BREAKTHROUGH REPORT

Soybean DICER-LIKE2 Regulates Seed Coat Color via Production of Primary 22-Nucleotide Small Interfering RNAs from Long Inverted Repeats

Jinbu Jia,a,b,c,1 Ronghuan Ji,d,1 Zhuowen Li,a,b,c Yiming Yu,a,b,c Mayumi Nakano,e Yanping Long,a,b,c Li Feng,a,b,c Chao Qin,d Dongdong Lu,a,b,c Junpeng Zhan,a,b,c,e Rui Xia,f Blake C. Meyers,e,g Bin Liu,d,2 and Jixian Zhaia,b,c,2

a Department of Biology, Southern University of Science and Technology, Shenzhen 518055, China
b Institute of Plant and Food Science, Southern University of Science and Technology, Shenzhen 518055, China
c Key Laboratory of Molecular Design for Plant Cell Factory of Guangdong Higher Education Institutes, Southern University of Science and Technology, Shenzhen 518055, China
d Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China
e Donald Danforth Plant Science Center, St. Louis, Missouri 63132
f College of Horticulture, South China Agricultural University, Guangzhou 510642, China
g Division of Plant Sciences, University of Missouri, Columbia, Missouri 65211

ORCID IDs: 0000-0002-3305-7893 (Y.Y.); 0000-0002-9919-7282 (M.N.); 0000-0001-9317-4509 (M.L.); 0000-0003-4512-0594 (Y.L.); 0000-0003-3636-2962 (F.L.); 0000-0002-7278-6057 (C.Q.); 0000-0001-5163-0115 (D.L.); 0000-0001-7353-7608 (J.Z.); 0000-0003-2409-1181 (R.X.); 0000-0002-3436-6097 (B.CM.); 0000-0002-5836-2333 (B.L.); 0000-0002-0217-0666 (J.Z.);

In plants, 22-nucleotide small RNAs trigger the production of secondary small interfering RNAs (siRNAs) and enhance silencing. DICER-LIKE2 (DCL2)-dependent 22-nucleotide siRNAs are rare in Arabidopsis (Arabidopsis thaliana) and are thought to function mainly during viral infection; by contrast, these siRNAs are abundant in many crops such as soybean (Glycine max) and maize (Zea mays). Here, we studied soybean 22-nucleotide siRNAs by applying CRISPR-Cas9 to simultaneously knock out the two copies of soybean DCL2, GmDCL2a and GmDCL2b, in the Tianlong1 cultivar. Small RNA sequencing revealed that most 22-nucleotide siRNAs are derived from long inverted repeats (LIRs) and disappeared in the Gmdcl2a/2b double mutant. De novo assembly of a Tianlong1 reference genome and transcriptome profiling identified an intronic LIR formed by the chalcone synthase (CHS) genes CHS1 and CHS3. This LIR is the source of primary 22-nucleotide siRNAs that target other CHS genes and trigger the production of secondary 21-nucleotide siRNAs. Disruption of this process in Gmdcl2a/2b mutants substantially increased CHS mRNA levels in the seed coat, thus changing the coat color from yellow to brown. Our results demonstrated that endogenous LIR-derived transcripts in soybean are predominantly processed by GmDCL2 into 22-nucleotide siRNAs and uncovered a role for DCL2 in regulating natural traits.

INTRODUCTION

In plants, microRNAs (miRNAs) and small interfering RNAs (siRNAs) trigger cleavage of their complementary mRNA targets via the slicer activity of ARGONAUTE (AGO) proteins (Rogers and Chen, 2013; Song et al., 2019). In Arabidopsis (Arabidopsis thaliana), while the majority of the sliced mRNA targets are degraded after miRNA-mediated cleavage, a small subset of the cleaved targets generate secondary, trans-acting siRNAs that enhance the silencing cascade (Allen et al., 2005; Chen et al., 2007; Howell et al., 2007). Two major models account for the initiation of the production of trans-acting siRNAs: the “two-hit” model for the TAS3 locus, which requires the function of miR390 and AGO7 (Axtell et al., 2006; Xia et al., 2017); and the “one-hit” model, which requires a miRNA trigger to be precisely 22 nucleotides in length (Chen et al., 2010; Cuperus et al., 2010). The TAS2 locus in Arabidopsis mostly generates secondary 21-nucleotide siRNAs triggered by miR173, but the 22-nucleotide tasiR2140 from TAS2 3’ D6(-) processed by DICER-LIKE4 (DCL4) triggers tertiary siRNAs, although its functional significance is still unclear (Chen et al., 2010). The 22-nucleotide miRNAs can be generated by DCL1 directly processing precursors containing an asymmetric bulge (Manavella et al., 2012) or by nucleotidyltransferase-mediated 3’ terminal extension of one nucleotide on certain 21-nucleotide miRNAs (Zhai et al., 2013; Fei et al., 2018). In addition, 22-nucleotide noncanonical miRNAs can be produced by SdDCL2 in tomato (Solanum lycopersicum); these miRNAs in turn target...
SIDCL2 mRNAs, forming a feedback loop (Wang et al., 2018b). In *Medicago truncatula*, soybean (*Glycine max*), tomato, and many other plant species, 22-nucleotide miRNAs act as master regulators via the production of phased, secondary siRNAs (Fei et al., 2013) to target gene families with hundreds of members, such as the gene families encoding pentatricopeptide repeat proteins, nucleotide binding-leucine-rich repeat proteins, and MYB transcription factors (Zhai et al., 2011; Li et al., 2012; Shivprasad et al., 2012; Xia et al., 2013; Aritik et al., 2014). Moreover, the 22-nucleotide miR2118 family can target thousands of noncoding transcripts in reproductive tissues of monocots (Johnson et al., 2009; Song et al., 2012; Zhai et al., 2015).

Compared with miRNAs, 22-nucleotide siRNAs in plants are primarily produced by DCL2 and play essential roles in viral defense and transgene silencing (Gascioli et al., 2005; Ding and Voinnet, 2007; Parent et al., 2015; Taochy et al., 2017; Wu et al., 2017; Wang et al., 2018b). Recent reports also found that 22-nucleotide siRNAs derived from endogenous genes can be induced by nutrient deficiency and may contribute to plant adaptation to environmental stresses (Wu et al., 2020). However, the function of endogenous 22-nucleotide siRNAs under normal growth conditions is still largely unexplored, partially because they are rare in wild-type Arabidopsis, and loss-of-function *dcl2* mutants in Arabidopsis and tomato have no visible developmental defects (Henderson et al., 2006; Wang et al., 2018a, 2018b). Loss of function of DCL4 in Arabidopsis can induce the production of a large amount of DCL2-dependent 22-nucleotide siRNAs from endogenous genes and cause silencing of their targets, and as a result, plants exhibit various developmental defects (Bouché et al., 2006; Zhang et al., 2015; Wu et al., 2017, 2020). However, endogenous 22-nucleotide siRNAs are abundant in many major crops such as soybean, rice (*Oryza sativa*), and maize (*Zea mays*), indicating their potential importance (Lunardon et al., 2020). For example, 22-nucleotide siRNAs from transposable element (TE) regions are abundant in maize, especially in the mediator of paramutation 1-1 background, in which 24-nucleotide siRNAs disappear (Nobuta et al., 2008). Yet, their functions remain unclear, partly due to a lack of mutants that disrupt the production of 22-nucleotide siRNAs in major crops. Here, we take advantage of the completed soybean reference genome (Schmutz et al., 2010; Liu et al., 2020) and recent advances in crop genome-editing technologies (Shan et al., 2013) to directly investigate the function of DCL2-dependent 22-nucleotide siRNAs in soybean by studying *Gmdc2* clustered regularly interspaced short palindromic repeats (CRISPR)-Cas9 lines with frame-shift mutations.

## RESULTS

**CRISPR-Cas9 Editing of DCL2a/2b in Soybean**

To gain insight into the role of 22-nucleotide endogenous siRNAs in soybean, we applied the CRISPR-Cas9 technology in soybean cv Tianlong1 to obtain loss-of-function mutants of the *GmDCL2* genes. The soybean genome encodes two copies of *DCL2*, *GmDCL2a* (*Glyma.09G025300*) and *GmDCL2b* (*Glyma.09G025400*); these have high similarity (80.4%) in the encoded protein sequences (Supplemental Figures 1A and 1B; Supplemental Files 1 and 2). These two genes are adjacent to each other on chromosome 9, indicating that they are derived from a recent tandem duplication and thus have a high likelihood of being functionally redundant. Taking advantage of their high sequence similarity, we designed a single guide RNA (gRNA) to simultaneously target a conserved coding region on both genes (Figure 1A), and we performed CRISPR-Cas9-mediated editing through tissue culture (see Methods for details).

From multiple independent mutant lines, we selected three homozygous double mutant lines, *Gmdc2a/2b-8*, *Gmdc2a/2b-9*, and *Gmdc2a/2b-16*, for further studies (Figure 1A). All three lines harbor deletions around the target site that result in a frame shift and premature stop codon in the coding region of *GmDCL2a/b* (Figure 1A). The mutant lines showed no obvious abnormality in plant architecture except for being slightly dwarf (Supplemental Figure 1C; Supplemental File 3). The most dramatic phenotypic change is that the seed coat of all three lines was brown in contrast to the yellow seed coat of wild-type Tianlong1 (Figure 1B; Supplemental File 3). In our later work, this phenotype was of particular interest for studying the function of DCL2.

### Impacts of DCL2 Loss of Function on Small RNA Biogenesis in Soybean

To study the impacts of *GmDCL2a/b* loss of function on small RNA (sRNA) biogenesis and gene regulation, we performed sRNA sequencing (sRNA-seq) and mRNA sequencing (mRNA-seq) in both the wild type (*Tianlong1*) and the *Gmdc2a/2b-8* mutant. Both the leaf and seed coat tissues were examined with three biological replicates for each sample. The sRNA-seq results showed that wild-type soybean produced abundant 22-nucleotide siRNAs, which decreased substantially in the *Gmdc2a/2b* mutant (Figure 2A).

To further analyze these DCL2-dependent siRNAs, we classified sRNA clusters based on the dominant size class in each cluster using ShortStack (Axtell, 2013; Lunardon et al., 2020). Consistent with previous analyses of sRNAs in soybean (Aritik et al., 2014; Lunardon et al., 2020), the majority of sRNA loci in soybean features 24-nucleotide siRNAs, but 21- and 22-nucleotide loci demonstrated higher sRNA abundances (Figure 2B). In wild-type libraries, our analysis identified 1304 22-nucleotide sRNA loci from the seed coat and 1371 from leaf sample, which accounted for only ~3% of the total number of loci but over 20% of total sRNA abundance (Figure 2B). Compared with the wild type, siRNAs from 22-nucleotide sRNA loci decreased substantially in the *Gmdc2a/2b* mutant (Figures 2B and 2C). For example, over 70% of the 22-nucleotide sRNA loci showed at least a 10-fold decrease in abundance in the seed coat (Figure 2D). Also, siRNAs from some 21-nucleotide sRNA loci, especially those that share sequence homology with 22-nucleotide sRNA loci, were decreased in the mutant (Figure 2E), suggesting that DCL2-dependent 22-nucleotide siRNAs are capable of targeting in trans and triggering the biogenesis of secondary 21-nucleotide siRNAs. These results indicated that, compared with the low level of 22-nucleotide siRNAs found in Arabidopsis, soybean accumulates a large amount of endogenous 22-nucleotide siRNAs whose biogenesis requires DCL2.
We noticed that a handful of 22-nucleotide siRNA loci remained abundant in the Gmdcl2a/2b mutant (Supplemental Table 1), and a further investigation found that these loci exhibited miRNA-like characteristics. For example, one locus contains two homologous Gypsy TEs in the fifth intron of a clathrin gene (Glyma.14G058300); an RNA from this locus is predicted to form a stem-loop-like structure with bulges and mismatches that resemble a miRNA precursor (Figure 2F). While sRNA-B and sRNA-C nearly disappeared in the Gmdcl2a/2b mutant, the 22-nucleotide sRNA-A was surprisingly even more abundant in the Gmdcl2a/2b mutant (Figure 2F), suggesting that this stem-loop can be processed not only by DCL2 but also by other DCL proteins, most likely by DCL1. This finding indicates that DCL2-dependent 22-nucleotide siRNAs have the potential to evolve into DCL1-dependent 22-nucleotide miRNAs, thus shedding light on a potential path for the evolution of 22-nucleotide miRNA via DCL2.

A Large Number of TEs Are Targeted by DCL2-Dependent 22-Nucleotide siRNAs in Soybean

In addition to their relatively large number, 22-nucleotide siRNA loci in soybean have another unusual feature compared with Arabidopsis: ~70% of these 22-nucleotide siRNA loci are overlapping with TEs (Figure 3A; Supplemental Figure 2A), suggesting a potential role in TE silencing. In particular, instead of a typical 24-nucleotide length, the siRNAs from the Caulimovirus class of TEs are mainly 22 nucleotides in length; we found that other TE families, such as CMC-Enhancer/Suppressor-mutator (CMC-EnSpm, CMC represents the three conserved nucleotides in terminal inverted repeat region, M = A or C), also produced a large amount of 22-nucleotide siRNAs (Figure 3B; Supplemental Figure 2B). These TE-derived 22-nucleotide siRNAs are mostly strand-specific, and their abundances sharply decreased in the Gmdcl2a/2b mutant, as exemplified by one Caulimovirus TE locus and one CMC-EnSpm locus (Figures 3B and 3C; Supplemental Figures 2B and 2C). In addition to 22-nucleotide loci, we also identified dozens of highly abundant 22-nucleotide loci that are TE-related, and their sRNA accumulation also decreased in the Gmdcl2a/2b mutant (Figure 3D; Supplemental Figure 2D), such as one CMC-EnSpm locus that is homologous with other 22-nucleotide siRNA generating the CMC-EnSpm family of TEs (Figure 3C; Supplemental Figure 2C). These findings are consistent with the hypothesis that 22-nucleotide TE-derived siRNAs may target the homologous TE transcripts and trigger the generation of secondary 21-nucleotide siRNAs to enhance their silencing. In the Gmdcl2a/2b mutant, however, we did not observe an elevated level of TE mRNA accumulation (Figure 3D; Supplemental Figure 2E; with examples shown in Figure 3C and Supplemental Figure 2C), suggesting that there are other redundant mechanisms to suppress TE activities in soybean.

Figure 1. CRISPR/Cas9-Engineered Mutations in GmDCL2a/2b Result in a Brown Seed Coat.

(A) Deletion mutations at target sites in the Gmodcl2a/2b mutants. The gRNA target and PAM sequences are highlighted with boldface in black and blue, respectively. The gene structures of GmDCL2a and GmDCL2b are shown above. The coding sequence is shown as a dark box, the untranslated region is shown as a gray box, and the intron region is shown as a line.

(B) Seed coat color phenotype of the Gmdcl2a/2b mutants. Bar = 1 cm. WT, wild type.
Figure 2. The Accumulation Levels of 22-Nucleotide siRNAs Are Sharply Decreased in the Gmdcl2a/2b Mutant.

(A) The length distribution of siRNAs in the wild type (WT) and the Gmdcl2a/2b mutant. nt, nucleotide; TPM, transcripts per million.

(B) Proportions of total locus count and siRNA accumulation of different kinds of loci. The numbers above each column represent the number of siRNA loci identified in different samples. nt, nucleotide; WT, wild type.

(C) The length distribution of sRNAs from loci predominantly generating 21-, 22-, or 24-nucleotide (nt) siRNAs. TPM, transcripts per million; WT, wild type.

(D) The comparison of sRNA accumulation levels of 22-nucleotide (nt) siRNA loci in seed coat between the wild type (WT) and the Gmdcl2a/2b mutant. The arrow and circle mark a miRNA-like locus shown in (F). The dashed lines indicate an sRNA fold change of 10. TPM, transcripts per million.

(E) The comparison of sRNA accumulation levels of 21-nucleotide siRNA loci in seed coat between the wild type and the Gmdcl2a/2b mutant. The dashed lines indicate an sRNA fold change of 10.

(F) A miRNA-like locus. The potential miRNA/miRNA* pairs and their expression levels in the wild type (WT) and the Gmdcl2a/2b (dcl2) mutant are shown in the left panel. The depth of uniquely mapped mRNA reads and the accumulation levels of uniquely mapped sRNA reads in seed coat are shown in the right panel. nt, nucleotide.
The DNA of TE regions in plant genomes is often highly methylated in CG and non-CG contexts (Law and Jacobsen, 2010). We tested whether these TE-derived DCL2-dependent 22-nucleotide siRNAs are capable of triggering DNA methylation, like the well-established role of 24-nucleotide siRNAs in RNA-directed DNA methylation by methylome profiling of the leaf tissues. From our whole-genome bisulfite sequencing (WGBS) data, we found that the methylation level at these 22-nucleotide TE loci is similar between the wild type and the Gmdcl2a/2b mutant (Supplemental Figure 2C). Thus, we conclude that 22-nucleotide siRNAs are not required for the maintenance of DNA methylation at their targeted TEs in leaves.

GmDCL2 Favors Long Inverted Repeats as Its Substrates

Next, knowing that RNA secondary structure is a key feature of at least DCL1 processing, we examined the structural features of 22-nucleotide siRNAs. We found that inverted repeats (IRs) are enriched at a much higher proportion at 22-nucleotide loci compared with 21- or 24-nucleotide loci. For example, ~80% of 22-nucleotide siRNA loci with an sRNA accumulation level higher than 50 TPM overlapped with IRs, compared with only 12% of 21-nucleotide siRNA loci with similar abundance overlapping with IRs (Figure 4A). Our poly(A)-selected mRNA-seq data further showed that, for those IRs that are overlapping with 22-nucleotide siRNA loci and have relatively high sRNA accumulation levels (above 10 TPM), 90% of them are transcribed by Pol II (Figure 4B). In contrast, IRs overlapping with 24-nucleotide siRNA loci have almost no detectable poly(A) signal (Figure 4B), as most 24-nucleotide siRNAs are derived from Pol IV transcripts that do not have a poly(A) tail (Kuo et al., 2017). We also found that a smaller number of IRs mainly produced 21-nucleotide siRNAs and are also transcribed by Pol II (Figures 4A and 4B).

A previous study showed that DCL4 prefers double-stranded RNA (dsRNA) substrates that are over 100 nucleotides in length (Nagano et al., 2014). In addition, in a number of monocot species, 24-nucleotide reproductive phased siRNAs are derived from many IR precursors (Kakrana et al., 2018). However, the substrate preference for DCL2 remains unknown. Our analysis found that the median length of 22-nucleotide siRNA-generating IRs is much larger than that of 21-nucleotide siRNA-generating IRs, especially for loci with high sRNA abundance. For those loci with TPM higher than 50, the median lengths of 21- and 22-nucleotide siRNA-generating IRs were 254 and 1289 nucleotides, respectively (Figure 4C). For example, the eighth intron of gene Glyma.15G177500 contained a long IR consisting of Copia TEs. The repeat region of this IR was ~2800 nucleotides and produced a large number of 22-nucleotide siRNAs from only the sense strand in a DCL2-dependent manner (Figure 4D). In contrast, the pre-mRNAs of gene Glyma.14G055600 contained a short IR whose repeat region is 240 nucleotides and...
produced a series of 21-nucleotide sense siRNAs that were unaffected by Gmdcl2a/2b mutation (Figure 4D). These results suggested that long IRs transcribed by Pol II are preferentially processed by DCL2 in soybean. In contrast, shorter IRs (typically 300 nucleotides and shorter) are favored by DCL4 to produce 21-nucleotide siRNAs.

DCL2-Dependent 22-nucleotide siRNAs Regulate the Seed Coat Color in Soybean

Chalcone synthase (CHS) is a key enzyme for the biosynthesis of flavonoids, which cause black/brown pigmentation in the soybean seed coat (Tuteja et al., 2009). Wild soybean accessions have black or brown seed coats, whereas commercial soybean cultivars are yellow due to the presence of a dominant allele of the I locus. There are four genotypes of the I locus (I, ii, ik, and i, with that order of dominance). Seeds of the dominant I allele have no pigment either on the hilum or on seed coat proper; the ii allele exhibits pigment on the hilum but not on the seed coat proper; the ik allele shows a two-colored saddle pattern on the seed coat; and the recessive i alleles feature fully pigmented seed coat (Tuteja et al., 2004; Cho et al., 2019). The i allele contains two similar IR clusters, and each cluster contains three CHS genes (CHS1, CHS3, and CHS4 [Tuteja et al., 2004, 2009]; illustrated in Figure 5A). This locus is the source of siRNAs that target and silence other CHS genes in the genome in trans and result in a yellow seed coat (Tuteja et al., 2004, 2009). Large deletions, some of which arise by homologous recombination within the CHS clusters, lead to missing CHS genes in the clusters and thereby can abolish the production of the CHS-derived siRNAs and reverse the seed coat color back to black/brown (Tuteja et al., 2009; Cho et al., 2019). However, the genome sequencing of black W05 wild soybeans showed that it also contains this large CHS cluster, suggesting that the cluster itself is not sufficient to trigger siRNA-mediated silencing (Xie et al., 2019; Liu et al., 2020).

Comparative genome analysis found that a complex inversion and gene duplication event adjacent to CHS gene clusters occurs in most cultivars, such as Williams 82 and ZH13, and results in the promoter as well as the first four exons/introns of a subtilisin gene inserted to the upstream of the CHS gene cluster (Figure 5A; Shen et al., 2018; Xie et al., 2019; Liu et al., 2020). This leads to a model proposing that the CHS1 antisense transcripts driven by the promoter of the subtilisin gene can base pair with other CHS1 sense transcripts to form dsRNAs that are further processed into 21-nucleotide siRNAs to silence other CHS genes (Xie et al., 2019). Recent studies found that AGO5 is involved in this process, as
Figure 5. The 22-Nucleotide siRNAs Suppress the Expression of CHS Genes That Control the Color of the Seed Coat.

(A) The gene structures of the I locus in the wild soybean W05 genome as well as cultivated soybean ZH13 and Tianlong1 genomes. The mRNA-seq and sRNA-seq results of the Tianlong1 seed coat are shown below the gene structures. *, Due to the duplication of CHS1 and CHS3 genes, the reads specifically mapped to CHS1 or CHS3, but not to other CHS genes, are also shown as “Uniquely mapped.” The junction read counts (n) are indicated. nt, nucleotide; WT, wild type.

(B) The length distribution of CHS siRNAs in the wild type (WT). nt, nucleotide; TPM, transcripts per million.

(C) The sRNA and mRNA accumulation levels of CHS genes in the Gmdcl2a/2b mutant and the wild type (WT). FPKM, fragments per kilobase of exon model per million mapped reads. TPM, transcripts per million.

(D) A model of DCL2-dependent 22-nucleotide (nt) siRNAs regulating the expression levels of CHS genes.
DCL2 is required to maintain the silencing of the CHS gene family, as Gmdcl2a/2b mutation alters the seed coat color of Tianlong1 (Figure 1B). Our results demonstrated that the DCL2-dependent 22-nucleotide siRNAs are required to maintain the silencing of the CHS gene family and regulate the color of the seed coat in soybean (Figure 5D). In addition, by comparing mRNA-seq data of the wild type and the Gmdcl2a2b mutant, we identified 381 genes differentially accumulated in leaf (Supplemental Data Set 1) and 1912 in seed coat (including multiple CHS family members; Supplemental Data Set 2), suggesting that GmDCL2-dependent 22-nucleotide siRNAs can regulate the expression of genes besides the CHS family.

**DISCUSSION**

While our results showed that DCL2 can act on the transcripts containing long CHS IRs from the I locus (ii allele in cv Tianlong1), one needs to be extra cautious about the simple hypothesis that transcription originating at the partial subtilisin fragment is the main or only driver for the production of the CHS dsRNAs at the I locus, given its complicated structure. In addition, several lines of evidence suggest a more complicated situation for the regulation of the production of the CHS dsRNAs as well as the CHS siRNAs: (1) according to the recent soybean pan-genome analysis, 7 out of the 24 modern cultivars have a haplotype 2 of CHS arrangement that do not have the partial subtilisin fragment at the I locus yet still have a yellow seed coat (Liu et al., 2020); (2) although the wild species W05 lacks the partial subtilisin promoter fragment and has a black seed coat, it is unclear whether the potential mutations of other modifying factors, such as DCL2, AGOS, or other genes involved in the silencing pathway, result in the inability to produce CHS siRNAs in W05 (Cho et al., 2017); and (3) the rearranged subtilisin promoter hypothesis cannot explain the two-color pattern observed in the i and ii alleles, as both silencing (yellow) and nonsilencing (pigmented) regions should have the same genomic structure (Cho et al., 2017). Thus, future systemic studies using naturally occurring spontaneous mutations or transgenic lines that have closely related genetic backgrounds are best for investigating the mechanism of such a complex locus in diverse wild and cultivated soybeans.

The precursor of 22-nucleotide siRNAs in soybean resembles the precursor of miRNAs in many ways: a single-stranded RNA transcribed by Pol II forms a hairpin-like structure and does not require the activity of RNA-dependent RNA polymerase. Such types of IR precursors for siRNAs have also been observed for 24-nucleotide reproductive phased siRNAs in some monocot species, including asparagus (Asparagus officinalis), lily (Lilium maculatum), and daylily (Hemerocallis lilioasphodelus; Kakrana et al., 2018). Moreover, 22-nucleotide siRNAs in soybean can also trigger the production of secondary 21-nucleotide siRNAs, consistent with recent reports in Arabidopsis on the role of 22-nucleotide siRNAs in amplifying silencing signals (Wu et al., 2017, 2020). The parallels in the structure of the precursor and the mode of function between the 22-nucleotide miRNA and LIR-derived 22-nucleotide siRNA imply a potential mechanism for the direct evolutionary descent of 22-nucleotide miRNAs from 22-nucleotide siRNAs. Our findings also support the previously proposed model that miRNA precursors can originate from local duplication that forms an IR structure (Allen et al., 2004; Cuperus et al., 2011). Therefore, the large number of LIR loci that generate 22-nucleotide siRNAs in soybean could provide a rich foundation for miRNA origin.

To understand the role of DCL2 in regulating seed coat color, we de novo assembled the I locus in Tianlong1 (i allele, yellow seed coat with light brown hilum) using a combination of PacBio long-read sequencing, Illumina sequencing, and chromosome conformation capture sequencing (Hi-C; see Methods for details). We found that the I locus in Tianlong1 has the same inversion and duplication of the subtilisin gene fragment as previously shown in other cultivars (Figure 5A). We hypothesize that the transcription potentially driven by the promoter of this subtilisin gene traverses the antisense strand of CHS1 and the sense strand of CHS3 in Tianlong1 (Figure 5A). This chimeric transcript has been previously discovered in ESTs in the Williams 82 cultivar (Clough et al., 2004) and confirmed here by a large number of splicing junction mRNA-seq reads spanning the entire antisense-CHS1-sense-CHS3 IR region, connecting the last exon of the inserted subtilisin fragment and the exon located downstream of this IR. The antisense CHS1 region and the sense CHS3 region in the subtilisin-antisense-CHS1-sense-CHS3 chimeric transcripts can form a long inverted repeat (LIR) because of the close sequence similarity between CHS1 and CHS3. Consistent with our finding that GmDCL2 favors LIRs as its substrates, we detected a series of siRNAs from the antisense CHS1 region and sense CHS3 region that are mainly 22 nucleotides in length (Figures 5A and 5B) in the wild type and disappeared in the Gmdcl2a2b mutant (Figures 5A and 5C).

Besides CHS1 and CHS3, we also detected a large number of secondary siRNAs from other CHS genes, such as CHS2, CHS7, and CHS8 (Figure 5B). These siRNAs are mainly 21 nucleotides in length and generated from both the sense and antisense strands (Figure 5B), indicating the possible involvement of an RNA-dependent RNA polymerase in generating the dsRNA precursors. Interestingly, these 21-nucleotide siRNAs also disappeared in the Gmdcl2a2b mutant (Figure 5C), suggesting that they are likely secondary siRNAs triggered by DCL2-dependent 22-nucleotide CHS siRNAs (Figure 5D). This is consistent with a previous study that investigated siRNA profiles at 10 different stages of seed coat development and found that the 22-nucleotide CHS siRNAs are more prevalent than the 21-nucleotide siRNAs at the very early stages of seed coat development, including 4 to 24 d after flowering, whereas 21-nucleotide CHS siRNAs become dominant at the later stages of seed coat development. The transition from 22-nucleotide siRNAs at CHS1/3 (I locus) to predominantly 21-nucleotide secondary siRNAs at CHS7/8 likely occurs at the 5- to 6-mg stages (Cho et al., 2013). In the Gmdcl2a2b mutant, the mRNAs of CHS2, CHS7, and CHS8 genes accumulate at much higher levels compared with the wild type (Figure 5C), and the seed coat color changed from yellow to brown (Figure 1B).
While wild-type Arabidopsis generates few 22-nucleotide siRNAs, we found that a large number of 22-nucleotide siRNAs are produced mostly from LIR regions in soybean. These 22-mers are capable of triggering the production of secondary 21-nucleotide siRNAs to further silence TEs or protein-coding genes. A recent study of sRNAs from 47 diverse plant species showed that many major crops and vegetables produce 22-nucleotide siRNAs at much higher levels compared with Arabidopsis (Lunardon et al., 2020). Our results here demonstrated that DCL2-dependent 22-nucleotide siRNAs are capable of regulating natural traits, and future investigation in more species could help to expand our understanding about the possibly underappreciated function of 22-nucleotide siRNAs in wild-type plants.

METHODS

Plant Materials and Growth Conditions

For preparation of soybean (Glycine max) leaf samples, the wild-type Tianlong1 and CRISPR/Cas9-engineered mutant lines were grown under short-day conditions (10 h of 300 \mu mol m^{-2} s^{-1} white light/14 h of dark, 26°C) in the greenhouse, and the leaves of 10-d-old seedlings were collected at Zeitgeber time 8 (8 h after light exposure). For preparation of seed coat samples, the soybean plants were grown under long-day conditions (16 h of 300 \mu mol m^{-2} s^{-1} white light/8 h of dark, 26°C), and the seed coats were dissected from the fresh seeds with a weight of 50 to 75 mg (Tuteja et al., 2009) at Zeitgeber time 11 (11 h after light exposure). The RNA and DNA used in mRNA-seq, siRNA-seq, and WGBS were isolated from tissues of the wild type (Tianlong1) and the GmDcl2a/b-8 mutant. The seeds can be obtained by contacting the group of Bin Liu at the Chinese Agricultural Academy of Sciences.

Targeted Mutagenesis of GmDCL2 Genes

The gRNAs targeting GmDCL2 genes were designed by the CRISPR-P online tool (http://crispr.pku.edu.cn/CRISPR2/; Lei et al., 2014). The GmU6 promoter was amplified with a pair of primers, GmU6-F and GmU6-R, and the gRNA scaffold was amplified with a pair of primers, gRNA-F and Scaffold-R, using the plasmid pU3-gRNA as a template. The GmU6:gRNA cassette was constructed by overlapping PCR with the GmU6-F and Scaffold-R primers and then inserted into the 35S-Cas9-Bar vector between the XbaI and BgiII sites by infusion technology (Clontech). The constructed GmDCL2-gRNA binary vector was introduced into Agrobacterium tumefaciens strain EHA105 and transformed into wild-type soybean Tianlong1 (Paz et al., 2006). The T0 transgenic plants were regenerated on the medium under the selection of 8 mg/L glufosinate ammonium. For characterizing the targeted mutations, the genomic sequences of GmDCL2a and GmDCL2b were amplified and sequenced using the primer pairs GmDCL2a/SF/SR and GmDCL2b/SF/SR, respectively. Primers are listed in Supplementary Table 2.

Biallelic (both alleles are mutated but have different mutations), heterozygous (one allele is mutated, one is wild type), and chimeric (wild-type allele, and multiple different mutated alleles) mutations frequently occur in the soybean CRISPR/Cas9 transgenic lines. The T0 transgenic line #8 (brown seed coat) is biallelic in both GmDCL2a and GmDCL2b loci, based on the genotyping of 23 T1 transgenic lines (all have brown seed coat): 10 lines were homozygous but in two kinds of genotypes (6 lines:4 lines), and 13 are still biallelic. Both of the T0 transgenic lines #9 and #16 (yellow seed coat) are complex chimeric lines. We genotyped the offspring of transgenic plants to identify homozygous double mutants and only selected one mutant obtained from each line for further analysis; we named them GmDcl2a/2b-8, GmDcl2a/2b-9, and GmDcl2a/2b-16.

RNA Extraction and Sequencing of mRNAs and sRNAs

The RNA samples of the wild type (Tianlong1) and the GmDcl2a/b-8 mutant were prepared by the Spectrum Plant Total RNA Kit (Sigma-Aldrich, STRN50) according to the manufacturer’s instructions. The strand-specific mRNA and sRNA libraries were prepared and sequenced at BGI-Shenzhen.

WGBS and Analysis

Extraction of genomic DNA and the construction of WGBS libraries were performed by BGI-Shenzhen. After sequencing, the mapping of reads and the calculation of DNA methylation ratios in single-base levels were performed using BSMAP (v2.90; Xi and Li, 2009) with default parameters.

Bioinformatics Analysis of sRNAs

The adapters of raw reads were removed using cutadapt (v2.9; Martin, 2011), and the trimmed reads of 20 to 30 nucleotides in length were mapped to the Tianlong1 genome using bowtie (v1.2.2; Langmead et al., 2009) with the parameters -v 0 -a -S. After filtering out the reads matching to rRNAs, tRNAs, as well as the mitochondrial and chloroplast genome (allowing two mismatches), the clean reads of each library were individually mapped to the Tianlong1 genome by ShortStack (v3.8.5; Axtell, 2013) using the parameters -mismatches 0 -ranmax 5 -mincov 2rpm. The resulting files were used to identify siRNA and miRNA loci as previously reported Lunardon et al. (2020). The predicted RNA secondary structure of the miRNA-like locus was visualized using strucVis (Michael J. Axtell at https://github.com/MikeAxtell/strucVis). For CHS siRNA analysis, the siRNAs mapped to a specific CHS gene, but not to other CHS genes, were defined as siRNAs derived from this CHS gene.

Bioinformatics Analysis of mRNA-Seq

The paired-end reads of mRNA-seq were mapped to the Tianlong1 genome using STAR (v2.7.2a; Dobin et al., 2013) with the parameters --outFilterMultimapScoreRange 0 --outSAMattributes Standard --alignIntronMin 20 --alignIntronMax 12,000. The PCR duplication reads were removed by Picard (v2.18.22-SNAPSHOT). The uniquely mapped reads of 20 to 30 nucleotides in length were de-methylated using Bismark (v0.19.0; Liao et al., 2014) with parameters -O -p. Due to the duplication of CHS genes, the reads specifically mapped to a specific CHS gene, but not to other CHS genes, were defined as reads derived from this CHS gene. The segments per kilobase of exon model per million mapped reads values were calculated based on read counts by the edgeR package (v3.22.5; Robinson et al., 2010).

De Novo Assembly of the Tianlong1 Genome

Genomic DNA was isolated and sequenced by BGI-Shenzhen. The 20-kb library was sequenced on the PacBio Sequel II System with 80× coverage. The 270- and 500-bp paired-end Illumina libraries were constructed by BGI and sequenced using the HiSeq X Ten platform with 100× coverage. A total of 0.5 g of leaves of 40-d-old plants were collected and used for in situ Hi-C library construction, following the procedure previously described, with some modification, by Moissiard et al. (2012). In brief, the homogenized tissues were fixed in 1% (v/v) formaldehyde and stopped with 0.125 M Gly. Then, homogenate was filtered through two layers of Miracloth. After centrifugation, the nuclei were digested and ligated in situ as previously described. After reverse-cross-link and DNA purification, the DNA fragmentation was performed by Tn5 using the Vazyme kit (TruePrep DNA Library Prep Kit V2 for Illumina, TD501) and stopped with 0.2% (w/v) SDS at
55°C for 15 min. Biotinylated DNA fragments were pulled down and amplified. The Hi-C library was sequenced on the HiSeq X Ten platform using 150-bp paired-end mode.

De novo assembly was conducted according to a reported pipeline (Shen et al., 2018) with minor modifications. In brief, the PacBio reads were assembled to primary contigs using Canu (v1.7.1; Koren et al., 2017). The primary contigs were polished using SMRT Link (v6.0.0) and were corrected by whole-genome resequencing data using Pilon (v1.23; Walker et al., 2014). The contigs had an N50 (given a set of contigs arranged from long to short, N50 is defined as the sequence length of the n-th contig, where the total length of the first n contigs is equal to half of the total length of all contigs) up to 4.34 Mb and were further assembled into chromosome with Hi-C data using 3D DNA find (Dudchenko et al., 2017) and Juicebox Assembly Tools (Durand et al., 2016).

IR and TE Identification of the Tianlong1 Genome
IR identification was performed by running einverted (EMBOSS toolkit, v6.6.0) with parameters -threshold 150 -match 3 -mismatch 4 -gap 12 -max-repeat 8500. TEs were identified using RepeatMasker (A.F.A. Smii, R. Hubley, and P. Green at http://repeatmasker.org) with the parameter -nolow -no_is and a Glycine max-specific library of repeat sequences.

Accession Numbers
The genome assembly and annotation of Tianlong1 (v1.0) are available at the National Center for Biotechnology Information with accession number PRJNA645754 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA645754). The sRNA-seq, mRNA-seq, and WGBS data generated in this study were deposited at the National Center for Biotechnology Information under accession number PRJNA644259 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA644259) and also hosted at http://ipf.sustech.edu.cn/priv/Tianlong1_public with Tianlong1 as reference. The sRNA data mapped to the Williams 82 genome is also available at the MPSS plant sRNA website: https://mpss.danforthcenter.org/dbs/index.php?SITE=soy_sRNA_ddc2.

Supplemental Data
Supplemental Figure 1. The phylogenetic analysis of DCL proteins and the growth phenotype of the Gmdcl2a/2b mutants.
Supplemental Figure 2. TE-derived 22-nucleotide siRNAs in leaf.
Supplemental Table 1. The list of genes showing differential mRNA expression between wild type and Gmdcl2a/2b-8 mutant in seed coat.
Supplemental Table 2. Primer sequences.
Supplemental Data Set 1. The list of genes showing differential mRNA expression between wild type and Gmdcl2a/2b-8 mutant in seed coat.
Supplemental Data Set 2. The list of genes showing differential mRNA expression between wild type and Gmdcl2a/2b-8 mutant in leaf.
Supplemental File 1. Multiple sequence alignment fasta file for Supplemental Figure 1A.
Supplemental File 2. Newick tree file for Supplemental Figure 1A.
Supplemental File 3. The unprocessed images for Figure 1B and Supplemental Figure 1C.

ACKNOWLEDGMENTS
We thank the Cryo-EM center of Southern University of Science and Technology for providing server used for our data analysis work. We are grateful for the useful comments and edits suggested by the anonymous reviewers. This work was supported by the National Key R&D Program of China (grant 2019YFA0093003), the Program for Guangdong Introducing Innovative and Entrepreneurial Teams (grant 2016ZT06S172), the Shenzhen Sci-Tech Fund (grant KTYDPT20181011104005), the Key Laboratory of Molecular Design for Plant Cell Factory of Guangdong Higher Education Institutes (grant 2019KSYS006), the National Natural Science Foundation of China (grant 31871234 to J.Z.), the Agricultural Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences, and the Central Public-Interest Scientific Institution Basal Research Fund.

REFERENCES
Allen, E., Xie, Z., Gustafson, A.M., and Carrington, J.C. (2005). MicroRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell 121:207–221.
Allen, E., Xie, Z., Gustafson, A.M., Sung, G.-H., Spatafora, J.W., and Carrington, J.C. (2004). Evolution of microRNA genes by inverted duplication of target gene sequences in Arabidopsis thaliana. Nat. Genet. 36:1282–1290.
Ariket, S., Xia, R., Kakrania, A., Huang, K., Zhai, J., Yan, Z., Valdés-López, O., Prince, S., Musket, T.A., Nguyen, H.T., Stacey, G., and Meyers, B.C. (2014). An atlas of soybean small RNAs identifies phased siRNAs from hundreds of coding genes. Plant Cell 28:4584–4601.
Axtell, M.J. (2013). ShortStack: Comprehensive annotation and quantification of small RNA genes. RNA 19:740–751.
Axtell, M.J., Jan, C., Rajagopal, R., and Bartel, D.P. (2006). A two-hit trigger for siRNA biogenesis in plants. Cell 127:556–577.
Bouché, N., Lauressergues, D., Gasciolli, V., and Vaucheret, H. (2006). An antagonistic function for Arabidopsis DCL2 in development and a new function for DCL4 in generating viral siRNAs. EMBO J. 25:3347–3356.
Chen, H.-M., Chen, L.-T., Patel, K., Li, Y.-H., Baulcombe, D.C., and Wu, S.-H. (2010). 22-Nucleotide RNAs trigger secondary siRNA biogenesis in plants. Proc. Natl. Acad. Sci. USA 107:15269–15274.
Chen, H.-M., Li, Y.-H., and Wu, S.-H. (2007). Bioinformatic prediction and experimental validation of a microRNA-directed tandem trans-acting siRNA cascade in Arabidopsis. Proc. Natl. Acad. Sci. USA 104:3318–3323.
Cho, Y.B., Jones, S.I., and Vodkin, L. (2013). The transition from primary siRNAs to amplified secondary siRNAs that regulate chalcone synthase during development of Glycine max seed coats. PLoS One 8:e76954.
Cho, Y.B., Jones, S.I., and Vodkin, L.O. (2017). Mutations in Argo-naut6 illuminate epistatic interactions of the K7 and f loci leading to saddle seed color patterns in Glycine max. Plant Cell 29:708–725.
Cho, Y.B., Jones, S.I., and Vodkin, L.O. (2019). Nonallelic homologous recombination events responsible for copy number variation within an RNA silencing locus. Plant Direct 3:e00162.
Clough, S.J., Tuteja, J.H., Li, M., Marek, L.F., Shoemaker, R.C., and Vodkin, L.O. (2004). Features of a 103-kb gene-rich region in soybean include an inverted perfect repeat cluster of CHS genes comprising the l locus. Genome 47: 819–831.

Cuperus, J.T., Carbonell, A., Fahlgren, N., Garcia-Ruiz, H., Burke, R.T., Takeda, A., Sullivan, C.M., Gilbert, S.D., Montgomery, T.A., and Carrington, J.C. (2010). Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in Arabidopsis. Nat. Struct. Mol. Biol. 17: 997–1003.

Cuperus, J.T., Fahlgren, N., and Carrington, J.C. (2011). Evolution and functional diversification of MIRNA genes. Plant Cell 23: 431–442.

Ding, S.-W., and Voinnet, O. (2007). Antiviral immunity directed by the MERASE6/DICER-LIKE4 pathway in Arabidopsis reveals de novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science 356: 92–95.

Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., and Aiden, E.L. (2017). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 3: 95–98.

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

Fei, Q., Yu, Y., Liu, L., Zhang, Y., Baldrich, P., Dai, Q., Chen, X., and Meyers, B.C. (2018). Biogenesis of a 22-nt microRNA in Phaseolus species by precursor-programmed uridylation. Proc. Natl. Acad. Sci. USA 115: 8037–8042.

Gasciollii, V., Mallory, A.C., Bartel, D.P., and Vaucheret, H. (2005). Partially redundant functions of Arabidopsis DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs. Curr. Biol. 15: 1494–1500.

Henderson, I.R., Zhang, X., Lu, C., Johnson, L., Meyers, B.C., Green, P.J., and Jacobsen, S.E. (2006). Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. Nat. Genet. 38: 721–725.

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

Johnson, C., Kasprzewska, A., Tennessen, K., Fernandes, J., Nan, G.-L., Walbot, V., Sundaresan, V., Vance, V., and Bowman, L.H. (2009). Clusters and superclusters of phased small RNAs in the developing inflorescence of rice. Genome Res. 19: 1429–1440.

Kakrana, A., et al. (2018). Plant 24-nt reproductive phasiRNAs from intramolecular duplex mRNAs in diverse monochots. Genome Res. 28: 1333–1344.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27: 722–736.

Kuo, H.Y., Jacobsen, E.L., Long, Y., Chen, X., and Zhai, J. (2017). Characteristics and processing of Pol IV–dependent transcripts in Arabidopsis. J. Genet. Genomics 44: 3–6.

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

Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. 11: 204–220.

Lei, Y., Lu, L., Liu, H.-Y., Li, S., Xing, F., and Chen, L.-L. (2014). CRISPR-P: A web tool for synthetic single-guide RNA design of CRISPR-system in plants. Mol. Plant 7: 1494–1496.

Li, F., Pignatta, D., Bendix, C., Brunkard, J.O., Cohn, M.M., Tung, J., Sun, H., Kumar, P., and Baker, B. (2012). MicroRNA regulation of plant innate immune receptors. Proc. Natl. Acad. Sci. USA 109: 1790–1795.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30: 923–930.

Liu, Y., et al. (2020). Pan-genome of wild and cultivated soybeans. Cell 182: 162–176.e13.

Lunardon, A., Johnson, N.R., Hagerott, E., Phifer, T., Polodyre, S., Coruh, C., and Axtell, M.J. (2020). Integrated annotations and analyses of small RNA-producing loci from 47 diverse plants. Genome Res. 30: 497–513.

Manavella, P.A., Koenig, D., and Weigel, D. (2012). Plant secondary siRNA production determined by microRNA-duplex structure. Proc. Natl. Acad. Sci. USA 109: 2461–2466.

Martin, M. (2011). Cutadapt reduces adapter sequences from high-throughput sequencing reads. EMBNet. J. 17: 10–12.

Moissiard, G., et al. (2012). MORC family ATPases required for heterochromatin condensation and gene silencing. Science 336: 1448–1451.

Nagano, H., Fukudome, A., Hiraguri, A., Moriyama, H., and Fukuhara, T. (2014). Distinct substrate specificities of Arabidopsis DCL3 and DCL4. Nucleic Acids Res. 42: 1845–1856.

Nobuta, K., et al. (2008). Distinct size distribution of endogenous siRNAs in maize: Evidence from deep sequencing in the mop1-1 mutant. Proc. Natl. Acad. Sci. USA 105: 14958–14963.

Parent, J.-S., Bouteiller, N., Elmayan, T., and Vaucheret, H. (2015). Respective contributions of Arabidopsis DCL2 and DCL4 to RNA silencing. Plant J. 81: 223–232.

Paz, M.M., Martinez, J.C., Kalvig, A.B., Fonger, T.M., and Wang, K. (2006). Improved cotyledonary node method using an alternative explant derived from mature seed for efficient Agrobacterium-mediated soybean transformation. Plant Cell Rep. 25: 206–213.

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

Rogers, K., and Chen, X. (2013). Biogenesis, turnover, and mode of action of plant microRNAs. Plant Cell 25: 2383–2399.

Schmutz, J., et al. (2010). Genome sequence of the palaeopolyploid soybean. Nature 463: 178–183.

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J.J., Qiu, J.-L., and Gao, C. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. Nat. Biotechnol. 31: 686–688.

Shen, Y., Liu, J., Geng, H., Zhang, J., Liu, Y., Zhang, H., Xing, S., Du, J., Ma, S., and Tian, Z. (2018). De novo assembly of a Chinese soybean genome. Sci. China Life Sci. 61: 871–884.

Shivaprasad, P.V., Chen, H.-M., Patel, K., Bond, D.M., Santos, B.A.C.M., and Baulcombe, D.C. (2012). A microRNA superfamily regulates nucleotide binding site-leucine-rich repeats and other mRNAs. Plant Cell 24: 859–874.

Song, X., et al. (2012). Roles of DCL4 and DCL3b in rice phased small RNA biogenesis. Plant J. 69: 462–474.

Song, X., Li, Y., Cao, X., and Qi, Y. (2019). MicroRNAs and their regulatory roles in plant-environment interactions. Annu. Rev. Plant Biol. 70: 489–529.
Taochy, C., Gursanscky, N.R., Cao, J., Fletcher, S.J., Dressel, U., Mitter, N., Tucker, M.R., Kolatunow, A.M.G., Bowman, J.L., Vaucheret, H., and Carroll, B.J. (2017). A genetic screen for impaired systemic RNAi highlights the crucial role of DICER-LIKE 2. Plant Physiol. 175: 1424–1437.

Tuteja, J.H., Clough, S.J., Chan, W.-C., and Vodkin, L.O. (2004). Tissue-specific gene silencing mediated by a naturally occurring chalcone synthase gene cluster in Glycine max. Plant Cell 16: 819–835.

Tuteja, J.H., Zabala, G., Varala, K., Hudson, M., and Vodkin, L.O. (2009). Endogenous, tissue-specific short interfering RNAs silence the chalcone synthase gene family in Glycine max seed coats. Plant Cell 21: 3063–3077.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakhikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., and Earl, A.M. (2014). Pilon: An integrated tool for comprehensive microvariant detection and genome assembly improvement. PLoS One 9: e112963.

Wang, T., et al. (2018a). Tomato DCL2b is required for the biosynthesis of 22-nt small RNAs, the resulting secondary siRNAs, and the host defense against ToMV. Hortic. Res. 5: 62.

Wang, Z., Hardcastle, T.J., Canto Pastor, A., Yip, W.H., Tang, S., and Baulcombe, D.C. (2018b). A novel DCL2-dependent miRNA pathway in tomato affects susceptibility to RNA viruses. Genes Dev. 32: 1155–1160.

Wu, H., et al. (2020). Plant 22-nt siRNAs mediate translational repression and stress adaptation. Nature 581: 89–93.

Wu, Y.-Y., Hou, B.-H., Lee, W.-C., Lu, S.-H., Yang, C.-J., Vaucheret, H., and Chen, H.-M. (2017). DCL2- and RDR6-dependent transitive silencing of SMXL4 and SMXL5 in Arabidopsis dcl4 mutants causes defective phloem transport and carbohydrate over-accumulation. Plant J. 90: 1064–1078.

Xi, Y., and Li, W. (2009). BSMAP: Whole genome bisulfite sequence MAPping program. BMC Bioinformatics 10: 232.

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

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

Xue, M., et al. (2019). A reference-grade wild soybean genome. Nat. Commun. 10: 1216.

Zhai, J., et al. (2011). MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. Genes Dev. 25: 2540–2553.

Zhai, J., Zhang, H., Arikit, S., Huang, K., Nan, G.-L., Walbot, V., and Meyers, B.C. (2015). Spatiotemporally dynamic, cell-type-dependent premeiotic and meiotic phased siRNAs in maize anthers. Proc. Natl. Acad. Sci. USA 112: 3146–3151.

Zhai, J., et al. (2013). Plant microRNAs display differential 3’ truncation and tailing modifications that are ARGONAUTE1 dependent and conserved across species. Plant Cell 25: 2417–2428.

Zhang, X., et al. (2015). Suppression of endogenous gene silencing by bidirectional cytoplasmic RNA decay in Arabidopsis. Science 348: 120–123.