Received: 5 February 2024 | Accepted: 16 May 2024

DOI: 10.1111/pce.14974

ORIGINAL ARTICLE

Plant, Cell & Environment | WILEY

Genome-wide identification of phasiRNAs in *Arabidopsis thaliana*, and insights into biogenesis, temperature sensitivity, and organ specificity

Zedi Feng[1] | Xiaoxia Ma[2] | Xiaomei Wu[1] | Wenyuan Wu[3,4] | Bo Shen[1] | Shaolei Li[1] | Yinju Tang[1] | JiaCen Wang[1] | Chaogang Shao[5] | Yijun Meng[1]

[1]College of Life and Environmental Sciences, Hangzhou Normal University, Hangzhou, China

[2]School of Pharmacy, Hangzhou Normal University, Hangzhou, China

[3]School of Information Science and Technology, Hangzhou Normal University, Hangzhou, China

[4]Zhejiang Provincial Key Laboratory of Urban Wetlands and Regional Change, Hangzhou Normal University, Hangzhou, China

[5]College of Life Sciences, Huzhou University, Huzhou, China

Correspondence

Xiaoxia Ma, School of Pharmacy, Hangzhou Normal University, Hangzhou 311121, China.
Email: xiaoma_0910@163.com

Chaogang Shao, College of Life Sciences, Huzhou University, Huzhou 313000, China.
Email: shaocg@zjhu.edu.cn

Yijun Meng, College of Life and Environmental Sciences, Hangzhou Normal University, Hangzhou 311121, China.
Email: mengyijun@zju.edu.cn

Funding information

National Natural Science Foundation of China

Abstract

The knowledge of biogenesis and target regulation of the phased small interfering RNAs (phasiRNAs) needs continuous update, since the phasiRNA loci are dynamically evolved in plants. Here, hundreds of phasiRNA loci of *Arabidopsis thaliana* were identified in distinct tissues and under different temperature. In flowers, most of the 24-nt loci are RNA-dependent RNA polymerase 2 (RDR2)-dependent, while the 21-nt loci are RDR6-dependent. Among the RDR-dependent loci, a significant portion is Dicer-like 1-dependent, indicating the involvement of microRNAs in their expression. Besides, two *TAS* candidates were discovered. Some interesting features of the phasiRNA loci were observed, such as the strong strand bias of phasiRNA generation, and the capacity of one locus for producing phasiRNAs by different increments. Both organ specificity and temperature sensitivity were observed for phasiRNA expression. In leaves, the *TAS* genes are highly activated under low temperature. Several *trans*-acting siRNA—target pairs are also temperature-sensitive. In many cases, the phasiRNA expression patterns correlate well with those of the processing signals. Analysis of the rRNA-depleted degradome uncovered several phasiRNA loci to be RNA polymerase II-independent. Our results should advance the understanding on phasiRNA biogenesis and regulation in plants.

KEYWORDS

expression pattern, phased small interfering RNA, processing, ta-siRNA, transcription

1 | INTRODUCTION

Small RNAs (sRNAs) form an important group of noncoding RNAs with critical biological roles in plants (Chen, 2009). In addition to microRNAs (miRNAs) and heterochromatic small interfering RNAs (siRNAs), secondary phased small interfering RNAs (phasiRNAs) are another attractive subgroup of the plant sRNAs (Zhan & Meyers, 2023). In the pioneer studies, the secondary phasiRNAs were defined as a class of 21-nt sRNAs whose processing was triggered by miRNA-mediated cleavages and was dependent on Dicer-like 4 (DCL4). In *Arabidopsis* (*Arabidopsis thaliana*), *trans*-acting small interfering RNAs (ta-siRNAs) are the representative class of the secondary phasiRNAs. According to the TAIR10 (The *Arabidopsis* Information Resource, release 10) annotations (Swarbreck et al., 2008), there are 11 *TAS* genes including three *TAS1*s, *TAS2*, three *TAS3*s, *TAS4* and three unclassified ta-siRNA loci. For most of the *TAS* genes, RNA polymerase II (Pol II) is considered to be the principal transcription machinery (Zhan & Meyers, 2023). The ta-siRNAs are produced through an RNA-dependent RNA polymerase 6

Zedi Feng and Xiaoxia Ma contributed equally to this study.

(RDR6)- and DCL4-dependent pathway (Chen, 2009). Several models, including "One-hit," "two-hit," and "updated one-hit," have been proposed to depict the miRNA-mediated phase initiation for different *TAS* genes respectively (Deng et al., 2018). Notably, in the "updated one-hit" model, miRNA binding, although in the absence of cleavages, is sufficient for phasiRNA production. More recently, the secondary phasiRNA population has been greatly expanded. Both 21- and 24-nt phasiRNAs have been discovered in the reproductive tissues of angiosperms. In the male germ cells of rice (*Oryza sativa*), most of the 21-nt phasiRNA loci were predicted to be targeted by miR2118, and more than half of the 24-nt loci were targeted by miR2275 (Jiang et al., 2020). In Xia and his colleagues' study (2019), a broad survey on the 24-nt reproductive phasiRNAs shows that although the miR2275-initiated biogenesis pathway of these phasiRNAs is widely existed in eudicots, exceptional cases have been observed in legumes and *Arabidopsis*, indicating that the miRNA—phasiRNA pathway is under highly dynamic evolution (Xia et al., 2019). However, the knowledge of the mechanisms underlying phasiRNA transcription and processing is still lacking, especially for the dynamically evolved phasiRNA loci.

PhasiRNAs play essential regulatory roles in plant growth, development, response and signaling. In *Arabidopsis*, for example, *TAS3* is implicated in auxin signaling related to leaf development (Adenot et al., 2006) and juvenile-to-adult transition (Fahlgren et al., 2006). The ta-siRNAs from *TAS4* target the MYB transcription factors, thus functioning in anthocyanin biosynthesis (Luo et al., 2012) and trichome development (Guan et al., 2014). More interestingly, in addition to *TAS1* in *Arabidopsis* (Kume et al., 2010; Li et al., 2014), many other phasiRNAs have been demonstrated to play their regulatory roles through a temperature-sensitive way in rice (Shi et al., 2023; Si et al., 2023) and flax (*Linum usitatissimum*) (Pokhrel & Meyers, 2022).

Ongoing efforts have been devoted to expanding the phasiRNA population and revealing the diversified mechanisms of their transcription, processing, evolution and target regulation. Here, genome-wide identification of the phasiRNAs was performed in *Arabidopsis*. Combinatory use of the small RNA sequencing (sRNA-seq) data enabled us to search for the phasiRNA loci in distinct tissues (flowers, leaves, roots, and seedlings) and under two temperature treatments (15°C and 27°C). In flowers, most of the 24-nt loci were assigned to the RDR2-dependent biogenesis pathway, while nearly all of the 21-nt loci were specifically assigned to the RDR6-dependent pathway. Among these RDR-dependent loci, a significant portion is DCL1-dependent, indicating the involvement of miRNAs in their biogenesis. Two novel ta-siRNA loci were discovered in flowers. Deep investigations unveiled some interesting features of certain phasiRNA loci, such as the strong strand bias of phasiRNA generation, and the capacity of one locus for producing phasiRNAs by different increments. Based on expression pattern analysis, organ specificity and temperature sensitivity were observed for many phasiRNA loci. Specifically, the *TAS* genes in leaves are highly activated under the low-temperature treatment. The regulation of several ta-siRNA targets is also enhanced under low temperature. On the other hand, degradome sequencing (degradome-seq) data showed a great potential for tracing the processing signals of the phasiRNAs. In some cases, the organ specificity or the temperature sensitivity of the phasiRNAs correlates well with that of the processing signals. Different strategies for degradome library construction were employed, with the aim to distinguish the processing signals between polyA- and non-polyA-tailed transcripts. In seedlings, several phasiRNA loci including a ta-siRNA locus (*AT1G63130*) were suggested to be transcribed through a Pol II-independent pathway. Summarily, our results presented here should promote the discovery of novel phasiRNA classes, and could advance the understanding on expression and target regulation of the phasiRNAs.

2 | MATERIALS AND METHODS

2.1 | Plant growth, sequencing library preparation and sequencing data processing

The Col-0 ecotype plants of *Arabidopsis* were grown on B5 plates under sterile conditions. Plates were kept at 4°C for 2 days to synchronize germination, and then transferred to the plant growth chamber at 22°C with 12-h light and 12-h dark (12 L:12D). For sRNA-seq, the 5-day-old seedlings were transferred to the plant growth chamber at 15°C (12 L:12D) and 30°C (12 L:12D) respectively, and grown for 10 days. Finally, the 15-day-old seedlings were collected. For degradome-seq, the 15-day-old seedlings grown at 22°C (12 L:12D) were collected. For sequencing experiments, there are three biological replicates for each sample.

TruSeq® Small RNA Sample Prep Kit (Illumina) was used for sRNA-seq library construction according to the manufacturer's manual. Briefly, sRNAs (<200 nt) were separated from 1 μg total RNAs by polyacrylamide gel electrophoresis, and were ligated to the 5′ and the 3′ adapters by T4 RNA ligase. By using SuperScript™ II Reverse Transcriptase (Invitrogen), the sRNAs with adapters were transcribed to single-stranded cDNAs which were then treated as the PCR templates to generate double-stranded cDNAs. The sRNA-seq library was purified by 16% TBE gel electrophoresis. Sequencing was performed by LC-Bio with the Illumina HiSeq.2500 platform following the vendor's recommended protocol.

Two types of degradome libraries, that is, polyA-enriched ("polyA") and rRNA-depleted ("rRNA$^-$"), were constructed. Starting from 20 μg of total RNAs, the "polyA" library was constructed by using 150 ng poly(A)-tailed RNAs which were obtained through two-round purification by using polyT-attached magnetic beads. However, the "rRNA$^-$" library was constructed based on the 20 μg total RNAs after two-round rRNA removal. The follow-up steps are identical for both library types. Specifically, the 5' adapters were ligated to the RNA remnants by RNA ligase. Reverse transcription was performed to generate the first strands of cDNAs with a random primer. After size selection with AMPureXP beads, cDNAs were PCR amplified. Single-end sequencing (50 bp) was performed by LC-Bio with the

Illumina HiSeq.2500 platform following the vendor's recommended protocol.

FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) was used for the pretreatment of the sRNA- and degradome-seq reads, to trim the adapters and remove the low-quality reads. The processed reads ranging from 15 to 35 nt were used. For a sRNA- or degradome-seq data set, the expression level of one sequence was normalized by dividing the raw read count of this sequence by the total raw read counts of this data set, and then multiplied by 10^6. The normalized levels (in RPM, reads per million) of the sRNAs and the degradome signatures were used for further analyses, such as heatmap-based expression pattern analysis and degradome signal intensity comparison. Heatmaps were drawn by using the online tool "Omicstudio" (https://www.omicstudio.cn/).

2.2 | PhasiRNA loci identification

The sRNA-seq data sets with raw counts were submitted to PhaseTank v1.0 (Guo et al., 2015) for phasiRNA locus predictions, and the default normalization parameter CP20M (counts per 20 million) was used (see Figure S1 for the detailed descriptions of data format, analytical steps, and parameters). The *Arabidopsis* genome available in TAIR10 (Lamesch et al., 2012) was used for the whole-genome predictions. As required by PhaseTank v1.0, Bowtie v1.3.1 (Langmead et al., 2009) was installed for sRNA mapping. The sRNA-seq data sets (Table S1) retrieved from SRA (https://www.ncbi.nlm.nih.gov/sra/) were used for the predictions in two ways (Figure S2A). Briefly, in the tissue-specific way, the data sets of the same tissue under different temperature were combinatorially used for a prediction. For instance, SRR6041096 (replicate 1 of flowers at 15°C), SRR6041095 (replicate 2 of flowers at 15°C), SRR6041100 (replicate 1 of flowers at 27°C) and SRR6041099 (replicate 2 of flowers at 27°C) were submitted for the phasiRNA locus prediction in flowers. In the temperature-specific way, the data sets from flowers, leaves and roots under the same temperature were combinatorially used for a prediction. For instance, SRR6041096 (replicate 1 of flowers at 15°C), SRR6041095 (replicate 2 of flowers at 15°C), SRR6041098 (replicate 1 of leaves at 15°C), SRR6041103 (replicate 2 of leaves at 15°C), SRR6041107 (replicate 1 of roots at 15°C), and SRR6041106 (replicate 2 of roots at 15°C) were submitted for the phasiRNA locus prediction under 15°C. Only the loci with 21- to 24-nt phased increments were retained.

2.3 | Argonaute (AGO) enrichment analysis of the phasiRNAs

The public sequencing data sets of AGO1- and AGO4-associated sRNAs from different tissues (GEO accession IDs: from GSM707678 to GSM707689; see detail in Table S1) were used for the analysis. In a specific tissue, a sRNA would be regarded to be AGO-enriched, if: (1) the expression of the sRNA in the control sample was undetectable, while its level in AGO was ≥1 rpm; or (2) the level of sRNA in AGO was ≥3 times higher than that in the control sample.

2.4 | Analysis of the RDR and DCL dependence of the phasiRNA loci

All of the phasiRNAs identified from flowers were analyzed. The sRNA-seq data sets from the flowers of the *dcl1*, *dcl234*, *rdr2* and *rdr6* mutants, and the Col-0 wild type (WT) plants were retrieved from GEO (https://www.ncbi.nlm.nih.gov/geo/) (Table S1). A phasiRNA will be regarded to be RDR- or DCL-dependent, if its summed RPM value in WT (two replicates) is ≥2 times higher compared to that in the *dcl* or *rdr* mutant (two replicates). Additionally, the functional modules "estimateDisp" and "glmFit" of edgeR (Robinson et al., 2010) (http://bioconductor.org/packages/release/bioc/html/edgeR.html) were employed for statistical evaluation. *P* values were calculated for each of the RDR- or DCL-dependent phasiRNAs.

Then, the RDR/DCL-dependent phasiRNAs were assigned to their loci. A locus will be regarded to be RDR- or DCL-dependent, if >50% of the phasiRNAs within the locus are RDR- or DCL-dependent.

2.5 | Phylogenetic analysis and target identification of the *TAS* loci

A total of 13 *TAS* loci, including 11 TAIR10-annotated *TAS* genes (*AT2G27400*: *TAS1a*; *AT1G50055*: *TAS1b*; *AT2G39675*: *TAS1c*; *AT2G39681*: *TAS2*; *AT3G17185*: *TAS3*; *AT5G49615*: *TAS3b*; *AT5G57735*: *TAS3c*; *AT3G25795*: *TAS4*; *AT1G63080*: a ta-siRNA locus; *AT1G63130*: a ta-siRNA locus; and *AT1G63150*: a ta-siRNA locus) and two *TAS* candidates discovered in flowers (Chr1-3582-21nt-F and Chr5-2685-21nt-F) were analyzed. The multiple sequence alignment of the genomic sequences of the above loci was performed using MAFFT v7.520 (Katoh & Standley, 2013) with the option "FFT-NS-2". The phylogenetic tree was constructed using FastTree v2.1.11 (Price et al., 2009) with the "Jukes-Cantor + CAT" model. An irrelevant gene, *AT1G03650* (encoding an Acyl-CoA N-acyltransferases superfamily protein), was selected as the outgroup to root the tree.

The 21-nt ta-siRNAs were subject to the transcriptome-wide target prediction based on the miRU algorithm (Dai et al., 2018; Zhang, 2005). The predicted targets were subject to degradome data-based validation as described in our previous studies (Meng et al., 2011; Shao et al., 2013).

2.6 | Screening for the temperature-sensitive and organ-specific phasiRNA loci

For the phasiRNA loci identified from a specific tissue (i.e. flowers, leaves, roots or seedlings), the temperature-sensitive screening was

performed. Briefly, for one locus, the RPM values of all phasiRNAs (two biological replicates) under 15°C and 27°C were summed respectively. If the summed RPM value under 15°C is ≥3 times higher than that under 27°C, the locus will be considered to be 15°C-sensitive. Conversely, if the summed value under 27°C is ≥3 times higher than that under 15°C, the locus will be considered to be 27°C-sensitive.

For each phasiRNA encoded by a 15°C-sensitive locus, it will be regarded as a 15°C-sensitive phasiRNA if its summed RPM value under 15°C (two biological replicates) is ≥3 times higher than that under 27°C. For a 15°C-sensitive locus, the percentage of the 15°C-sensitive phasiRNAs was calculated as follows:

Percentage = the number of the 15°C-sensitive phasiRNAs/the total number of the phasiRNAs on the locus.

The same algorithm was also applied for the 27°C-sensitive loci.

For the phasiRNA loci identified under 15°C or 27°C, the organ-specific screening was performed. Briefly, for one locus, the RPM values of all phasiRNAs (two biological replicates) in flowers, leaves and roots were summed respectively. If the summed RPM value in a specific organ, flowers for example, is ≥3 times higher than those in the other two organs respectively, the locus will be considered to be flower-specific. The same algorithm was also applied for the identification of the leaf- and root-specific loci. For each phasiRNA encoded by an organ-specific (e.g., flower-specific) locus, it will be regarded as a flower-specific phasiRNA if its summed RPM value in flowers (two biological replicates) is ≥3 times higher than those in leaves and roots respectively. For a flower-specific locus, the percentage of the flower-specific phasiRNAs was calculated as following:

Percentage = the number of flower-specific phasiRNAs/the total number of the phasiRNAs on the locus

The same algorithm was also applied for the leaf- and root-specific loci.

Additionally, the functional modules "estimateDisp" and "glmFit" of edgeR (Robinson et al., 2010) were employed for statistical evaluation. *P* values were calculated for each phasiRNA encoded by the temperature-sensitive or organ-specific loci.

3 | RESULTS

3.1 | Genome-wide identification of the phasiRNA loci in *Arabidopsis*

The sRNA-seq data sets from four different tissues of *Arabidopsis* under two temperature treatments (Table S1) were utilized for phasiRNA locus identification by employing PhaseTank v1.0 (Guo et al., 2015). Different combinations of these data sets allowed us to search for the phasiRNA loci in the four tissues (i.e. flowers, leaves, roots, and seedlings) and under 15°C and 27°C, respectively (see Section 2, and Figure S2A). Only the loci with 21- to 24-nt phased increments were retained for the following analyses. As a result, hundreds of loci were discovered. Briefly, 393, 131, 107, and 154 loci were identified from flowers, leaves, roots, and seedlings, respectively, and 213 and 206 loci were discovered under 15°C and 27°C respectively (Data S1 and Table S2). Among these loci, the 24-nt (from 82.44% to 92.23%) and the 21-nt ones (3.88% to 11.45%) are the two most abundant types. A total of 2132, 705, 440, and 671 phasiRNAs were identified from flowers, leaves, roots, and seedlings respectively, and 1114 and 1000 phasiRNAs were discovered under 15°C and 27°C respectively (Data S2 and Table S2).

Further analysis provided some deep insights into the sequence features of the loci identified above. First, for a large portion of the loci, the numbers of phases range from five to 10 (Figure 1a,b). In other words, the genomic length of many loci ranges from 100 to 250 bp, partially reflecting their noncoding feature. Second, more than 94% of the loci can produce phasiRNAs on both strands (Figure 1c,d), increasing the possibility that most phasiRNAs were generated through the RDR-dependent pathway (s). Finally, the 24-nt phasiRNAs, followed by the 21-nt ones, are the most abundant classes (Figure 1e,f). However, some subtle but significant differences were observed for the phasiRNAs in flowers. Specifically, the percentages of the 24-nt phasiRNAs identified from the other three tissues range from 63.69% to 73.92%, while it reaches to 87.29% in flowers. The 24-nt phasiRNAs have been reported by several groups in the reproductive organs of rice (Johnson et al., 2009; Song et al., 2012a, 2012b). Thus, the above observation may not be surprising. Besides, more than 20% of the phasiRNAs identified from leaves, roots, and seedlings are 21 nt in length, in contrast to 5.02% in flowers (Figure 1e). The analysis of the 5′ terminal nucleotides shows that the 5′-A-started (>44%) and the 5′-U-started (>21%) phasiRNAs are the two most abundant classes (Figure 1e,f).

Argonaute (AGO) protein association may provide some valuable hints related to the biological roles of the sRNAs. In this regard, the public sequencing data sets of AGO1- and AGO4-associated sRNAs from different tissues (Table S1) were used for AGO enrichment analysis. Briefly, 2.53% and 11.96% of the phasiRNAs in flowers were identified to be enriched in AGO1 and AGO4 respectively, 4.68% and 17.16% of the phasiRNAs in leaves were enriched in AGO1 and AGO4 respectively, 4.32% and 17.73% of the phasiRNAs in roots were AGO1- and AGO4-enriched respectively, and 8.79% and 11.33% of the phasiRNAs in seedlings were AGO1- and AGO4-enriched respectively (Table S3). We also noticed that most of the AGO1-enriched phasiRNAs were 21 nt in length while the AGO4-enriched ones were mostly 24 nt (Figure S3), which was consistent with the features reported by the previous study (Mi et al., 2008). However, an exceptional case was observed in flowers. Among the 54 AGO1-enriched phasiRNAs identified from flowers, only 20 phasiRNAs were 21 nt, and a total of 23 phasiRNAs were 24 nt. Besides, to our knowledge, whether phasiRNAs can guide DNA methylation remains elusive (Deng et al., 2018). It is intriguing to investigate the potential involvement of the AGO4-enriched phasiRNAs in DNA methylation.

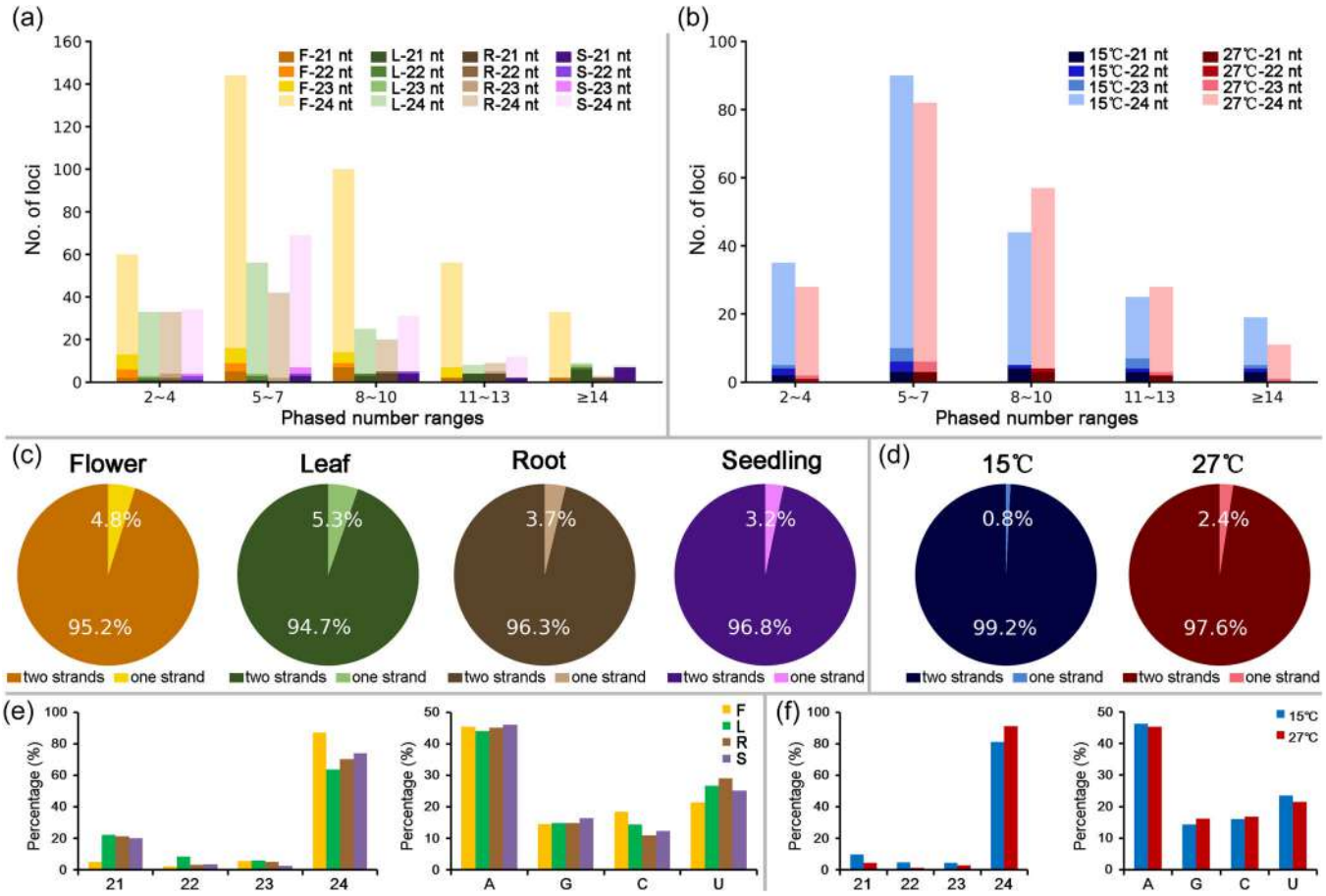

FIGURE 1 Sequence features of the phasiRNA loci identified in *Arabidopsis*. (a) Phased number distribution patterns of the 21- to 24-nt loci were identified from four tissues (F: flower, L: leaf, R: root, and S: seedling). (b) Phased number distribution patterns of the 21- to 24-nt loci identified under 15°C and 27°C. (c) Statistics of the loci with a double ("two strands")- or single ("one strand")-stranded mode for phasiRNA production in different tissues. (d) Statistics of the loci with a double- or single-stranded mode for phasiRNA production under 15°C and 27°C. (e) Sequence features (left: sequence length, right: 5′ terminal nucleotides) of the phasiRNAs identified from different tissues. (f) Sequence features of the phasiRNAs identified under 15°C and 27°C. [Color figure can be viewed at wileyonlinelibrary.com]

3.2 | DCL/RDR-dependent biogenesis of the phasiRNAs

The RDR6- and DCL4-dependent pathway is well-known for the plant phasiRNA biogenesis (Fei et al., 2013). Thus, it will be interesting to investigate the RDR and DCL dependence of the phasiRNAs identified in our study. Since the sRNA-seq data from the floral organ of the *dcl1*, *dcl234* (a triple mutant), *rdr2*, and *rdr6* mutants was publicly available, we examined the 2,132 phasiRNAs encoded by 393 loci identified in flowers. As a result, 38, 206, 207, and 38 phasiRNAs were observed to be significantly repressed in the *dcl1*, *dcl234*, *rdr2*, and *rdr6* mutants. Notably, most of the RDR/DCL-dependent phasiRNAs are 24 nt in length. Moreover, most of the 21-nt RDR/DCL-dependent phasiRNAs showed the specific dependence on DCL1 and RDR6 (Figure S4). Then, all of the RDR/DCL-dependent phasiRNAs were allocated to their origins to search for the RDR/DCL-dependent loci (see Section 2). Finally, 229 of 393 loci (58.27%), including 13 loci with 21-nt phased increments, eight with 22-nt increments, 14 with 23-nt increments and 194 with 24-nt increments, were identified to be RDR/DCL-dependent in flowers (Figure 2a). It is worth noting that the activities of eight loci with 21-nt increments are intensively repressed in the *dcl234* and *rdr6* mutants. Seven of these loci are also DCL1-dependent (Figure 2a and Table S4). Besides, the activities of 167 loci with 24-nt increments are remarkably repressed in the *dcl234* and *rdr2* mutants. And, 22 of these loci are also DCL1-dependent (Figure 2a and Table S4).

The TAIR10 annotations show that six of the eight 21-nt loci dependent on DCL234 and RDR2 overlap with the *TAS* genes (*TAS1a*, *TAS1b*, *TAS1c*, *TAS2*, *TAS3*, and *AT1G63150*). The remaining two 21-nt loci (Chr1-3582-21nt-F and Chr5-2685-21nt-F) were considered to be the novel *TAS* candidates. Phylogenetic analysis shows that the *TAS* gene family might be divided into four subfamilies in *Arabidopsis*. And, the candidate locus Chr5-2685-21nt-F may be classified into the *TAS3* subgroup, while the candidate Chr1-3582-21nt-F may be assigned to a newly undefined ta-siRNA locus subgroup (Figure 2b). Chr1-3582-21nt-F overlaps with *AT1G62910* encoding an RNA processing factor, and Chr5-2685-21nt-F overlaps with *AT5G39370*. Interestingly, the *TAS* candidate Chr5-2685-21nt-F also partially overlaps with miR3933

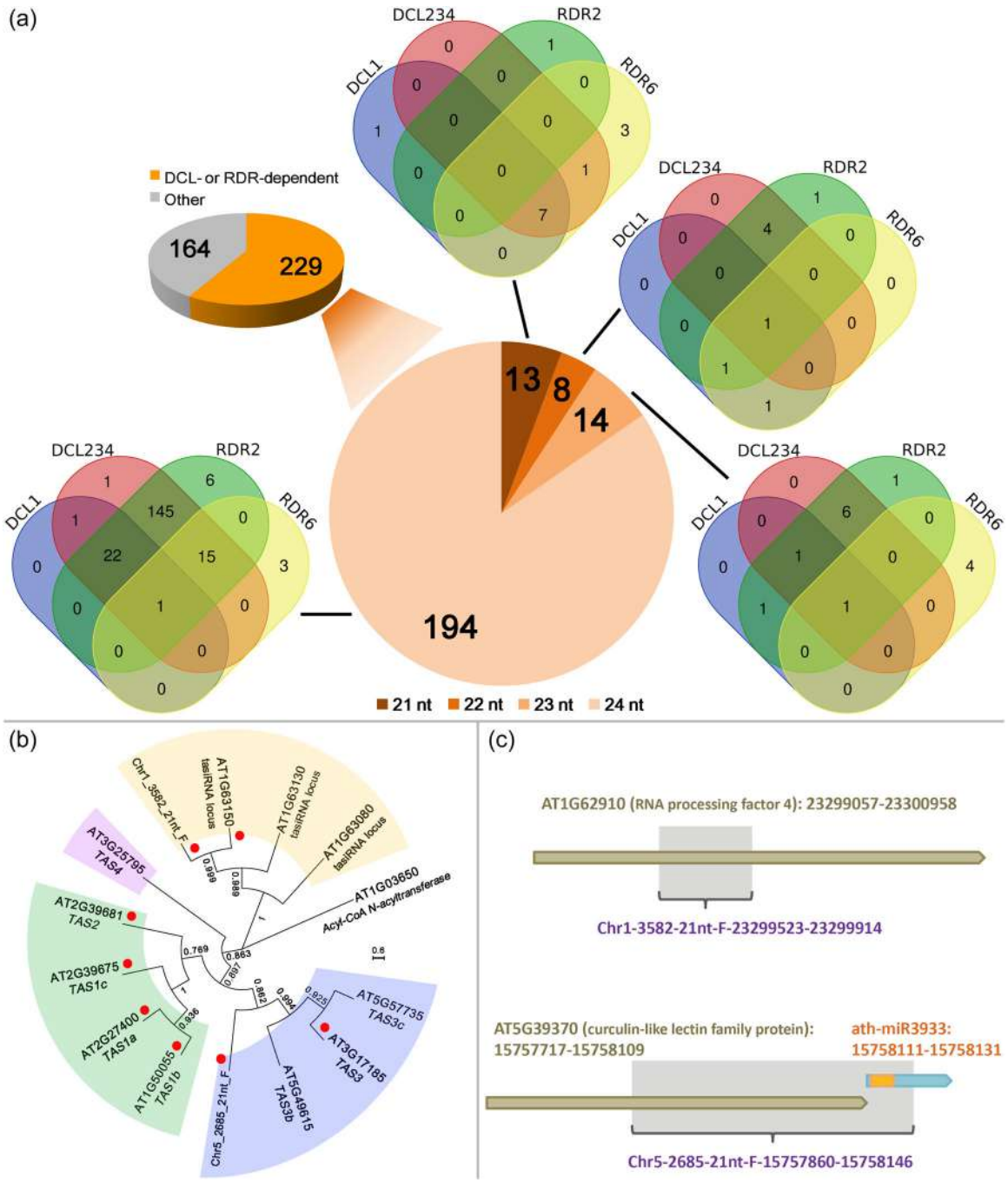

FIGURE 2 Analysis of the RDR/DCL-dependence of the phasiRNA loci in flowers. (a) Among the 393 loci analyzed, 229 (13, eight, 14, and 194 loci encoding 21- to 24-nt phasiRNAs, respectively) show their RDR/DCL dependence. (b) Phylogeny of the *Arabidopsis TAS* genes. A total of 11 *TAS* genes available in TAIR10 were included, and six (marked by red dots) were detected in flowers in this study. Two novel *TAS* candidates (Chr1-3582-21nt-F and Chr5-2685-21nt-F, also marked by red dots) were discovered in flowers. An irrelevant gene *AT1G03650* was selected as the outgroup to root the phylogenetic tree. The scale bar indicates the number of nucleotide substitutions per site. The bootstrap values calculated by FastTree v2.1.11 are shown beside selected major nodes. (c) Genomic positions of the two *TAS* candidates. One locus overlaps with *AT1G62910* encoding an RNA processing factor. Another overlaps with *AT5G39370* and a miRNA gene (cyan: miRNA precursor, orange: mature miRNA). [Color figure can be viewed at wileyonlinelibrary.com]

(Figure 2c), which may create a perfect binding site for "one-hit" model-based ta-siRNA production (Deng et al., 2018).

Target prediction and degradome-based validation were performed for the 21-nt phasiRNAs encoded by the two *TAS* candidates identified from flowers. Unfortunately, no target was validated. However, one *TAS* candidate Chr1-3582-21nt-F was observed to be expressed in leaves (Chr1-2460-21nt-L) and seedlings (Chr1-3120-21nt-S) (Table S5). Fortunately, based on degradome-seq data, targets of this *TAS* candidate were identified in both leaves and seedlings (Figure S5 and Table S6). Two targets (AT1G63130.1 and

AT1G63150.1) are encoded by the ta-siRNA generating loci. For the remaining seven targets, five are transcribed from the *PPR* (pentatricopeptide repeat) gene family. The *PPR* genes are the targets of *TAS1* and *TAS2* in plants (Chen et al., 2007; Felippes & Weigel, 2009). More interestingly, the *PPR* genes can also be targeted by miRNAs to trigger the production of *TAS*-like phasiRNAs in eudicots (Xia et al., 2013). Considering the activity of ta-siRNAs in secondary siRNA production (Deng et al., 2018), the *TAS* candidate Chr1-3582-21nt-F identified in this study is likely to be a true *TAS* gene.

3.3 | Pol II-dependent or independent transcription: Hints obtained from the degradome data

Our previous study showed the utility of degradome-seq data in tracing the RNA processing signals (Ma et al., 2020). In this regard, the degradome data of four different tissues under two temperature treatments was mapped onto the phasiRNA loci identified from flowers, leaves, roots, and seedlings, respectively. And, the degradome data of flowers, leaves, and roots under 15°C was mapped onto the loci identified under 15°C, and the data of the three organs under 27°C was mapped onto the loci identified under 27°C (Table S1 and Figure S2B). The degradome signatures were regarded as the processing signals if their 5′ ends could be mapped to the 5′ ends or the "3′ ends + 1 nt" of the phasiRNAs. The result shows that the percentages of the loci containing degradome-supported processing signals range from 20.61% to 35.68% (Figure 3a). Notably, the processing signals were searched for the phasiRNAs detected by the sRNA-seq data. On one locus, many phasiRNA-coding positions could not be detected by the sRNA-seq data possibly due to their low expression levels or poor stability. If the degradome signatures mapped to these positions are included, the above percentages will be elevated significantly.

The above result demonstrates that the processing signals of the phasiRNAs can be partially traced by degradome. Another question regarding the phasiRNA biogenesis is how the loci are transcribed. According to the well-established models, the plant *TAS* genes are transcribed by RNA Pol II, while the heterochromatic loci mostly producing 24-nt siRNAs are transcribed by Pol IV (Chen, 2009). Here, the rRNA-depleted ("rRNA$^-$") degradome library of *Arabidopsis* seedlings was constructed in addition to the canonical polyA-enriched library. Theoretically, the "rRNA$^-$" library not only contains the processing remnants from the polyA-tailed transcripts, but also includes those from the non-polyA transcripts. Genome-wide distribution patterns of the degradome signatures were obtained for the comparison between the two degradome libraries (Figure 3b). Although nearly 80% of the degradome reads of both libraries were mapped to the CDSs (coding sequences), significant differences were observed for the intronic and the intergenic regions. Specifically, 3.37% and 1.90% of the degradome reads from the "polyA" library were mapped to the intronic and the intergenic regions respectively, while 8.36% and 3.05% from the "rRNA$^-$" library were mapped to the two regions respectively. Another discrepancy was observed that much more degradome reads from the "polyA" library (7.71%) were mapped to the 3′ UTRs (untranslated regions) when compared to the "rRNA$^-$" libraries (3.73%). It is reasonable since the "polyA" library is more specific for tracing the 3′ polyA-tailed remnants in which the 3′-UTRs are included.

Next, both the "polyA" and the "rRNA$^-$" degradome reads were mapped to the phasiRNA loci identified from seedlings. As a result, 28 of 154 loci (18.18%) contain the "rRNA$^-$" degradome-supported processing signals, and 23 loci (14.94%) contain the "polyA" signals (Figure 3a). For the 23 loci with the "polyA" processing signals, six are assigned to the intergenic regions and another six overlap with the protein-coding genes. Among the remaining 11 loci mapped to the noncoding regions, eight overlap with five *TAS* genes (*TAS1b*, *TAS1c*, *TAS2*, *TAS3*, and *TAS4*) (Figure 3c). Among the 28 loci with the "rRNA$^-$" processing signals, nine are specifically supported by the "rRNA$^-$" degradome, indicating that these loci likely encode non-polyA transcripts. Based on the annotations, two loci overlap with the protein-coding genes, two overlap with the pseudogenes, and three were mapped to the intronic regions. For the remaining two mapped to the noncoding regions, one locus overlaps with a ta-siRNA locus (*AT1G63130*) (Figure 3c).

3.4 | Temperature sensitivity of the phasiRNAs

The temperature sensitivity of phasiRNAs was reported to be important for the reproductive development in rice (Shi et al., 2022, 2023; Si et al., 2023) and flax (Kume et al., 2010). Here, we performed a screening to search for the temperature-sensitive phasiRNA loci in different tissues of *Arabidopsis* (see Section 2, and Figure S2A). As a result, 49 of 393 loci (12.47%) were discovered to be sensitive to the 15°C or 27°C treatment in flowers, eight of 107 loci (7.48%) were discovered to be temperature-sensitive in roots, and 16 of 154 loci (10.39%) were temperature-sensitive in seedlings (Table S2 and Data S3). The discovery rate (51 of 131 loci, i.e., 38.93%) of the temperature-sensitive loci is especially high in leaves, indicating that the phasiRNAs in leaves is more susceptible to the ambient temperature changes. Deep analysis shows that for most of the temperature-sensitive loci identified in the four tissues, >50% of the phasiRNAs display consistent temperature sensitivity on each locus (Figure 4a, and see e.g., Figure 4b). Besides, among the temperature-sensitive loci discovered in leaves, roots and seedlings, a dominant portion (42 of 51, seven of eight and 13 of 16, respectively) are sensitive to 15°C. In contrast, the amount of the 27°C-sensitive loci is higher than that of the 15°C-sensitive ones in flowers (30 loci vs. 19 loci) (Figure 4a). Interestingly, in all tissues investigated, most of the 27°C-sensitive loci are phased in 24-nt increments, while no 21-nt locus has been observed.

A brief overview of the phasiRNA distributions on the temperature-sensitive loci shows that most of these loci can produce phasiRNAs on both strands (93.88% of the temperature-sensitive loci in flowers, 94.12%

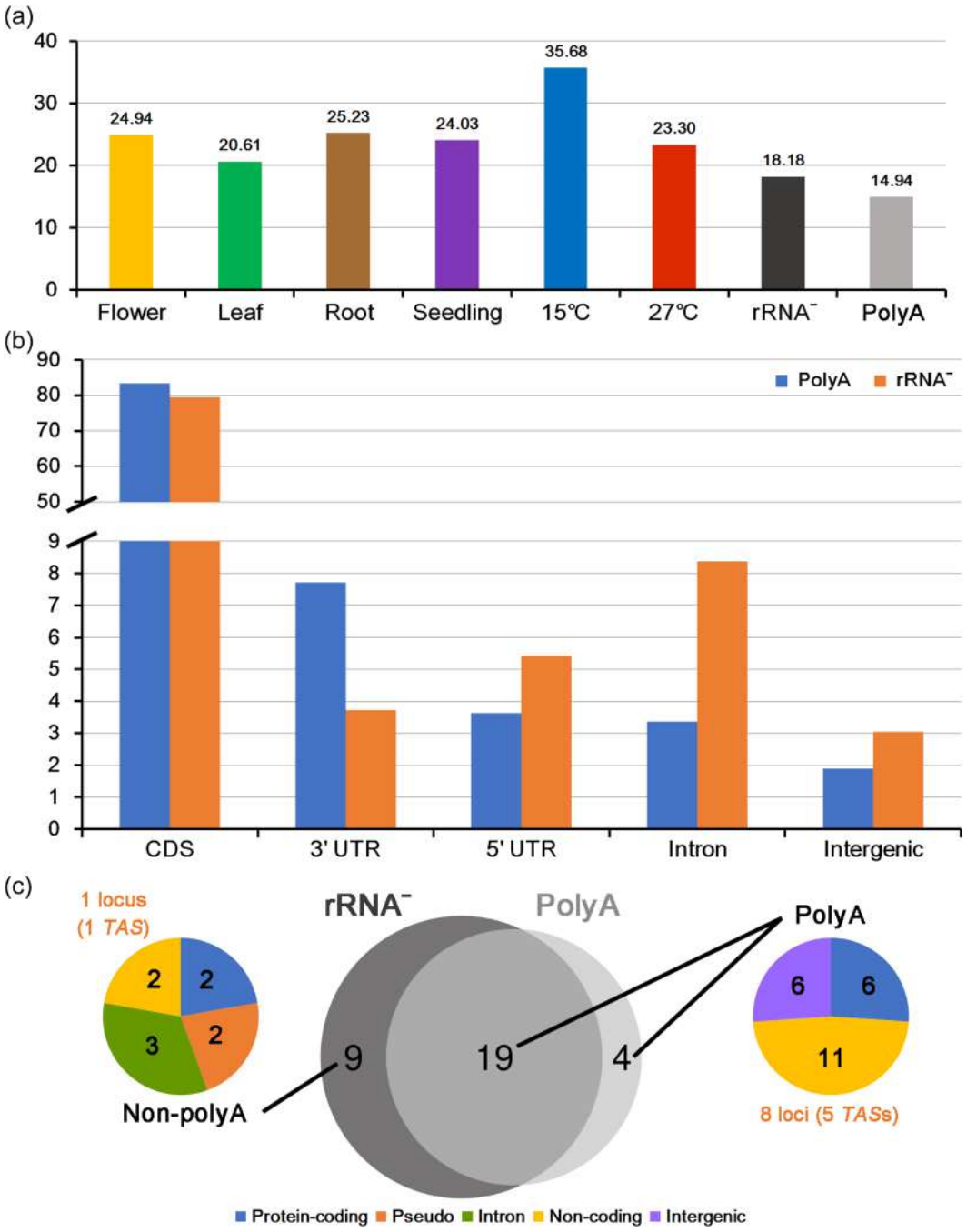

FIGURE 3 Degradome-based investigation of the transcription of the phasiRNA loci. (a) Percentages of the loci with degradome-supported processing signals. See Figure S2B for the detailed workflow of this analysis. (b) Different genome-wide distribution patterns between the degradome libraries prepared from polyA-tailed ("polyA") and rRNA-depleted ("rRNA⁻") remnants. (c) Detailed information of the phasiRNA loci traced by the "polyA" and/or the "rRNA⁻" degradome signatures. Notably, eight loci overlapping with five *TAS* genes (*AT1G50055*: *TAS1b*; *AT2G39675*: *TAS1c*; *AT2G39681*: *TAS2*; *AT3G17185*: *TAS3*; *AT3G25795*: *TAS4*) were indicated to encode polyA-tailed transcripts, while one ta-siRNA locus (*AT1G63130*) possibly encode a non-polyA transcript. [Color figure can be viewed at wileyonlinelibrary.com]

in leaves, 100% in roots, and 100% in seedlings, respectively) (Figure S6A). However, significant strand bias for phasiRNA production was observed for a large portion of the temperature-sensitive loci. "Significant strand bias" was defined for a locus as the number of phasiRNAs detected on one strand was two times or more than the other strand. As a result, 46.94% of the loci in flowers, 54.90% in leaves, 62.50% in roots, and 43.75% in seedlings were observed to have significant strand bias, respectively (Figure S6B).

Another interesting observation is that phasiRNAs in distinct lengths can be generated from a highly overlapped genomic region. For instance, Chr2-3-23nt-L and Chr2-3-24nt-L are two highly overlapped loci sensitive to 27°C in leaves (Figure 4c). Chr2-3-23nt-L can produce

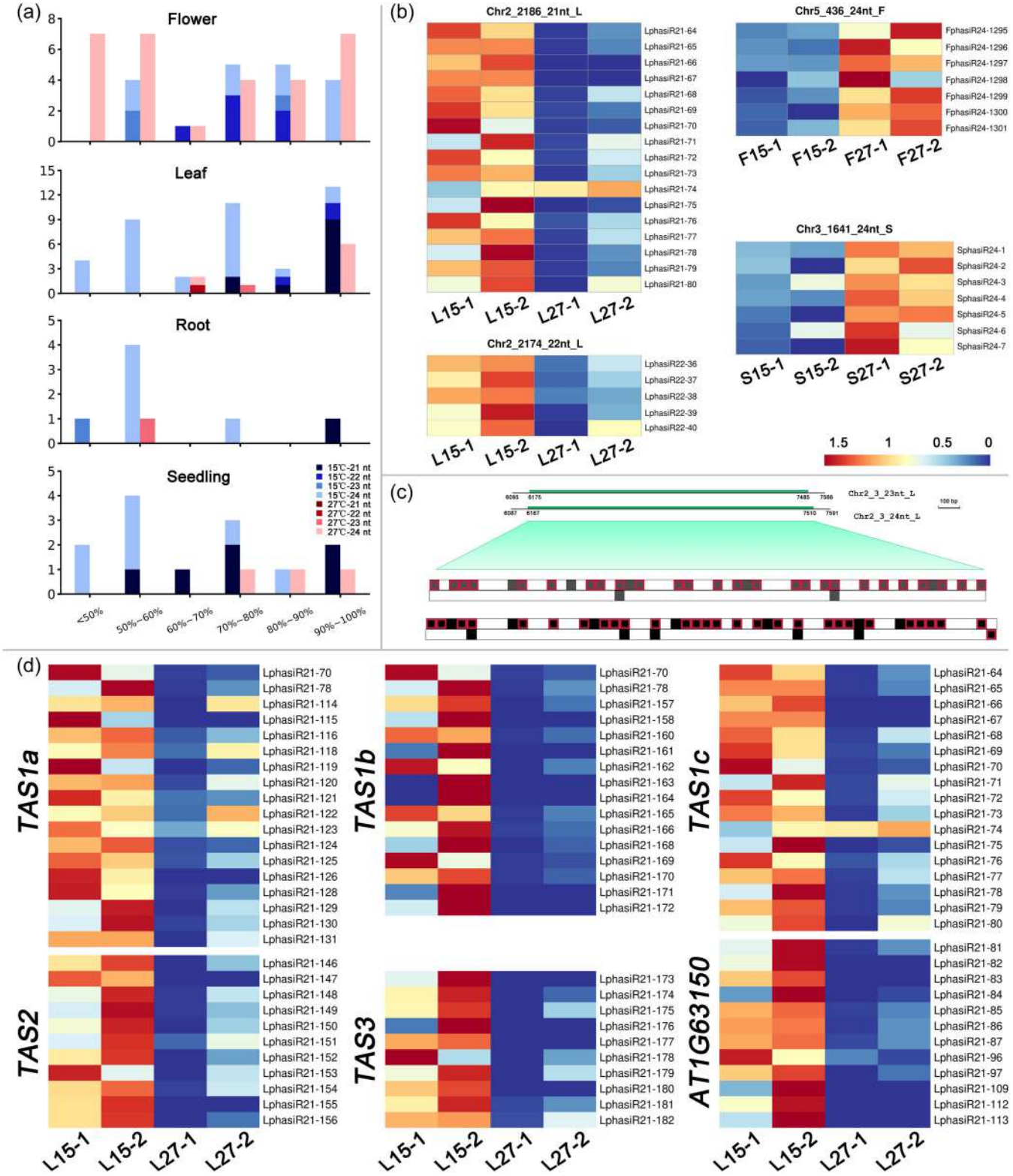

FIGURE 4 (See caption on next page).

23-nt phasiRNAs, and Chr2-3-24nt-L can produce 24-nt ones. Notably, strong strand bias was observed for both loci. For Chr2-3-23nt-L, 29 phasiRNAs including 24 activated under 27°C are produced from one strand, while only two phasiRNAs not significantly activated under 27°C are produced from the other strand. For Chr2-3-24nt-L, 35 phasiRNAs including 28 activated under 27°C are produced from one strand, while only six phasiRNAs including one activated under 27°C are produced from the other strand (Figure 4c).

3.5 | Temperature sensitivity of the *TAS* loci

As described above, eight *TAS* loci including two novel candidates were identified in flowers. Here, we also searched for the *TAS* loci in the other three tissues. As a result, a total of seven, eight, and nine *TAS* loci were identified from leaves, roots, and seedlings, respectively (Table S5). Notably, Chr1-3582-21nt-F, a *TAS* candidate locus identified in flowers, was also expressed in the other tissues. Since the phasiRNA loci in leaves were observed to be more susceptible to temperature changes (see the above section), the temperature sensitivity of the *TAS* loci in leaves was investigated firstly. Surprisingly, all of the seven loci were shown to be activated by the 15°C treatment (Table S5 and see e.g., Figure 4d). Also in seedlings, four of the nine *TAS* loci were observed to be activated under 15°C. We performed a sRNA-seq experiment for the seedlings under 15°C and 30°C respectively. Based on the data, the temperature sensitivity of some *TAS* loci and some other phasiRNA loci (such as Chr3-1641-24nt-S shown in Figure 4b) was verified (see e.g., Figure S7). It is also worth noting that only one *TAS* locus is activated under 15°C in roots, and none has been detected in flowers (Table S5). Another intriguing observation is that for all of the tissues investigated, none of the *TAS* loci has been detected to be activated under the high-temperature treatment.

The temperature-sensitive expression patterns of the *TAS* loci, especially in leaves, drove us to investigate the regulatory mode of the ta-siRNAs through a temperature-specific way. Firstly, all of the 21-nt ta-siRNAs generated from the seven *TAS* loci identified from leaves were subject to the transcriptome-wide target prediction and degradome-based target validation. It was observed that the regulation of several targets was notably reinforced under 15°C (Figure S8). For example, AT1G63150.1, transcribed from a ta-siRNA locus, was targeted by LphasiR21-101. This phasiRNA is encoded by the locus Chr1-2460-21nt-L which is overlapped with Chr1-3582-21nt-F, a *TAS* candidate locus identified from flowers. Based on the degradome signal intensity, the regulation of AT1G63150.1 by LphasiR21-101 is enhanced under 15°C in leaves (Figure S8A). Both LphasiR21-148 and LphasiR21-149 are encoded by the locus Chr2-2187-21nt-L which is overlapped with *TAS2*. Similar to the above case, the regulatory pairs LphasiR21-148—AT1G12300.1 and LphasiR21-149—AT1G63080.1 become much more active with stronger cleavage signals under 15°C in leaves (Figure S8B,C). *AT1G63080* is a ta-siRNA generating locus targeted by *TAS2*-derived ta-siR2140 (5′-AUAUCCCAUUUCUACCAUCUG-3′) for secondary ta-siRNA production (Chen et al., 2007). Here, LphasiR21-149 (5′-UUUGCAUAUACUCGAAUACCU-3′), also derived from *TAS2*, is identified to be another upstream regulator of *AT1G63080*. LphasiR21-149 is likely to play a role in secondary ta-siRNA amplification through a temperature-dependent way. Besides, both *AT1G12300* and *AT1G63080* are annotated to encode pentatricopeptide repeat proteins. In this regard, *AT1G12300*, targeted by LphasiR21-148, has a great potential of evolving into a *TAS* locus. Finally, all of the three ta-siRNAs reported here are highly activated under 15°C when compared to the 27°C treatment (Figure S8D,F). Taken together, our results show that some of the ta-siRNAs not only have temperature-sensitive expression patterns, but also play regulatory roles in a temperature-dependent manner.

3.6 | Organ specificity of the phasiRNAs

We were also interested in the expression of the phasiRNAs in different organs. To this end, we performed a screening to search for the organ-specific phasiRNA loci identified under the 15°C and 27°C treatments (see Section 2, and Figure S2A). Among the 213 loci identified under 15°C, 111 were flower-specific, 13 were leaf-specific, and none was discovered for roots. Among the 206 loci identified under 27°C, 125 were flower-specific, and none was discovered for the other two organs (Table S2 and Data S3). Summarily, the flower-specific loci encoding 24-nt phasiRNAs are the most abundant ones under both temperature treatments, and the leaf-specific loci can only be detected under 15°C. Besides, for most of the organ-specific loci, >50% of the phasiRNAs display consistent organ specificity on each locus (Figure 5a, and see e.g., Figure 5b).

PhasiRNA distribution analysis shows that most of the organ-specific loci can produce phasiRNAs on both strands (99.19% under 15°C, and 97.60% under 27°C) (Figure S6A). However, significant strand bias for phasiRNA generation was observed for a large portion of the organ-specific loci (37.90% under 15°C, and 46.40% under

FIGURE 4 Temperature sensitivity of the phasiRNA loci. (a) Statistics of the percentages of the phasiRNAs with a concordant expression pattern on each temperature-sensitive locus. For each tissue, all of the 15°C- and 27°C-sensitive loci with different phased increments (21 to 24 nt) were included. The percentages were assigned to six ranges (*x* axes). The *y* axes measure the numbers of the loci assigned to a specific percentage range. (b) Examples of the temperature-sensitive loci producing the temperature-sensitive phasiRNAs with high percentages. Two loci identified from leaves (Chr2-2186-21nt-L and Chr2-2174-22nt-L) are shown to be activated under 15°C. Chr5-436-24nt-F from flowers and Chr3-1641-24nt-S from seedlings are activated under 27°C. For each locus, all of the detected phasiRNAs are included in the heatmap-based expression pattern analysis. (c) Example of two highly overlapped loci producing 23- and 24-nt phasiRNAs, both of which are activated under 27°C. The genomic positions of the loci on chromosome 2 are indicated by gray lines on the top. The green bars mark the regions producing phasiRNAs. The details of the phasiRNA-producing regions are shown at the bottom. The gray and black boxes indicate the 23- and 24-nt phasiRNAs, respectively. The boxes with red borderlines indicate those phasiRNAs activated under 27°C. (d) The 15°C-sensitive *TAS* genes in leaves. Six *TAS* loci are taken for examples. For phasiRNA expression heatmaps, two biological replicates are shown for each sample. L15: leaves at 15°C; L27: leaves at 27°C; F15: flowers at 15°C; F27: flowers at 27°C; S15: seedlings at 15°C; S27: seedlings at 27°C. The color bar of all heatmaps is available in (b). [Color figure can be viewed at wileyonlinelibrary.com]

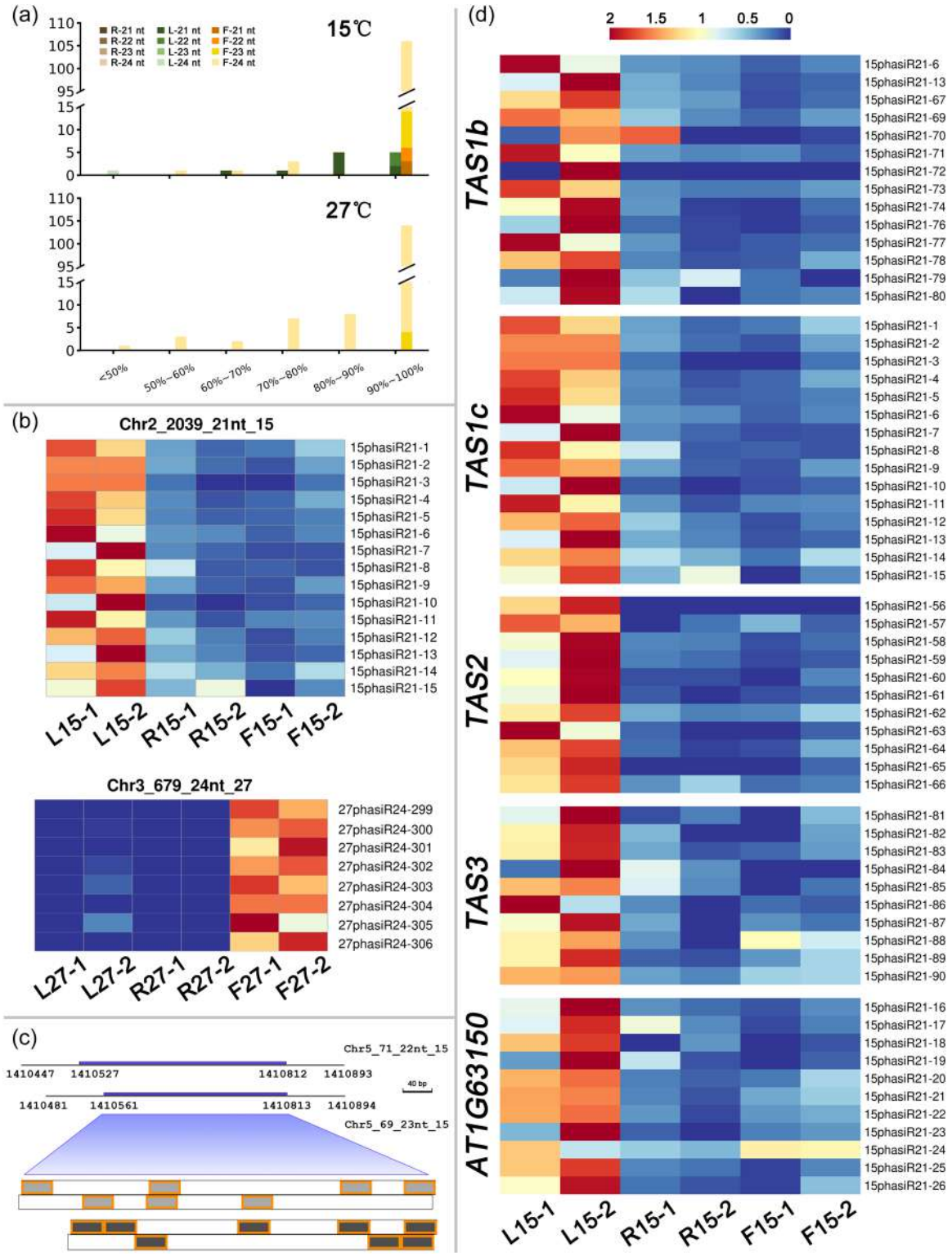

FIGURE 5 (See caption on next page).

27°C) (Figure S6B). Some loci with highly overlapped genomic regions were observed to produce phasiRNAs by different increments. For example, Chr5-71-22nt-15 and Chr5-69-23nt-15 are flower-specific loci identified under 15°C. These two loci share a highly overlapped genomic region on chromosome 5. Interestingly, Chr5-71-22nt-15 can produce 22-nt phasiRNAs, and Chr5-69-23nt-15 can produce 23-nt ones (Figure 5c).

Expression pattern analysis of the *TAS* genes shows that they are highly activated in leaves under 15°C (Figure 5d and Table S5). However, no obvious organ specificity was observed for the *TAS* loci under 27°C.

3.7 | Correlation between degradome-based processing signals and phasiRNA expression patterns

Our result above shows that the power of the degradome data in tracing the processing signals from the phasiRNA loci (Figure 3a). On the other hand, many loci are temperature-sensitive or organ-specific. Thus, we are curious about the relationship between the processing signals and the expression patterns of the phasiRNAs. Several loci were identified to display a well correlation between processing and expression (Figures 6 and S9). For example, based on the sRNA-seq data, nearly all of the phasiRNAs detected on the loci Chr5-3341-24nt-F (Figure 6a), Chr3-529-21nt-R (overlapped with *TAS4*) (Figure 6b), Chr1-3005-22nt-F (Figure S9A), Chr5-1893-21nt-L (Figure S9B), and Chr3-451-21nt-S (overlapped with *TAS4*) (Figure S9C) are expressed at much higher levels under 15°C than 27°C. Accordingly, all of the degradome signals mapped to the phasiRNA processing sites are much stronger under 15°C. Another case was observed for Chr2-2025-22nt-15 (Figure 6c). The phasiRNAs on this locus are highly expressed in leaves when compared to the other two organs. And, for three of the four detected processing sites, the degradome signals are stronger in leaves. Moreover, the abundances of the phasiRNAs from Chr5-551-24nt-27 (Figure 6d) and Chr2-964-24nt-15 (Figure S9D) are much higher in flowers, and the detected processing signals are consistently stronger in this organ. Taken together, our observation indicates that, at least for some loci, the specific processing patterns can partially contribute to the specific expression patterns of the phasiRNAs. And, the specific pattern is quite consistent among the phasiRNAs produced from one locus.

4 | DISCUSSION

4.1 | Divergent pathways for phasiRNA biogenesis

In this study, a genome-wide identification of the phasiRNA loci was performed in *Arabidopsis*. Hundreds of loci were discovered in different tissues and under different temperature treatments. As introduced above, the miRNA-triggered processing pathway of phasiRNAs is dynamically evolved (Deng et al., 2018; Xia et al., 2019). Some of the phasiRNA loci might be still young, resulting in a lack of the well-known features of the canonical loci. Thus, with the aim to expand the phasiRNA population, miRNA binding sites were not predicted for the phasiRNA loci identified in the present study. In other words, we did not merely focus on the secondary phasiRNAs. Instead, all of the loci with 21- to 24-nt phased increments were analyzed.

Analysis of the sRNA-seq data of the *rdr* and the *dcl* mutants of *Arabidopsis* showed that a large portion of the phasiRNA loci were RDR- and DCL-dependent (Table S4). Among the 192 RDR-dependent loci phased in 24-nt increment, 173 (90.10%) are specifically dependent on RDR2, while only three loci are specifically dependent on RDR6 (Figure 2a). For the 12 RDR-dependent loci with 21-nt increment, 11 (91.67%) are specifically dependent on RDR6, while only one is specifically dependent on RDR2 (Figure 2a). On the other hand, the activities of most phasiRNA loci are repressed in the triple mutant *dcl234*. Based on the current models of phasiRNA processing (Chen, 2009), it can be inferred that the 24-nt phasiRNAs are DCL3-dependent, while the 21-nt ones are DCL4-dependent. Taken together, the RDR- and DCL-dependent processing pathways could be partially distinguished by the phased increments of the loci. Another interesting observation is that phasiRNAs in different lengths could be produced from highly overlapped genomic regions (Figures 4c and 5c). It implies that some of the precursors might be compatible with distinct processing systems for phasiRNA production with different increments.

For more than 90% of the loci identified, phasiRNAs can be detected on both strands (Figure 1c,d), which correlates well with their RDR dependence. Intriguingly, deep investigation shows that nearly half of the loci produce phasiRNAs with significant strand bias. Notably, this kind of bias has no obvious organ specificity and temperature sensitivity (Figure S6B). Thus, it will be interesting to uncover the underlying reasons in further studies. On the other hand, for most of the organ-specific and temperature-sensitive loci, >50% of the phasiRNAs generated from the same locus exhibit a concordant expression pattern (Figures 4a and 5a), indicating a well-orchestrated regulation of phasiRNA expression and stability maintenance in specific organs or under specific conditions.

FIGURE 5 Organ specificity of the phasiRNAs. (a) Statistics of the percentages of the phasiRNAs with a concordant expression pattern on each organ-specific locus. For each temperature treatment (15°C or 27°C), all of the organ-specific (i.e., root (R)-, leaf (L)- or flower (F)-specific) loci with different phased increments (21 to 24 nt) were included. The percentages were assigned to six ranges (*x* axes). The *y* axes measure the numbers of the loci assigned to a specific percentage range. (b) Examples of the organ-specific loci producing the organ-specific phasiRNAs with high percentages. One locus identified at 15°C (Chr2-2039-21nt-15) is highly active in leaves, while another locus identified at 27°C (Chr3-679-24nt-27) show its flower specificity. For each locus, all of the detected phasiRNAs are included in the heatmap-based expression pattern analysis. (c) Example of two highly overlapped loci producing 22- and 23-nt phasiRNAs, both of which are activated in flowers. The genomic positions of the loci on chromosome 5 are indicated by gray lines on the top. The blue bars mark the regions producing phasiRNAs. The details of the phasiRNA-producing regions are shown at the bottom. The light and dark gray boxes indicate the 22- and 23-nt phasiRNAs, respectively. The boxes with orange borderlines indicate those phasiRNAs specifically expressed in flowers. (d) The leaf-specific *TAS* genes detected under 15°C. Five *TAS* loci are taken for examples. For phasiRNA expression heatmaps, two biological replicates are shown for each sample. L15: leaves at 15°C; L27: leaves at 27°C; F15: flowers at 15°C; F27: flowers at 27°C; R15: roots at 15°C; R27: roots at 27°C. The color bar of all heatmaps is available in (d). [Color figure can be viewed at wileyonlinelibrary.com]

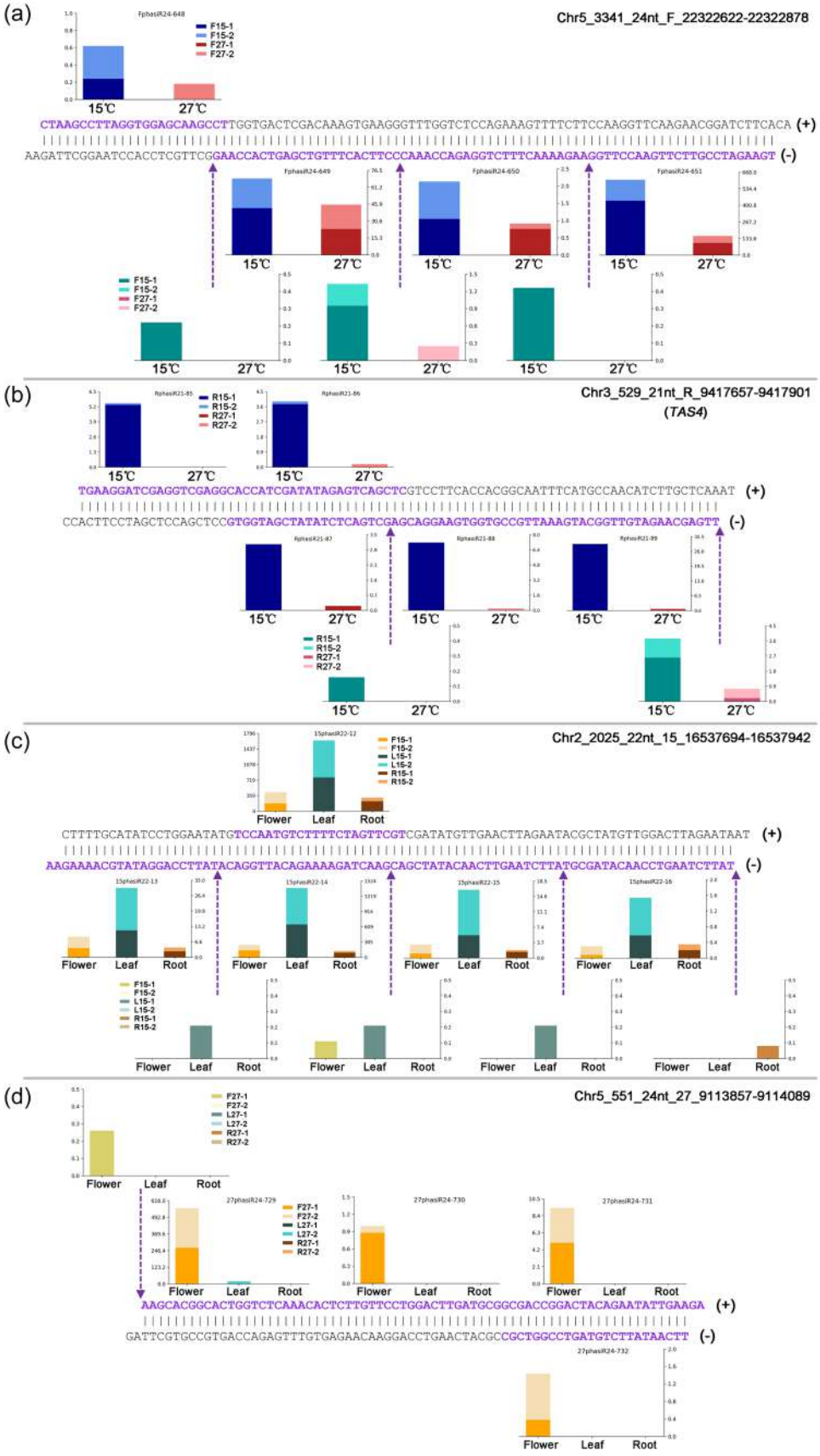

FIGURE 6 (See caption on next page).

In flowers, a total of eight 21-nt loci were identified to be repressed in *rdr6* and *dcl234* (Figure 2a). Among these loci, six have been annotated as the *TAS* loci in TAIR10. Phylogenetic analysis indicates that the remaining two loci might be novel *TAS* candidates. In plants, different models, such as "one-hit", "two-hit" and "updated one-hit", have been introduced to depict the initiation of ta-siRNA processing (Deng et al., 2018). However, in any model, miRNA binding is essential for triggering ta-siRNA production. Indeed, it was observed that seven of the eight loci, including the two novel *TAS* candidates, were also dependent on DCL1 which was essential for miRNA maturation (Figure 2a). Additionally, the genomic region of one *TAS* candidate (Chr5-2685-21nt-F) is overlapped with the region encoding ath-miR3933 (Figure 2c). From this point of view, miR3933 has a complementary binding site on the *TAS* candidate, which can be or will evolve to be the trigger for ta-siRNA production. Another interesting observation is that a total of 22 RDR2-dependent loci with 24-nt increment are also DCL1-dependent (Figure 2a), pointing to the potential involvement of miRNAs in the processing of these loci.

In this study, degradome data showed their power in tracing the residual signals during phasiRNA processing. Two strategies, "polyA" and "rRNA$^-$," were employed for degradome library construction, in an attempt to discriminate between the polyA- and the non-polyA-tailed phasiRNA precursors. As a result, a total of nine loci whose processing signals could be detected by the "rRNA$^-$" but not the "polyA" degradome were suggested to be transcribed through a non-Pol II-dependent pathway. Among the nine loci, six produce 24-nt phasiRNAs and three loci are phased in 21-nt increment. Considering the fact that the 24-nt heterochromatic loci are transcribed by Pol IV (Chen, 2009), the above observation increases the possibility that the loci specifically traced by the "rRNA$^-$" degradome are Pol II-independent. On the other hand, among the 23 loci could be traced by the "polyA" degradome, only three loci can produce 24-nt phasiRNAs and 14 loci are phased in 21-nt increment. Thus, in many cases, the Pol II and the non-Pol II transcription systems might be employed by the 21-nt and the 24-nt phasiRNA loci, respectively. However, exceptional cases were also observed. Although eight loci overlapped with five *TAS* genes could be traced by the "polyA" degradome, one ta-siRNA locus (*AT1G63130*) could only be traced by the "rRNA$^-$" degradome. There are three possibilities. First, the processing signals of this locus were not included in the "polyA" degradome due to the random factor during library construction. Second, the annotation of this *TAS* locus needs to be re-examined. Third, not all of the *TAS* loci are transcribed by Pol II. In other words, different transcription mechanisms might be applied for the *TAS* genes.

4.2 | PhasiRNA expression and target regulation: Temperature sensitivity and organ specificity

Our results show that many phasiRNA loci are expressed with temperature-sensitive or organ-specific patterns. For instance, in leaves, roots and seedlings, much more phasiRNA loci are highly activated under 15°C compared to 27°C (Figure 4a). On the other hand, under both 15°C and 27°C, much more loci are highly expressed in flowers than in the other organs (Figure 5a).

One important discovery in this study is related to the expression of the *TAS* genes. In leaves, seven *TAS* loci have been identified. All of these loci are activated under 15°C. In seedlings, there are nine *TAS* loci, four of which are 15°C-sensitive. Contrarily, although eight *TAS* loci have been identified in roots and flowers respectively, only one in roots and none in flowers is 15°C-sensitive. For all of the *TAS* loci, none was observed to be activated by 27°C (Table S5). Previously, it was reported in *Arabidopsis* that *TAS1* was repressed under the heat shock treatment, resulting in the upregulation of several heat stress transcription factors and the enhancement of plant heat tolerance (Li et al., 2014). Our results show that the temperature sensitivity of the *TAS* loci is strikingly different in distinct tissues. The expression of the *TAS* genes in both roots and flowers seems to be much less sensitive to the ambient temperature changes. For all tissues investigated, not only *TAS1*, but also other *TAS* genes (e.g., *TAS2*, *TAS3*, *TAS4*, and the *TAS* candidate identified in this study) are highly activated under the low temperature treatment (Table S5). Furthermore, the *TAS* genes exhibit strong organ specificity under 15°C they are highly expressed

FIGURE 6 Examples indicating the high correlation between the levels of phasiRNAs and their processing signal intensities. (a) The phasiRNAs produced from Chr5-3341-24nt-F (identified from flowers; ranging from 22,322,622 to 22,322,878 on chromosome 5) are expressed at higher levels under 15°C (refer to the histograms marked by different phasiRNA IDs; the phasiRNA sequences are highlighted by purple; similarly hereinafter). Accordingly, the degradome signals mapped to the processing sites are stronger under 15°C (refer to the histograms with the purple dotted arrows indicating the processing sites, similarly hereinafter). For all histograms, the *y* axes measure the RPM (reads per million) levels, similarly hereinafter. (b) The phasiRNAs produced from Chr3-529-21nt-R (identified from roots; ranging from 9,417,657 to 9,417,901 on chromosome 3) are expressed at higher levels under 15°C. Accordingly, the degradome signals mapped to the processing sites are stronger under 15°C. *Note*: Chr3-529-21nt-R overlaps with *TAS4* (*AT3G25795*). (c) The phasiRNAs produced from Chr2-2025-22nt-15 (identified under 15°C; ranging from 16,537,694 to 16,537,942 on chromosome 2) are expressed at higher levels in leaves. Accordingly, the degradome signals mapped to the processing sites are stronger in leaves. (d) The phasiRNAs produced from Chr5-551-24nt-27 (identified under 27°C; ranging from 9,113,857 to 9,114,089 on chromosome 5) are expressed at higher levels in flowers. Accordingly, the degradome signals mapped to the processing sites are stronger in flowers. For both phasiRNA and degradome levels, two biological replicates of each sample were summed in RPM for histogram presentations. L15: leaves at 15°C; L27: leaves at 27°C; F15: flowers at 15°C; F27: flowers at 27°C; R15: roots at 15°C; R27: roots at 27°C. [Color figure can be viewed at wileyonlinelibrary.com]

in leaves compared to flowers and roots. However, no such organ specificity was observed under 27°C (Table S5).

Further analysis shows that the temperature sensitivity of the *TAS* genes in leaves may result in temperature-dependent regulation of the ta-siRNA targets. Three ta-siRNAs were observed to be activated by the 15°C treatment. Accordingly, the regulatory intensities of the ta-siRNAs on their targets, judged by degradome-based cleavage signals, are also increased under low temperature (Figure S8). Interestingly, all of the ta-siRNA targets are suggested to be involved in secondary ta-siRNA production. Thus, it will be interesting to investigate the biological meanings of the temperature-dependent pathway for ta-siRNA amplification.

5 | CONCLUSION

In this study, combinatory use of sRNA-seq data of *Arabidopsis thaliana* enabled us to discover hundreds of phasiRNA loci in distinct tissues and under different temperatures. In flowers, most of the 24-nt loci were demonstrated to be RDR2-dependent, while the 21-nt loci were RDR6-dependent. Notably, among the RDR-dependent loci, a significant portion was also DCL1-dependent, indicating the involvement of miRNAs in phasiRNA processing. Besides, two novel *TAS* candidate genes were discovered in flowers. Deep investigations unveiled some interesting features of the phasiRNA loci, such as the strong strand bias of phasiRNA generation, and the capacity of one locus for producing phasiRNAs by different increments. More importantly, both organ specificity and temperature sensitivity were observed for phasiRNA expression. In leaves, roots, and seedlings, much more phasiRNA loci are 15°C-sensitive compared to the 27°C-sensitive ones. Especially in leaves, the *TAS* genes are highly activated under the low-temperature treatment. Besides, the regulation intensities of several ta-siRNA—target pairs are influenced by temperature. On the other hand, degradome sequencing showed a great potential for tracing the processing signals of the phasiRNAs. In many cases, the organ-specific or temperature-sensitive expression patterns of the phasiRNAs correlate well with those of the processing signals. Analysis of the degradome containing the remnants from non-polyA-tailed transcripts uncovered several phasiRNA loci, including a ta-siRNA locus, to be transcribed through a Pol II-independent pathway. Summarily, our results should advance the understanding of the plant phasiRNAs, such as transcription, processing, and organ-specific or temperature-sensitive expression and regulation.

ACKNOWLEDGMENTS

This work was funded by the National Natural Science Foundation of China [32070655], [32370707] and [31771457].

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The degradome-seq and the sRNA-seq data sets reported in this study are available in SRA (https://www.ncbi.nlm.nih.gov/sra/) under the accession IDs PRJNA1072254 and PRJNA1072263 respectively. See Table S1 for the details of all sequencing data sets used in this study. All of the key results reported by this study are publicly available at https://github.com/HydroSteed/AT_phasiRNA.

ORCID

Yijun Meng http://orcid.org/0000-0002-4654-0622

REFERENCES

Adenot, X., Elmayan, T., Lauressergues, D., Boutet, S., Bouché, N., Gasciolli, V. et al. (2006) DRB4-dependent TAS3 *trans*-acting siRNAs control leaf morphology through AGO7. *Current Biology*, 16, 927–932.

Chen, H.M., Li, Y.H. & Wu, S.H. (2007) Bioinformatic prediction and experimental validation of a microRNA-directed tandem trans-acting siRNA cascade in arabidopsis. *Proceedings of the National Academy of Sciences*, 104, 3318–3323.

Chen, X. (2009) Small RNAs and their roles in plant development. *Annual Review of Cell and Developmental Biology*, 25, 21–44.

Dai, X., Zhuang, Z. & Zhao, P.X. (2018) psRNATarget: a plant small RNA target analysis server (2017 release). *Nucleic Acids Research*, 46, W49–W54.

Deng, P., Muhammad, S., Cao, M. & Wu, L. (2018) Biogenesis and regulatory hierarchy of phased small interfering RNAs in plants. *Plant Biotechnology Journal*, 16, 965–975.

Fahlgren, N., Montgomery, T.A., Howell, M.D., Allen, E., Dvorak, S.K., Alexander, A.L. et al. (2006) Regulation of AUXIN RESPONSE FACTOR3 by *TAS3* ta-siRNA affects developmental timing and patterning in *Arabidopsis*. *Current Biology*, 16, 939–944.

Fei, Q., Xia, R. & Meyers, B.C. (2013) Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks. *The Plant Cell*, 25, 2400–2415.

Felippes, F.F. & Weigel, D. (2009) Triggering the formation of tasiRNAs in *Arabidopsis thaliana*: the role of microRNA miR173. *EMBO Reports*, 10, 264–270.

Guan, X., Pang, M., Nah, G., Shi, X., Ye, W., Stelly, D.M. et al. (2014) miR828 and miR858 regulate homoeologous *MYB2* gene functions in *Arabidopsis* trichome and cotton fibre development. *Nature Communications*, 5, 3050.

Guo, Q., Qu, X. & Jin, W. (2015) PhaseTank: genome-wide computational identification of phasiRNAs and their regulatory cascades. *Bioinformatics*, 31, 284–286.

Jiang, P.F., Lian, B., Liu, C.Z., Fu, Z.Y., Shen, Y., Cheng, Z.K. et al. (2020) 21-nt phasiRNAs direct target mRNA cleavage in rice male germ cells. *Nature Communications*, 11, 5191.

Johnson, C., Kasprzewska, A., Tennessen, K., Fernandes, J., Nan, G.L., Walbot, V. et al. (2009) Clusters and superclusters of phased small RNAs in the developing inflorescence of rice. *Genome Research*, 19, 1429–1440.

Katoh, K. & Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, 30, 772–780.

Kume, K., Tsutsumi, K. & Saitoh, Y. (2010) TAS1 trans-acting siRNA targets are differentially regulated at low temperature, and TAS1 trans-acting siRNA mediates temperature-controlled At1g51670 expression. *Bioscience, Biotechnology, and Biochemistry*, 74, 1435–1440.

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 Research*, 40, D1202–D1210.

Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10, R25.

Li, S., Liu, J., Liu, Z., Li, X., Wu, F. & He, Y. (2014) Heat-induced TAS1 TARGET1 mediates thermotolerance via heat stress transcription factor A1a–directed pathways in arabidopsis. *The Plant Cell*, 26, 1764–1780.

Luo, Q.J., Mittal, A., Jia, F. & Rock, C.D. (2012) An autoregulatory feedback loop involving *PAP1* and *TAS4* in response to sugars in Arabidopsis. *Plant Molecular Biology*, 80, 117–129.

Ma, X., Yin, X., Tang, Z., Ito, H., Shao, C., Meng, Y. et al. (2020) The RNA degradome: a precious resource for deciphering RNA processing and regulation codes in plants. *RNA Biology*, 17, 1223–1227.

Meng, Y., Shao, C. & Chen, M. (2011) Toward microRNA-mediated gene regulatory networks in plants. *Briefings in Bioinformatics*, 12, 645–659.

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.

Pokhrel, S. & Meyers, B.C. (2022) Heat-responsive microRNAs and phased small interfering RNAs in reproductive development of flax. *Plant Direct*, 6, e385.

Price, M.N., Dehal, P.S. & Arkin, A.P. (2009) FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular Biology and Evolution*, 26, 1641–1650.

Robinson, M.D., McCarthy, D.J. & Smyth, G.K. (2010) edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26, 139–140.

Shao, C., Chen, M. & Meng, Y. (2013) A reversed framework for the identification of microRNA-target pairs in plants. *Briefings in Bioinformatics*, 14, 293–301.

Shi, C., Zhang, J., Wu, B., Jouni, R., Yu, C., Meyers, B.C. et al. (2022) Temperature-sensitive male sterility in rice determined by the roles of AGO1d in reproductive phasiRNA biogenesis and function. *New Phytologist*, 236, 1529–1544.

Shi, C.L., Zou, W.L., Zhu, Y.W., Zhang, J., Teng, C., Wei, H. et al. (2023) mRNA cleavage by 21-nucleotide phasiRNAs determines temperature-sensitive male sterility in rice. *Plant Physiology*.

Si, F., Luo, H., Yang, C., Gong, J., Yan, B., Liu, C. et al. (2023) Mobile ARGONAUTE 1d binds 22-nt miRNAs to generate phasiRNAs important for low-temperature male fertility in rice. *Science China Life Sciences*, 66, 197–208.

Song, X., Li, P., Zhai, J., Zhou, M., Ma, L., Liu, B. et al. (2012a) Roles of DCL4 and DCL3b in rice phased small RNA biogenesis. *The Plant Journal*, 69, 462–474.

Song, X., Wang, D., Ma, L., Chen, Z., Li, P., Cui, X. et al. (2012b) Rice RNA-dependent RNA polymerase 6 acts in small RNA biogenesis and spikelet development. *The Plant Journal*, 71, 378–389.

Swarbreck, D., Wilks, C., Lamesch, P., Berardini, T.Z., Garcia-Hernandez, M., Foerster, H. et al. (2007) The arabidopsis information resource (TAIR): gene structure and function annotation. *Nucleic Acids Research*, 36, D1009–D1014.

Xia, R., Chen, C., Pokhrel, S., Ma, W., Huang, K., Patel, P. et al. (2019) 24-nt reproductive phasiRNAs are broadly present in angiosperms. *Nature Communications*, 10, 627.

Xia, R., Meyers, B.C., Liu, Z., Beers, E.P., Ye, S. & Liu, Z. (2013) MicroRNA superfamilies descended from miR390 and their roles in secondary small interfering RNA biogenesis in eudicots. *The Plant Cell*, 25, 1555–1572.

Zhan, J. & Meyers, B.C. (2023) Plant small RNAs: their biogenesis, regulatory roles, and functions. *Annual Review of Plant Biology*, 74, 21–51.

Zhang, Y. (2005) miRU: an automated plant miRNA target prediction server. *Nucleic Acids Research*, 33, W701–W704.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Feng, Z., Ma, X., Wu, X., Wu, W., Shen, B., Li, S. et al. (2024) Genome-wide identification of phasiRNAs in *Arabidopsis thaliana*, and insights into biogenesis, temperature sensitivity, and organ specificity. *Plant, Cell & Environment*, 1–16. https://doi.org/10.1111/pce.14974