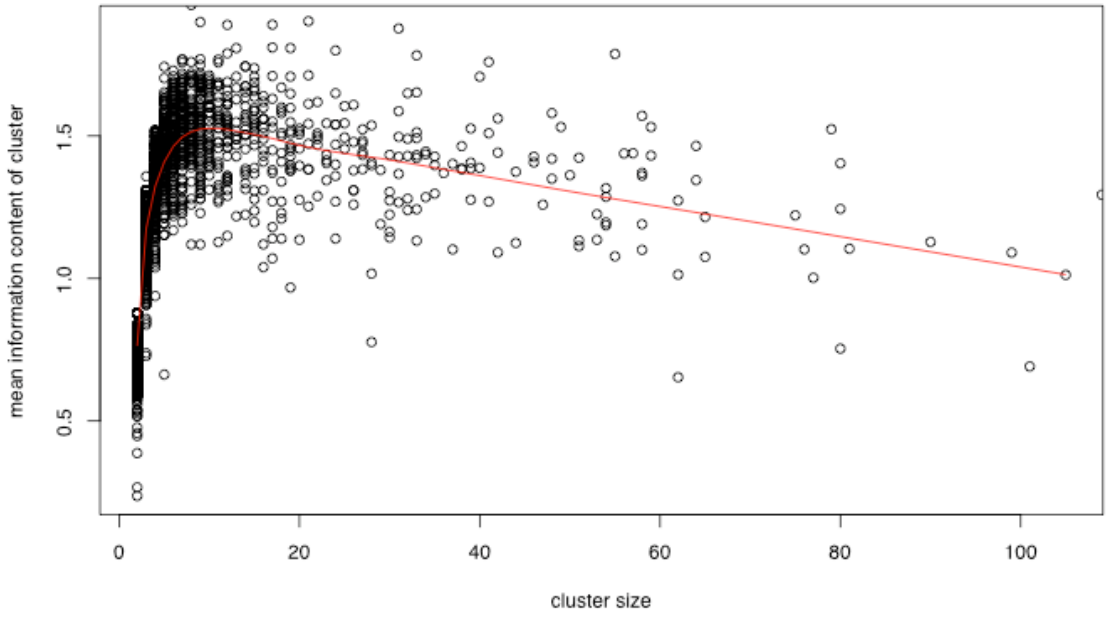
Supplementary Figure 1: The number of perfect alignments to the rice genome differs for the different classes of small RNAs in both *P. contorta* (dark grey) and *O. sativa* (light grey), ignoring sequences shorter than 18-nt (to reduce noise). **a)** Sequences derived from genomic repeats for both organisms have a very heavy tail as a result of the high (and variable) copy number of these elements in the genome. **b)** miRNAs have an overall lower number of loci, with very few miRNAs having greater than 15 genomic sites. **c)** rRNA-derived sequences show a distinct bimodal trend, with the lower lobe representing unique or near-unique sequences and the second lobe comprising sequences with between 11 and 19 perfect matches in the rice genome. **d)** tRNA fragments, like the rRNA-derived sequences, have more perfect hits in the genome than the miRNAs, which likely represents the frequency of these genes in the *O. sativa* genome **e)** The un-annotated sequences have a higher number of unique genomic loci than any other class (off-scale value for *O. sativa* is 555 sequences). The shape of this distribution suggests a more random process producing many of these sequences, with some potentially un-recognized repeat-derived and rRNA-derived sequences.

Supplementary Figure 2

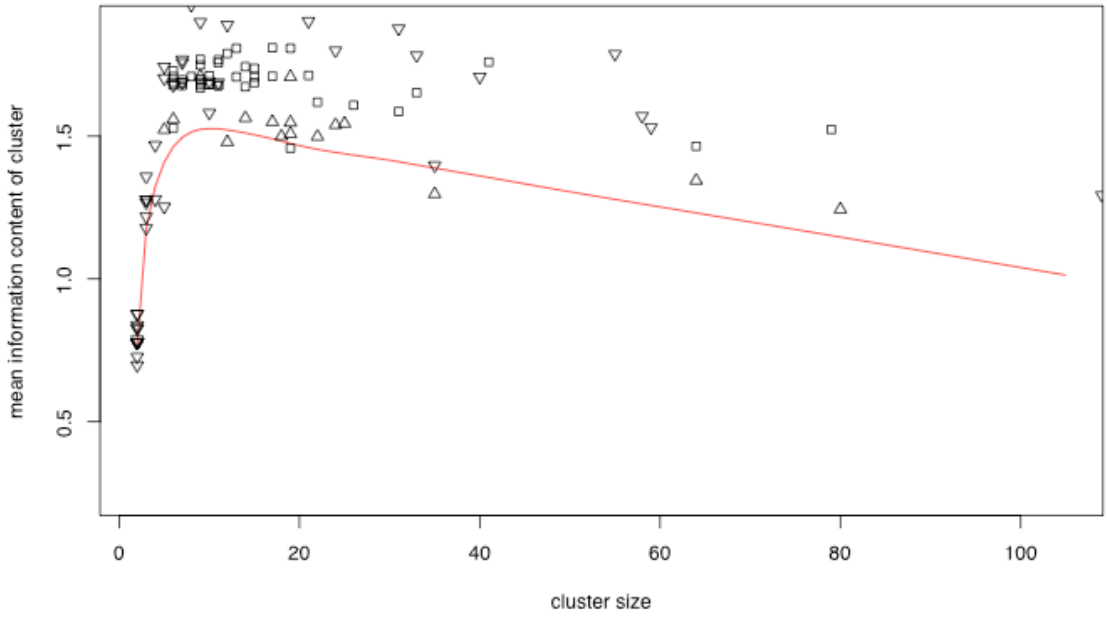
Supplementary Figure 2: Demonstrating the distinctive alignment patterns of degradation products versus microRNAs. **a)** *P. contorta* (green) small RNAs, which aligned perfectly to a partial 16s-rRNA gene (orange bar) in the *O. sativa* genome. The alignment signature displays a trend common to degradation fragments. All alignments are to the strand opposite to the direction of transcription, suggesting they may not be degradation products, but rather are siRNAs. **b)** *P. contorta* (green) and *O. sativa* (red) Small RNAs that aligned perfectly to a conserved miRNA gene (blue bar) in the *O. sativa* genome. The two sets of aligned sequences represent the miRNA and the miRNA*. Each bar represents an isomiR of this miRNA.

Supplementary Figure 3

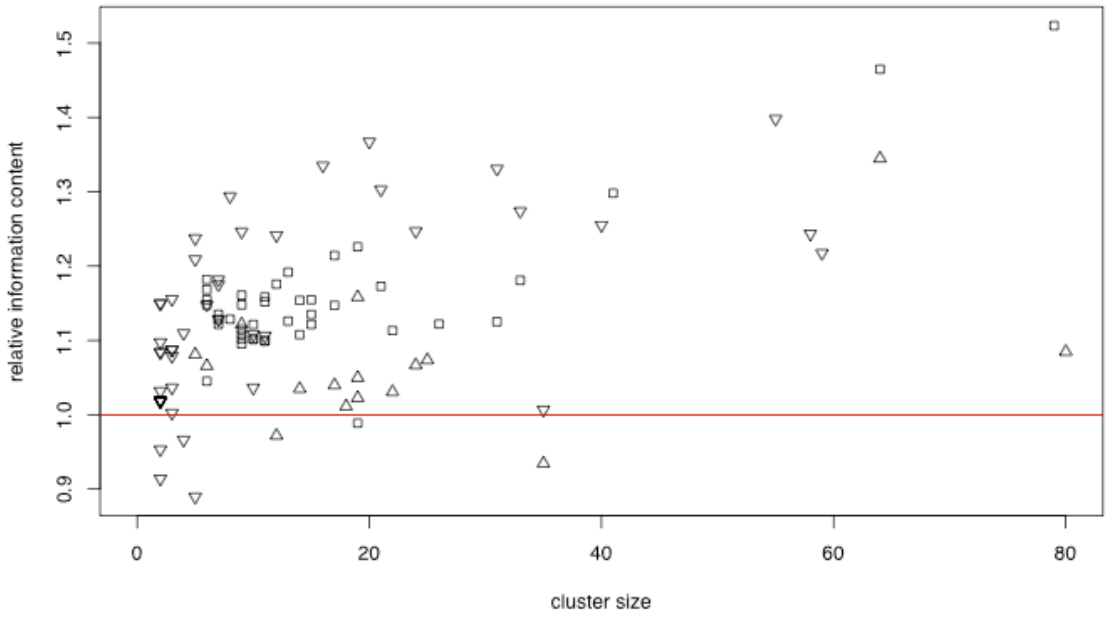
Supplementary Figure 3a



Supplementary Figure 3b

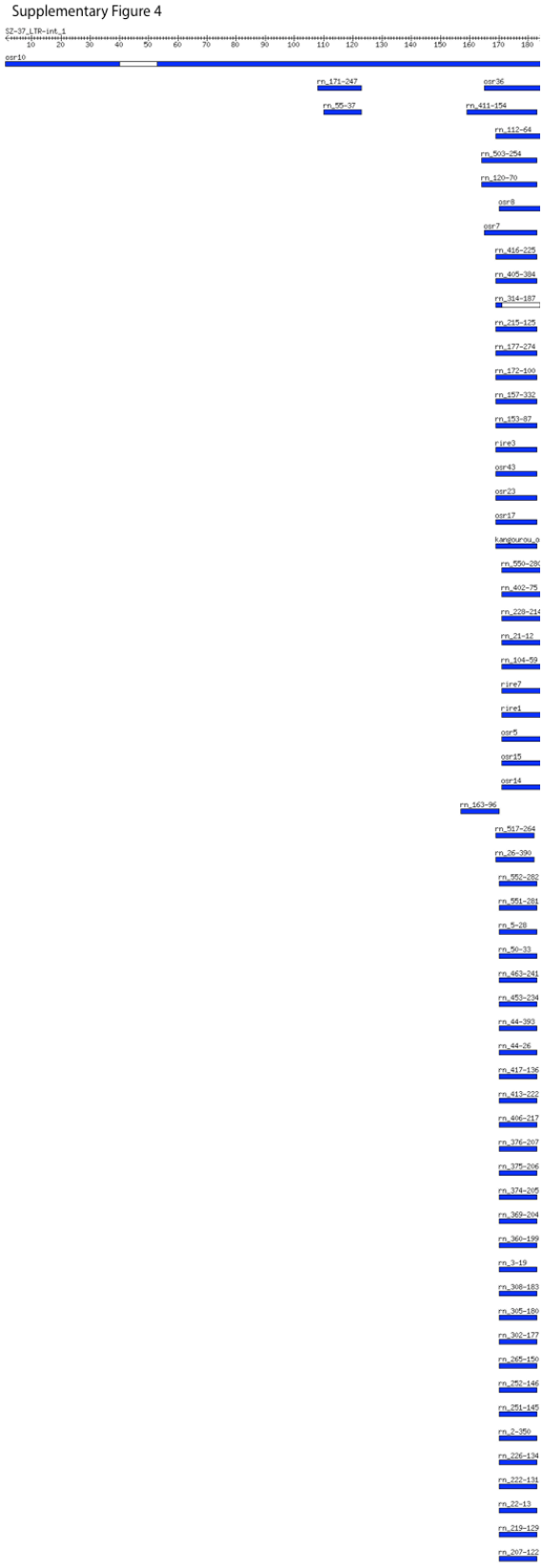


Supplementary Figure 3c



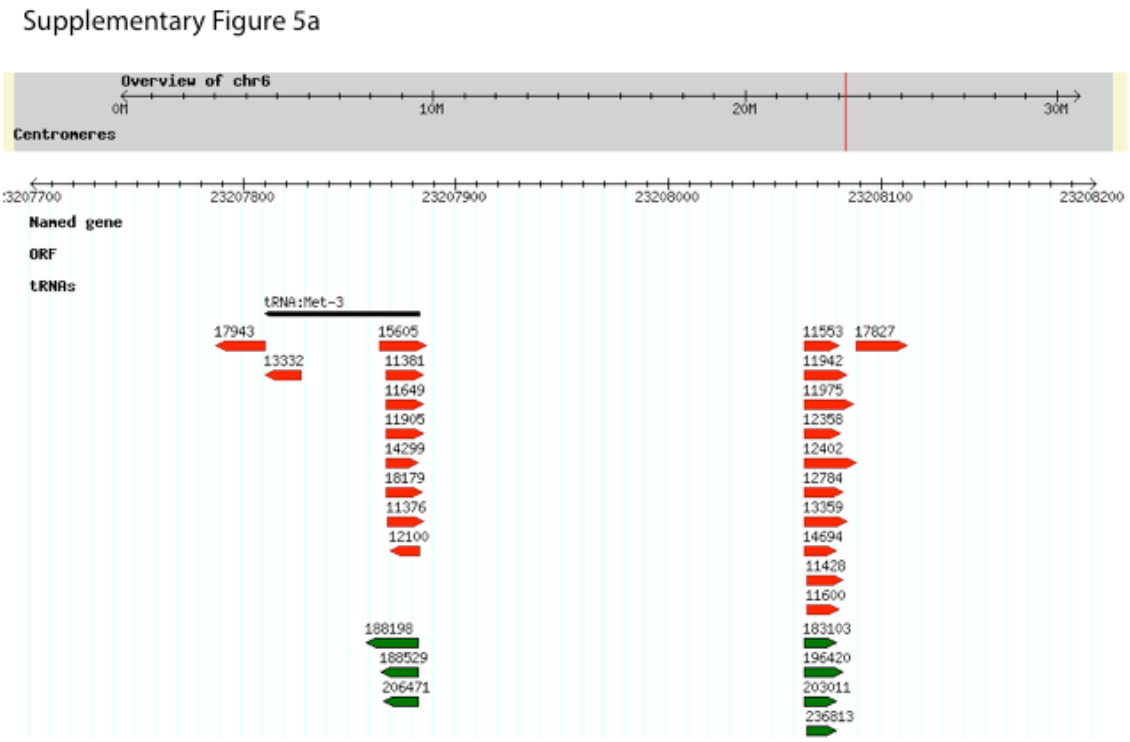
Supplementary Figure 3: Scatter plot of the mean Information content of sequence clusters.
a) All sequence clusters showing a rough correlation between cluster size and mean information content. Smoothing of this plot with the Lowess function (shown in red) provides a local average trend line. **b)** Sequence clusters representing known and novel microRNA families identified in this study. Inverted triangles represent clusters comprising known miRNA sequences (as identified by high similarity to miRBase sequences). Triangles represent clusters containing novel miRNAs summarized in Table 1. Squares represent the clusters identified by the SVM classifier as microRNA-like, summarized in Table 2. **c)** The same data as in b with each data point divided by the lowess value (local average) to demonstrate its relative placement above or below the local median. Points showing a 'relative information content' value above 1.0 (red line) show an above-average information content relative to other clusters of the same size.

Supplementary Figure 4

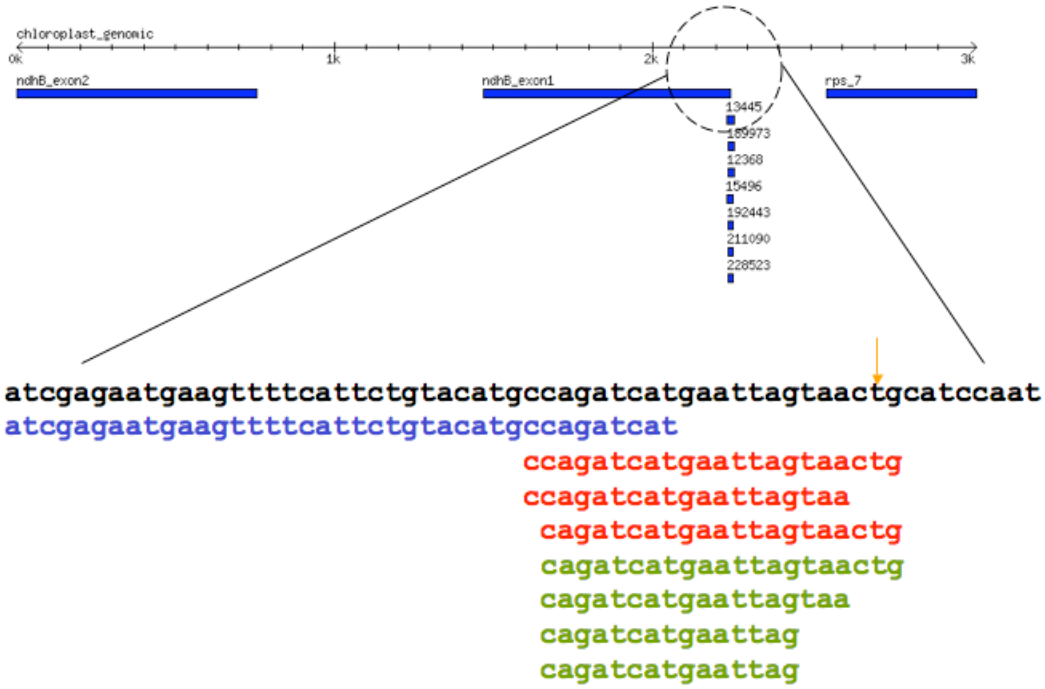


Supplementary Figure 4: Highlighting conserved region of a subclass of rice Long Terminal Repeat transposable element (LTR). All repeats from the rice repeat database RetrOryza (Chaparro et al. 2007) were aligned to a representative SZ-37 LTR sequence. One short motif is found in many of the other repeats. Perfectly aligned regions of other repeat sequences from RetrOryza are solid blue and gaps are white. This is the same region that yields 2 rice small RNA sequences and 9 *P. contorta* small RNA sequences, based on their alignment to the rice genome. The core sequence of this motif is 5'-AACCTGGCTCTGATACCA-3'.

Supplementary Figure 5

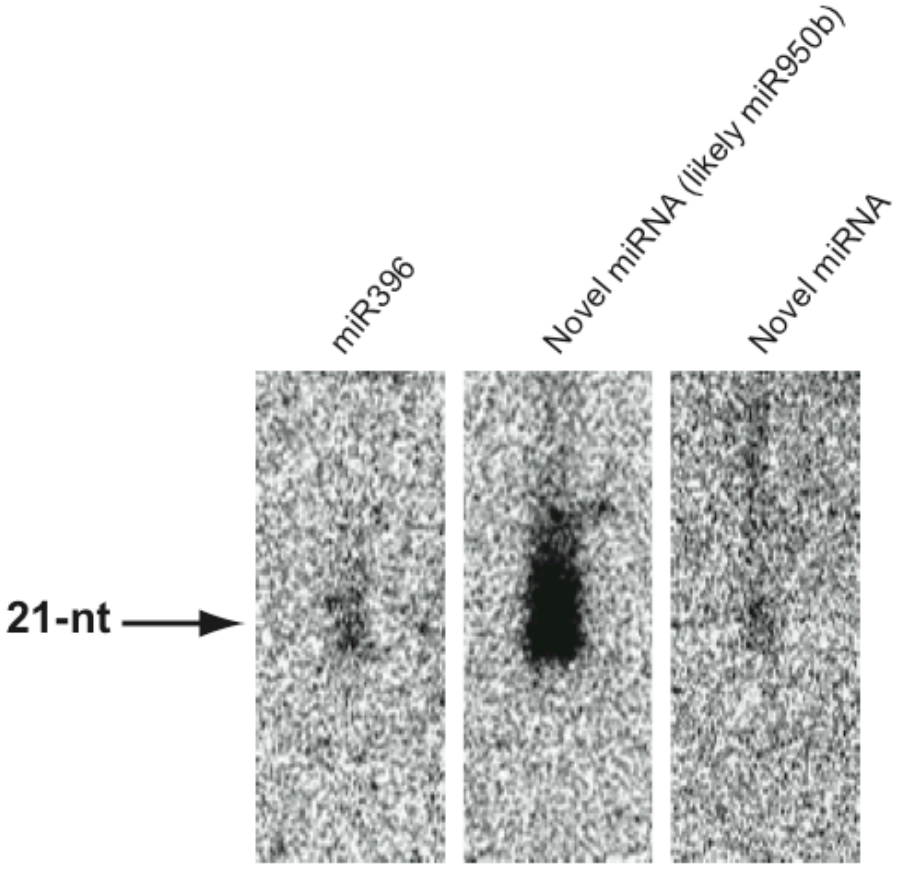


Supplementary Figure 5b



Supplementary Figure 5: Snapshots of the small RNAs derived from chloroplast transcripts or their nuclear counterparts. a) Showing *P. contorta* (green) and *O. sativa* (red) small RNAs in relation to a tRNA-met gene (black bar) in the rice genome. Most of the sequences derive from the strand opposing the annotated tRNA direction of transcription. b) Relative position of the small RNA sequences in relation to the ndhB/rps7 genes (all alignments are on the '-' strand relative to the rice genomic sequence). All of the small RNAs that align to this transcript (named by integers) align across the site of enzymatic cleavage (orange arrow), upstream of the ndhB start codon in exon 1 (coding DNA segment is shown in blue, start codon to the extreme right). A zoom of this region highlights the perfect conservation between some of the *O. sativa* (red) and *P. contorta* (green) small RNA sequences.

Supplementary Figure 6



Supplementary Figure 6: Northern blot detection of three miRNA sequences. Three predicted RNA clusters were tested for expression using a Northern blot. miR396, a miRNA perfectly conserved between gymnosperms and angiosperms, a small RNA with EST support (likely related to miR950b) and a novel small RNA that we predict to be a miRNA but which lacks any EST support.