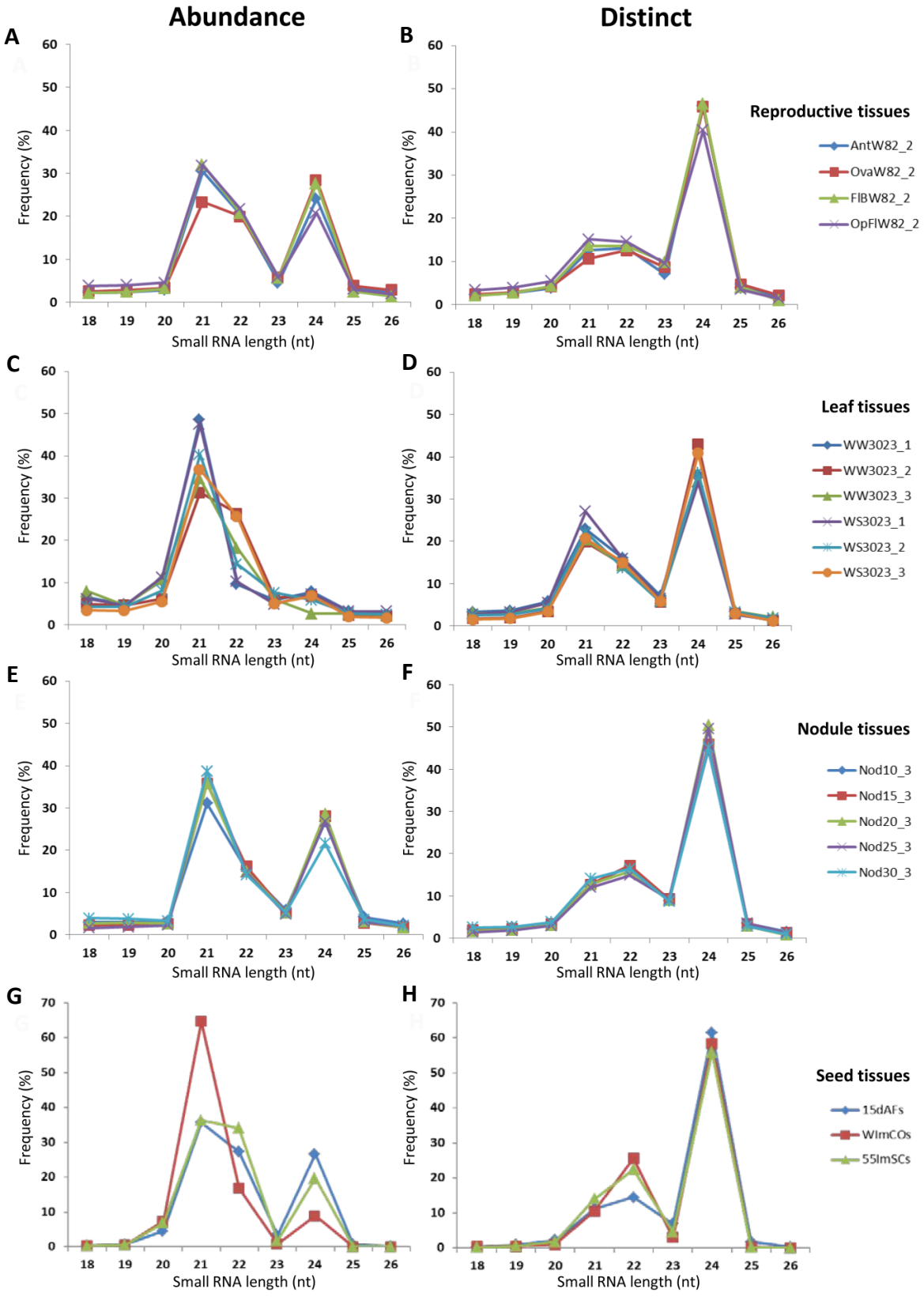
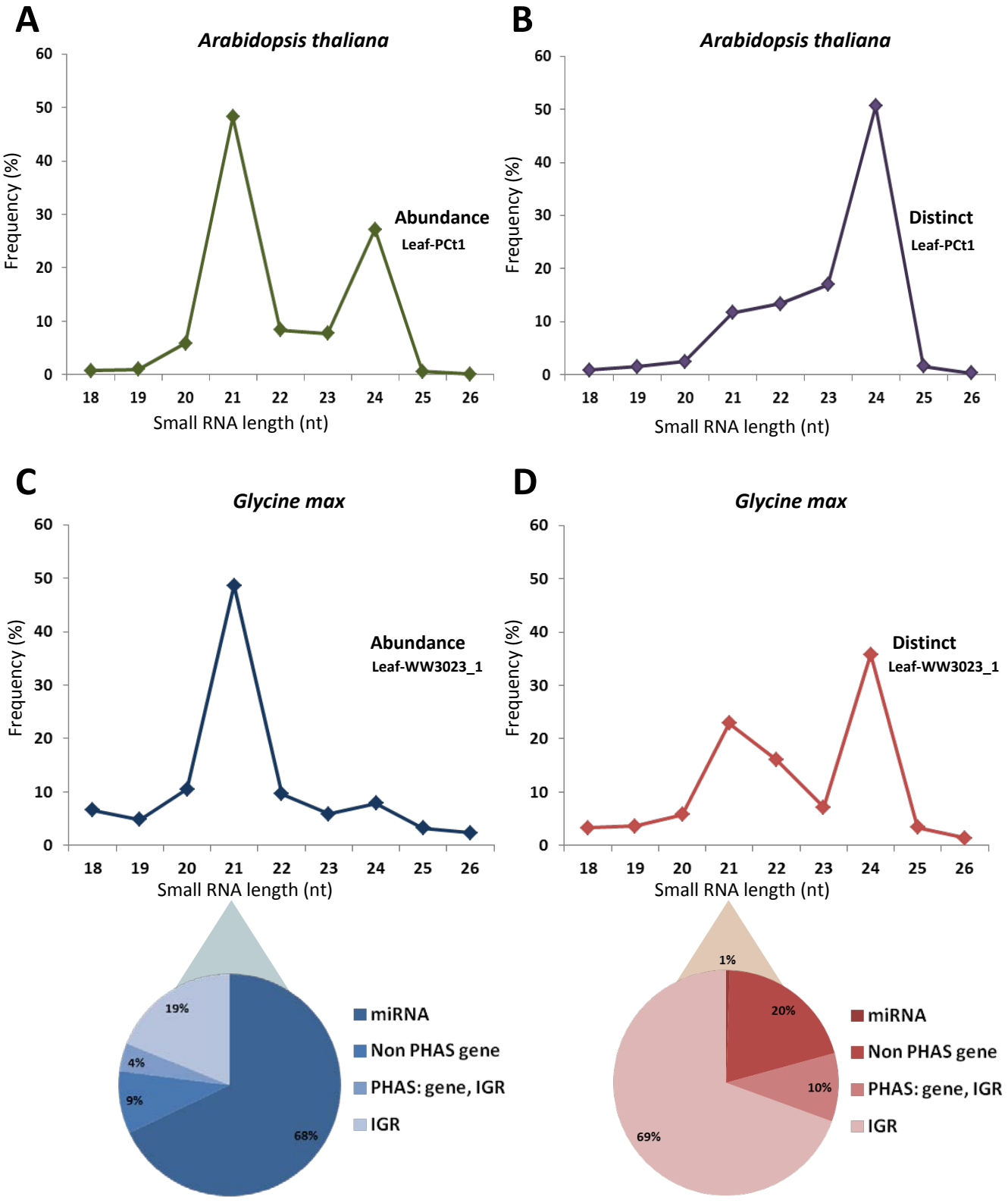
Supplemental Figure 1. Small RNA size distribution from different soybean tissues.

The size of small RNAs was plotted versus frequency (percentage) among total sequences (A, C, E and G) or distinct sequences (B, D, F, and H). Library codes and detailed information about each library are indicated in Supplemental Data set 1A.



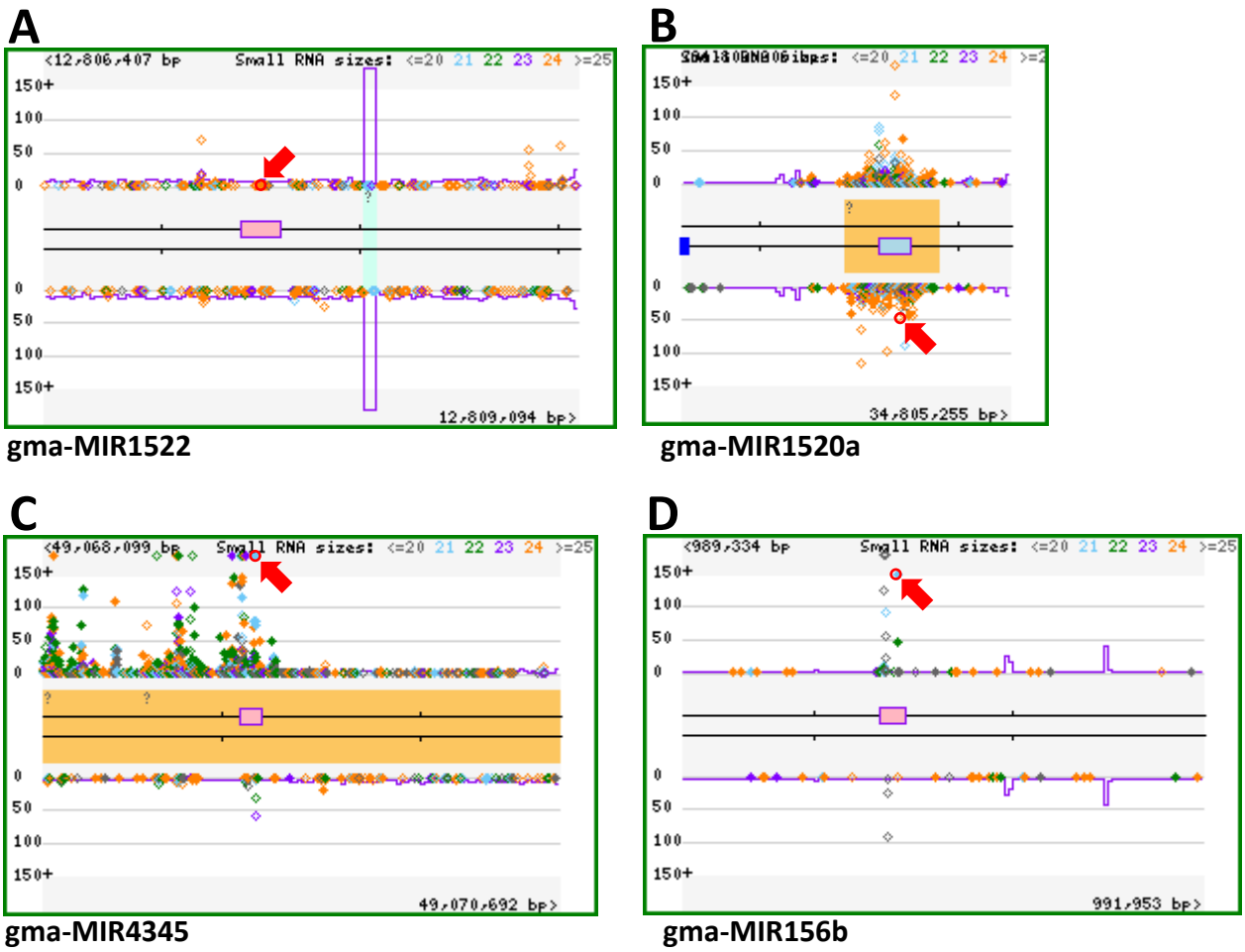
Supplemental Figure 2. Leaf small RNA size distributions in Arabidopsis and soybean.

Size distribution of total small RNAs ([A], [C]) and distinct small RNAs ([B], [D]) in Arabidopsis and soybean. The bottom panels of [C] and [D] provide details of the categories of 21-nt small RNAs in soybean, in the form of a pie chart.

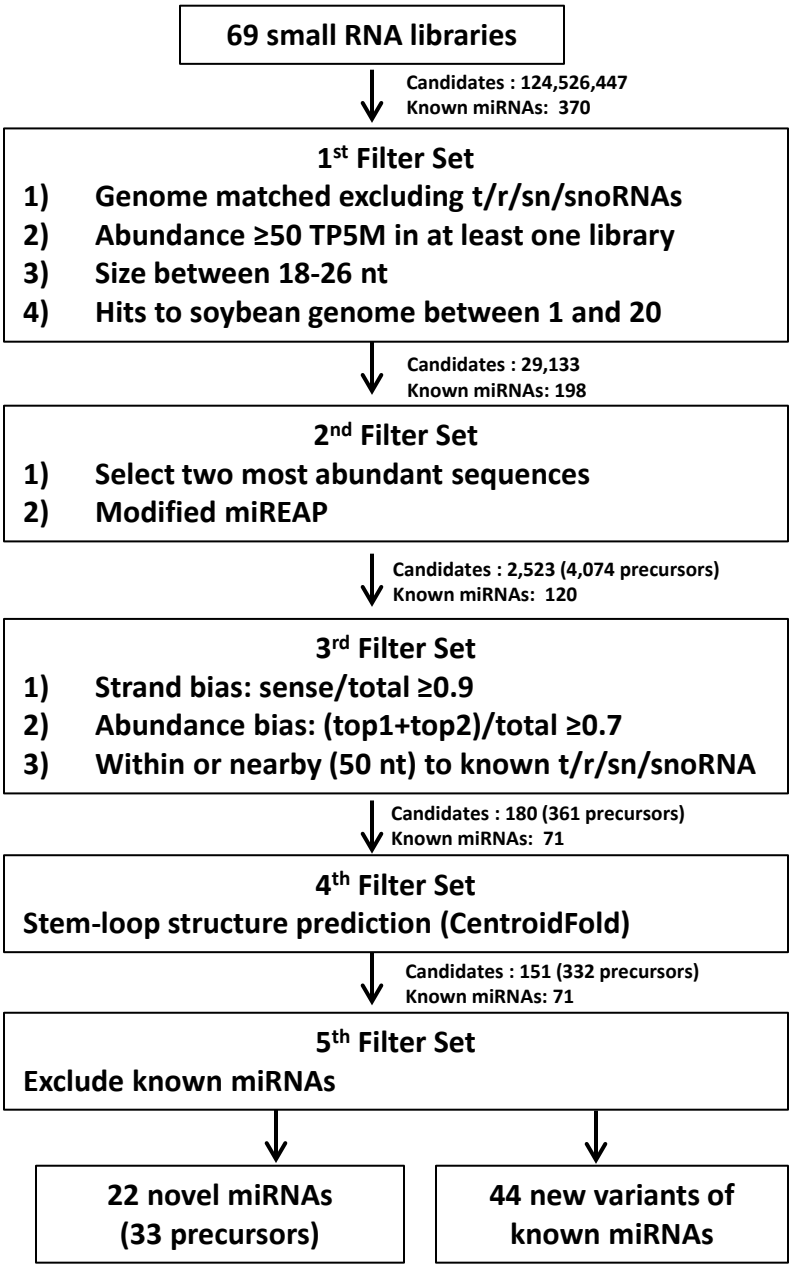


Supplemental Figure 3. Examples of miRBase-annotated miRNAs in our four categories of assessment.

The miRBase-annotated miRNA of interest is circled in red, with a red arrowhead for additional emphasis. One example of each type that we identified is shown, including the following: (A) Weakly-expressed loci. (B) siRNA-like miRNA loci. (C) Marginal miRNA loci. (D) Typical miRNA loci.

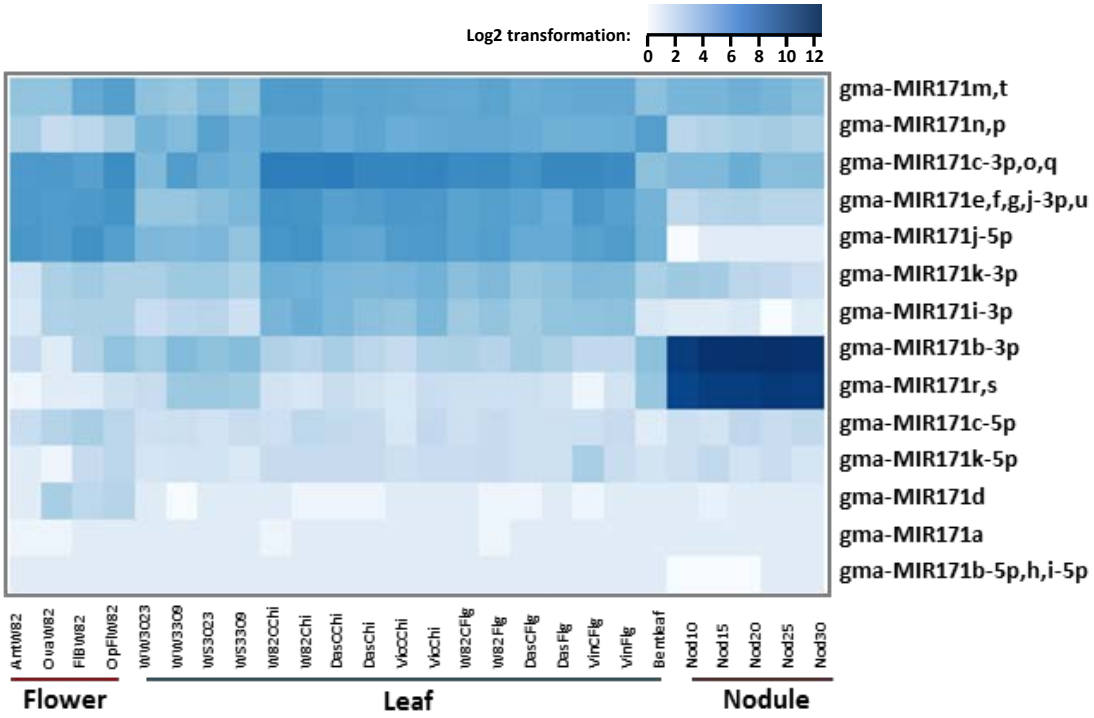


Supplemental Figure 4. The pipeline for the identification of novel miRNA and miRNA variants.



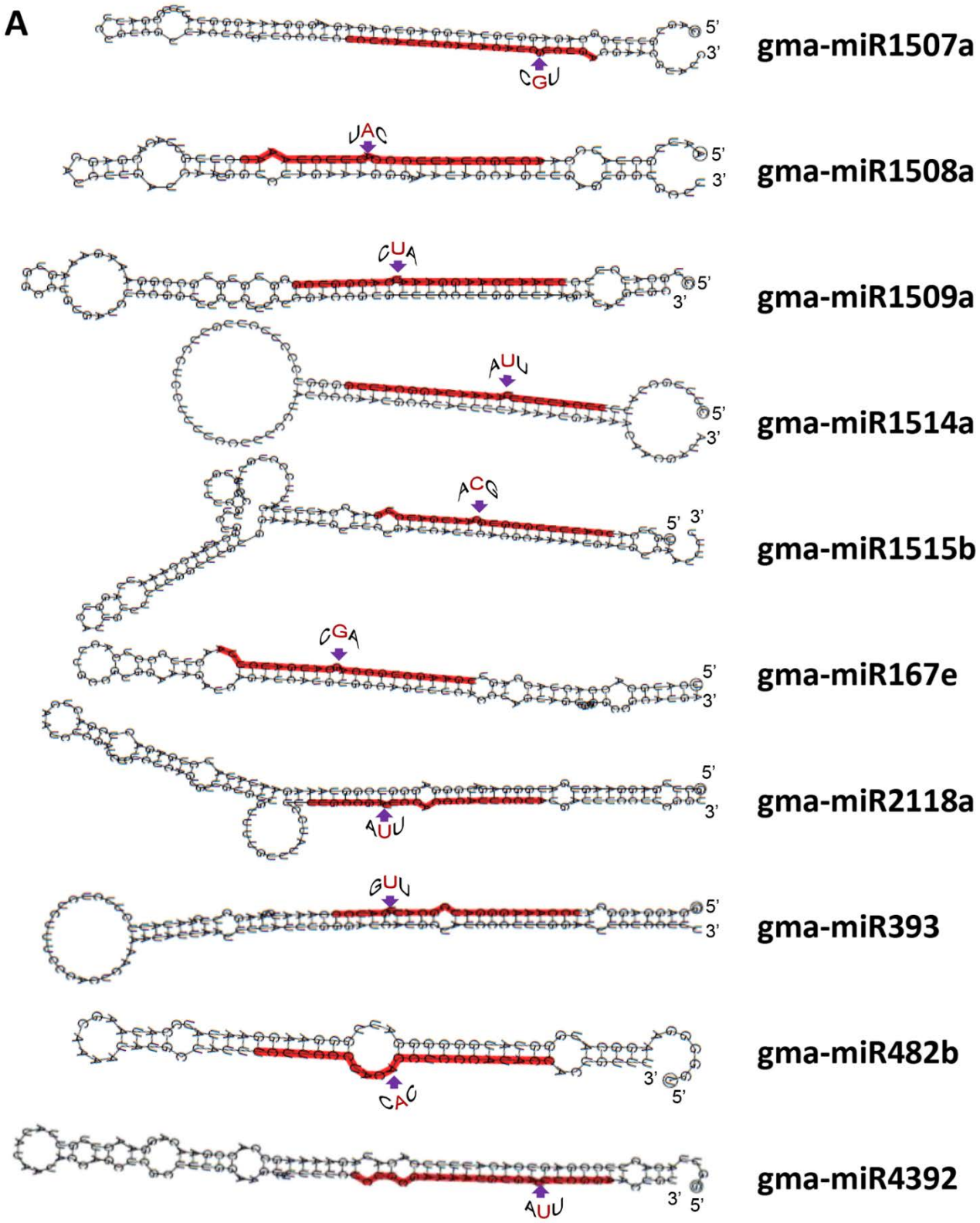
Supplemental Figure 5. Abundance profiles of the miR171 family.

The log2-transformed values of normalized read abundance (TP5M) in flower, leaf and nodule tissues are displayed as a heat map. The miRNA members with the same sequences were collapsed into one row.

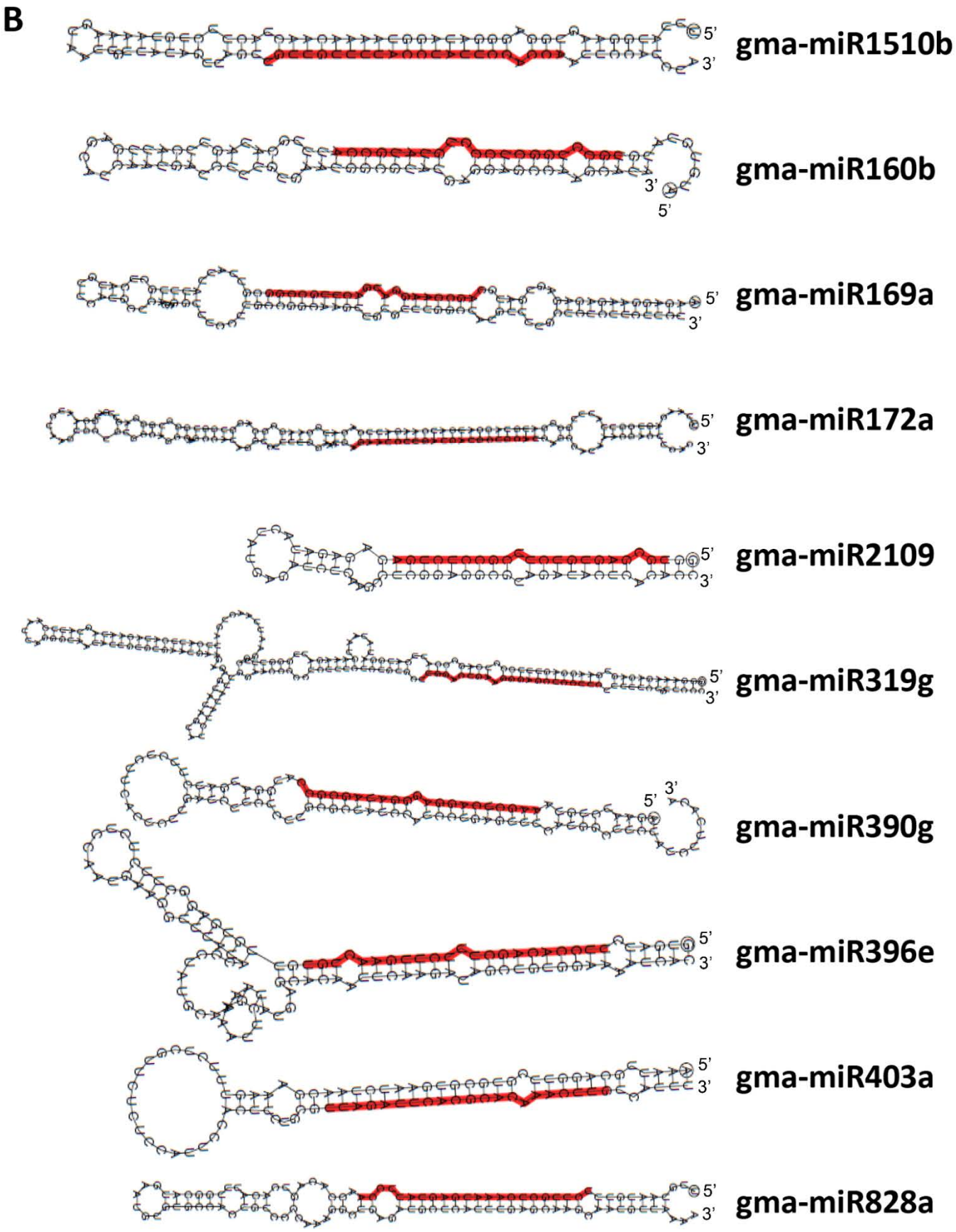


Supplemental Figure 6. Secondary structures for the stem-loop region of precursors of the soybean miRNAs that are triggers of *PHAS* loci.

Sequences highlighted in red are the mature miRNA sequence.
(A) miRNA triggers with asymmetric bulges in the precursors; the asymmetric bulge in the stem-loop is indicated by a purple arrow, combined with a “zoomed-in” view of the nucleotides found at the bulge.
(B – next page) miRNA triggers lacking an asymmetric bulge in the precursor.

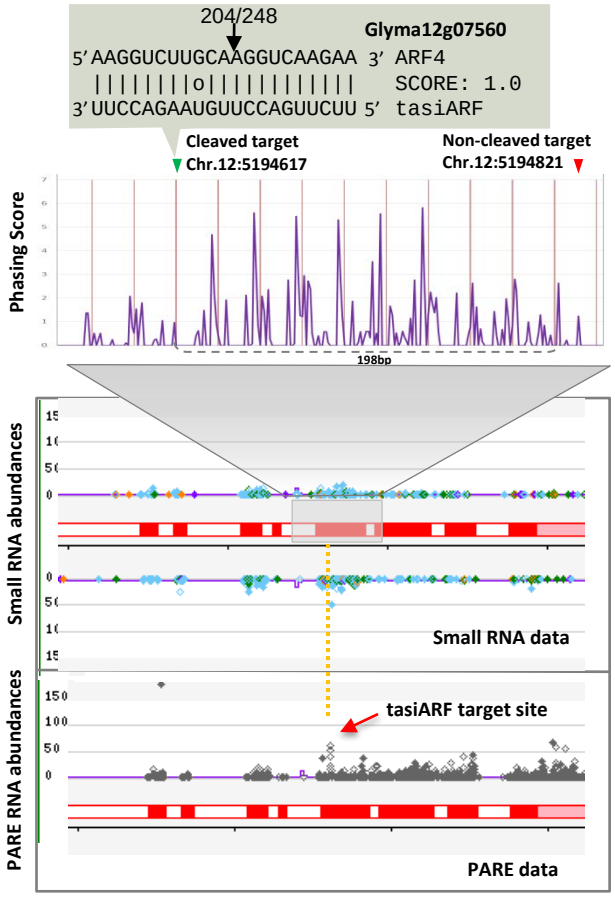


Supplemental Figure 6. (continued)



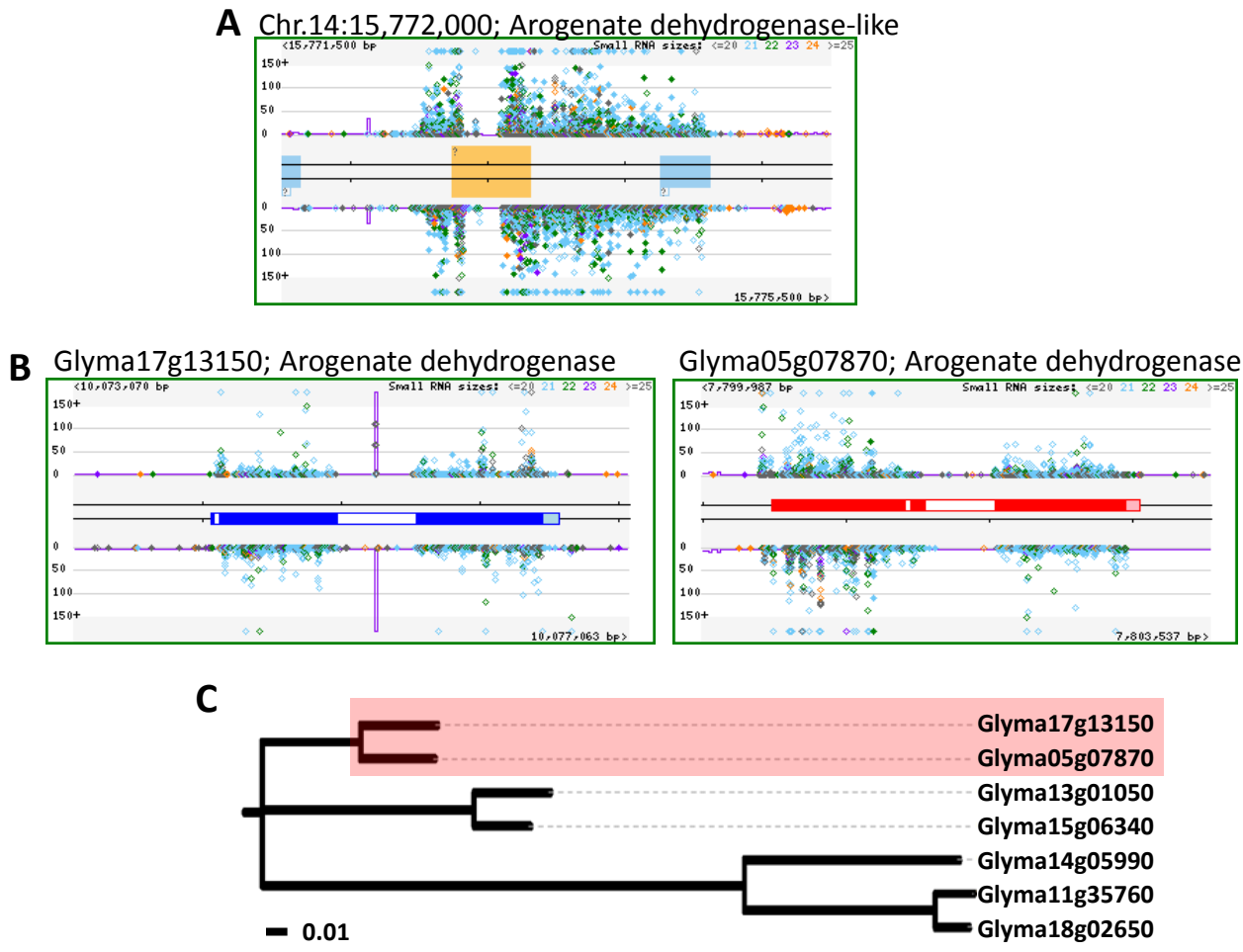
Supplemental Figure 7. PhasiRNAs from the *ARF4* locus are triggered by tasiARFs.

Small RNA and PARE data demonstrate that like *ARF3/ETT* (Figure 4), the *ARF4* locus is targeted by 21-nt tasiARFs at both cleaved and non-cleaved sites, triggering secondary siRNAs biogenesis via the two-hit mechanism. At the top, the cleavage position is indicated with a black arrow in the *tasiARF-ARF4* sequence alignment. See the legend of Figure 4 for a full description of the images and symbols.



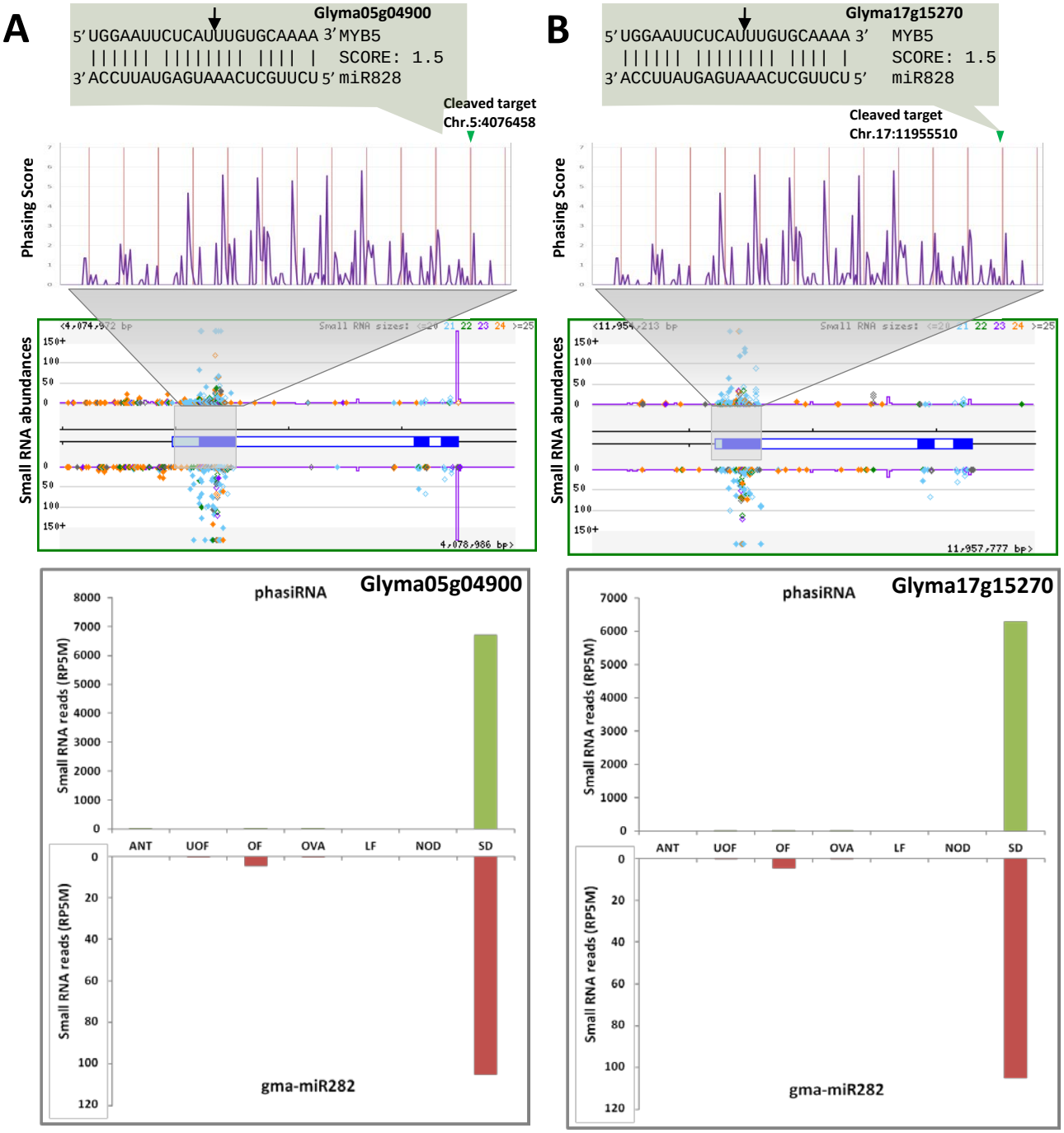
Supplemental Figure 8. Several soybean loci related to arogenate dehydrogenase-encoding transcripts are abundant sources of 21-nt phasiRNAs.

(A) An intergenic (unannotated) region on chromosome 14 approximately between positions 15,772,000 and 15,775,000 generates an abundant set of secondary siRNAs (blue dots are 21-mers, green are 22-mers). This sequence is predicted to form a long hairpin structure (data not shown); this is recognizable by the presence of an inverted repeat at the loop region (orange box). The blue boxes in the center indicate tandem repeats. Other features are as described in the legend of Figure 4.
(B) Two arogenate dehydrogenase-encoding genes in soybean are *PHAS* loci.
(C) A tree of the arogenate dehydrogenase proteins encoded in the soybean genome, with gene identifiers at right. The clade highlighted in pink contains the two loci producing phasiRNAs. The tree was produced using from the alignment generated by CLUSTALW.



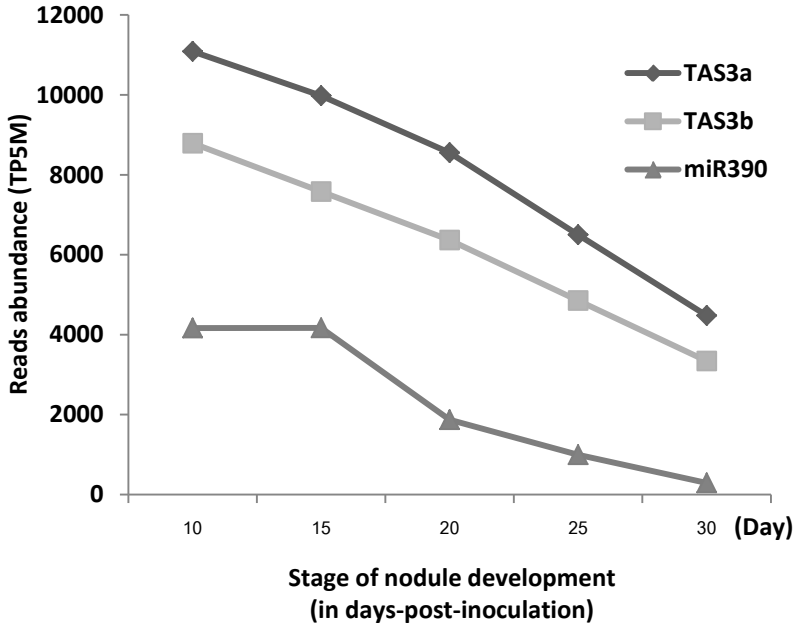
Supplemental Figure 9. Seed-specific *PHAS* loci from *MYB5* genes.

Small RNA plots and phasing scores of *MYB5* phasiRNA precursors ([A] is Glyma05g04900, [B] is Glyma17g15270). The target binding sites for the miRNA triggers are indicated with green triangles and the cleavage positions are indicated with black arrows in the sequence alignment at the top. In the lower panels, the green and red bars show the levels of normalized read abundances for phasiRNAs and miRNA triggers, respectively, in different tissues for these two *MYB5* genes.



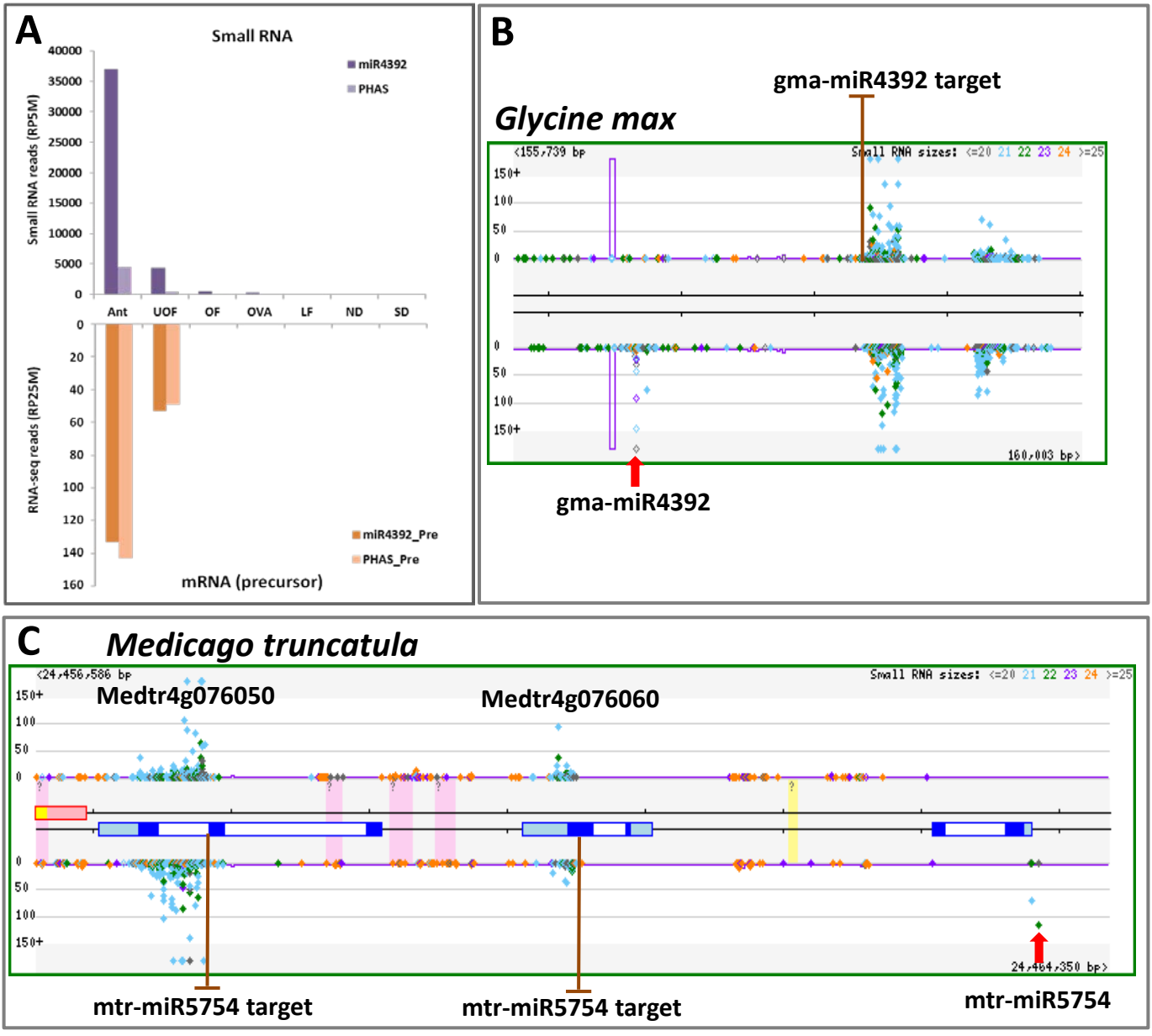
Supplemental Figure 10. Decrease of miR390 and *TAS3a* and *TAS3b* read abundances during nodule development.

Normalized read abundances of miR390 and 21-nt phasiRNAs generated from *TAS3a* and *TAS3b* in developing nodules, from 10 to 30 days after inoculation, measured over an interval of five days.



Supplemental Figure 11. Co-localization of legume anther- or flower-specific *PHAS* loci and their miRNA triggers.

(A) Small RNA abundances of miR4392 and phasiRNAs generated from the adjacent *PHAS* locus; both sets of small RNAs accumulate exclusively in soybean flower tissues, particularly in anthers. The RNA-seq read abundances of the miRNA and PHAS precursors confirm the exclusive expression of both loci in flower tissues.
(B) The target site of miR4392 on the soybean chromosome is indicated with brown lines, and the position of the miR4392 locus and the mature miRNA is indicated with a red arrow.
(C) The target site of miR5754 on Medicago chromosome 4 is indicated in brown. The locus generating miR5754 is nearby, with the mature miRNA indicated with a red arrow.



Supplemental Figure 12. A model of transposable element control in legume pollen by pollen-specific phasiRNAs.
The MIRNA gene and the PHAS genes localized in the adjacent region are shown as blue and green horizontal bars, respectively. The long black line represents a portion of chromosome region. The MIRNA gene transcribes a miRNA precursor, which will be subsequently processed to the 22-nt miRNA by a dicer-like protein (DCL, normally DCL1). The 22-nt miRNA will bind to an Argonaute (AGO), targeting the transcript produced from the PHAS gene. A set of 21-nt phasiRNAs will then be generated from the PHAS gene transcript. These phasiRNAs, mediated by AGO, will target hundreds of transposable elements (TEs) genome-widely and probably control those TEs via transcription (TGS) or post-transcriptional gene silencing (PTGS) mechanisms. The twenty soybean chromosomes are shown in green vertical bars and the locations of predicted TE targets are indicated as red triangles.