



# Systematic Characterization of MicroRNA Processing Modes in Plants With Parallel Amplification of RNA Ends

Ning Li and Guodong Ren\*

State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, School of Life Sciences, Institute of Plant Biology, Fudan University, Shanghai, China

## OPEN ACCESS

### Edited by:

Xiaozeng Yang,  
Beijing Academy of Agricultural  
and Forestry Sciences, China

### Reviewed by:

Andrzej Mirosław Pacak,  
Adam Mickiewicz University, Poland  
Jose Luis Reyes,  
National Autonomous University  
of Mexico, Mexico

### \*Correspondence:

Guodong Ren  
gdren@fudan.edu.cn

### Specialty section:

This article was submitted to  
Plant Physiology,  
a section of the journal  
Frontiers in Plant Science

**Received:** 12 October 2021

**Accepted:** 19 November 2021

**Published:** 07 December 2021

### Citation:

Li N and Ren G (2021) Systematic  
Characterization of MicroRNA  
Processing Modes in Plants With  
Parallel Amplification of RNA Ends.  
*Front. Plant Sci.* 12:793549.  
doi: 10.3389/fpls.2021.793549

In plants, the RNase III-type enzyme Dicer-like 1 (DCL1) processes most microRNAs (miRNAs) from their primary transcripts called pri-miRNAs. Four distinct processing modes (i.e., short base to loop, sequential base to loop, short loop to base, and sequential loop to base) have been characterized in Arabidopsis, mainly by the Specific Parallel Amplification of RNA Ends (SPARE) approach. However, SPARE is a targeted cloning method which requires optimization of cloning efficiency and specificity for each target. PARE (Parallel Amplification of RNA Ends) is an untargeted method *per se* and is widely used to identify miRNA mediated target slicing events. A major concern with PARE in characterizing miRNA processing modes is the potential contamination of mature miRNAs. Here, we provide a method to estimate miRNA contamination levels and showed that most publicly available PARE libraries have negligible miRNA contamination. Both the numbers and processing modes detected by PARE were similar to those identified by SPARE in Arabidopsis. PARE also determined the processing modes of 36 Arabidopsis miRNAs that were unexplored by SPARE, suggesting that it can complement the SPARE approach. Using publicly available PARE datasets, we identified the processing modes of 36, 91, 90, and 54 miRNAs in maize, rice, soybean, and tomato, respectively, and demonstrated that the processing mode was conserved overall within each miRNA family. Through its power of tracking miRNA processing remnants, PARE also facilitated miRNA characterization and annotation.

**Keywords:** microRNA, plants, degradome, parallel analysis of RNA ends (PARE), DCL1

## INTRODUCTION

MiRNAs are a class of endogenous small non-coding RNAs that direct post-transcriptional gene silencing through bases complementary with their target genes. Most miRNAs are transcribed by DNA-dependent RNA polymerase II (Pol II); they are folded into imperfect stem-loop structures, and undergo processing by the RNase III enzyme-containing microprocessor (Song et al., 2019; Wang et al., 2019). Precise releasing of miRNA/miRNA\* duplexes from their precursor RNAs is vital for miRNA biogenesis with minimal off targets. Understanding the molecular features of miRNA processing is beneficial for designing artificial miRNAs with enhanced efficiency.

In animals, pri-miRNAs exhibit rigid secondary structures that consist of a ~35-bp stem, a terminal loop, and long single-stranded RNAs (ssRNAs) flanking the fold-back region (Han et al., 2006). Pri-miRNAs are diced by Droscha at the lower stem ~11-bp away from the ssRNA-dsRNA junction, releasing pre-miRNAs with 2 nucleotides (nt) overhangs (Han et al., 2006; Kwon et al., 2016). After export to the cytoplasm, a second cut is performed by Dicer at ~22-bp upstream of the Droscha cut site, yielding mature miRNA-5p/miRNA-3p (miR-5p/miR-3p) duplexes (Grishok et al., 2001; Ketting et al., 2001; Treiber et al., 2019). In plants, however, pri-miRNAs are more heterogeneous in size, with fold-back lengths varying from 60 to over 500 nt (Reinhart et al., 2002; Zhang et al., 2006; Cuperus et al., 2011). Consequently, four different processing modes have been described: (a) short base to loop; (b) short loop to base; (c) sequential base to loop; and (d) sequential loop to base (Bologna and Voinnet, 2014). In the canonical short base-to-loop mode, an internal bulge at 15–17 bp below the miR-5p/miR-3p region guides the first cut at the loop-distal miRNA site (Cuperus et al., 2010; Mateos et al., 2010; Bologna et al., 2013; Moro et al., 2018). Whereas in the loop-to-base mode, a terminal loop or bulge at 15–17 bp above the miR-5p/miR-3p region directs DCL1 processing from the loop-proximal cleavage site (Bologna et al., 2009, 2013; Axtell et al., 2011; Moro et al., 2018). In both cases, the 15–17 bp lower or upper stem tends to be conserved with paired status at different taxonomic levels (Chorostecki et al., 2017). Longer stems may cause additional cuts, which are termed sequential base-to-loop or loop-to-base processing modes, according to their processing direction.

Specific Parallel Amplification of RNA Ends (SPARE) is a modified 5' RACE (rapid amplification of cDNA ends) with pri-miRNA specific primers for targeted cloning of 3' DCL processing remnants. It provides a high-throughput solution for the identification of miRNA processing modes (Bologna et al., 2013). Although the method provides high precision and sensitivity, it only clones DCL processing remnants with pre-designed primers. Parallel Amplification of RNA Ends [PARE; also referred to as degradome sequencing and GMUCT (genome-wide mapping of uncapped transcripts)] is a well-established method that captures 5' termini of uncapped and polyadenylated RNA fragments. It has been widely used to track miRNA mediated target slicing events and to estimate RNA degradation levels (Addo-Quaye et al., 2008; German et al., 2008; Gregory et al., 2008). In principle, PARE could also detect 3' dicing products and infer miRNA processing modes in an untargeted manner. In fact, it has been employed to discover the conserved loop-first processing of miR319 precursors in plants (Addo-Quaye et al., 2009). Concern over mature miRNAs contamination limits its use in probing miRNA processing (Lu et al., 2009). However, there have been no rigorously tests to determine whether PARE tags matching mature miRNA sequences are a result of contamination or not.

Using SPARE results in *Arabidopsis* as a benchmark, we here provided a solution for the estimation of mature miRNA contamination of PARE data. We found that most PARE tags matching mature miRNA sequences were likely not miRNA

contaminations. Overall, PARE had comparable accuracy and sensitivity to SPARE and could be used as a complementary method. Using publicly available PARE data, we systematically identified miRNA processing modes in four crop species and found that miRNAs within the same family tended to share the same processing modes across species. We also showed that PARE could provide independent supporting evidence during miRNA annotation when combined with known miRNA prediction/annotation tools, which are largely relied on small RNA sequencing (sRNA-seq) data.

## MATERIALS AND METHODS

### Data Sources

PARE and sRNA-seq datasets from *Arabidopsis thaliana*, *Zea mays*, *Oryza sativa*, *Glycine max*, and *Solanum lycopersicum* were downloaded from the Sequence Read Archive (SRA) database (Leinonen et al., 2011). The accession numbers are listed in **Supplementary Table 1**. The SPARE datasets of *Arabidopsis thaliana* used in this study are stored under the accession numbers SRP021538 and SRP137005.

The miRNA annotation information for *Arabidopsis*, maize, and rice were retrieved from miRbase v.22 (Kozomara and Griffiths-Jones, 2011). The miRNA annotation information for soybean and tomato were obtained from the Plant small RNA genes website (Lunardon et al., 2020). The genome sequences of different species were retrieved from TAIR (TAIR10) (Lamesch et al., 2012), Ensembl Plants (B73\_RefGen\_v4, IRGSP-1.0, SL2.50) (Bolser et al., 2016), and the Plant small RNA genes (*Glycine\_max\_V1-0*) (Lunardon et al., 2020).

### Parallel Amplification of RNA Ends, Specific Parallel Amplification of RNA Ends, and sRNA-Seq Analysis

Fastq-dump was used to convert SRA format to fastq format (Leinonen et al., 2011). FastQC<sup>1</sup> and fastp (Chen et al., 2018) were used for quality evaluation and adapter trimming, respectively. Reads with 20 nt (PARE), 18–51 nt (SPARE) and 17–26 nt (sRNA) were kept for further analysis. ShortStack (Axtell, 2013) was used to assign multiple-mapped PARE and sRNA-seq reads to the genome. Bowtie (Langmead, 2010) was used to map SPARE reads to the genome.

### miR-3p Information

For miR-3p with poor or no annotation, after loading merged sRNA-seq alignment bam files (**Supplementary Table 1**) with Integrative Genomics Viewer (IGV)-sRNA genome browser, we folded the precursor using the built-in RNAfold program and determined the miR-5p/miR-3p positions according to the most abundant reads on one side (usually the miR-5p in this case) and deduced the other side by the 2 nt 3' overhang rule.

<sup>1</sup><https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

## miRNA Processing Modes Characterization

Only precursors with 10 reads in at least one sample, or five reads in at least two samples within a  $\pm 50$  nt window surrounding miR-3p were kept. Deduction of miRNA processing modes from patterns of PARE signatures is depicted in **Figure 1A** and is determined manually.

## RESULTS

### Deducing miRNA Processing Modes From Parallel Amplification of RNA Ends Patterns Around miR-3p

In PARE, polyadenylated RNAs are isolated, and an RNA adaptor containing an *MmeI* enzyme recognition site is ligated to the 5' end of uncapped RNA fragments. *MmeI* cuts reverse transcribed DNA  $\sim 20$  nt downstream of its recognition site. The cleaved 20-nt tags are cloned and sequenced to track the 5' end of uncapped RNAs (German et al., 2008). In principle, for miRNA processed from loop to base, two (short loop to base) or more (sequential loop to base) 3' remnants of DCL cleavage products will be cloned. Meanwhile, for miRNA processed from base to loop, only one 3' remnant of the DCL processing products downstream of the first cleavage site carries the poly(A) tail, and consequently will be cloned. **Figure 1A** shows schematic PARE patterns for different processing modes.

To test whether PARE is suitable for the characterization of miRNA processing modes, we collected 24 Arabidopsis PARE datasets from public databases (**Supplementary Table 1**). After removing adapters, the 20-nt of degradome tags were kept for further analysis. ShortStack software was employed to allocate multi-mapping reads based on the number of unique mapped reads on different loci (Johnson et al., 2016).

We selected four miRNAs with different processing modes pre-determined by SPARE and checked whether they displayed expected PARE patterns. For the short loop-to-base processed miR162a, we detected a major PARE peak at miR162a-3p and a minor peak afterward (**Figure 1B**). For the sequentially loop-to-base processed miR319a, two additional PARE peaks upstream of miR319a-3p were found (**Figure 1C**). For the short base-to-loop processed miR168a, one major peak right after miR168a-3p was observed (**Figure 1D**), and finally, for the sequentially base-to-loop processed miR169b, a distinct peak 21 nt downstream of miR169b-3p was detected (**Figure 1E**). These data indicate the robustness of PARE in miRNA processing mechanism recognition.

### Quality Control for miRNAs Contamination

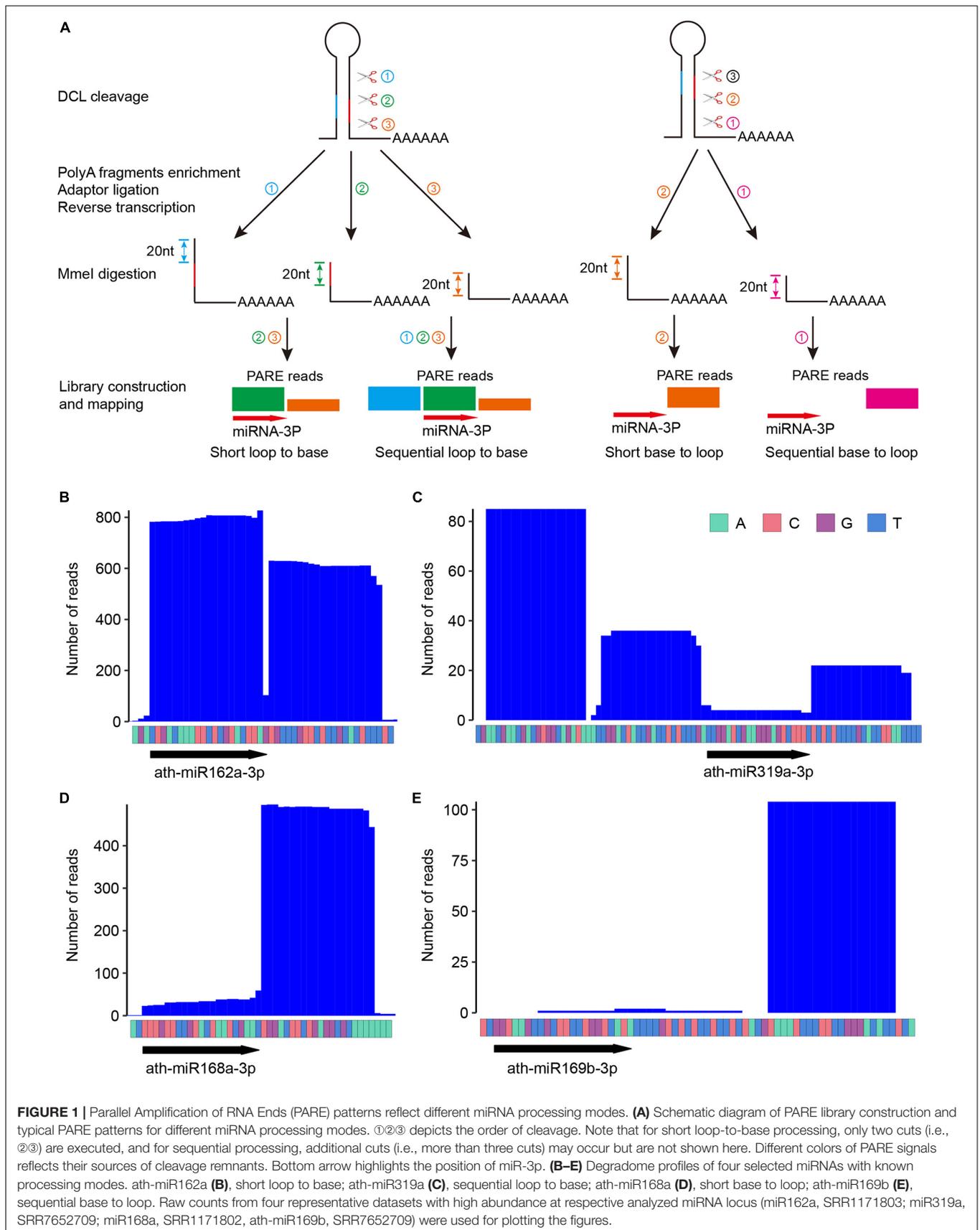
Because only 20-nt degradome tags were retrieved after *MmeI* digestion, one major concern regarding the use of PARE in characterizing miRNA processing modes was the contamination of mature miRNAs (Lu et al., 2009). As such, a quality control step for each PARE dataset was crucial to avoid misidentification. As described above, base-to-loop processed miRNAs should have

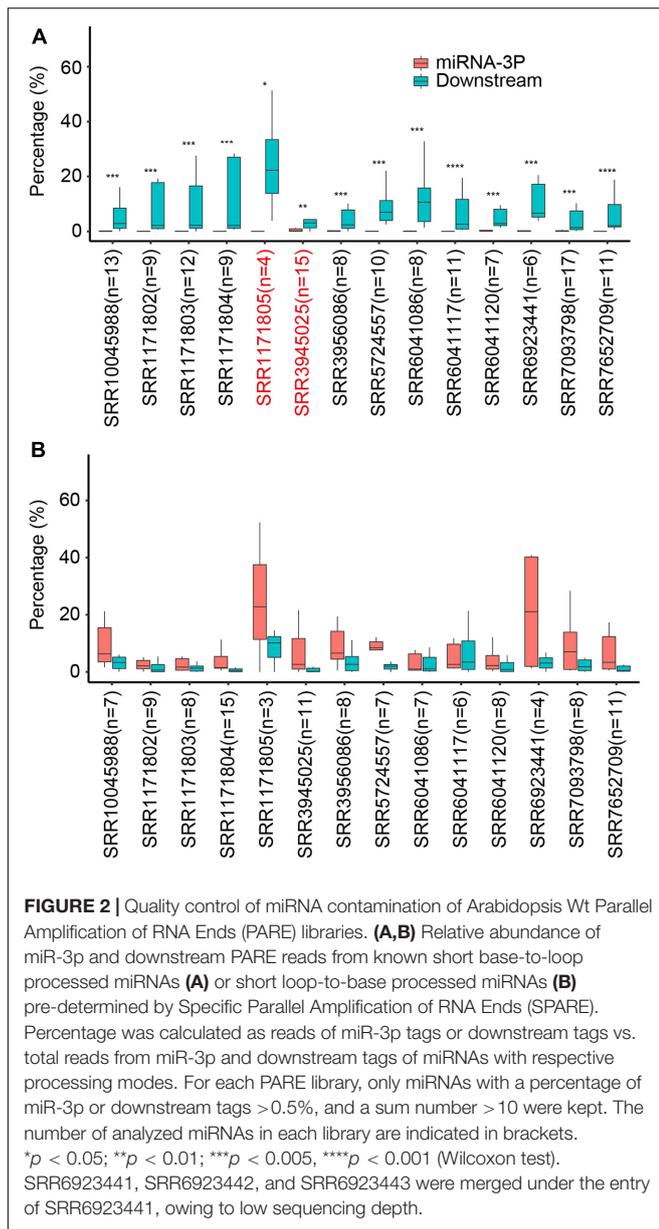
no miR-3p tags if there is no contamination; this could therefore be used to estimate the miRNAs contamination levels of PARE data. 16 wild-type (Wt) PARE libraries (Three of them from a same experiment were combined due to low sequencing depth) were analyzed for possible miRNA contamination. Remarkably, for miRNAs with known short base-to-loop processing modes, distinct PARE tags were observed right after miR-3ps, with almost no miR-3ps detected in most analyzed libraries (**Figure 2A**). In sharp contrast, for miRNAs with known short loop-to-base processing modes, significantly higher miR-3p tags than downstream tags were detected (**Figure 2B**). This strongly suggested that in most PARE libraries the contamination of mature miRNAs was low and had negligible impact on miRNA processing modes determination. Twelve out of 14 libraries passed the quality control step with a stringent *p*-value cut-off (0.005).

### Parallel Amplification of RNA Ends as a Promising Tool in Systematic Characterization of miRNA Processing Modes

We next determined the miRNA processing modes in Arabidopsis based on PARE patterns and compared them with SPARE results. We not only validated most of the results (71/107, 66%) by SPARE, but, more importantly, we determined the processing modes of 36 additional miRNAs (**Supplementary Table 2** and **Figures 3A,B**). In particular, PARE showed comparable first-cut accuracy as SPARE (**Figure 3C**; Moro et al., 2018), indicating the robustness of PARE in the characterization of miRNA processing modes. The processing modes of 49 miRNAs from the two unpassed libraries were identified, and surprisingly all were concordant with those from passed libraries (**Supplementary Figure 1**).

Notably, six miRNAs (miR158a, miR390a, miR390b, miR391, miR396a, and miR396b) displayed inconsistent processing modes between PARE and SPARE (**Figure 3A**). For miR391, the conclusion by PARE may have been inaccurate as only a few degradome tags were obtained (**Supplementary Figure 2**). For the remaining five miRNAs, SPARE annotated them as short base-to-loop, but all PARE data treated them as short loop-to-base. We thus reanalyzed the SPARE signals around these miRNAs. For miR158a, miR390b, miR396a, and miR396b, robust signals were only detected in the *fiery1* (*fry1*) mutant but not in the Wt (**Figures 4A,B** and **Supplementary Figures 3B,C**). For miR390a, robust and consistent signals were detected in both *fry1* and Wt (**Supplementary Figure 3A**). *fry1* has been frequently used in PARE and SPARE experiments because the *FRY1* mutation attenuates RNA degradation from 5' to 3' by XRN3, thereby accumulating more miRNA targets cleavage remnants and miRNA processing intermediates (Gy et al., 2007). A comparison between Wt and *fry1* revealed additional inconsistent patterns, including miR400 and miR408 (**Figures 4C,D**). More strikingly, in another amino acid substitution allele of *FRY1*, *sal1/alx8* (Wilson et al., 2009), the processing of more miRNAs changed from short base-to-loop to short/sequential loop-to-base, including miR166b,





miR167d, miR168a, miR403, miR841a, miR841b, and miR850 (Figures 4E–G and Supplementary Figures 3D–G), implying that the SAL1/FRY1 mutation may influence the characterization of miRNA processing and should be used carefully.

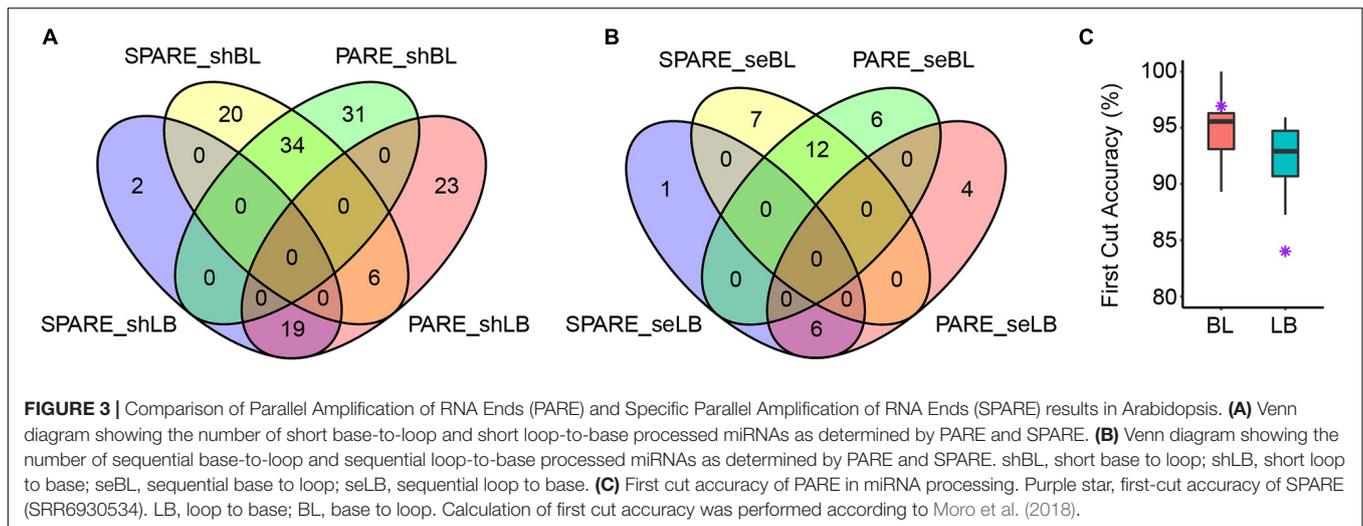
## Identification of miRNA Processing Modes in Four Crops

Though SPARE results from crops are lacking, evolutionary conservation analysis reveals that miRNAs within a same family tend to share conserved processing modes across angiosperms (Chorostecki et al., 2017). We retrieved 6, 16 (Four of them from a same experiment were combined due to low sequencing depth), 14, and 13 PARE datasets of *Zea mays*, *Oryza sativa*, *Glycine max* and *Solanum lycopersicum* from public databases

(Supplementary Table 1). MiRNAs whose counterparts in Arabidopsis are processed in short base-to-loop were used to evaluate the miRNA contamination levels in each library (Supplementary Figure 4). After filtering with a  $p$ -value cut-off of 0.05, two, nine, four, and seven libraries from *Zea mays*, *Oryza sativa*, *Glycine max* and *Solanum lycopersicum*, respectively, were kept for further analysis. For miRNAs lacking miR-3p annotation, sRNA-seq data (Supplementary Table 1) and RNA folding structures were used to infer their positions based on the 2-nt overhang rule of the miR-5p/miR-3p duplexes (Supplementary Table 3). Following the procedures described above, we obtained the degradome profiles around miR-3ps. We determined the processing modes of 36, 91, 90, and 54 miRNAs in *Zea mays*, *Oryza sativa*, *Glycine max*, and *Solanum lycopersicum*, respectively (Supplementary Table 2). The relatively fewer numbers in crops compared with Arabidopsis are likely owing to poor annotation of miRNAs and low or even no expression because of the limited amount of sequencing data. Consistent with previous notions (Chorostecki et al., 2017), an inspection of processing modes identified in at least three species revealed that 14 miRNA families (i.e., miR159, miR162, miR164, miR167, miR168, miR169, miR172, miR319, miR393, miR394, miR398, miR399, miR408, and miR2118) shared conserved processing modes (Figure 5). Members of the miR171 family in Arabidopsis utilize different processing modes (Bologna et al., 2013). Here, we also detected different processing modes in five miRNA families (i.e., miR156, miR160, miR166, miR396, and miR827) at different degrees (Figure 5). These analyses suggest that although members in the same miRNA family tend to share the same processing mode, differential processing modes may occur at intraspecific or interspecific levels.

## Parallel Amplification of RNA Ends Assists miRNA Annotation

Accurate detection of miRNA processing remnants suggests that PARE is a useful tool for assisting miRNA annotation and/or prediction. We re-examined the annotation of miRBase and corrected 43 records (Supplementary Table 4). For instances, the Arabidopsis miR169 family has 14 members with miR169a being the most abundant (Bologna et al., 2013). According to the annotation from miRBase, the length of ath-miR169a-3p is 20-nt with the sequence being GGCAAGUUGUCCUUGGCUAC. The 3' end of the ath-miR169a-3p reads from sRNA-seq data is ragged. In sharp contrast, we detected a sharp degradome signal starting 1 nt downstream of the end of the miRBase annotation (Figure 6A), revealing that the correct ath-miR169a-3p sequence should be 21-nt in size (i.e., GGCAAGUUGUCCUUGGCUACA). Gma-miR408c belongs to the conserved miR408 family and is processed in a short loop-to-base direction. The miRBase annotation shows that the sequence of gma-miR408c-3p is AUGCACUGCCUCUCCCCUGGC. We detected a major degradome peak that began 1 nt downstream of the above annotation (Figure 6B), suggesting that the majority of Gma-miR408c-3p started 1 nt after the annotated start site (i.e., UGCACUGCCUCUCCCCUGGCU). In both cases, the corrected version but not the miRBase annotation meets



the 2-nt overhang rule (Supplementary Figure 5). We also predicted the targets of different versions of *ath*-miR169a-3p and *gma*-miR408c-3p with TarHunter and determined their cleavage site using PARE datasets (Ma et al., 2017) (Supplementary Figure 6 and Supplementary Table 5). For *ath*-miR169a-3p, one same potential target was retrieved, which showed a weak target plot (T-plot) signal. This could be due to the fact that *ath*-miR169a-3p is the passenger strand of the miR169a/\* duplex. For *gma*-miR408c-3p, multiple conserved copper-related targets were identified with strong T-plots signals for both versions (Ma et al., 2015). An additional target (EBP1) was retrieved only for the corrected version (Supplementary Figure 6 and Supplementary Table 5). Importantly, all the cleavage sites were located at positions complementary to the 10th and the 11th nucleotides of the corrected *gma*-miR408c-3p, which is canonical for miRNA-guided slicing (Kasschau et al., 2003).

Taken together, our study suggested that PARE was effective for comprehensively analyzing miRNA processing modes and assisting miRNA annotation in an untargeted manner with satisfactory accuracy.

## DISCUSSION

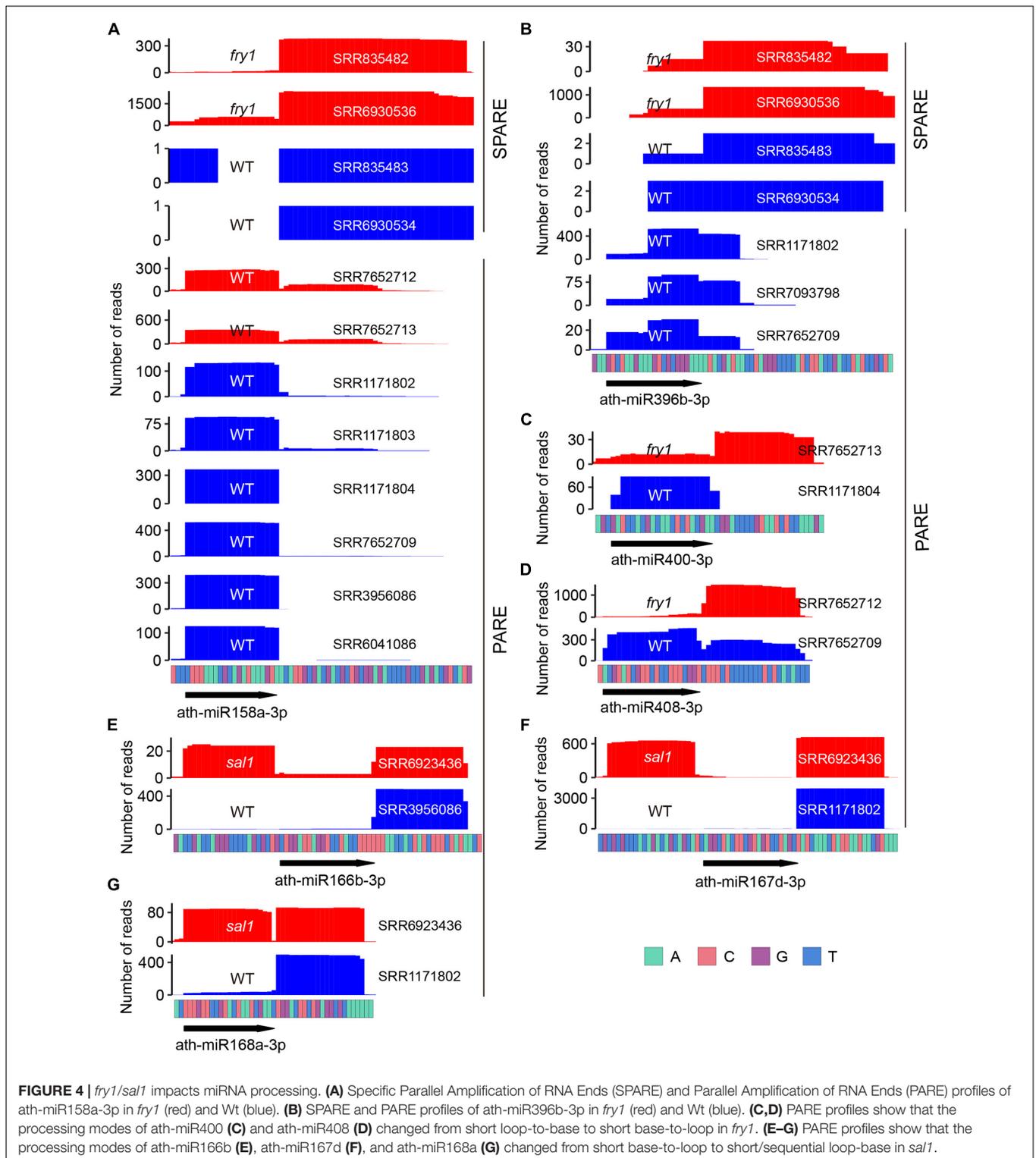
Having more variable hairpin sizes and fold-back structures than their animal counterparts, plants evolve more complicated miRNA processing modes. Hitherto, systematic investigation of miRNA processing modes has only been conducted in the model plant *Arabidopsis* using SPARE. Though the same miRNA families across different species tend to share conserved processing modes, exceptions are frequently observed, and as such, experimental approaches for the characterization of miRNA processing modes are indispensable.

Although effective, SPARE is a targeted approach that only captures selected DCL processing remnants and is time-consuming. By contrast, PARE is an untargeted approach that captures 5' end uncapped and 3' end polyadenylated RNAs. More

importantly, PARE is commercialized and has been widely used to determine the miRNA targets in various plant species and many PARE datasets are publicly available. In principle, PARE is capable of tracking miRNA processing intermediates and has been used to determine the processing modes of miR319 and miR159 (Addo-Quaye et al., 2009; Li et al., 2011). Yet, concerns of potential miRNA contamination impede its application in the systematic characterization of miRNA processing modes (Lu et al., 2009). Here, we provide a solution for evaluating miRNA contamination and demonstrate that PARE can be used to dissect miRNA processing modes globally. As an untargeted approach, PARE can identify miRNA processing modes in an unbiased manner. On the other hand, PARE has less sensitivity and requires higher sequencing depth. Moreover, some PARE libraries may have higher miRNA contamination that influence the prediction (Figure 2 and Supplementary Figure 4). Thus, a quality control step for miRNA contamination is crucial for accuracy. Alternatively, replacing *MmeI* with *EcoP15I*, which produces 27-nt tags will effectively overcome this defect (Addo-Quaye et al., 2009; Li et al., 2019). Overall, SPARE and PARE can complement each other in dissecting miRNA processing modes in plants.

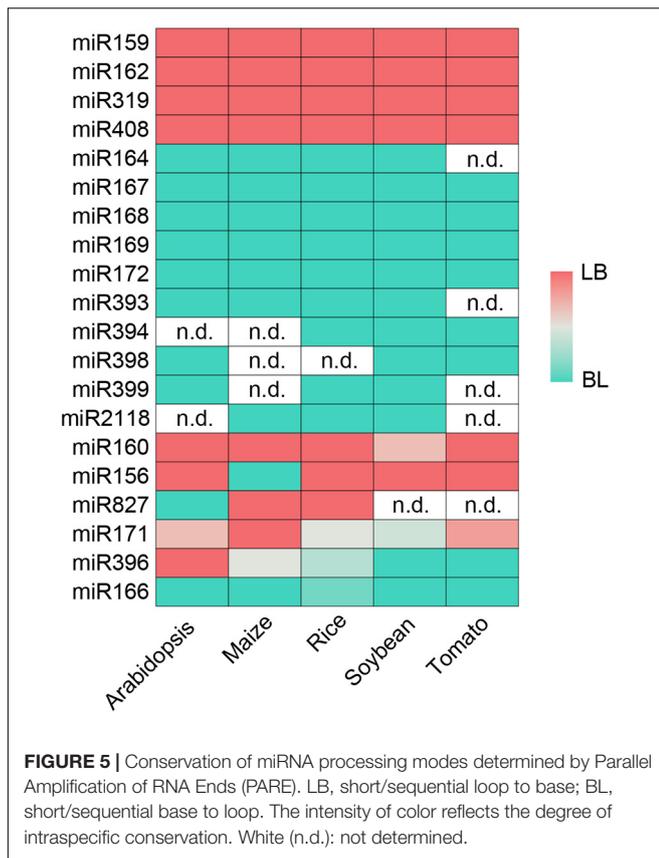
During our analysis, we also frequently detected obvious degradome signals at the beginning of the miR-5p or internal of miRNAs (Supplementary Figure 7). This could be owing to partial or misprocessing by DCL1, or slicing by ARGONAUTE 1 (AGO1) (Bologna et al., 2013). Other possible causes include: (i) 5' processing remnants are re-polyadenylated and captured during library construction (Chekanova et al., 2007); (ii) the 5' end of 3' processing remnants are trimmed *in vivo* or during library construction *in vitro*; and (iii), multiple overlapped miRNAs/siRNAs exist in the same primary transcript that generates complex processing remnants (Reinhart et al., 2002; Allen et al., 2004; German et al., 2008).

*FRY1/SAL1* encodes a phosphatase with dual activities, including converting 3'-phosphoadenosine 5'-phosphosulfate (PAPS) into adenosyl phosphosulfate (APS) and dephosphorylates 3'-phosphoadenosine 5'-phosphate (PAP)



to adenosine 5'-phosphate (AMP) (Quintero et al., 1996; Gil-Mascarell et al., 1999). *FRY1* plays important roles in multiple biological processes including post-transcriptional gene silencing (PTGS) (Gy et al., 2007) and RNA quality control (You et al., 2019). In the *fry1* mutant, the accumulation of toxic

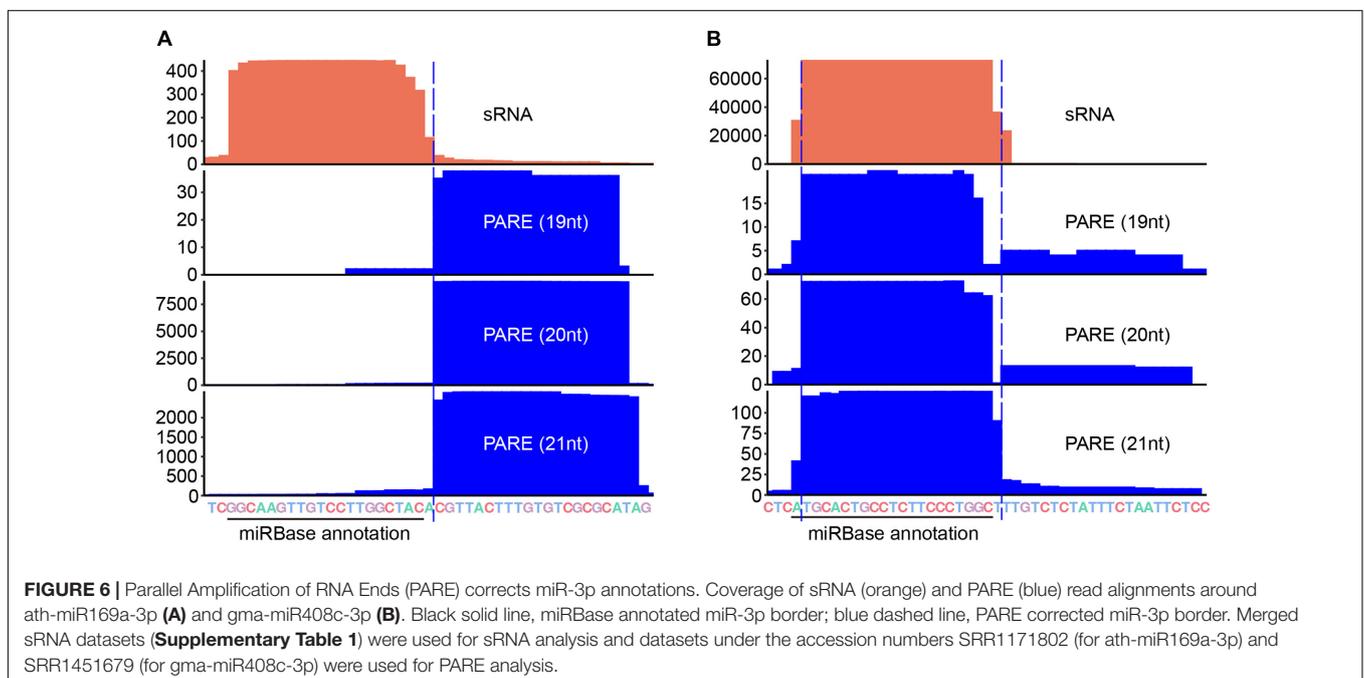
PAP in *fry1* impaired the activity of 5' to 3' exoribonucleases (XRNs). Consequently, *MIRNA*-derived loop and 3' products become over-accumulated owing to the inhibition activities of XRN2 and XRN3 (Gy et al., 2007). It has been reported that 90% of degradome tags of loop-to-base processed



miRNAs in *fry1* correspond to the position of the second cleavage site (Moro et al., 2018), indicating that *FRY1* may influence miRNA processing. Strikingly, opposite processing

modes were frequently detected between Wt and *fry1/sal1* (Figure 4 and Supplementary Figure 3). Moreover, dramatic elevated partial cleavage was observed only in *fry1/sal1* (Supplementary Figure 8). Thus, *fry1* should be used only cautiously in the identification of processing mode. We analyzed the *fry1* small RNA sequencing data and found that misprocessed miRNA ratios were slightly elevated in *fry1-6* and *fry1-8* (Supplementary Figure 9), suggesting that *FRY1* may also affect DCL1 processing. Nuclear exosome components HUA ENHANCER 2 (HEN2) and SUPPRESSOR OF PAS2 1 (SOP1) act on miRNA precursor degradation with selective impacts on loop-to-base miRNA processing when HYPOPLASTIC LEAVES 1 (HYL1) is impaired (Gao et al., 2020). It will be interesting to investigate the relationship between 5' to 3' and 3' to 5' degradation pathways on miRNA processing.

In plants, exhaustive efforts have been paid to miRNA prediction and annotation with a huge number of sRNA-seq datasets and different prediction tools. Although miRNA isoforms are frequently reported, the heterogeneous ends of sRNA reads from sRNA-seq data can lead to misannotation, which may cause mistaken inferences about their AGO sorting and/or target identification (Mi et al., 2008; Zhang et al., 2014; Supplementary Figure 6). As an independent method, we showed that PARE can assist with accurate miRNA annotation (Supplementary Table 4 and Figure 6). Moreover, owing to the complexity of small RNA compositions in plants, false positives lead to many questionable miRNA annotations, which have now become a major concern to the community (Axtell and Meyers, 2018). To solve this issue, the Axtell group developed ShortStack, which has high precision and near-zero false positives; however, it shows limited sensitivity and high false negatives. By analyzing 28 Arabidopsis sRNA-seq libraries, only 143 out of 325 entries in miRBase were designated as *bona fide*



miRNAs by ShortStack (Lunardon et al., 2020). Unpassed entries were largely due to imprecise processing, unpaired bases, bulges limitation and undetected miRNA\*. We suggest that PARE may also be used to help with miRNA characterization by tracking processing remnants. Our preliminary analysis revealed that 113 entries had robust PARE support, including 30 miRNAs that were not designated as miRNAs by ShortStack (**Supplementary Figure 10**). We believe that integration of degradome signatures into miRNA prediction tools will improve both accuracy and sensitivity.

In conclusion, we provided a solution for the estimation of miRNA contamination and demonstrated the capacity of PARE in characterizing miRNA processing and miRNA annotation.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

## REFERENCES

- Addo-Quaye, C., Eshoo, T. W., Bartel, D. P., and Axtell, M. J. (2008). Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. *Curr. Biol.* 18, 758–762. doi: 10.1016/j.cub.2008.04.042
- Addo-Quaye, C., Snyder, J. A., Park, Y. B., Li, Y. F., Sunkar, R., and Axtell, M. J. (2009). Sliced microRNA targets and precise loop-first processing of MIR319 hairpins revealed by analysis of the *Physcomitrella patens* degradome. *RNA* 15, 2112–2121. doi: 10.1261/rna.1774909
- 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. doi: 10.1038/ng1478
- Axtell, M. J. (2013). ShortStack: comprehensive annotation and quantification of small RNA genes. *RNA* 19, 740–751. doi: 10.1261/rna.035279.112
- Axtell, M. J., and Meyers, B. C. (2018). Revisiting criteria for Plant MicroRNA annotation in the era of big data. *Plant Cell* 30, 272–284.
- Axtell, M. J., Westholm, J. O., and Lai, E. C. (2011). Vive la difference: biogenesis and evolution of microRNAs in plants and animals. *Genome Biol.* 12:221. doi: 10.1186/gb-2011-12-4-221
- Bologna, N. G., Mateos, J. L., Bresso, E. G., and Palatnik, J. F. (2009). A loop-to-base processing mechanism underlies the biogenesis of plant microRNAs miR319 and miR159. *EMBO J.* 28, 3646–3656. doi: 10.1038/emboj.2009.292
- Bologna, N. G., Schapire, A. L., Zhai, J., Chorostecki, U., Boisbouvier, J., Meyers, B. C., et al. (2013). Multiple RNA recognition patterns during microRNA biogenesis in plants. *Genome Res.* 23, 1675–1689. doi: 10.1101/gr.153387.112
- Bologna, N. G., and Voinnet, O. (2014). The diversity, biogenesis, and activities of endogenous silencing small RNAs in Arabidopsis. *Annu. Rev. Plant Biol.* 65, 473–503. doi: 10.1146/annurev-arplant-050213-035728
- Bolser, D., Staines, D. M., Pritchard, E., and Kersey, P. (2016). Ensembl plants: integrating tools for visualizing, mining, and analyzing plant genomics data. *Methods Mol. Biol.* 1374, 115–140. doi: 10.1007/978-1-4939-3167-5\_6
- Chekanova, J. A., Gregory, B. D., Reverdatto, S. V., Chen, H., Kumar, R., Hooker, T., et al. (2007). Genome-wide high-resolution mapping of exosome substrates reveals hidden features in the Arabidopsis transcriptome. *Cell* 131, 1340–1353. doi: 10.1016/j.cell.2007.10.056
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560
- Chorostecki, U., Moro, B., Rojas, A. M. L., Debernardi, J. M., Schapire, A. L., Notredame, C., et al. (2017). Evolutionary footprints reveal insights into

## AUTHOR CONTRIBUTIONS

GR and NL conceived the idea and wrote the manuscript. NL conducted the analysis. Both authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by grants from the National Natural Science Foundation of China (31970275) and the “Shuguang Program” sponsored by Shanghai Education Development Foundation and Shanghai Municipal Education Commission (20SG02).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.793549/full#supplementary-material>

- plant MicroRNA biogenesis. *Plant Cell* 29, 1248–1261. doi: 10.1105/tpc.17.0.0272
- Cuperus, J. T., Fahlgren, N., and Carrington, J. C. (2011). Evolution and functional diversification of MIRNA genes. *Plant Cell* 23, 431–442. doi: 10.1105/tpc.110.082784
- Cuperus, J. T., Montgomery, T. A., Fahlgren, N., Burke, R. T., Townsend, T., Sullivan, C. M., et al. (2010). Identification of MIR390a precursor processing-defective mutants in Arabidopsis by direct genome sequencing. *Proc. Natl. Acad. Sci. U.S.A.* 107, 466–471. doi: 10.1073/pnas.0913203107
- Gao, S., Wang, J., Jiang, N., Zhang, S., Wang, Y., Zhang, J., et al. (2020). Hyponastic Leaves 1 protects pri-miRNAs from nuclear exosome attack. *Proc. Natl. Acad. Sci. U.S.A.* 117, 17429–17437. doi: 10.1073/pnas.2007203117
- German, M. A., Pillay, M., Jeong, D. H., Hetawal, A., Luo, S., Janardhanan, P., et al. (2008). Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nat. Biotechnol.* 26, 941–946. doi: 10.1038/nbt1417
- Gil-Mascarell, R., Lopez-Coronado, J. M., Belles, J. M., Serrano, R., and Rodriguez, P. L. (1999). The Arabidopsis HAL2-like gene family includes a novel sodium-sensitive phosphatase. *Plant J.* 17, 373–383. doi: 10.1046/j.1365-313x.1999.00385.x
- Gregory, B. D., O'Malley, R. C., Lister, R., Urich, M. A., Tonti-Filippini, J., Chen, H., et al. (2008). A link between RNA metabolism and silencing affecting Arabidopsis development. *Dev. Cell* 14, 854–866. doi: 10.1016/j.devcel.2008.04.005
- Grishok, A., Pasquinelli, A. E., Conte, D., Li, N., Parrish, S., Ha, I., et al. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* 106, 23–34. doi: 10.1016/s0092-8674(01)00431-7
- Gy, I., Gascioli, V., Laressergues, D., Morel, J. B., Gombert, J., Proux, F., et al. (2007). Arabidopsis FIERY1, XRN2, and XRN3 are endogenous RNA silencing suppressors. *Plant Cell* 19, 3451–3461. doi: 10.1105/tpc.107.05.5319
- Han, J., Lee, Y., Yeom, K. H., Nam, J. W., Heo, I., Rhee, J. K., et al. (2006). Molecular basis for the recognition of primary microRNAs by the Drosophila DGCR8 complex. *Cell* 125, 887–901. doi: 10.1016/j.cell.2006.03.043
- Johnson, N. R., Yeoh, J. M., Coruh, C., and Axtell, M. J. (2016). Improved Placement of Multi-mapping Small RNAs. *G3 (Bethesda)* 6, 2103–2111. doi: 10.1534/g3.116.030452
- Kasschau, K. D., Xie, Z. X., Allen, E., Llave, C., Chapman, E. J., Krizan, K. A., et al. (2003). P1/HC-Pro, a viral suppressor of RNA silencing, interferes with Arabidopsis development and miRNA function. *Dev. Cell* 4, 205–217. doi: 10.1016/S1534-5807(03)00025-X

- Ketting, R. F., Fischer, S. E., Bernstein, E., Sijen, T., Hannon, G. J., and Plasterk, R. H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.* 15, 2654–2659. doi: 10.1101/gad.927801
- Kozomara, A., and Griffiths-Jones, S. (2011). miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* 39, D152–D157. doi: 10.1093/nar/gkq1027
- Kwon, S. C., Nguyen, T. A., Choi, Y. G., Jo, M. H., Hohng, S., Kim, V. N., et al. (2016). Structure of human DROSHA. *Cell* 164, 81–90. doi: 10.1016/j.cell.2015.12.019
- Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., et al. (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 40, D1202–D1210. doi: 10.1093/nar/gkr1090
- Langmead, B. (2010). Aligning short sequencing reads with Bowtie. *Curr. Protoc. Bioinform.* Chapter 11, Unit 11.7. doi: 10.1002/0471250953.bi1107s32
- Leinonen, R., Sugawara, H., Shumway, M., International Nucleotide Sequence Database Collaboration (2011). The sequence read archive. *Nucleic Acids Res.* 39, D19–D21. doi: 10.1093/nar/gkq1019
- Li, Y., Li, C., Ding, G., and Jin, Y. (2011). Evolution of MIR159/319 microRNA genes and their post-transcriptional regulatory link to siRNA pathways. *BMC Evol. Biol.* 11:122. doi: 10.1186/1471-2148-11-122
- Li, Y. F., Zhao, M., Wang, M., Guo, J., Wang, L., Ji, J., et al. (2019). An improved method of constructing degradome library suitable for sequencing using Illumina platform. *Plant Methods* 15:134. doi: 10.1186/s13007-019-0524-7
- Lu, S. F., Sun, Y. H., and Chiang, V. L. (2009). Adenylation of plant miRNAs. *Nucleic Acids Res.* 37, 1878–1885. doi: 10.1093/nar/gkp031
- Lunardon, A., Johnson, N. R., Hagerott, E., Phifer, T., Polydore, S., Coruh, C., et al. (2020). Integrated annotations and analyses of small RNA-producing loci from 47 diverse plants. *Genome Res.* 30, 497–513. doi: 10.1101/gr.256750.119
- Ma, C., Burd, S., and Lers, A. (2015). miR408 is involved in abiotic stress responses in Arabidopsis. *Plant J.* 84, 169–187. doi: 10.1111/tpj.12999
- Ma, X., Liu, C., Gu, L., Mo, B., Cao, X., and Chen, X. (2017). TarHunter, a tool for predicting conserved microRNA targets and target mimics in plants. *Bioinformatics* 34, 1574–1576. doi: 10.1093/bioinformatics/btx797
- Mateos, J. L., Bologna, N. G., Chorostecki, U., and Palatnik, J. F. (2010). Identification of microRNA processing determinants by random mutagenesis of Arabidopsis MIR172a precursor. *Curr. Biol.* 20, 49–54. doi: 10.1016/j.cub.2009.10.072
- Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., et al. (2008). Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. *Cell* 133, 116–127. doi: 10.1016/j.cell.2008.02.034
- Moro, B., Chorostecki, U., Arikait, S., Suarez, I. P., Hobartner, C., Rasia, R. M., et al. (2018). Efficiency and precision of microRNA biogenesis modes in plants. *Nucleic Acids Res.* 46, 10709–10723. doi: 10.1093/nar/gky853
- Quintero, F. J., Garcíadeblas, B., and Rodríguez-Navarro, A. (1996). The SAL1 gene of Arabidopsis, encoding an enzyme with 3'(2'),5'-bisphosphate nucleotidase and inositol polyphosphate 1-phosphatase activities, increases salt tolerance in yeast. *Plant Cell* 8, 529–537. doi: 10.1105/tpc.8.3.529
- Reinhart, B. J., Weinstein, E. G., Rhoades, M. W., Bartel, B., and Bartel, D. P. (2002). MicroRNAs in plants. *Genes Dev.* 16, 1616–1626. doi: 10.1101/gad.1004402
- Song, X., Li, Y., Cao, X., and Qi, Y. (2019). MicroRNAs and Their regulatory roles in plant-environment interactions. *Annu. Rev. Plant Biol.* 70, 489–525. doi: 10.1146/annurev-arplant-050718-100334
- Treiber, T., Treiber, N., and Meister, G. (2019). Regulation of microRNA biogenesis and its crosstalk with other cellular pathways. *Nat. Rev. Mol. Cell Biol.* 20, 5–20. doi: 10.1038/s41580-018-0059-1
- Wang, J. L., Mei, J., and Ren, G. D. (2019). Plant microRNAs: biogenesis, homeostasis, and degradation. *Front. Plant Sci.* 10:360. doi: 10.3389/fpls.2019.00360
- Wilson, P. B., Estavillo, G. M., Field, K. J., Pornsiriwong, W., Carroll, A. J., Howell, K. A., et al. (2009). The nucleotidase/phosphatase SAL1 is a negative regulator of drought tolerance in Arabidopsis. *Plant J.* 58, 299–317. doi: 10.1111/j.1365-313X.2008.03780.x
- You, C., He, W., Hang, R., Zhang, C., Cao, X., Guo, H., et al. (2019). FIERY1 promotes microRNA accumulation by suppressing rRNA-derived small interfering RNAs in Arabidopsis. *Nat. Commun.* 10:4424. doi: 10.1038/s41467-019-12379-z
- Zhang, B. H., Pan, X. P., Cannon, C. H., Cobb, G. P., and Anderson, T. A. (2006). Conservation and divergence of plant microRNA genes. *Plant J.* 46, 243–259. doi: 10.1111/j.1365-313X.2006.02697.X
- Zhang, X., Niu, D., Carbonell, A., Wang, A., Lee, A., Tun, V., et al. (2014). ARGONAUTE PIWI domain and microRNA duplex structure regulate small RNA sorting in Arabidopsis. *Nat. Commun.* 5:5468. doi: 10.1038/ncomms6468

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Li and Ren. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.