Indian Himalayan natural *Arabidopsis thaliana* accessions with abolished miR158 levels exhibit robust miR173-initiated *trans*-acting cascade silencing

Abhinandan Mani Tripathi[1,2,‡], Rajneesh Singh[1,2,¶], Ashwani Kumar Verma[1,2,¶], Akanksha Singh[1,2,†], Parneeta Mishra[1,2], Varun Dwivedi[1], Shiv Narayan[2,3], Vivek Hari Sundar Gandhivel[4], Pramod Arvind Shirke[2,3], Padubidri V. Shivaprasad[4] [iD] and Sribash Roy[1,2,*] [iD]

[1]*Molecular Biology and Biotechnology Division, CSIR-National Botanical Research Institute, Lucknow 226001, India,*
[2]*Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India,*
[3]*Plant Physiology Laboratory, CSIR-National Botanical Research Institute, Lucknow 226001, India,*
[4]*National Centre for Biological Sciences, Tata Institute of Fundamental Research, GKVK Campus, Bangalore 560065, India*

SUMMARY

Small RNAs (sRNAs) such as microRNAs (miRNAs) and small interfering RNAs (siRNAs) are short 20–24-nucleotide non-coding RNAs. They are key regulators of gene expression in plants and other organisms. Several 22-nucleotide miRNAs trigger biogenesis cascades of *trans*-acting secondary siRNAs, which are involved in various developmental and stress responses. Here we show that Himalayan *Arabidopsis thaliana* accessions having natural mutations in the miR158 locus exhibit robust cascade silencing of the pentatrico-peptide repeat (*PPR*)-like locus. Furthermore, we show that these cascade sRNAs trigger tertiary silencing of a gene involved in transpiration and stomatal opening. The natural deletions or insertions in *MIR158* led to improper processing of miR158 precursors, thereby blocking synthesis of mature miR158. Reduced miR158 levels led to increased levels of its target, a pseudo-*PPR* gene that is targeted by tasiRNAs generated by the miR173 cascade in other accessions. Using sRNA datasets derived from Indian Himalayan accessions, as well as overexpression and knockout lines of miR158, we show that absence of miR158 led to buildup of pseudo-*PPR*-derived tertiary sRNAs. These tertiary sRNAs mediated robust silencing of a gene involved in stomatal closure in Himalayan accessions lacking miR158 expression. We functionally validated the tertiary phasiRNA that targets *NHX2*, which encodes a Na^+-K^+/H^+ antiporter protein, thereby regulating transpiration and stomatal conductance. Overall, we report the role of the miRNA–*TAS*–siRNA–pseudogene–tertiary phasiRNA–*NHX2* pathway in plant adaptation.

INTRODUCTION

Small RNAs (sRNAs) are typically 20–24-nucleotide non-coding RNAs that act as important gene and chromatin regulators in plants. miRNAs are a class of sRNAs that negatively regulate gene expression, either by homology-mediated cleavage of mRNA or by translational inhibition, affecting many aspects of plant development (Bartel, 2004). They originate from miRNA genes (*MIRNAs*) transcribed by RNA polymerase II. The primary transcripts (pri-miRNAs) have characteristic stem-loop-like secondary structures of various lengths and complementary regions (Hirsch et al., 2006; Xie et al., 2005). The appropriate secondary structures are then processed by the RNase III enzyme Dicer-like1 (DCL1) (Bologna et al., 2009; Zhang et al., 2017) to precursor miRNA (pre-miRNA) and finally to mature miRNA duplexes (Voinnet, 2009). One strand of the mature miRNA duplex gets incorporated into the ARGO-NAUTE 1 (AGO1) protein to form an RNA-induced silencing complex (RISC) and carry out either post-transcriptional gene silencing or translational inhibition (Achkar et al., 2016; Yu et al., 2017).

Besides target degradation or translational inhibition, some miRNAs also act as a trigger for biogenesis of another class of sRNAs, namely the *trans*-acting small interfering RNAs (tasiRNAs) or phased small interfering RNAs (phasiRNAs), to target downstream genes (Allen et al., 2005;

Yoshikawa et al., 2005). The loci that generate such siRNAs are known as *PHAS* loci (Liu et al., 2020). The triggering at the target *PHAS* locus can be mediated either by a single 22-nucleotide miRNA (one-hit model) or by two 21-nucleotide miRNAs (two-hit model) (Axtell et al., 2006; Fei et al., 2013). After cleavage, the 3′ fragment is converted to double-stranded RNA (dsRNA) by RNA-DEPENDENT RNA POLYMERASE6 (RDR6) with the assistance of SUPPRESSOR OF GENE SILENCING3 (SGS3). This dsRNA is then sequentially processed by DCLs with the help of DOUBLE-STRANDED RNA BINDING FACTORS into 21- or 24-nucleotide secondary siRNAs from the 5′ end of the template strand (Adenot et al., 2006; Vazquez et al., 2004). In the two-hit model, which is well studied in Arabidopsis *TAS3*, two miR390 molecules bind to the *TAS3* locus, but only the 3′ proximal end is cleaved while the 5′ binding site defines the boundary, making tasiRNA production very efficient, precise, and accurate (Axtell et al., 2006; Liu et al., 2020). However, in gymnosperms and eudicots, the 5′ miR390 target sites are likely cleaved, generating tasiRNAs in the 5′ to 3′ direction (de Felippes et al., 2017). Besides, in the two-hit mechanism, two cleavages in one target occur; this mechanism has been reported in plants such as *Medicago truncatula* and strawberry (*Fragaria vesca*) (Xia et al., 2015; Zhai et al., 2011). The bidirectional processing mechanism at two cleaved ends has not been reported in Arabidopsis (Xia et al., 2017). Subsequently, using artificial tasiRNA constructs, the double cleavage and bidirectional processing was reported in Arabidopsis but the phasing was found to be poor (de Felippes et al., 2017). Thus although mechanisms of phasiRNA biogenesis are well reported, negative regulation of phasiRNA biogenesis is not known.

PHAS loci can be located within both protein-coding and non-coding genes (Fei et al., 2013). Among the non-coding *PHAS* loci, the *TAS* loci, their targets, and derived phasiRNAs have been well studied. In the model plant *A. thaliana*, *TAS1a*, *TAS1b*, *TAS1c*, and *TAS2* are targeted by miR173, and their 3′ cleavage products are either identical or very similar tasiRNAs (Yoshikawa et al., 2005). *TAS3* and *TAS4* are targeted by miR390 and miR828, respectively, and derived tasiRNAs target three auxin response factors, namely ARF2, ARF3, and ARF4 (Allen et al., 2005; Williams et al., 2005), as well as MYB transcription factors (TFs) (Xu et al., 2012). Non-coding *PHAS* loci have been reported in other plant species as well (Huang et al., 2019; Xia et al., 2015). *PHAS* loci within protein-coding genes have been reported in the leucine-rich repeat (*NLR*) genes (Zhai et al., 2011), MYB TF genes (Xia et al., 2013), auxin response factor (*ARF*) genes (Xia et al., 2017), NAC TF genes (Sosa-Valencia et al., 2017), and pentatricopeptide repeat (*PPR*) genes (Chen et al., 2007; Howell et al., 2007). Importantly, some *PPRs* also produce tertiary siRNAs whose production is triggered by tasiRNAs derived from *TAS* or *TAS*-like genes (Howell et al., 2007; Xia et al., 2013).

This miRNA–*TASL*–*PPR*–siRNA pathway was reported to be conserved in many plant species (Xia et al., 2013). In the above examples, cascade silencing produced abundant secondary siRNAs that targeted other RNAs *in trans*. Since the number of secondary silencing loci in a given species is not very high, it is possible that plants also evolved negative regulators of such cascade silencing to avoid unintended regulation. However, such negative regulators of phasiRNA biogenesis to stop unintended robust cascade silencing are yet to be identified. In addition, although the biological functions of different phasiRNAs are well documented (Marin et al., 2010; Shahid et al., 2018; Zhou et al., 2013), the functional significance of tertiary silencing is yet to be identified (Jia et al., 2020).

miR158 is among the first identified plant miRNAs (Reinhart et al., 2002). It is a non-conserved young miRNA, found only among the members of Brassicaceae (Figure S1a) (Fahlgren et al., 2007; Rajagopalan et al., 2006). Later it was identified in several cruciferous plants such as *Arabidopsis lyrata* (Fahlgren et al., 2007), *Brassica oleracea* (Wang et al., 2012), *Raphanus sativus* (Wang, Liu, et al., 2015), and *Brassica napus* (Xu et al., 2012). In *A. thaliana*, it has two family members: *ath-miR158a* and *ath-miR158b* (Zhao et al., 2018) (Figure S1b). More recently it was shown that miR158 plays a role in pollen development by regulating PPR in *Brassica campestris* (Ma et al., 2017).

Arabidopsis thaliana harbors wide genetic variations among wild types that have been explored extensively over the past decades (Alonso-Blanco & Koornneef, 2000; Mitchell-Olds & Schmitt, 2006). Effects of single-nucleotide polymorphisms in *MIRNA* sequences and their implications have been reported (Ryan et al., 2010; Todesco et al., 2012). In our previous study, we observed negligible expression of miR158 in a particular *A. thaliana* population of the Indian west Himalayas under different growing conditions (Tripathi et al., 2019). These populations, which originate from an elevation range of 700 to 3400 m above mean sea level (amsl), having varied climatic conditions, were reported to be genetically and morphologically distinct from each other (Singh et al., 2015; Singh & Roy, 2017; Tyagi, Singh, et al., 2016). Here, we show that negligible or moderate expression of miR158 is due to deletion and insertion events, respectively, in the *MIR158* gene, affecting its processing to mature miRNA. We validated that miR158 targets a pseudogene, pseudo-*PPR*, which acts as a *PHAS* locus and is a member of the *PPR* gene family. Additionally, we found that absence of miR158 expression in these Himalayan accessions led to enhanced tertiary phasiRNA biogenesis from pseudo-*PPR* initiated by *TAS2*-derived secondary siRNAs. Two such *TAS2*-derived siRNAs triggered a 'two-hit two-cleavage' mode of silencing to induce the production of tertiary phasiRNA. We also show that one of the tertiary phasiRNA
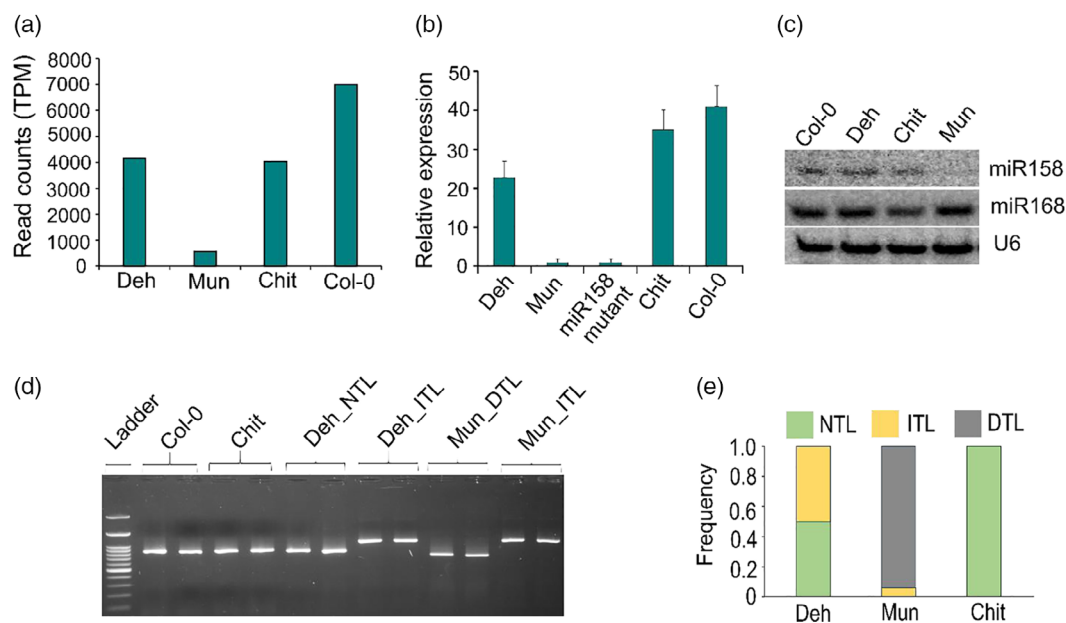
Figure 1. Expression pattern and length polymorphisms of the miR158 locus in west Himalayan *Arabidopsis thaliana* populations.

(a) The expression values of miR158 in transcripts per million (TPM) in different Arabidopsis populations as determined from small RNA sequencing data. Data are shown as the average of two independent biological replicates of each population.

(b) Stem-loop qRT-PCR analysis of miR158 expression in different populations as well as the miR158 SALK mutant and Col-0. The data are presented as the mean ± SD of four independent biological replicates.

(c) Expression pattern of miR158 in different populations as revealed by Northern blot. U6 and miR168 served as loading control and positive control, respectively.

(d) Representative agarose gel image depicting the length polymorphisms of the miR158 locus from a few individuals of each of the three populations. The Deh population harbored both insertion type (ITL) and normal type (NTL), the Mun population harbored deleted (DTL) and insertion type, and the Chit population harbored only normal type length polymorphisms as compared to the reference Col-0.

(e) Frequency of occurrence of NTL, ITL, and DTL plants in each of the three populations as determined from 50 plants of each population.

targeted *NHX2*, thereby regulating physiological processes and prompting plants to complete the life cycle early, presumably as a strategy of local adaptation in Himalayan environments. These results suggest the occurrence of molecular innovations in secondary silencing mechanisms playing crucial roles in plant adaptation.

RESULTS

Population-specific expression of miR158

Our previous miRNA analysis (Tripathi et al., 2019) and the present data consistently indicate low expression of miR158 in one of the Indian Himalayan populations of *A. thaliana*, Mun, as compared to the other two populations, Deh and Chit (Figure 1a, Table S1). To reconfirm this, we analyzed its expression under native, controlled, and common garden conditions along with a miR158 SALK mutant by stem-loop quantitative reverse transcriptase-PCR (qRT-PCR). As expected, the expression of miR158 was consistently negligible in Mun under controlled (Figure 1b,c), native, and common garden conditions, as that of the miR158 mutant (Figure S2a,b), while its expression was high in the other two populations. Among the multiple forms of miR158 that are expressed in plants, the

relative expression of miR158a and miR158a-3p was significantly higher compared to miR158b and miR158a-5p in all populations (Table S1; Figure S2c,d).

Polymorphisms in *MIRNA158* are unique to the west Himalayan populations

To find out probable reasons behind the population-specific expression of miR158, we first analyzed the putative promoter sequence of miR158 (1 kb upstream of the transcription start site [TSS]) and observed no significant sequence polymorphisms among the three populations (GenBank# OQ267618–OQ267620), ruling out the possibility that the low expression of miR158 in Mun was due to differential promoter activity (Figure S3). However, PCR amplification of the *MIRNA158 (MIRNA158a)* locus using genomic DNA from a few plants of each of the three populations as a template followed by sequencing showed length polymorphisms (Figure 1d,e, GenBank# OQ267615–OQ267617). Subsequently, approximately 50 plants from each of the three populations were analyzed (Figure S4b–d). The results indicated that while the normal (standard) 535-bp *MIRNA158* of reference Col-0 (Szarzynska et al., 2009) was present in all Chit plants, individuals of the Mun population harbored either a 437-bp locus having a discontinuous

deletion of 98 bp (starting at approximately 10 bp downstream of the TSS), including four nucleotides of the 5′ precursor, or a 760-bp locus with an insertion of 225 bp (approximately 10 bp downstream of the TSS); Deh harbored both the normal type and the insertion type locus (Figure 1d and Figures S4 and S5). Hereafter, we refer to any normal type locus as NTL, to the deleted type locus as DTL, and to the insertion type locus as ITL. Notably, the frequency of occurrence of DTL in Mun was 0.94, the remaining being ITL; and Deh harbored both NTL and ITL, having a frequency of approximately 0.5 each (Figure 1e). Hereafter, all experiments were carried out considering individuals of Deh having an NTL type miR158 locus as NTL, and individuals of Mun harboring ITL or DTL type loci were considered as ITL and DTL plants, respectively. To examine if the insertion/deletion polymorphisms were present in the primary transcript as well, we amplified the locus by PCR using cDNA as a template. Amplified products exhibited similar length polymorphisms as observed with genomic DNA, ruling out the possibility that the observed polymorphisms are due to any splice variants (Figure S6). We also checked if these polymorphisms are unique to the west Himalayan populations by comparing the *MIRNA158* sequences from available 1001 genomes of *A. thaliana* (Alonso-Blanco et al., 2016). The observed insertion/deletion polymorphisms were found to be unique to the west Himalayan populations. Further, the inserted fragment was not a transposable element (TE), as a BLAST search against the TAIR database did not reveal any TEs. However, the inserted fragment showed 97% similarity with a sequence of chromosome 4, while *MIRNA158* is located on chromosome 3 of *A. thaliana*. Thus, the insertion event might be due to inter-chromosomal translocation.

Polymorphisms in *MIRNA158* affect its expression pattern

We checked the expression pattern of miR158 in NTL (Deh), ITL (Mun), and DTL (Mun) by qRT-PCR. The expression of miR158 was the highest, intermediate, and negligible in NTL, ITL, and DTL, respectively (Figure 2a; Table S1). In addition, we determined the expression level of both primary transcripts (pri-miR158) and the precursor of miR158 (pre-miR158). As expected, the level of pri-miR158 was more or less similar in all three variants (Figure 2b). However, the pre-miR158 level was high, intermediate, and low in NTL, ITL, and DTL, respectively, as observed with mature miR158 (Figure 2c). This further suggests that the variations in the expression of miR158 were not due to promoter activity, but presumably due to differential processing of primary transcripts to precursors. We evaluated the secondary structures of all three different primary transcript sequences from NTL, ITL, and DTL plants. The secondary structure prediction of all the variants revealed that the lower stem and their internal loops in DTL plants were structurally different (disturbed) from the canonical ones,

and so was the large internal loop in ITL plants, but these were intact in NTL plants (Figure 2d–f). However, there were no such structural anomalies in the predicted secondary structures of precursor sequences of all the three variants (Figure S7). Any anomaly in the lower stem and internal loop structure can lead to sloppy primary transcript processing, ultimately influencing the abundance of mature miRNAs (Mateos et al., 2010; Zhu et al., 2013). Further, to examine if insertion or deletion affected the stability of pri-miR158 and thus led to lower levels of miR158, an RNA decay assay was performed. The insignificant differences in accumulation of pri-miR158 between NTL, ITL, and DTL indicate that lower expression of miR158 may not be due to the stability of pri-miR158 but improper processing of pri-transcripts to precursor (Figure S8a). This was further validated by transient expression analysis of pri-*MIR158* and miR158 in *Nicotiana benthamiana*. The level of primary transcripts was more or less similar in the leaves infiltrated with any of the three constructs. On the other hand, negligible and intermediate expression of miR158 was observed in the leaves of DTL and ITL infiltrated constructs, respectively, as compared to NTL (Figure S8b,c). Finally, we reconfirmed the expression pattern of mature miR158 in NTL, ITL, DTL, miR158 OE, and its mutant line by Northern hybridization (Figure 2g). Furthermore, as expected, we observed significantly reduced expression of miR158 in two CRISPR–Cas9 edited lines as well (Figure S9a,b). It is to be noted here that the nucleotide deletion was found in the immediate 5′ end (lower stem) of the mature miR158, leading to disturbed secondary structures (Figure S9c) as observed in the DTL population.

miR158 targets a pseudogene, pseudo-*PPR*

The target of miR158 was predicted by using psRNATarget (Dai & Zhao, 2011). Four targets, namely, pseudo-*PPR*, tetratricopeptide repeat (*TPR*), transducing/WD40 repeat (*BUB3.2*), and *PPR*, were identified as the putative targets of miR158 (Table S3). We attempted to validate the predicted targets by modified 5′ RLM-RACE, and pseudo-*PPR* (AT1G62860.1) was identified as the only authentic target of miR158 (20/20 clones) (Figure 3a). Quantification of the pseudo-*PPR* transcripts by qRT-PCR also showed an inverse expression pattern with miR158 expression (Figure 3b), confirming it is a target of miR158. Further, analysis of available degradome data (GSE55151) (Thatcher et al., 2015) also revealed the identified pseudo-*PPR* as the target of miR158 ($P < 0.0001$) (Figure 3c). *PPRs* are mostly known to be involved in RNA editing (Yagi et al., 2013), and some act as *PHAS* loci (Howell et al., 2007). *PPRs* are separated into two major classes based on the nature of their PPR motifs, the P and PLS classes (Schmitz-Linneweber & Small, 2008). Phylogenetic analysis of the target pseudo-*PPR* with 297 *A. thaliana* PPRs (Lurin et al., 2004) suggested that it is a P-type *PPR* (Figure S10a).
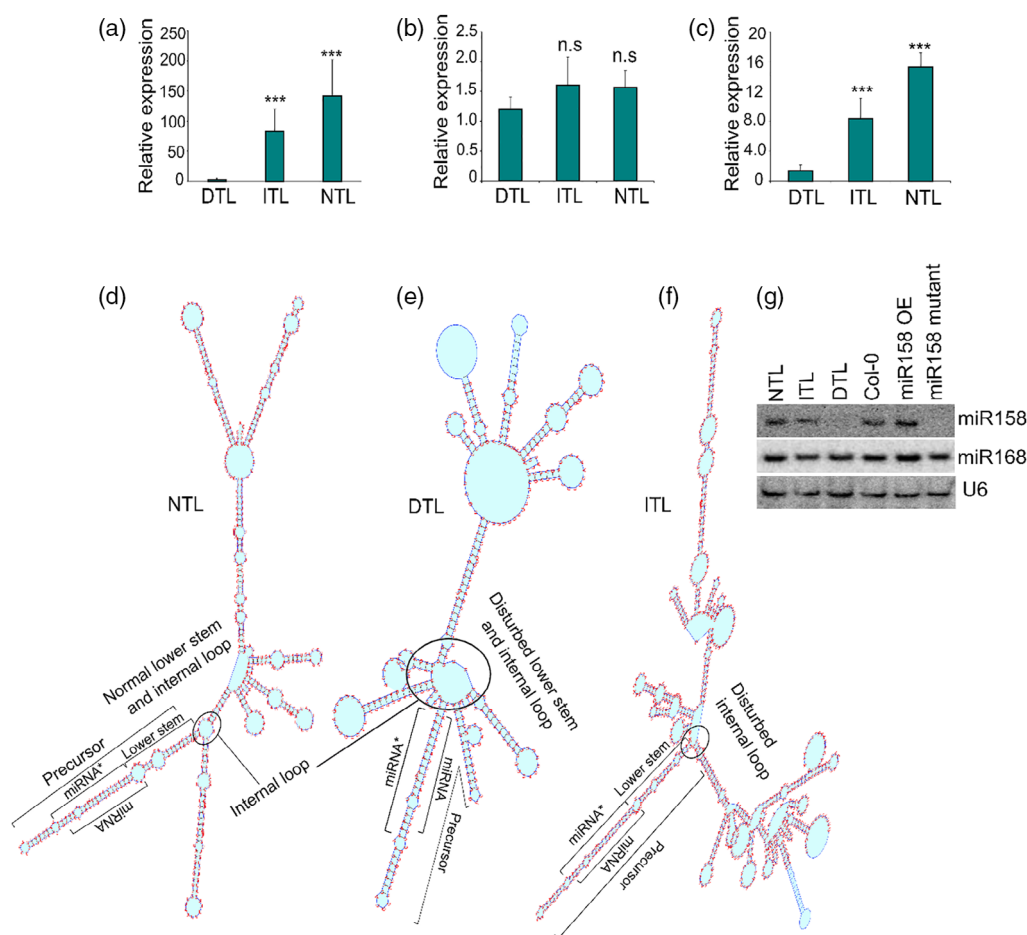
Figure 2. Expression patterns of mature miR158, pri-miR158, and pre-miR158 and predicted secondary structures of pri-miRNA158.

(a–c) Relative expression patterns of (a) miR158, (b) primary-miR158, and (c) precursor-miR158 in DTL, ITL, and NTL plants. The data are presented as the mean $\pm$ SD of three independent biological replicates. ***$P < 0.001$, **$P \leq 0.01$, *$P \leq 0.05$ (Student's t-test); ns, not significant.

(d–f) The predicted secondary structures of pri-miR158 in (d) NTL, (e) DTL, and (f) ITL plants, derived from the respective sequence using the RNAfold web server. The miRNA/miRNA*, precursor, and lower stem regions are shown along with disturbed large internal loops in DTL and ITL (encircled).

(g) Northern blot for miR158 in different lines. U6 and miR168 served as loading control and positive control, respectively.

The target pseudo-*PPR* acts as a *PHAS* locus in absence of miR158 activity

Pseudogenes have been reported to act as sponges for some miRNAs and thus protect the cognate target from degradation (Yu et al., 2014). However, our surprising observation that miR158 cleaves the pseudogene prompted us to conduct a more detailed investigation. The target pseudo-*PPR* (AT1G62860) is annotated as a pseudogene in the TAIR database, having several stop codons to limit protein production (Figure S10b). This pseudo-*PPR* shows 78% similarity with its parent *PPR* (AT1G64100) (Figure S10b). Previous computational prediction analyses suggested this pseudo-*PPR* is the target of two sRNAs, namely, TAS1b 3′D4(−) and miR161.1, thus indicating its potential as a *PHAS* locus (Axtell et al., 2006; Howell et al., 2007).

PHAS loci are targeted either by one 22-nucleotide sRNA (one-hit model) or by two 21-nucleotide sRNAs

(two-hit model) (Fei et al., 2013). We analyzed the available degradome data to determine the putative sRNAs that might cleave the pseudo-*PPR*. Our analysis indicated that apart from miR158, pseudo-*PPR* is also a putative target of tasiRNAs, atTAS2 (2)-siR9(−) (hereafter after referred to as siR9) and atTAS2(2)-siR12(−) (hereafter after referred to as siR12) derived from miR173-triggered TAS2 transcripts (Figure 3c) (Felippes & Weigel, 2009). We further validated the target cleavage of both tasiRNAs by using modified 5′ RLM-RACE and detected the expected cleaved ends (Figure 3d,e). We observed neither targeting nor cleavage of the pseudo-*PPR* by the previously identified TAS1b 3′D4 (−) and miR161.1 (Axtell et al., 2006; Howell et al., 2007).

Further, to understand the nature of pseudo-*PPR* targeting by various sRNAs, we pooled sRNA libraries from Deh, Mun, and Chit and mapped them against the pseudo-*PPR*. A large number of sRNAs mapped to pseudo-*PPR*;
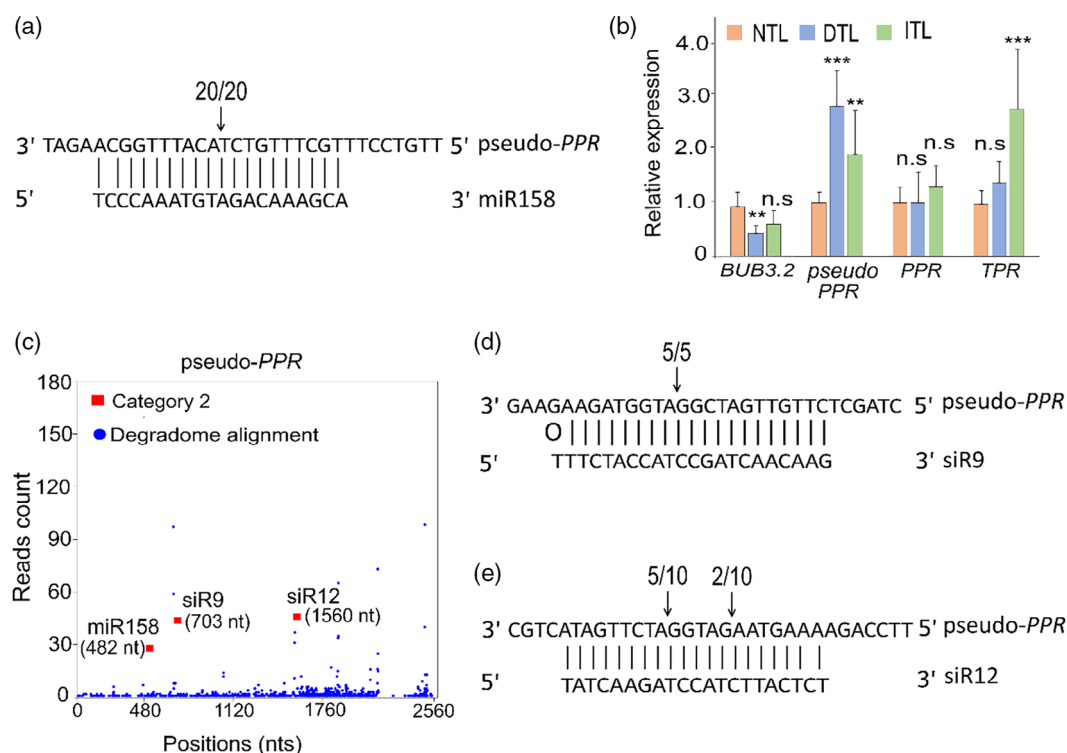
Figure 3. Validation and expression analysis of target gene.
(a) Pseudo-*PPR* was validated as the target of miR158 by modified 5′ RLM-RACE. The number above each arrow indicates the positive clones at this cleavage site/total number of colonies sequenced.
(b) The expression patterns of all the predicted targets of miR158 as evaluated by qRT-PCR. The data are presented as the mean ± SD of three independent biological replicates. ***$P < 0.001$, **$P \leq 0.01$, *$P \leq 0.05$ (Student's t-test); ns, not significant.
(c) Target plots (T-plots) of miR158, siR9, and siR12 based on degradome data. The cleavage sites of pseudo-*PPR* by miR158 and tasiRNAs are shown in red squares. The blue circles represent the fragments mapped on the pseudo-*PPR*. The x- and y-axis represent the nucleotide position of pseudo-*PPR* and the read abundance of cleaved transcripts detected in degradome data, respectively. Degradome category 2 is defined as a signal having more than one raw read mapped at the position, with abundance at the position less than the maximum but higher than the median abundance for that transcript.
(d, e) Pseudo-*PPR* was validated as the target of (d) siR9 and (e) siR12 by modified 5′ RLM-RACE. The numbers above arrows indicate the positive clones at this cleavage site/total number of colonies sequenced.

they actually mapped between the siR9 and siR12 binding sites, while no sRNA mapped near miR158 binding sites (Figure 4a). Moreover, these sRNAs were in phase with phasing scores of 1.38–35.0 ($P < 0.0001$) (Figure 4b), starting with the cleavage site of siR9 (702 nucleotides, from the 5′ end of pseudo-*PPR*) (Figure 4c). Although siR9 is reported to be a 21-nucleotide sRNA (Howell et al., 2007; Zhang et al., 2014; Zhong et al., 2013), surprisingly our sRNA analysis suggested higher expression of the 22-nucleotide siR9 rather than the 21-nucleotide ones (Figure 4d and Table S4). Further, if miR158 targeting altered the abundance of these siRNAs, then the populations must show differences in the accumulation of pseudo-*PPR*-derived siRNAs. In order to check this, we mapped sRNAs of each of the populations individually against the pseudo-*PPR* and observed that a higher number of them mapped in Mun as compared to the other two populations (Figure S11a,b). Additionally, out of 3138, a total of 830 non-redundant phasiRNAs derived from the pseudo-*PPR* (hereafter referred to as PDP) from all the

populations were identified (Table S5). A majority of the PDPs were 21 nucleotides long, as expected, while a small fraction was 22 nucleotides long (Figure S11c). Further, around 45% of these PDPs had a characteristic 5′ terminal uridine, indicating their potential to get associated with AGO1 (Figure S11c). Since many *PPRs* are known to act as *PHAS* loci and phasiRNAs produced from them might show redundancy due to repetitive domains, we speculated that the PDPs might be identical with the other *PPR*-derived phasiRNAs. However, we found only 8% of the pseudo-*PPR*-derived sRNAs to be similar to sRNAs derived from the 23 known (Howell et al., 2007) *PPR PHAS* loci (Table S6). These results show that pseudo-*PPR* acts as a *PHAS* locus, where siR9 (22 nucleotides) acts as a trigger. It was clearly evident that the abundance of PDPs was more in Mun than in Deh and Chit ($P = 0.005645$, $F = 4.32423$) (Table 1; Figures S12a). Since the above sRNA libraries were prepared from pooled leaf samples of several plants of each population (Tripathi et al., 2019), we re-examined the data using sRNA sequencing data from
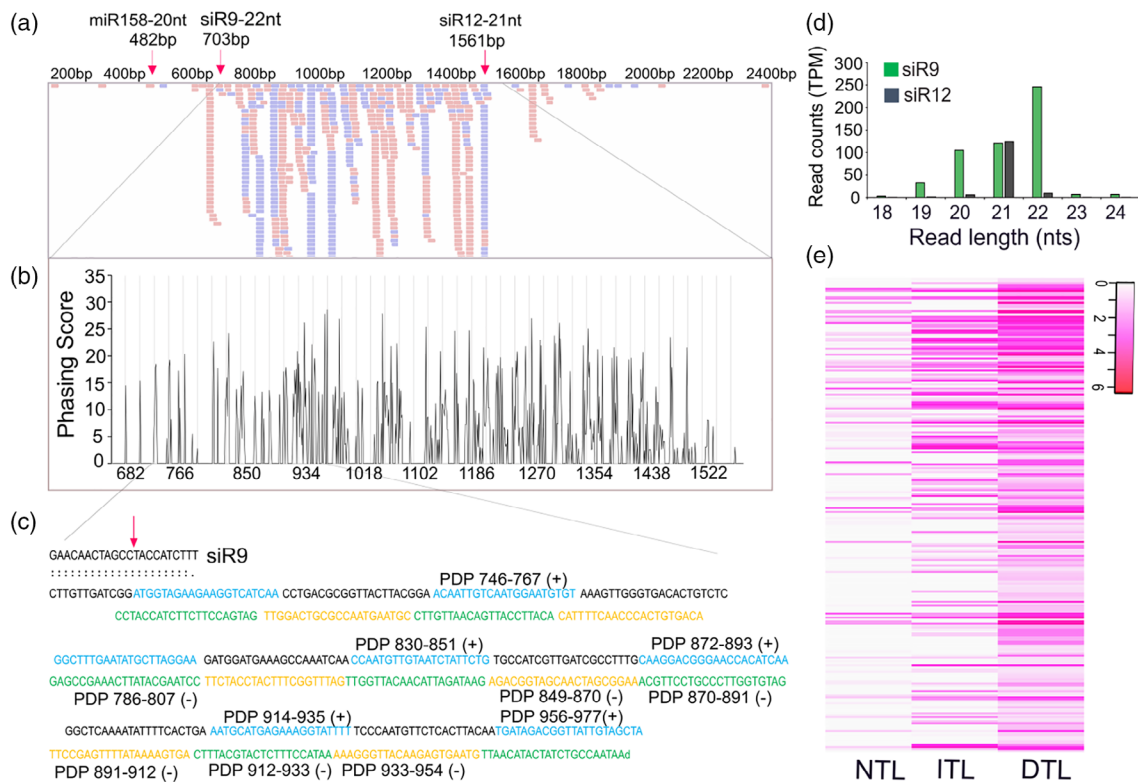
Figure 4. PhasiRNA analysis and their expression.

(a) The small RNAs were mapped against pseudo-*PPR* shown as a horizontal line along with nucleotide positions and cleavage sites (arrow) of miR158, siR9, and siR12. The red blocks depict reads aligned in forward orientation and blue blocks depict reads aligned in reverse orientation (visualized via Integrative Genomics Viewer).

(b) The phasing scores calculated using small RNAs mapped at nucleotide positions 600–1600 of pseudo-*PPR*. The x- and y-axes represent the phasing score and nucleotide positions of pseudo-*PPR*, respectively. The 21-nucleotide phasing intervals are shown by vertical gray lines.

(c) The phasing pattern of a significant cluster of phasiRNA derived from pseudo-*PPR* (PDP) initiated at the siR9 cleavage site (marked by arrow). Blue and black sequences indicate the phasiRNAs (PDPs) derived from the sense (+) strand while yellow and green sequences represent phasiRNAs derived from the antisense (−) strand.

(d) Read length distribution of siR9 and siR12 in Col-0 showing that the expression of 22-nucleotide siR9 is higher than that of 21-nucleotide siR9.

(e) Heatmap depicting expression patterns of phasiRNAs in NTL, ITL, and DTL ($n = 238$). The log$_2$(TPM) values of PDP read counts are represented by the color intensity scale. NTL, normal type locus; ITL, inserted type locus; DTL, deleted type locus.

randomly selected individual plants of each variant type. sRNA data from individual DTL, ITL, and NTL plants confirmed the above observations (Table 1; Figure 4e). Similarly, PDP analyses of miR158 mutant, pseudo-*PPR* mutant, and Col-0 plants confirmed the higher expression of PDPs in the miR158 mutant and lower in the pseudo-*PPR* mutant (Table 1; Figure S12b). Moreover, overexpression lines of both miR158 (miR158 OE) and pseudo-*PPR* (pseudo-*PPR* OE) resulted in lower and higher expression of PDP, respectively, as compared to Col-0 (Table 1; Figure S12c). Further, to rule out the possibility that the differences in targeting of pseudo-*PPR* and thereby PDP biogenesis in the DTL, ITL, and NTL plants were not due to differences in the expression of miR173, siR9, siR12, or *TAS2*, which are involved in PDP biogenesis, we checked their expression levels in these plants. Both sRNA and qRT-PCR analyses did not reveal any significant differences in their expression levels in DTL, ITL, and NTL, except siR12

(Figure S13a–d, Table S7). Similarly, there were no polymorphisms in the miR173 and *TAS2* loci of DTL, ITL, and NTL (GenBank# OQ267609–OQ267614) (Figure S13f–i). Overall, these analyses suggest miR158 negatively regulates PDP biogenesis. This adds an additional layer of regulation of secondary siRNAs.

The highly expressed PDP targets *NHX2*

Most of the 21-nucleotide siRNAs are known to be involved in AGO1-mediated cleavage of target mRNA (Qi et al., 2005). In order to determine if any of the PDPs were associated with AGO proteins, we took advantage of previously published AGO1- and AGO4-bound sRNA datasets (GSE28591) (Wang et al., 2011). We identified a total of 200 PDPs overlapping with AGO1-associated sRNAs, while as expected only a small fraction of them (25 PDPs) overlapped with AGO4 immunoprecipitation datasets (Table S8). This indicates the possibility that a large

Table 1 Comparative read counts of PDPs and related small RNAs in different plants

	Deh		Mun		Chit		Col-0	
	Rep1	Rep2	Rep 1	Rep 2	Rep 1	Rep 2	Rep 1	Rep 2
miR166e-3p	1768.80	1029.94	1567.28	1289.56	1743.09	2119.24	1327.27	1507.60
miR158	6683.79	1576.04	576.62	552.27	3567.53	4469.36	6070.95	7749.56
miR173	89.03	27.92	41.66	34.85	87.95	78.77	47.14	49.10
siR12	98.46	19.97	82.12	67.93	73.68	67.89	57.23	81.55
siR9	403.07	411.02	414.49	358.79	281.51	357.60	250.80	350.86
PDPs	19.09	9.49	41.81	33.79	9.21	12.36	9.22	12.74

	NTL		DTL	ITL		Col-0	miR158 mutant		Pseudo-PPR mutant		miR158 OE	Pseudo-PPR OE
	Rep 1	Rep 2	Rep 1	Rep 1	Rep 2	Rep 1	Rep 1	Rep 2	Rep 1	Rep 2	Rep 1	Rep 1
miR166e-3p	1945.29	1830.50	1548.18	2505.44	1523.02	3557.76	1909.96	1669.16	1651.97	1767.17	2256.39	2140.78
miR158	27267.35	17607.99	57.95	2653.43	4424.90	19942.47	341.52	256.19	23300.67	34848.24	192265.2	19942.29
siR9	664.21	344.52	405.54	459.21	264.66	539.26	527.39	528.19	422.76	557.04	403.50	298.87
siR12	89.10	27.56	51.37	43.54	6.29	64.16	130.41	74.80	47.58	90.38	65.30	52.53
PDPs	21.12	19.66	46.56	42.07	40.47	14.76	25.66	23.80	11.44	13.06	10.86	84.10

Expression level of different small RNAs (TPM) in west Himalayan populations and in Col-0 (upper panel) as well as in DTL, ITL, NTL, and different mutant and overexpression (OE) lines (lower panel). The expression of miR166 was considered as a technical standard being not associated with PDPs and related siRNA pathways. PDPs, pseudo-*PPR*-derived phasiRNAs.

number of the PDPs function via target cleavage. For further characterization of PDPs, we focused on 11 PDPs which were derived from the most abundant clusters (start: 746; end: 997; $P < 0.00001$, $n = 11$, $k = 236$) (Table S9) and were in phase with the siR9 cleavage site. To find out the cleaved ends of mRNAs by these PDPs, available degradome data were analyzed. We identified three different mRNAs belonging to degradome category 2 ($P < 0.05$) (more than one raw read mapped at the position and abundance at the position is less than the maximum but higher than the median for the transcript) (Table S10). These three mRNAs were putative targets of one of the highly expressed PDPs, PDP956-977(+) (Table S10). The best target identified by this analysis was *NHX2* (Figure S14), which is expressed in stomatal guard cells and maintains turgor to regulate transpiration (Andrés et al., 2014; Barragán et al., 2012; Bassil et al., 2011). Two other targets identified were DNAJ heat shock N-terminal domain-containing proteins (AT2G05250.1 and AT2G05230.1). Target validation using modified 5′ RLM-RACE also identified *NHX2* as a target of PDP956-977(+) (Figure 5a). Subsequently, we checked the expression of *NHX2* in DTL and miR158 mutant plants by qRT-PCR. As expected, the expression of *NHX2* in miR158 mutant and DTL plants was lower than in other lines (Figure 5b, Figure S15a). Notably, the expression of this gene in Mun was also found to be lower in our earlier global RNA sequencing data (Tyagi, Yadav, et al., 2016).

NHX1 and *NHX2* are members of the *NHX* gene family and are involved in regulating stomatal functions (Barragán et al., 2012; Bassil et al., 2011). It was reported that *NHX2* exhibits a diurnal expression pattern and that expression of *NHX1* is increased in *nhx2* to counterbalance the reduced NHX2 activity (Andrés et al., 2014). We quantified the expression of *NHX1* and *NHX2* at different time points of the day (Figure S15b,c). In contrast to earlier observations, we did not observe increased expression of *NHX1* in miR158 mutants, as expected with *nhx2*. This may be due to knockout effects of *NHX2* as compared to knockdown in the miR158 mutant. Further, maximum expression of *NHX2* was observed after 4 h of light in Col-0 and miR158 OE, while in the miR158 mutant its expression was the highest after 8 h of light (Figure S15d). We speculated that the observed expression pattern of *NHX2* might be due to diurnal variation in expression of either miR158 or its target pseudo-*PPR*. However, though there was no variation in expression of miR158 at different time points of the day, pseudo-*PPR* exhibited diurnal variations with maximum expression after 8 h and gradually decreasing expression thereafter (Figure S15d). Our results indicate that the diurnal expression pattern of both pseudo-*PPR* and *NHX2* might be due to diurnal variation of the stoichiometric ratio between miR158 and pseudo-*PPR*.

Downregulation of *NHX2* by PDP affects physiological processes

NHXs, mainly NHX1 and NHX2, are essential for turgor regulation and stomatal functions (Barragán et al., 2012; Bassil et al., 2011). Therefore, to examine if PDP956-977(+)-accumulating plants having altered *NHX2* expression also exhibit impaired stomatal functioning, we measured stomatal conductance, transpiration, and water use efficiency
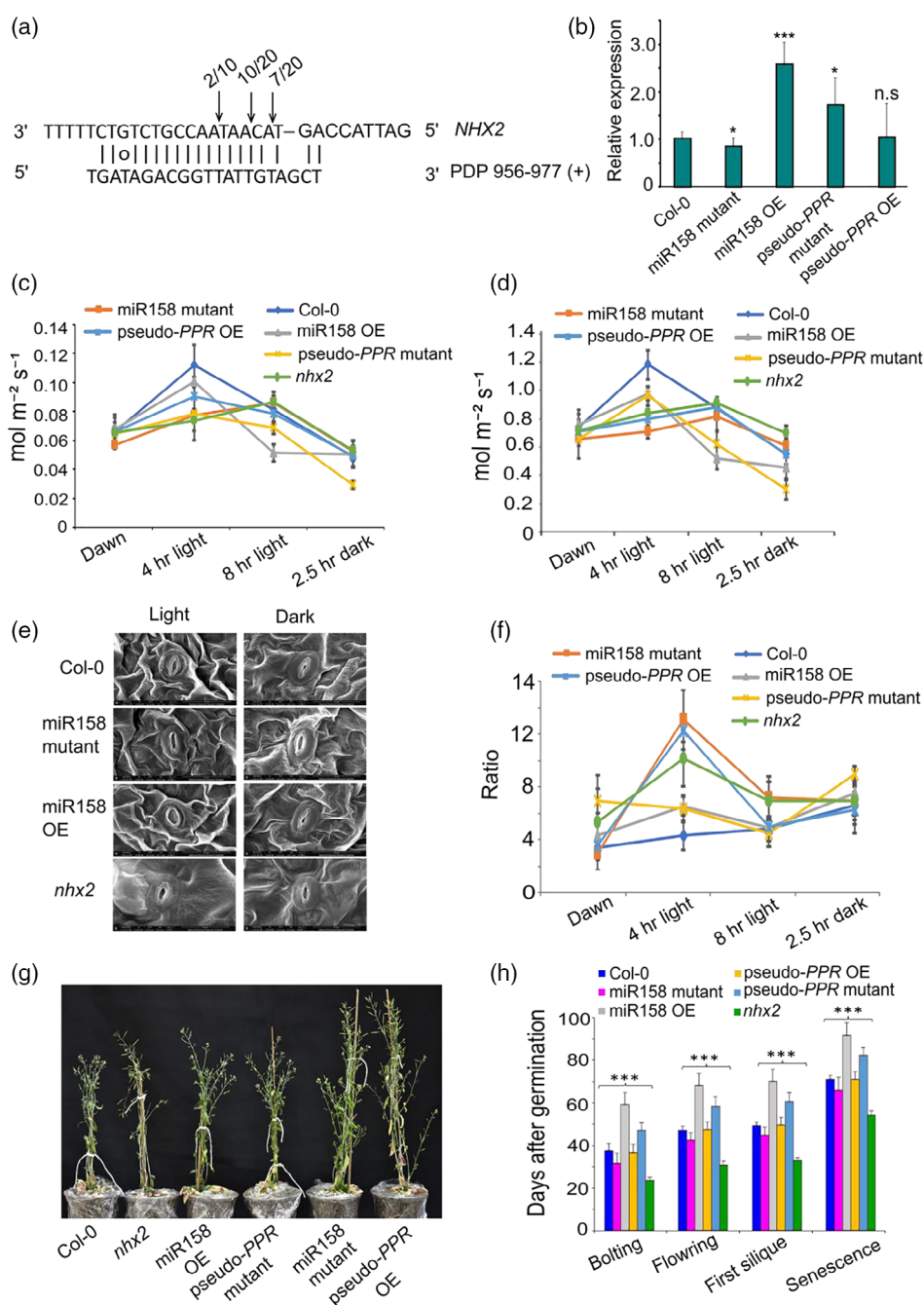
Figure 5. PDP target validation by modified 5′ RLM-RACE and their functional characterization.

(a) The complementary pairing of PDP956-977(+) with its target *NHX2* is shown with arrows indicating the cleaved positions with the frequency of cloned RACE products above it.

(b) Expression patterns of *NHX2* in different transgenic and mutant lines as determined by qRT-PCR. The data are presented as the mean ± SD of three independent biological replicates. ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$ (Student's *t*-test); ns, not significant.

(c–f) Physiological parameters of mutants and transgenic lines in the Col-0 background measured at different time points: (c) stomatal conductance, (d) transpiration rate, and (e) scanning electron microscope images of stomata at two time points (in light and after 150 min in dark) taken at 8000×. (f) Water use efficiency. The data are presented as the mean ± SD of five biological replicates.

(g) Images of miR158 mutant, miR158 OE, pseudo-*PPR* mutant, pseudo-*PPR* OE, *nhx2*, and Col-0 plants taken on day 56 post-germination. Plants were grown with normal watering until day 21 and then water was withheld for 35 days after pots were saturated with water.

(h) Measurement of various life cycle-related parameters in mutant and transgenic lines in a Col-0 background. Life cycle parameters, namely, bolting, first flower opening, emergence of first silique, and senescence, were measured in terms of days after germination. Data are presented as the average ± SD of 14–15 plants per genotype. Asterisks above the graph show significance (***$P < 0.001$, **$P < 0.01$, *$P < 0.05$; ns, not significant).

(WUE) of miR158 mutant, miR158 OE, pseudo-*PPR* mutant, pseudo-*PPR* OE, and *nhx2* plants at different time points of the day. All lines showed reduced stomatal conductance and transpiration at different time points compared to the control, except the miR158 mutant and *nhx2*, whose transpiration rates were slightly higher after 8 h (Figure 5c,d, Table S11). Interestingly, pseudo-*PPR* OE also exhibited lower stomatal conductance and transpiration rates compared to the control, presumably due to the higher stoichiometric ratio in favor of pseudo-*PPR* over miR158. Further, stomatal conductance and the transpiration rate were maximum after 4 h and gradually decreased thereafter in all plants except the miR158 mutant, *nhx2*, and pseudo-*PPR* OE, and finally all reached the lowest levels in the dark. The impaired stomatal activity in the miR158 mutant, *nhx2*, and miR158 OE was further examined by imaging the stomatal aperture of leaves collected from light- or dark-treated plants under a scanning electron microscope. It is evident that stomatal closure of these lines was delayed under dark conditions as compared to the control (Figure 5e; Figure S16a) and they exhibited significantly higher WUE (Figure 5f) and lower rates of water loss after 165 min as compared to control plants (Figure S16b, Table S11). Accordingly, when subjected to water withholding for 35 days, the miR158 mutant, pseudo-*PPR* OE, and *nhx2* plants could complete their life cycle, while the growth of Col-0, miR158 OE, and the pseudo-*PPR* mutant was severely affected, bearing no or very few siliques and seeds (Figure 5g). Similar observations were made with NTL, ITL, and DTL (Figure S16c–h) and edited lines (data not shown). These results indicate that in absence of miR158 activity, one of the PDPs, PDP956-977(+), derived from the miR158 target, a pseudo-*PPR*, regulates transpiration and stomatal conductance via *NHX2*.

Finally, we evaluated the effects of these altered physiological processes on plant growth and development by recording various life cycle and morphological parameters (Figure 5h; Figure S17, Table S11). Data suggested the miR158 mutant possessed elongated leaves and petioles with a larger rosette area (Figures S17 and S18) and bolted 6 days ($\pm$1 day) earlier as compared to the Col-0 (Figure 5h). Bolting was delayed by 21 days ($\pm$2 days) in overexpression lines compared with Col-0. Similarly, pseudo-*PPR* OE plants matured 11 days ($\pm$1 days) earlier than pseudo-*PPR* mutant plants (Figure 5h and Figure S18). Overall, both physiological and morphological data indicate that miR158 regulates growth and development of *A. thaliana* by regulating *NHX2* via PDPs.

DISCUSSION

Natural variations in *MIR158* affect pri-miR158 processing

We identified natural variants of *MIR158* in west Himalayan populations of *A. thaliana*. The insertion/deletion events are unique to these populations and have not been observed in any other accessions. Though both insertions and deletions were observed in these populations, notable is the discontinuous deletion with a very high frequency of occurrence (0.94) in a particular population (Mun). Subsequently, we identified some other populations besides Deh, Mun, and Chit in the region and screened *MIR158* in these new populations by PCR as well. Interestingly, we could not identify any NTL plants in any of the populations, including newly screened ones, except Chit and Deh. It is noteworthy that all these populations are from the same geographical continuity along the Himalayan elevation (700 to 2000 m amsl), except Chit, which is at the extreme elevation of approximately 3400 m amsl and lies hundreds of kilometers apart. It remains to be seen if abolition of miR158 activity in the Mun population is under purifying selection, aiding in local adaptation. Intuitively we would guess that this is indeed the case, as there is a gradual decrease in the frequency of NTL plants as one moves from lower elevation (Deh), with almost equal frequencies of NTL and ITL, to higher elevation (Mun), with a high frequency of DTL plants (0.94) and no NTL plants.

The insertions/deletions altered the secondary structures of pri-miR158 (Figure 2d–f), affecting their processing efficiency and/or accuracy by plant miRNA processing machineries, such as DCL1, HYL1, or SERRATE (Kurihara & Watanabe, 2004; Reinhart et al., 2002). Particularly, the positioning of the initial DCL1 processing event in many precursors is dependent on a lower stem of approximately 15 bp below the miRNA/miRNA* duplex followed by a large internal loop (Cuperus et al., 2010; Mateos et al., 2010; Werner et al., 2010). These critical features of secondary structures, particularly the lower stem structure and internal loop, are severely affected in DTL, while in ITL, the internal loop is absent (Figure 2d–f). Thus, the miRNA processing machineries completely or partially fail to recognize the inappropriate secondary structures (Figure 2d–f), resulting in negligible or moderate expression in DTL and ITL, respectively (Figure 2a). Earlier, a 139-bp insertion in primary transcripts of miR163 in *Arabidopsis arenosa* was shown to affect its processing, leading to reduced expression (Ng et al., 2011). Alternatively, expression of miR158 might be affected due to instability of pri-miR158 because of the insertion or deletion in ITL and DTL. However, our RNA decay assay showed almost equal amounts of pri-miR158 accumulation in NTL, ITL, and DTL but a higher level of pre-miR158 in NTL and ITL than in DTL. Thus, the possibility of infective processing of pri-miR158 leading to aberrant RNA production and subsequent degradation cannot be ruled out. Different miRNAs may be processed by different mechanisms, depending on the sequential direction of the miRNA processing machineries. miR158 is reported to be processed by a canonical base-to-loop mechanism, where the first cut is made by counting a 15-nucleotide stem region above the internal loop

(Bologna et al., 2013). Overall, the insertion/deletion results in secondary structures of pri-miR158 that are inappropriate for plant RNA processing machineries, leading to reduced or negligible expression of miR158. We further validated this result by generating two edited lines containing a one- and two-nucleotide deletion in the lower stem, which resulted in a disturbed secondary structure and suppressed miR158 expression (Figure S9a–c).

Abolished miR158 activity triggers PDP biogenesis following the 'two-hit, two-cleavage' model

Earlier it was reported that miR158 targets a *PPR* and thereby regulates pollen development in *Brassica campestris* (Ma et al., 2017). However, our repeated attempts failed to identify *PPR* as a target of miR158 in *A. thaliana*. Subsequently, based on our *in silico* target prediction results, we validated a pseudo-*PPR* as its target. Though many pseudogenes are reported to generate siRNAs in both plants and animals (Guo et al., 2009; Thibaud-Nissen et al., 2009) or act as sponges for miRNA (An et al., 2017; Yu et al., 2014), our findings that miR158 targets and cleaves a pseudogene may widen the range of biological functions of plant pseudogenes. Indeed, in absence of miR158 activity, the *TAS2*-derived siRNAs siR9 and siR12 bind to the miR158 target pseudo-*PPR* and are subsequently processed to phasiRNAs (Table 1). Thus, a large number of PDPs were generated in DTL and miR158 mutant plants as compared to NTL and ITL. Conversely, we detected low PDP levels in pseudo-*PPR* mutant and miR158 OE lines. Although it has been reported that *TAS2*-derived siRNAs act as a trigger of tasiRNA biogenesis (e.g., tasiRNA2140) from *PPR*, (Chen et al., 2007; Howell et al., 2007), there is scanty of reports showing cleavage of target by both the tasiRNAs for biogenesis of tertiary phasiRNAs. Moreover, considering the one-hit model is the most prevalent pathway for phasiRNA production and the two-hit model has been primarily reported for *TAS3*, the observed pathway of phasiRNA production implicates diversity in their biogenesis. More importantly, we observed phasiRNA production from both 5′ and 3′ cleaved ends of the target, albeit in higher number from the 5′ cleaved end than from the 3′ cleaved end. Recently, similar two-end target cleavage and phasiRNA production from downstream of both cleavage sites were reported in wild strawberry (*F. vesca*) *TAS3S* (Xia et al., 2015) and *M. truncatula* (Zhai et al., 2011). The above study together with the present report suggest that in addition to the prevalent 'two-hit, one-cleavage' configuration for tasiRNA production (Liu et al., 2020), a 'two-hit, two-cleavage' configuration may be operating as well, at least in some plant species including *A. thaliana*.

A new module of phasiRNA regulation

Our observation that pseudogenes act as templates for phasiRNA biogenesis implicates a larger role for pseudogenes in plants. Firstly, pseudogenes are known to function as templates for biogenesis of 24-nucleotide siRNAs, which are primarily involved in RNA-directed DNA methylation (Guo et al., 2009). In contrast, we observed a large number of 21-nucleotide PDPs which were associated with AGO1 (Table S8), indicating probable involvement in cleavage. Secondly, to the best of our knowledge, pseudogene-derived siRNAs targeting protein-coding genes have not been previously reported in plants. Here, we show that one of the PDP targets is *NHX2*, a member of the *NHX* gene family involved in vacuolar K^+ homeostasis (Barragán et al., 2012). Further, in other organisms, most of the pseudogene-derived siRNAs are known to target their parent protein-coding genes (Tam et al., 2008; Watanabe et al., 2008) rather than non-parental genes. The PDPs neither showed significant similarity with parent *PPR* (AT1G64100) nor could any *PPR* target be identified in degradome data. The targeting of *NHX2* by one of the PDPs also contradicts an earlier report that argued against secondary targeting by clad-derived siRNAs on non-clad target transcripts, especially in the case of *PPR*-derived phasiRNAs (Howell et al., 2007). It is noteworthy that phasiRNA production is restricted to P-type *PPRs* in Arabidopsis (Howell et al., 2007; Xia et al., 2013) and phylogenetic analysis of the template pseudo-*PPR* shows that it belongs to the P-type *PPR* clad (Figure S10a). The secondary siRNA biogenesis pathway involving *PPRs*, that is, miRNA–*TAS*-like gene–*PPR*–siRNA module (Xia et al., 2013), and the other most common pathway, that is, miRNA–*TAS*–siRNA pathway are mainly triggered by 22-nucleotide siRNAs (a few are 21 nucleotides, e.g., miR156 and miR172) following the 'two-hit, one-cleavage' model (Zhai et al., 2011). We demonstrate here another layer of siRNA guided gene regulation following a miRNA–*TAS*–siRNA–pseudo-*PPR*–tertiary phasiRNA–*NHX2* module. We term this module a 'three-hit, three-cleavage' model, where two secondary siRNAs of 22 and 21 nucleotides act as a trigger and the 20-nucleotide miR158 acts as a negative regulator (Figure 6). However, the dominant role of miR158-triggered cleavage of pseudo-*PPR* and subsequent degradation over siR9 and siR12 cleavage and PDP generation is very interesting. Intuitively though, it suggests that the 20-nucleotide miR158 may facilitate target cleavage and degradation over the 22-nucleotide siR9 or the 21-nucleotide siR12, considering that there is not much difference in the complementarity of these sRNAs with the target. Further, it is well reported that phasiRNA initiators are generally 22-nucleotide sRNAs. Thus, it is to be explored if a 21- or 22-nucleotide miR158 acts as a phasiRNA initiator rather than a target degrader.

PDP-mediated downregulation of *NHX2* affects physiological processes in *A. thaliana*

The downregulation of *NHX2* by a PDP (Figure 5b) led to reduced transpiration, reduced stomatal conductance, and better WUE (Figure 5c–g). The delayed response of
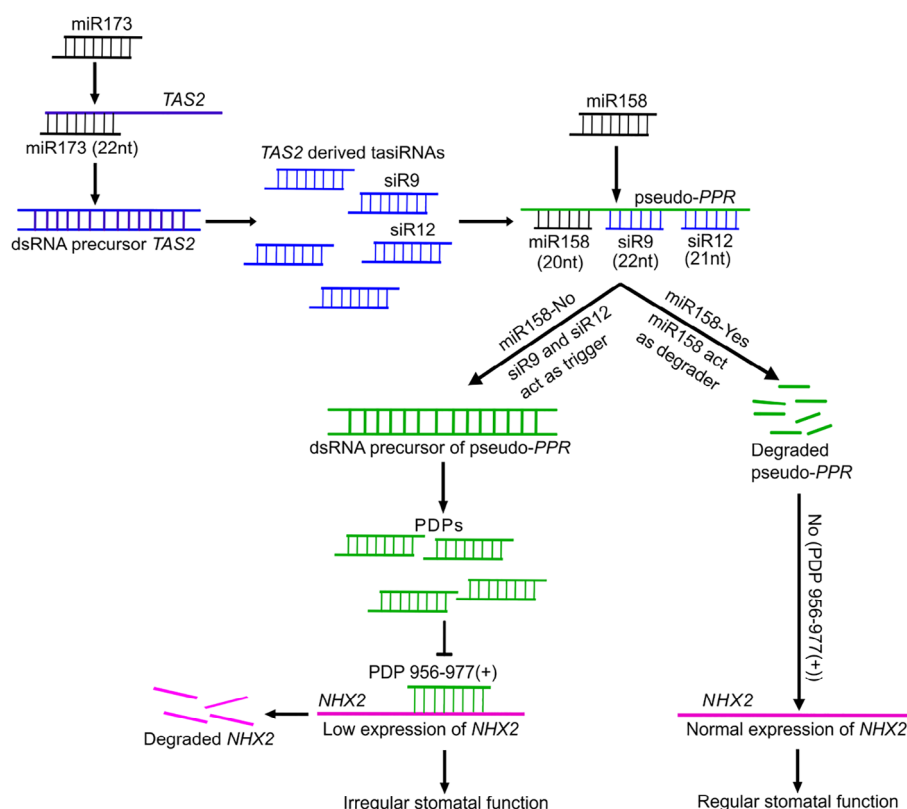
Figure 6. The 'three-hit three-cleavage' model of the function of miR158.
The model suggests that miR158 acts as a negative regulator of phasiRNA biogenesis from a pseudo-*PPR*. The pseudo-*PPR* is targeted by miR158 besides being targeted by two tasiRNAs, siR9 and siR12, derived from the *TAS2* locus. While miR158 activity leads to degradation of the pseudo-*PPR* and thus there is no biogenesis of phasiRNAs, null activity of miR158 makes the pseudo-*PPR* template available for targeting by the two tasiRNAs, leading to biogenesis of phasiRNAs. One of the phasiRNAs produced by this mechanism targets the *NHX2* gene for degradation, which in turn regulates stomatal functioning.

stomatal functioning has been attributed to NHX proteins which help to maintain turgor by regulating K$^+$ homeostasis and it agrees with the observed diurnal variations in *NHX2* expression in different lines (Figure S14b–d). Both opening and closure of stomata were reported to be affected in a *nhx1 nhx2* double knockout mutant in *A. thaliana*, and the *nhx2* single mutant was more affected in stomatal opening than closure (Andrés et al., 2014). It is evident that the stomatal opening was more impaired than closure in the miR158 mutant as compared to other lines, as the rates of transpiration and stomatal conductance did not fall rapidly as others as dusk approached. In fact, the response to stomatal closure was delayed showing higher rates at dusk (8 h) and even at dark (150 min). The reduced transpiration could not be attributed to the defective guard cells (Figure 5e). Notably, the miR158 mutant and to a lesser extent pseudo-*PPR* OE lines completed their life cycle earlier than control both under drought stress (Figure 5g,h) and under regular water supply conditions. However, in contrast to an earlier report where a double *nhx1 nhx2* knockout mutant produced phenotypes but not *nhx2*, we observed phenotypes with downregulated *NHX2*.

Notably, in an earlier report, knockout of *NHX2* led to upregulation of *NHX1* but not *vice versa* (Andrés et al., 2014). However, in our study we did not observe significant upregulation of *NHX1* with concomitant downregulation of *NHX2* at any time point (Figure S15b,c). As reported earlier (El-Brolosy & Stainier, 2017), this may be because genetic compensation by family members is different in knockout and knockdown (this study) mutants. Nevertheless, the overall performance of pseudo-*PPR* OE is in line with the observed phasiRNA generation and the probable stoichiometric ratio between pseudo-*PPR* and miR158 in favor of pseudo-*PPR* in this line. Thus, negligible expression of miR158 in a particular population (Mun, DTL) led to activation of an sRNA biogenesis cascade, ultimately leading to shortening of the life cycle of the plants. This points towards selective adaptation strategies of the population existing at high elevation (2000 m amsl) that might help to escape the onset of harsh winter and relatively dry conditions in elevated regions. However, it will be interesting to examine the evolutionary significance of this local adaptation strategy. Although favorable, it is absent in the Chit population, a site situated at even higher elevation

(approximately 3500 m amsl) than the Mun population (approximately 2000 m amsl). Notably, the Mun population spreads over a much larger area than Chit, which is found only in a small patch.

In summary, natural variants of *MIR158* led us to identify a new regulatory module of 21-nucleotide phasiRNA biogenesis. The identified pathway widens the scope of phasiRNA biogenesis and regulation beyond prevalent pathways. miRNA-driven cleavage of a pseudogene or regulation of a protein-coding gene by a pseudogene-derived phasiRNA highlights the significance of pseudogenes as well as tertiary phasiRNAs in plants. Understanding the evolutionary significance of such a pathway may shed light on siRNA-regulated local adaptation.

EXPERIMENTAL PROCEDURES

Plant materials and growth conditions

The seeds of different west Himalayan *A. thaliana* populations (Deh, Mun, Chit) and SALK mutant lines (miR158 mutant, SALK_025691; pseudo-*PPR* mutant, SALK_043644; *nhx2*, SALK_084844c) and Col-0 were grown under controlled conditions (temperature 22°C, light intensity 140 μmol m^{-2} sec^{-1}, and a 16/8-h day/night cycle). Besides, seeds were also grown in a common garden at CSIR-NBRI, Lucknow, Uttar Pradesh, India during the months of November to February to check the influence of environment on the expression of miR158. The homozygous lines of SALK mutants were selected as described in GABI (http://signal.salk.edu/tdnaprimers.2.html) using primers listed in Table S12.

RNA isolation, sRNA sequencing, and analysis

Total RNA was isolated from rosette leaves of 28-day-old plants using the mirVana™ miRNA isolation kit (Ambion, Austin, TX, USA) following the manufacturer's instructions. The quality of RNA was checked using both agarose gel electrophoresis and a bioanalyzer (Agilent, Santa Clara, CA, USA). RNA samples having RIN > 7 were used for sRNA library preparation. Four pooled sRNA libraries corresponding to Deh, Mun, Chit, and Col-0 were prepared and used as replicates of earlier sequenced data from the same populations (Tripathi et al., 2019). Further, for each plant type individual libraries were prepared and sequenced in replicates, except for DTL, miR158 OE, and pseudo-*PPR* OE (one of these libraries failed the quality control). Libraries were prepared using the NEB Next Small RNA Library Prep Set (New England Biolabs, Ipswich, MA, USA) and sequencing was carried out using an Illumina HiSeq 4000 platform in single-end mode by out sourcing (Agrigenome, Hyderabad, India). Thus, sequencing data from a total of 20 sRNA libraries were analyzed (Table S13). The raw sequences were analyzed using a UEA sRNA workbench (Version 4.5) (Stocks et al., 2018). Briefly, the adapters were removed from raw reads and size-filtered between 18–28 nucleotides using an adapter removal tool. The filtered reads were mapped onto the reference Arabidopsis TAIR10 genome using Bowtie2 with default parameters. The genome mapped reads were used for identification and quantification of miRNAs and tasiRNAs by using miRProf implemented in the UEA workbench with miRBase22 (http://www.mirbase.org/) and the tasiRNA database (http://bioinfo.jit.edu.cn/tasiRNADatabase) (Zhang et al., 2014) as background, respectively. All sRNA expression values (transcripts per million [TPM]) were normalized using transcripts that mapped on the reference Col-0 genome.

PhasiRNA identification and scoring

For identification of PDPs, the filtered sRNA reads were mapped onto the pseudo-*PPR* reference sequence using Bowtie 2.0 (zero mismatch) (Langmead & Salzberg, 2012). It is to be noted that there is a perfectly complementary region of 23 bp between *TAS2* (near the siR9 production site) and pseudo-*PPR* (near the siR9 binding site) (Figure S19a). Therefore, in order to exclude any ambiguous mapping of siR9 (higher expression as compared to PDPs) on pseudo-*PPR*, we masked (replaced by N) half of the complementary sites of pseudo-*PPR* (Figure 19b). This led us to specifically identify the PDPs.

To test whether this pseudo-*PPR* is a true *PHAS* locus, the *P*-value was calculated with the help of the tasi predictor tool implemented in the UEA sRNA Workbench using the pseudo-*PPR* sequence as a reference (Chen et al., 2007). Further, the phasing score was calculated using the 21-nucleotide phasing register in a 10-cycle window with $P > 0.001$. The phasing score was calculated following (De Paoli et al., 2009) as follows:

$$\text{phasing score} = ln\left[\left(1 + 10 \times \frac{\sum_{i=1}^{10} Pi}{1 + \sum U}\right)^{n-2}\right],$$

where *n* is the number of phase registers occupied by at least one unique 21-nucleotide sRNA within a 10-phase register window, *P* is the total number of reads for all 21-nucleotide sRNA reads falling into a given phase in a given window, and *U* is the total number of unique reads for all 21-nucleotide sRNAs falling out of a given phase.

Promoter analysis and identification of polymorphisms in *MIR158*

The putative promoter sequences of *MIR158* (1 kb upstream of the TSS) of Deh, Mun, and Chit were sequenced (GenBank# OQ267618–OQ267620) and checked for any sequence variations using the multiple sequence alignment tool implemented in MEGA 6.0 (Tamura et al., 2013).

In order to identify the polymorphisms, if any, the miR158 locus was amplified covering approximately 165 bp upstream of the TSS and approximately 165 bp downstream of the TTS of the primary transcript (using primer pair 1F/1R; Table S12 and Figure 4a). Initially, genomic DNA was extracted from a few individuals of the three populations and the locus was amplified and visualized on a 2% agarose gel. Thereafter, we screened approximately 50 individual plants from each population using the same set of primers (1F/1R). Subsequently the PCR products from a few individuals were sequenced and aligned using the multiple sequence alignment tool to identify the polymorphisms. To confirm the length polymorphisms at the transcript level, we amplified the pri-miR158 locus using cDNA as a template from a few individuals of the three populations. Since there were polymorphisms in the downstream of TSSs in each population, we designed NTL-, ITL-, and DTL-specific forward primers (2F) and a common reverse primer (2R) (Table S12 and Figure S4) to amplify the transcript. Subsequently, the PCR products were sequenced and aligned to observe polymorphisms.

Further, the miR158 loci of west Himalayan populations were compared with 1135 *A. thaliana* accessions as well using the POLYMORPH 1001 tool (https://tools.1001genomes.org/polymorph/) selecting the nucleotide position range 3 366 553–3 366 019 on chromosome 3 (range of the miR158a locus).

Secondary structure prediction of pri-miR158

The primary transcripts of miR158 amplified from each of the three populations were sequenced and the secondary structures of these primary transcripts were predicted using the RNAfold web server (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) with default parameters.

Target prediction and degradome analysis

Targets of miRNA, tasiRNA, and phasiRNA were predicted by psRNA target (Dai & Zhao, 2011) using an expectation value cutoff of <3. Available degradome data of *A. thaliana* (GSE55151) (Thatcher et al., 2015) were downloaded and analyzed using PAR-Esnip (Folkes et al., 2012) following stringent criteria. Briefly, mismatches on positions 10 and 11 were not allowed, the maximal number of mismatches was set to four, and two mismatches on adjacent positions were not allowed. Based on the degradome fragment abundance mapped on the target gene, PAREsnip categorizes miRNA targets into five categories, category 0 to category 4, according to signal strength and read abundance at the position. Category 0 indicates the highest confidence of target cleavage, and category 4 indicates the lowest. The targets with lower confidence like categories 4 and 3 were not considered in this study. The T-plots were generated by the VisSR tool (Stocks et al., 2018).

Target validation by modified 5′ RLM-RACE

To validate the putative targets of miRNA and phasiRNA, modified 5′RLM -RACE was carried out using the GeneRacer™ Kit (Ambion, Austin, TX, USA) following the user's manual. Briefly, RNA was isolated using the RNeasy Mini Kit (Qiagen, Hilden, Germany). Adapter was ligated to 4 μg freshly isolated RNA at 37°C for 1 h, followed by 16°C for 8 h. cDNA was synthesized using random primers with ligated RNA. Touchdown PCR amplification was performed using the adapter-specific GeneRacer 5′ primer and gene-specific outer primers with the following PCR program: 2 min at 94°C, followed by five cycles of 30 sec at 94°C and 90 sec at 72°C, five cycles of 30 sec at 94°C and 120 sec at 70°C, and 25 cycles of 30 sec at 94°C, 30 sec at 65°C, and 90 sec at 72°C, and a final extension step of 10 min at 72°C. Nested PCR amplification was performed using the adapter-specific GeneRacer 5′ nested primer and gene-specific inner primers and 2 μl template from the initial touchdown PCR with the following PCR program: 2 min at 94°C, followed by 25 cycles of 30 sec at 94°C, 30 sec at 65°C, and 2 min at 72°C, and a final extension step of 10 min at 72°C. The PCR product was run on a 2% agarose gel and the expected bands were eluted using the Wizard® SV Gel and PCR cleanup kit (Promega, Madison, WI, USA) and cloned in the pGEM-T easy vector (Promega), and five individual clones were sequenced for each target.

Expression analysis of miR158, its targets, and other sRNAs by qRT-PCR

Expression levels of miR158 primary transcripts, precursors, its target genes, *TAS2*, and *NHX2* were analyzed by qRT-PCR. cDNAs were prepared by the GoScript™ Reverse Transcription System following the manufacturer's protocol (Promega). All qRT-PCRs were performed in an ABI 7300 real-time PCR machine (Applied Biosystem, Foster City, CA, USA) using DyNAmo Flash SYBR Green (Thermo Fisher Scientific, Waltham, MA, USA) with the following cycling conditions: denaturation at 95°C for 10 min, followed by 40 cycles of denaturation at 95°C for 20 sec and annealing and extension together at 60°C for 60 sec. For quantification of pre-miR158 and pri-miR158, primer pairs 3F/3R and 4F/4R were used

(Figure S4a). Similarly, quantification of miR158, miR173, siR9, and siR12 was also carried out with specific primer sets. All primer sequences are listed in Table S12. Briefly, 100 ng of total RNA was reverse transcribed using respective stem-loop primers and Super-Script® III Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA) following the manufacturer's instructions. PCR was performed using 2 μl of cDNA as template and miR158-specific forward primer and stem-loop-specific reverse primer using the following thermocycle conditions: 94°C for 2 min, followed by 35 cycles of 94°C for 15 sec and 60°C for 1 min. The PCR product was checked on a 2% agarose gel. All reactions were performed with three to five biological and two technical replicates. 5s rRNA and the *actin* gene were used as endogenous controls for stem-loop PCR and target genes, respectively (Czechowski et al., 2005; May et al., 2013). The expression value of each sample was normalized using the expression value of the respective endogenous control and relative expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method (Schmittgen & Livak, 2008). Before conducting stem-loop qRT-PCR, the specificity of miR158 stem-loop primers was checked using end point PCR following (Varkonyi-Gasic et al., 2007) (Figure S20). The variation in expression of the primary transcripts of miR158a and miR158b was taken as a proxy to distinguish between the expression of miR158a and miR158b, because the precursor and mature sequences of miR158a and miR158b do not harbor sufficient variations (Figure S1b).

RNA decay assay

The RNA decay assay was performed as described in (Zhao et al., 2018). ITL, DTL, and NTL plants were grown for 2 weeks after 4 days of stratification of seeds at 4°C. Subsequently plants were transferred into half-strength MS medium with 100 μg ml⁻¹ actinomycin D (Sigma-Aldrich, Saint Louis, MO, USA). Whole seedlings were harvested after 1, 2, 4 and 6 h of treatment. Total RNA was extracted from these seedlings by TRIzol reagent and used for RT-qPCR assays.

Transient expression analysis by agro-infiltration

Primary transcripts of miR158 in ITL, DTL, and NTL were amplified by PCR and cloned in the pBI121 binary vector. The constructs were transformed into *Agrobacterium tumefaciens* GV3101 following (Kościańska et al., 2005) with minor modifications. Briefly, *Agrobacterium* was grown overnight in LB medium with rifampicin (25 μg ml⁻¹), gentamycin (30 μg ml⁻¹), and kanamycin (50 μg ml⁻¹). Bacterial cells were briefly spun down and washed with sterile water. The pellet was re-suspended in 10 mm MES, pH 5.6, 200 μm acetosyringone, 10 mm MgCl₂, and 3% sucrose solution to an OD of about 0.5 and incubated for 2 h at 28°C. Infiltration of bacterial cells into healthy leaves of 28-day-old *N. benthamiana* plants grown under controlled conditions of 24°C and a 16-h photoperiod was performed with a 1-ml syringe. At 72 h post-infiltration, total RNA was isolated from the infiltrated leaves using the mirVana™ miRNA isolation kit (Ambion) following the manufacturer's instructions. The expression of pri-miR158 and miR158 was checked using qRT-PCR analysis as described above.

sRNA gel blot analysis

Total RNA was extracted using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) as per the manufacturer's instructions. The seedling samples were pulverized in liquid nitrogen with chilled mortar and pestle. Isolated RNA was quantified using a Qubit Fluorimeter with RNA BR dye (Thermo Fisher Scientific). sRNA Northern hybridization was performed as described previously (Shivaprasad et al., 2012; Tirumalai et al., 2020). Roughly 10 μg of total RNA was electrophoresed at 80 V for 3 h in a 15% acrylamide gel (with 8 m urea

and acrylamide:bisacrylamide [19:1]). The gel was electroblotted onto a Hybond N+ membrane (GE Healthcare, Chicago, IL, USA) at 10 V overnight at 4°C. The RNA was cross-linked to the membrane with a UV crosslinker (UVP) and hybridization with the labeled probes was performed using UltraHyb-Oligo buffer (Ambion) at 35°C for 12 h. The probes were made by end labeling the oligos (Table S12) with T4 PNK (NEB, M0201) as per the manufacturer's instructions using gamma ^{32}P labeled ATP (BRIT, Mumbai, India). Labeled reaction mixture was purified using Illustra Microspin G-25 columns (GE Healthcare) and added to the hybridization buffer. Post-hybridization, the blot was washed twice with 2× SSC, 0.5% SDS buffer for 30 min each at 35°C. The blots were exposed to phosphor screen (GE Healthcare) and imaged on a Typhoon molecular imager (GE Healthcare). The membranes were stripped at 80°C, followed by repeated hybridization with other probes as described before (Tirumalai et al., 2020).

Phylogenetic analysis of *PPRs*

In total, 297 *PPRs* having conserved *PPR*-domains among 450 reported *A. thaliana PPRs* (Lurin et al., 2004) were considered for phylogenetic analysis. The sequences of the *PPRs* were downloaded from TAIR (www.arabidopsis.org) and aligned using CLUSTAL-W implemented in MEGA 6.0 (Tamura et al., 2013). The phylogenetic tree was constructed using the neighbor-joining method implemented in MEGA 6.0 with 1000 bootstrap replicates.

Cloning, generation of overexpression and edited transgenic plants

Pri-miR158 and pseudo-*PPR* fragments were amplified by PCR using respective gene-specific primers (Table S12) following standard PCR protocols. The amplified fragments were cloned in the pGEM-T cloning vector (Promega) and subsequently transferred into the pBI121 binary vector. *Arabidopsis thaliana* plants (Col-0 background) were transformed by the floral dip method. Transformants were screened on half-strength MS agar plates containing kanamycin (50 μg ml^{-1}) followed by PCR screening. Subsequently, T3 homozygous plants were selected for further analyses. In spite of repeated attempts, we failed to generate a complementation line of miR158 in DTL plants due to unknown reasons.

To generate CRISPR–Cas9-derived miR158 edited lines, a gRNA specific for miR158 (Table S12) was cloned into the binary vector pHEE401 (Wang, Xing, et al., 2015) using the *Bsa*I restriction site. The gRNA was designed using the CRISPR Arizona Tool (http://www.genome.arizona.edu/crispr). The constructs were further confirmed by sequencing and then transformed. The transformants were screened on half-strength MS agar plates containing hygromycin (20 μg ml^{-1}). The putative transgenic lines were screened for target edition by a T7 endonuclease assay using Guide-it Mutation Detection Kit following the manufacturer's instructions (Takara, Shiga, Japan). The edited lines were finally confirmed by sequencing of the target locus.

Measurement of morphological and physiological parameters

Various morphological and life cycle parameters of 10- and 5-week-old miR158 mutants, pseudo-*PPR* mutants, transgenic lines, and Col-0 plants were measured. Leaf area was estimated as the product of length and width. Leaf shape was represented as the ratio of length to width. Rosette area was measured as the product of the major and minor axes and rosette shape as the ratio of major to minor axis. Both leaf and rosette shape were considered as rounded when the ratio was close to 1 and elongated if the

ratio was more than 1. Life cycle parameters, namely, bolting, first flower opening, emergence of first silique, and senescence, were measured in terms of days after germination. Plants were monitored regularly and data were recorded every two days.

Stomatal conductance (g_s), the rate of transpiration (E), and WUE (as a ratio of photosynthetic rate, P_N to E) of plants at the bolting stage (principal growth phase 5.10; Boyes et al., 2001) were measured using a LICOR-6400 portable photosynthesis system (LI-COR, Lincoln, NE, USA) at four different time points: immediately after onset of light, after 4 and 8 h of exposure to light, and after 150 min of exposure to dark. Stomatal aperture was measured by peeling the epidermal layer of the leaf lower surface and mounting in glycerol and visualized under a LEICA DM2500 light microscope at 100×. Stomatal measurements were taken at two time points: after 4 h of exposure to light and after 150 min of exposure to dark. Additionally, the stomata from these two time points were also visualized and measured using a scanning electron microscope (FEI Quanta 250, Thermo Fisher Scientific). The leaf tissues were fixed and processed according to (Bomblies et al., 2008). Briefly, the tissue was fixed overnight at 4°C in FAA fixative (3.7% formaldehyde, 50% ethanol, 5% acetic acid) and then dehydrated serially in ethanol solutions of 30%, 50%, 70%, 90%, 95%, and 100% for about 1 h each. The tissue was kept at 4°C overnight in 100% ethanol and then processed for critical-point drying followed by gold coating and visualization at 8000X.

Drought assay

Ten plants each of the miR158 mutant, the miR158 OE line, the pseudo-*PPR* mutant, the pseudo-*PPR* OE line, *nhx2*, and Col-0 along with NTL, ITL, and DTL plants were grown under controlled conditions for 3 weeks with normal watering. On the last day of the third week, pots were saturated with water and soil surface was covered with plastic wrap. The plants were allowed to grow without additional water and monitored until they no longer survived. The photographs were taken after 35 days of water withholding.

Relative water content

The relative water content and water loss were measured following (Barrs & Weatherley, 1962). Briefly, rosette leaves were harvested and soaked in water for 30 min to make them completely turgid. The turgid weight of each leaf was recorded and then leaves were kept at room temperature for varying intervals of time. The weight of each leaf was measured after 5 min, 15 min, 30 min, 60 min, 105 min, and 165 min. After 165 min, the leaves were kept at 65°C in a dry oven for 72 h and then the dry weight was measured. The relative water content was calculated using the following formula:

$$\text{RWC} (\%) = [(\text{FW}-\text{DW})/(\text{TW}-\text{DW})] \times 100\%,$$

where RWC is the relative water content, FW is the fresh weight at different time points, DW is the dry weight, and TW is the turgid weight.

All morphological and physiological data were analyzed using one-way ANOVA in R. The multiple comparison analysis of the linear ANOVA model was conducted using the 'glht' function in the multcomp package in R.

ACCESSION NUMBERS AND DATA AVAILABILITY

All sRNA sequencing data generated in this study have been deposited in National Center for Biotechnology Information (NCBI) under BioProject PRJNA674752 with Sequence Read

Archive (SRA) accession# SRR12990663–SRR129990678. Whole genome sequence data of Deh (SRR14075201) and Chit (SRR14075202) and other relevant gene sequences reported here (GenBank# OQ267609–OQ267620) are available at NCBI. The Arabidopsis Genome Initiative locus identifiers for the genes mentioned in this article are *MIR158* (AT3G10745), pseudo-*PPR* (AT1G62860), *TAS2* (AT2G39681), *NHX2* (AT3G05030), and *NHX1* (AT5G27150). The T-DNA insertion mutants used included miR158 (SALK_025691), *nhx2* (SALK_084844c), and pseudo-*PPR* (SALK_043644).

AUTHOR CONTRIBUTIONS

AMT conducted experiments, analyzed sequence data, and drafted the manuscript; RS, AKV, PM, and VD conducted experiments; AKV, RS, and AS analyzed the data; SN and PAS conducted physiological experiments; PVS and VHS carried out Northern blot experiments; SR designed experiments, analyzed the data, and wrote the manuscript.

ACKNOWLEDGMENTS

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Sequence conservation of pre-miR158.

(a) Multiple sequence alignment of pre-miR158 from different members of the Brassicaceae family. The conserved nucleotides are shown in red and also shown are the regions of 5p and 3p arms with respective lower stem.

(b) Sequence alignment of precursor sequences of miR158a and miR158b in *A. thaliana*.

Figure S2. Relative expression of miR158 in different populations as determined by stem-loop qRT-PCR.

(a, b) Expression pattern of miR158 under (a) native field (west Himalayas) and (b) common garden conditions. The data are presented as the mean ± SD of five independent biological replicates.

(c, d) Comparative expression patterns of (c) miR158a and miR158b and (d) 158a-3p and 158a-5p. The data are presented as the mean ± SD of three independent biological replicates. The expression data were normalized against 5s rRNA expression, and relative expression was calculated using the $2^{-\Delta\Delta Ct}$ method.

Figure S3. Multiple sequence alignment of miR158 promoter sequences from different populations (Deh, Mun, Chit).

The 1-kb sequence upstream of the transcription starts site (TSS) of miR158 was extracted from the previously sequenced genomes of the three Himalayan populations (unpublished) as putative promoter regions. The gray shaded area indicates the conserved region. The TSS is indicated by an arrow. The miR158 gene sequences downstream of the TSS showing polymorphisms are shown in yellow.

Figure S4. Schematic diagram of the miR158 locus and its length polymorphisms in west Himalayan Arabidopsis populations.

(a) Schematic representation of the miR158 locus and positions of different primers used to amplify its genomic locus and for quantification of expression levels. The primer sequences are listed in Table S12.

(b–d) Variations in the size of the miR158 locus in different individuals of (b) Deh, (c) Mun, and (d) Chit populations. The first lane (marked with cyan) represents the size of the miR158 locus in Col-0 and the normal type locus (NTL); the inserted type locus (ITL) and deleted type locus (DTL) are indicated by arrows.

Figure S5. Multiple sequence alignment of the miR158 genomic locus showing length polymorphisms in different populations.

The miR158 loci of DTL, ITL, and NTL were amplified using genomic DNA as a template and primer pair 1F/1R and then sequenced (Table S12). The insertion sequence in ITL and the deleted sequence in DTL are shown in green and yellow boxes, respectively. The precursor and mature sequences of miR158 are shown in magenta and cyan boxes, respectively. The conserved regions are shaded in gray.

Figure S6. Length polymorphism in the pri-miR158 locus in NTL, DTL, and ITL plants.

Multiple sequence alignment of pri-miR158. Pri-miR158 of DTL, ITL, and NTL was amplified using cDNA as a template and a DTL-, ITL-, or NTL-specific forward primer (2F) and a common reverse primer (2R) (Figure S4 and Table S12). The insertion sequence in ITL and the deleted sequence in DTL are shown in green and yellow boxes, respectively. The precursor and mature sequences of miR158 are shown in magenta and cyan boxes, respectively. The conserved regions are shaded in gray.

Figure S7. Predicted secondary structures of miR158 precursors (a–c). The secondary structure of miR158 precursors in (a) NTL, (b) DTL, and (c) ITL. The miRNA and miRNA* sequences are shown in square brackets. The precursor sequence was annotated using the pre-miR158 sequence reported in miRBase21.

Figure S8. RNA decay assay of pri-miR158 in DTL, ITL, and NTL plants and transient expression of pri-miR158 and miR158 in *Nicotiana benthamiana*.

(a) Remaining levels of pri-miR158 at different time points post-actinomycin D (100 μg ml^{-1}) treatment in DTL, ITL, and NTL plants as determined by qRT-PCR using pri-miR158-specific primers. For transient expression analysis, constructs of the pri-miR158 locus of DTL, ITL, and NTL were transiently expressed. (b, c) Relative expression patterns of (b) primary miR158 and (c) mature miR158. The data are presented as the mean ± SD of three independent biological replicates. $***P < 0.001$, $**P \leq 0.01$, $*P \leq 0.05$ (Student's *t*-test); ns, not significant. EV, pBI121 empty vector negative control.

Figure S9. Effects of nucleotide deletion using CRISPR–Cas9 on miR158's lower stem.

(a) Two CRISPR–Cas9 edited lines in which the lower stem of miR158 was targeted were generated (miR158_CR_N-1 and miR158_CR_N-2). Sequencing of the targeted regions indicated one- and two-nucleotide deletions in the lower stem regions. The nucleotide sequences in magenta boxes represent miR158 precursors, yellow highlighted sequences represent mature miR158, arrows represent edited sequences, and the red color sequence represents the PAM region.

(b) Expression of miR158 in the two edited lines as determined by stem-loop qRT-PCR. The expression is severely affected in both edited lines.

(c) The predicted secondary structure of pri-miR158 in edited lines. The nucleotide deletion in the lower stem of miR158 pri-transcript generated disturbed secondary structures as compared to the Col-0 sequence. The yellow region indicates the precursor.

Figure S10. Phylogeny of pseudo-*PPR* along with other *PPR* genes and their protein coding potentials.

(a) Unrooted neighbor-joining tree of *PPRs* reported in *A. thaliana* generated using MEGA6.0 with default parameters.

(b) Protein coding sequences of the pseudo-*PPR* gene and its parent *PPR* as obtained using the Expasy web server with default parameters. Protein coding sequences are represented in magenta and stop codons are indicated in yellow.

Figure S11. Mapping of phasiRNA (PDPs) in different populations.

(a, b) The small RNA reads mapped onto pseudo-*PPR* from different populations (a) Deh, Mun, Chit, and Col-0 and (b) DTL, ITL, and NTL are shown in different tracks in Integrative Genomics Viewer. Red and blue blocks represent reads mapped in forward and reverse orientation, respectively.

(c) Bar plot representing length distribution of the PDPs. The first nucleotide of the PDPs of different categories is depicted with different colors. PDPs with 5′-uridine with a length of 21 nucleotides were predominant.

Figure S12. Expression of PDPs using $\log_2$ values of read counts.

(a–c) Heatmaps indicating differential expression of PDPs in (a) different west Himalayan populations ($n = 452$), (b) Col-0, pseudo-*PPR* mutant, and miR158 mutant ($n = 535$), and (c) Col-0, miR158 OE, and pseudo-*PPR* OE ($n = 1235$). The $\log_2$ values of average read counts (TPM) are represented by a color intensity scale.

Figure S13. Expression analysis of miR173, siR9, siR12, and *TAS2* and polymorphism analysis of pri-miR173 and *TAS2* in DTL, ITL, and NTL.

(a–c) Bar plots indicating expression of (a) miR173, (b) siR9, and (c) siR12. The left panel data indicate expression patterns as quantified from small RNA sequencing data and the corresponding right panel shows expression levels as quantified by stem-loop qRT-PCR.

(d) Expression of *TAS2* as determined by qRT-PCR.

(e) Northern blot of miR173 and siR9.

(f, g) Agarose gel image of PCR amplified fragments of (f) miR173 and (g) *TAS2* loci in different individuals. PCR was carried out to identify the length polymorphisms, if any, against reference Col-0.

(h, i) Multiple sequence alignment of (h) miR173 primary transcripts (i) *TAS2* loci as extracted from whole genome sequences of DTL, ITL, and NTL (unpublished).

Figure S14. Target plot (T-plot) of *NHX2*.

PDP956-977(+) and its variant PDP963-983(+) cleaved transcripts are represented by orange and red squares. The blue circles represent the fragments mapped on *NHX2*. The x- and y-axes represent nucleotide position on the target transcript and read abundance of cleaved transcript detected in degradome-seq, respectively. mRNA:phasiRNA alignments along with P-value, alignment score, and cleavage positions are also shown. Category 0 is defined as signals having more than one raw read mapped at the position, the abundance at the position equals the maximum on the transcript, and there is only one maximum. Degradome category 2 is defined as having more than one raw read at the position, with abundance at the position less than the maximum but higher than the median for the transcript.

Figure S15. Relative expression of *NHX2*, *NHX1*, pseudo-*PPR*, and miR158 as determined using qRT-PCR.

(a) Expression pattern of *NHX2* in west Himalayan Arabidopsis populations.

(b, c) Diurnal variation in relative expression of (b)*NHX1* and (c) *NHX2* in Col-0, miR158 mutant, and miR158 OE.

(d) Diurnal variation in relative expression patterns of miR158, pseudo-*PPR*, *NHX2*, and *NHX1* in Col-0. All data are presented as the mean ± SD of three independent biological replicates.

Figure S16. Physiological parameters of different plant samples.

(a) Stomatal aperture of mutant and transgenic lines in Col-0 measured under light and 150 min of dark conditions. Data are presented as the mean ± SD. Asterisks above the bars indicate statistical significance as determined using Students *t*-test.

(b) Relative water content measured at different time points (inset bar plot representing total loss of water content, calculated as initial RWC [5 min] – final RWC [165 min]). Asterisks above the graph indicate statistical significance, as determined by ANOVA (***$P < 0.001$). (c–e) Physiological parameters of NTL, ITL, and DTL plants measured at different time points. (c) Stomatal conductance. (d) Transpiration rate. (e) Water use efficiency. The data are presented as the mean ± SD of five biological replicates.

(f) Stomatal aperture of NTL, ITL, and DTL measured under light and 150 min of dark conditions. Data are presented as the mean ± SD. Asterisks above the bars indicate statistical significance as determined using Students *t*-test.

(g) Relative water content measured at different time points (inset bar plot representing total loss of water content, calculated as initial RWC [5 min] – final RWC [165 min]). Asterisks above the graph indicate statistical significance as determined using ANOVA (***$P < 0.001$).

(h) Image showing DTL, ITL, and NTL plants after 35 days of water withholding.

Figure S17. Growth and leaf size of different transgenic and west Himalayan populations.

(a) Images of miR158 and pseudo-*PPR* mutant plants and their over expression lines in a Col-0 background. Images were taken 45 days post-germination.

(b) Serially arranged leaves of mutants and transgenic lines in a Col-0 background to depict variations in leaf morphology. Images were taken 30 days post-germination.

(c) Images of NTL, ITL, and DTL plants at 30 days post-germination.

(d) Serially arranged leaves of NTL, ITL, and DTL plants to depict variations in leaf morphology. Images were taken 30 days post-germination.

Figure S18. Morphological variations in different transgenic lines and Col-0.

(a–c) Bar plots representing morphological variations in mutant and transgenic lines in a Col-0 background. (a) Leaf count at two stages. (b) Rosette area. (c) Other leaf traits. Data are presented as mean ± SD. The significance value is shown above the bars (***$P < 0.001$, **$P < 0.01$, *$P < 0.05$; ns, not significant).

Figure S19. Masked sequence of pseudo-*PPR* used for identifying pseudo-*PPR*-derived phasiRNAs (PDPs).

(a) Complementarity (23 bp) between pseudo-*PPR* and the *TAS2* gene.

(b) Masked reference sequence of pseudo-*PPR* used for identifying PDPs. The 23-bp complementary region between *TAS2* and pseudo-*PPR* was masked by replacing the nucleotides lying in the complementary region with N (shown as bold and underlined).

Figure S20. Gel image of endpoint PCR in the three populations.
Gel image showing the endpoint PCR results for miR158. This was used to check the specificity of the designed stem-loop primer of miR158. Equal amounts of amplified product were loaded onto a 2% agarose gel. The presence of single band in each PCR product confirms the specificity of the primers as well as the variation in the expression pattern of miR158.

Table S1. Expression (transcript per million [TPM] values) of miR-NAs, siR9, and siR12 in pooled accessions of Indian west Himalayan Arabidopsis populations.

Table S2. Expression (transcript per million [TPM] values) of miR-NAs, siR9, and siR12 in individual accessions of Indian west Himalayan Arabidopsis populations.

Table S3. List of putative targets of miR158 as predicted using psRNATarget.

Table S4. Length distribution of expressed forms of siR9 and siR12 in NTL, ITL, and DTL.

Table S5. Pseudo-*PPR*-derived phasiRNA (PDPs) and their expression.

Table S6. Comparison of phasiRNA derived from 23 reported *A. thaliana PPRs* and the PDPs along with their read counts and sequences.

Table S7. Raw qRT-PCR data of siR9 and siR12 in NTL, ITL, and DTL.

Table S8. AGO1- and AGO4-associated PDPs.

Table S9. Significant small RNA cluster ($P < 0.01$) on pseudo-*PPR*, phasiRNAs and their expression levels in different lines identified from highly significant clusters.

Table S10. List of targets of PDP956-977(+) identified using the PAREsnip tool.

Table S11. ANOVA for the measured physiological and morphological traits along with pairwise comparison.

Table S12. List of primers, probes, and their sequences used in the study.

Table S13. Sequencing and mapping read statistics of the small RNA libraries analyzed in the study.

REFERENCES

Achkar, N.P., Cambiagno, D.A. & Manavella, P.A. (2016) miRNA biogenesis: a dynamic pathway. *Trends in Plant Science*, **21**, 1034–1044.

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.

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

Alonso-Blanco, C., Andrade, J., Becker, C., Bemm, F., Bergelson, J., Borgwardt, K.M. et al. (2016) 1,135 genomes reveal the global pattern of polymorphism in Arabidopsis thaliana. *Cell*, **166**, 481–491.

Alonso-Blanco, C. & Koornneef, M. (2000) Naturally occurring variation in Arabidopsis: an underexploited resource for plant genetics. *Trends in Plant Science*, **5**, 22–29.

An, Y., Furber, K.L. & Ji, S. (2017) Pseudogenes regulate parental gene expression via ce RNA network. *Journal of Cellular and Molecular Medicine*, **21**, 185–192.

Andrés, Z., Pérez-Hormaeche, J., Leidi, E.O., Schlücking, K., Steinhorst, L., McLachlan, D.H. et al. (2014) Control of vacuolar dynamics and regulation of stomatal aperture by tonoplast potassium uptake. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, E1806–E1814.

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

Barragán, V., Leidi, E.O., Andrés, Z., Rubio, L., De Luca, A., Fernández, J.A. et al. (2012) Ion exchangers NHX1 and NHX2 mediate active potassium uptake into vacuoles to regulate cell turgor and stomatal function in Arabidopsis. *The Plant Cell*, **24**, 1127–1142.

Barrs, H. & Weatherley, P. (1962) A re-examination of the relative turgidity technique for estimating water deficits in leaves. *Australian Journal of Biological Sciences*, **15**, 413–428.

Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.

Bassil, E., Tajima, H., Liang, Y.-C., Ohto, M.-a., Ushijima, K., Nakano, R. et al. (2011) The Arabidopsis Na+/H+ antiporters NHX1 and NHX2 control vacuolar pH and K+ homeostasis to regulate growth, flower development, and reproduction. *The Plant Cell*, **23**, 3482–3497.

Bologna, N.G., Mateos, J.L., Bresso, E.G. & Palatnik, J.F. (2009) A loop-to-base processing mechanism underlies the biogenesis of plant microRNAs miR319 and miR159. *The EMBO Journal*, **28**, 3646–3656.

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 Research*, **23**, 1675–1689.

Bomblies, K., Shukla, V. & Graham, C. (2008) Scanning electron microscopy (SEM) of plant tissues. *CSH Protocols*, **2008**, pdb.prot4933.

Boyes, D.C., Zayed, A.M., Ascenzi, R., McCaskill, A.J., Hoffman, N.E., Davis, K.R. et al. (2001) Growth stage–based phenotypic analysis of Arabidopsis: a model for high throughput functional genomics in plants. *The Plant Cell*, **13**, 1499–1510.

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 of the United States of America*, **104**, 3318–3323.

Cuperus, J.T., Carbonell, A., Fahlgren, N., Garcia-Ruiz, H., Burke, R.T., Takeda, A. et al. (2010) Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in Arabidopsis. *Nature Structural & Molecular Biology*, **17**, 997–1003.

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M.K. & Scheible, W.-R. (2005) Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. *Plant Physiology*, **139**, 5–17.

Dai, X. & Zhao, P.X. (2011) psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Research*, **39**, W155–W159.

de Felippes, F.F., Marchais, A., Sarazin, A., Oberlin, S. & Voinnet, O. (2017) A single miR390 targeting event is sufficient for triggering TAS3-tasiRNA biogenesis in Arabidopsis. *Nucleic Acids Research*, **45**, 5539–5554.

De Paoli, E., Dorantes-Acosta, A., Zhai, J., Accerbi, M., Jeong, D.-H., Park, S. et al. (2009) Distinct extremely abundant siRNAs associated with cosuppression in petunia. *RNA*, **15**, 1965–1970.

El-Brolosy, M.A. & Stainier, D.Y. (2017) Genetic compensation: a phenomenon in search of mechanisms. *PLoS Genetics*, **13**, e1006780.

Fahlgren, N., Howell, M.D., Kasschau, K.D., Chapman, E.J., Sullivan, C.M., Cumbie, J.S. et al. (2007) High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. *PLoS One*, **2**, e219.

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.

Folkes, L., Moxon, S., Woolfenden, H.C., Stocks, M.B., Szittya, G., Dalmay, T. et al. (2012) PAREsnip: a tool for rapid genome-wide discovery of small RNA/target interactions evidenced through degradome sequencing. *Nucleic Acids Research*, **40**, e103.

Guo, X., Zhang, Z., Gerstein, M.B. & Zheng, D. (2009) Small RNAs originated from pseudogenes: cis-or trans-acting? *PLoS Computational Biology*, **5**, e1000449.

Hirsch, J., Lefort, V., Vankersschaver, M., Boualem, A., Lucas, A., Thermes, C. et al. (2006) Characterization of 43 non-protein-coding mRNA genes in Arabidopsis, including the MIR162a-derived transcripts. *Plant Physiology*, **140**, 1192–1204.

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

Huang, J., Wang, R., Dai, X., Feng, J., Zhang, H. & Zhao, P.X. (2019) A microRNA biogenesis-like pathway for producing phased small interfering RNA from a long non-coding RNA in rice. *Journal of Experimental Botany*, **70**, 1767–1774.

Jia, J., Ji, R., Li, Z., Yu, Y., Nakano, M., Long, Y. et al. (2020) Soybean DICER-LIKE2 regulates seed coat color via production of primary 22-nucleotide small interfering RNAs from long inverted repeats. *The Plant Cell*, **32**, 3662–3673.

Kościańska, E., Kalantidis, K., Wypijewski, K., Sadowski, J. & Tabler, M. (2005) Analysis of RNA silencing in agroinfiltrated leaves of Nicotiana benthamiana and Nicotiana tabacum. *Plant Molecular Biology*, **59**, 647–661.

Kurihara, Y. & Watanabe, Y. (2004) Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. *Proceedings of the National Academy of Sciences*, **101**, 12753–12758.

Langmead, B. & Salzberg, S.L. (2012) Fast gapped-read alignment with bowtie 2. *Nature Methods*, **9**, 357–359.

Liu, Y., Teng, C., Xia, R. & Meyers, B.C. (2020) PhasiRNAs in plants: their biogenesis, genic sources, and roles in stress responses, development, and reproduction. *Plant Cell*, **32**, 3059–3080.

Lurin, C., Andreés, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyère, C. et al. (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *The Plant Cell*, **16**, 2089–2103.

Ma, Z., Jiang, J., Hu, Z., Lyu, T., Yang, Y., Jiang, J. et al. (2017) Overexpression of miR158 causes pollen abortion in Brassica campestris ssp. chinensis. *Plant Molecular Biology*, **93**, 313–326.

Marin, E., Jouannet, V., Herz, A., Lokerse, A.S., Weijers, D., Vaucheret, H. et al. (2010) miR390, Arabidopsis TAS3 tasiRNAs, and their AUXIN RESPONSE FACTOR targets define an autoregulatory network quantitatively regulating lateral root growth. *The Plant Cell*, **22**, 1104–1117.

Mateos, J.L., Bologna, N.G., Chorostecki, U. & Palatnik, J.F. (2010) Identification of microRNA processing determinants by random mutagenesis of Arabidopsis MIR172a precursor. *Current Biology*, **20**, 49–54.

May, P., Liao, W., Wu, Y., Shuai, B., Richard McCombie, W., Zhang, M.Q. et al. (2013) The effects of carbon dioxide and temperature on microRNA expression in Arabidopsis development. *Nature Communications*, **4**, 1–11.

Mitchell-Olds, T. & Schmitt, J. (2006) Genetic mechanisms and evolutionary significance of natural variation in Arabidopsis. *Nature*, **441**, 947–952.

Ng, D.W., Zhang, C., Miller, M., Palmer, G., Whiteley, M., Tholl, D. et al. (2011) Cis-and trans-regulation of miR163 and target genes confers natural variation of secondary metabolites in two Arabidopsis species and their allopolyploids. *The Plant Cell*, **23**, 1729–1740.

Qi, Y., Denli, A.M. & Hannon, G.J. (2005) Biochemical specialization within Arabidopsis RNA silencing pathways. *Molecular Cell*, **19**, 421–428.

Rajagopalan, R., Vaucheret, H., Trejo, J. & Bartel, D.P. (2006) A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. *Genes & Development*, **20**, 3407–3425.

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B. & Bartel, D.P. (2002) MicroRNAs in plants. *Genes & Development*, **16**, 1616–1626.

Ryan, B.M., Robles, A.I. & Harris, C.C. (2010) Genetic variation in microRNA networks: the implications for cancer research. *Nature Reviews Cancer*, **10**, 389–402.

Schmittgen, T.D. & Livak, K.J. (2008) Analyzing real-time PCR data by the comparative CT method. *Nature Protocols*, **3**, 1101–1108.

Schmitz-Linneweber, C. & Small, I. (2008) Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends in Plant Science*, **13**, 663–670.

Shahid, S., Kim, G., Johnson, N.R., Wafula, E., Wang, F., Coruh, C. et al. (2018) MicroRNAs from the parasitic plant Cuscuta campestris target host messenger RNAs. *Nature*, **553**, 82–85.

Shivaprasad, P.V., Dunn, R.M., Santos, B.A., Bassett, A. & Baulcombe, D.C. (2012) Extraordinary transgressive phenotypes of hybrid tomato are influenced by epigenetics and small silencing RNAs. *The EMBO Journal*, **31**, 257–266.

Singh, A. & Roy, S. (2017) High altitude population of Arabidopsis thaliana is more plastic and adaptive under common garden than controlled condition. *BMC Ecology*, **17**, 1–16.

Singh, A., Tyagi, A., Tripathi, A.M., Gokhale, S.M., Singh, N. & Roy, S. (2015) Morphological trait variations in the west Himalayan (India) populations of Arabidopsis thaliana along altitudinal gradients. *Current Science*, **108**, 2213–2222.

Sosa-Valencia, G., Palomar, M., Covarrubias, A.A. & Reyes, J.L. (2017) The legume miR1514a modulates a NAC transcription factor transcript to trigger phasiRNA formation in response to drought. *Journal of Experimental Botany*, **68**, 2013–2026.

Stocks, M.B., Mohorianu, I., Beckers, M., Paicu, C., Moxon, S., Thody, J. et al. (2018) The UEA sRNA workbench (version 4.4): a comprehensive suite of tools for analyzing miRNAs and sRNAs. *Bioinformatics*, **34**, 3382–3384.

Szarzynska, B., Sobkowiak, L., Pant, B.D., Balazadeh, S., Scheible, W.-R., Mueller-Roeber, B. et al. (2009) Gene structures and processing of Arabidopsis thaliana HYL1-dependent pri-miRNAs. *Nucleic Acids Research*, **37**, 3083–3093.

Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S. et al. (2008) Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature*, **453**, 534–538.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, **30**, 2725–2729.

Thatcher, S.R., Burd, S., Wright, C., Lers, A. & Green, P.J. (2015) Differential expression of miRNAs and their target genes in senescing leaves and siliques: insights from deep sequencing of small RNAs and cleaved target RNAs. *Plant, Cell & Environment*, **38**, 188–200.

Thibaud-Nissen, F., Ouyang, S. & Buell, C.R. (2009) Identification and characterization of pseudogenes in the rice gene complement. *BMC Genomics*, **10**, 1–13.

Tirumalai, V., Prasad, M. & Shivaprasad, P. (2020) RNA blot analysis for the detection and quantification of plant micrornas. *Journal of Visualized Experiments*, **161**, e61394. Available from: https://doi.org/10.3791/61394

Todesco, M., Balasubramanian, S., Cao, J., Ott, F., Sureshkumar, S., Schneeberger, K. et al. (2012) Natural variation in biogenesis efficiency of individual Arabidopsis thaliana microRNAs. *Current Biology*, **22**, 166–170.

Tripathi, A.M., Singh, A., Singh, R., Verma, A.K. & Roy, S. (2019) Modulation of miRNA expression in natural populations of a. thaliana along a wide altitudinal gradient of Indian Himalayas. *Scientific Reports*, **9**, 1–16.

Tyagi, A., Singh, S., Mishra, P., Singh, A., Tripathi, A.M., Jena, S.N. et al. (2016) Genetic diversity and population structure of Arabidopsis thaliana along an altitudinal gradient. *AoB Plants*, **8**, plv145. Available from: https://doi.org/10.1093/aobpla/plv145

Tyagi, A., Yadav, A., Tripathi, A.M. & Roy, S. (2016) High light intensity plays a major role in emergence of population level variation in Arabidopsis thaliana along an altitudinal gradient. *Scientific Reports*, **6**, 1–13.

Varkonyi-Gasic, E., Wu, R., Wood, M., Walton, E.F. & Hellens, R.P. (2007) Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs. *Plant Methods*, **3**, 1–12.

Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gasciolli, V., Mallory, A.C. et al. (2004) Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. *Molecular Cell*, **16**, 69–79.

Voinnet, O. (2009) Origin, biogenesis, and activity of plant microRNAs. *Cell*, **136**, 669–687.

Wang, H., Zhang, X., Liu, J., Kiba, T., Woo, J., Ojo, T. et al. (2011) Deep sequencing of small RNAs specifically associated with Arabidopsis AGO1 and AGO4 uncovers new AGO functions. *The Plant Journal*, **67**, 292–304.

Wang, J., Yang, X., Xu, H., Chi, X., Zhang, M. & Hou, X. (2012) Identification and characterization of microRNAs and their target genes in Brassica oleracea. *Gene*, **505**, 300–308.

Wang, Y., Liu, W., Shen, H., Zhu, X., Zhai, L., Xu, L. et al. (2015) Identification of radish (Raphanus sativus L.) miRNAs and their target genes to explore miRNA-mediated regulatory networks in lead (Pb) stress responses by high-throughput sequencing and degradome analysis. *Plant Molecular Biology Reporter*, **33**, 358–376.

Wang, Z.-P., Xing, H.-L., Dong, L., Zhang, H.-Y., Han, C.-Y., Wang, X.-C. et al. (2015) Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in Arabidopsis in a single generation. *Genome Biology*, **16**, 1–12.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y. et al. (2008) Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature*, **453**, 539–543.

Werner, S., Wollmann, H., Schneeberger, K. & Weigel, D. (2010) Structure determinants for accurate processing of miR172a in Arabidopsis thaliana. *Current Biology*, **20**, 42–48.

Williams, L., Carles, C.C., Osmont, K.S. & Fletcher, J.C. (2005) A database analysis method identifies an endogenous trans-acting short-interfering RNA that targets the Arabidopsis ARF2, ARF3, and ARF4 genes. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 9703–9708.

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.

Xia, R., Xu, J. & Meyers, B.C. (2017) The emergence, evolution, and diversification of the miR390-TAS3-ARF pathway in land plants. *The Plant Cell*, **29**, 1232–1247.

Xia, R., Ye, S., Liu, Z., Meyers, B.C. & Liu, Z. (2015) Novel and recently evolved microRNA clusters regulate expansive F-BOX gene networks through phased small interfering RNAs in wild diploid strawberry. *Plant Physiology*, **169**, 594–610.

Xie, Z., Allen, E., Fahlgren, N., Calamar, A., Givan, S.A. & Carrington, J.C. (2005) Expression of Arabidopsis MIRNA genes. *Plant Physiology*, **138**, 2145–2154.

Xu, M.Y., Dong, Y., Zhang, Q.X., Zhang, L., Luo, Y.Z., Sun, J. et al. (2012) Identification of miRNAs and their targets from Brassica napus by high-throughput sequencing and degradome analysis. *BMC Genomics*, **13**, 1–15.

Yagi, Y., Tachikawa, M., Noguchi, H., Satoh, S., Obokata, J. & Nakamura, T. (2013) Pentatricopeptide repeat proteins involved in plant organellar RNA editing. *RNA Biology*, **10**, 1419–1425.

Yoshikawa, M., Peragine, A., Park, M.Y. & Poethig, R.S. (2005) A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. *Genes & Development*, **19**, 2164–2175.

Yu, G., Yao, W., Gumireddy, K., Li, A., Wang, J., Xiao, W. et al. (2014) Pseudogene PTENP1 functions as a competing endogenous RNA to suppress clear-cell renal cell carcinoma ProgressionPTENP1 as ceRNA in ccRCC. *Molecular Cancer Therapeutics*, **13**, 3086–3097.

Yu, Y., Jia, T. & Chen, X. (2017) The 'how' and 'where' of plant microRNAs. *New Phytologist*, **216**, 1002–1017.

Zhai, J., Jeong, D.-H., De Paoli, E., Park, S., Rosen, B.D., Li, Y. et al. (2011) MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. *Genes & Development*, **25**, 2540–2553.

Zhang, C., Li, G., Zhu, S., Zhang, S. & Fang, J. (2014) tasiRNAdb: a database of ta-siRNA regulatory pathways. *Bioinformatics*, **30**, 1045–1046.

Zhang, Z., Guo, X., Ge, C., Ma, Z., Jiang, M., Li, T. et al. (2017) KETCH1 imports HYL1 to nucleus for miRNA biogenesis in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, **114**, 4011–4016.

Zhao, X., Li, J., Lian, B., Gu, H., Li, Y. & Qi, Y. (2018) Global identification of Arabidopsis lncRNAs reveals the regulation of MAF4 by a natural antisense RNA. *Nature Communications*, **9**, 1–12.

Zhong, S.-H., Liu, J.-Z., Jin, H., Lin, L., Li, Q., Chen, Y. et al. (2013) Warm temperatures induce transgenerational epigenetic release of RNA silencing by inhibiting siRNA biogenesis in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 9171–9176.

Zhou, C., Han, L., Fu, C., Wen, J., Cheng, X., Nakashima, J. et al. (2013) The trans-acting short interfering RNA3 pathway and no apical meristem antagonistically regulate leaf margin development and lateral organ separation, as revealed by analysis of an argonaute7/lobed leaflet1 mutant in Medicago truncatula. *The Plant Cell*, **25**, 4845–4862.

Zhu, H., Zhou, Y., Castillo-González, C., Lu, A., Ge, C., Zhao, Y.-T. et al. (2013) Bidirectional processing of pri-miRNAs with branched terminal loops by Arabidopsis Dicer-like1. *Nature Structural & Molecular Biology*, **20**, 1106–1115.