Plant Cell Advance Publication. Published on October 19, 2020, doi:10.1105/tpc.20.00562

BREAKTHROUGH REPORT

Soybean Dicer-Like 2 Regulates Seed Coat Color via Production of Primary 22-nt Small Interfering RNAs from Long Inverted Repeats

Jinbu Jia[1,2,3†], Ronghuan Ji[4,†], Zhuowen Li[1,2,3], Yiming Yu[1,2,3], Mayumi Nakano[5], Yanping Long[1,2,3], Li Feng[1,2,3], Chao Qin[4], Dongdong Lu[1,2,3], Junpeng Zhan[1,2,3,5], Rui Xia[6], Blake C. Meyers[5,7], Bin Liu[4*], and Jixian Zhai[1,2,3*]

Affiliations:

[1] Department of Biology, Southern University of Science and Technology, Shenzhen 518055, China;

[2] Institute of Plant and Food Science, Southern University of Science and Technology, Shenzhen 518055, China;

[3] Key Laboratory of Molecular Design for Plant Cell Factory of Guangdong Higher Education Institutes, Southern University of Science and Technology, Shenzhen 518055, China;

[4] Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, 100081, China;

[5] Donald Danforth Plant Science Center, St. Louis, MO 63132, United States;

[6] College of Horticulture, South China Agricultural University, Guangzhou 510642, China;

[7] University of Missouri – Columbia, Division of Plant Sciences, Columbia, MO 65211, United States.

[†] These authors contributed equally to this work.
[*] Correspondence to: zhaijx@sustech.edu.cn and liubin05@caas.cn

Short title: Soybean DCL2 regulates seed coat color

One-sentence summary: Soybean DCL2 favors long inverted repeat-derived double-stranded RNA as its substrate, and generates primary 22-nt siRNAs to regulate seed coat color and silence transposable elements.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Bin Liu (liubin05@caas.cn).

ABSTRACT

In plants, 22-nucleotide (nt) small RNAs (sRNAs) trigger the production of secondary small interfering RNAs (siRNAs) and enhance silencing. DICER-LIKE 2 (DCL2)-dependent 22-nt siRNAs are rare in *Arabidopsis thaliana* and 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-nt siRNAs by applying CRISPR-Cas9 to simultaneously knock out the two copies of soybean *DCL2*, *GmDCL2a* and *GmDCL2b*, in the Tianlong1 cultivar. sRNA sequencing revealed that most 22-nt 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-nt siRNAs that target other *CHS* genes and trigger the production of secondary 21-nt 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 predominately processed by GmDCL2 into 22-nt 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 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 (ta-siRNAs) which 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 ta-siRNAs: the “two-hit” model for *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 s miRNA trigger to be precisely 22-nucleotide (nt) in length (Chen et al., 2010; Cuperus et al., 2010). The *TAS2* loci in Arabidopsis mostly generate secondary 21-nt siRNAs triggered by miR173, but the 22-nt tasiR2140 from *TAS2* 3’ D6(-) processed by DCL4 triggers tertiary siRNAs, although its functional significance is still unclear (Chen et al., 2010).

The 22-nt 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-nt miRNAs (Zhai et al., 2013; Fei et al., 2018). In addition, 22-nt noncanonical miRNAs can be produced by SlDCL2 in tomato (*Solanum lycopersicum*); these miRNAs in turn target *SlDCL2* mRNAs, forming a feedback loop (Wang et al., 2018b). In *Medicago truncatula*, soybean, tomato (*Solanum lycopersicum*) and many other plant species, 22-nt 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; Shivaprasad et al., 2012; Xia et al., 2013; Arikit et al., 2014). Moreover, the 22-nt miR2118 family can target thousands of non-coding transcripts in reproductive tissues of monocots (Johnson et al., 2009; Song et al., 2012; Zhai et al., 2015).

Compared to miRNAs, 22-nt siRNAs in plants are primarily produced by DCL2 and play essential roles in viral defense and transgene silencing (Gasciolli 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-nt 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-nt siRNAs under normal growth condition 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; Wang et al., 2018b). Loss-of-function of *DCL4* in Arabidopsis can

induce the production of a large amount of DCL2-dependent 22-nt 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; Wu et al., 2020). However, endogenous 22-nt siRNAs are abundant in many major crops such as soybean, rice, and maize, indicating their potential importance (Lunardon et al., 2020). For example, 22-nt siRNAs from transposable element (TE) regions are abundant in maize, especially in the *mediator of paramutation1-1* (*mop1-1*) background in which 24-nt siRNAs disappear (Nobuta et al., 2008). Yet, their functions remain unclear partly due to a lack of mutants that disrupt the production of 22-nt 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-nt siRNAs in soybean by studying *Gmdcl2* CRISPR-Cas9 lines with frame-shift mutations.

RESULTS

CRISPR-Cas9 editing of *DCL2a/2b* in *Glycine max*

To gain insight into the role of 22-nt endogenous siRNAs in *Glycine max*, we applied the CRISPR-Cas9 technology in soybean cultivar 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 sequence (Supplemental Figure 1A, 1B). These two genes are adjacent to each other on Chromosome 9, indicating 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 gRNA to simultaneously target a conserved coding

region on both genes (Figure 1A), and we carried out CRISPR-Cas9 mediated editing through tissue culture (see Methods for detail).

From multiple independent mutant lines, we selected three homozygous double-mutant lines, *Gmdcl2a/2b-8*, *Gmdcl2a/2b-9*, and *Gmdcl2a/2b-16*, for further studies (Figure 1A). All three lines harbor deletions around the target site that result in a frameshift and premature stop codon in the coding region of *GmDCL2a/2b* (Figure 1A). The mutant lines showed no obvious abnormality in plant architecture except for being slightly dwarf (Supplemental Figure 1C). 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). In our later work, this phenotype was of particular interest for studying the function of DCL2.

Impacts of DCL2 loss-of-function on sRNA biogenesis in soybean

To study the impacts of *GmDCL2a/2b* loss-of-function on sRNA biogenesis and gene regulation, we performed sRNA sequencing (sRNA-seq) and mRNA sequencing (mRNA-seq) in both wild type (Tianlong1) and the *Gmdcl2a/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-nt siRNAs, which decreased substantially in the *Gmdcl2a/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 analysis of sRNAs in soybean (Arikit et al., 2014; Lunardon et al., 2020),

the majority of siRNA loci in soybean features 24-nt siRNAs, but 21- and 22-nt loci demonstrated higher sRNA abundances (Figure 2B). In wild-type libraries, our analysis identified 1,304 22-nt siRNA loci from the seed coat and 1,371 from leaf sample, which accounted for only ~3% of the total number of loci but over 20% of total siRNA abundance (Figure 2B). Compared to the wild type, siRNAs from 22-nt siRNA loci decreased substantially in the *Gmdcl2a/2b* mutant (Figure 2B, 2C). For example, over 70% of the 22-nt siRNA loci showed at least a ten-fold decrease in abundance in the seed coat (Figure 2D). Also, siRNAs from some 21-nt siRNA loci, especially those that share sequence homology with 22-nt siRNA loci, were decreased in the mutant (Figure 2E), suggesting DCL2-dependent 22-nt sRNAs are capable of targeting in *trans* and triggering the biogenesis of secondary 21-nt siRNAs. These results indicated that, compared to the low level of 22-nt siRNAs found in Arabidopsis, soybean accumulates a large amount of endogenous 22-nt siRNAs whose biogenesis requires DCL2.

We noticed that a handful of 22-nt 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 transposable elements (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 resembles a miRNA precursor (Figure 2F). This locus produces three highly abundant sRNAs in wild-type (namely sRNA-A, sRNA-B, and sRNA-C in Figure 2F). While sRNA-B and sRNA-C nearly disappeared in the *Gmdcl2a/2b* mutant, the 22-nt 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-nt siRNAs have the potential to evolve into DCL1-dependent 22-nt miRNAs, thus shedding light on a potential path for the evolution of 22-nt miRNA via DCL2.

A large number of transposable elements are targeted by DCL2-dependent 22-nt siRNAs in soybean

In addition to their relatively large number, 22-nt siRNA loci in soybean have another unusual feature compared to Arabidopsis: that is, ~70% of these 22-nt siRNA loci are overlapping with transposable elements (TEs) (Figure 3A, Supplemental Figure 2A), suggesting a potential role in TE silencing. In particular, instead of a typical 24-nt length, the siRNAs from the Caulimovirus class of TEs are mainly 22-nt; we found that other TE families, such as CMC-EnSpm, also produced a large amount of 22-nt siRNAs (Figure 3B, Supplemental Figure 2B). These TE-derived 22-nt 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 (Figure 3B, 3C, Supplemental Figure 2B, 2C). In addition to 22-nt loci, we also identified dozens of highly abundant 21-nt 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-nt siRNA generating CMC-EnSpm family of TEs (Figure 3C, Supplemental Figure 2C). These findings are consistent with the hypothesis that 22-nt TE-derived siRNAs may target the homologous TE transcripts and trigger the generation of secondary 21-nt 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 there are other redundant mechanisms to suppress TE activities in soybean.

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-nt siRNAs are capable of triggering DNA methylation like the well-established role of 24-nt 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-nt TE loci is similar between wild-type and the *Gmdcl2a/2b* mutant (Supplemental Figure 3C). Thus, we conclude that 22-nt siRNAs are not required for the maintenance of DNA methylation at their targeted TEs in leaves.

GmDCL2 favors long inverted repeats (LIRs) 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-nt siRNA loci. We found that inverted repeats (IRs) are enriched at a much higher proportion at 22-nt loci, compared to 21- or 24-nt loci. For example, ~80% of 22-nt siRNA loci with a sRNA accumulation level higher than 50 TPM (transcripts per million) overlapped with IRs, compared to only 12% of 21-nt siRNA loci with similar abundances overlapping with IRs (Figure 4A). Our polyA-selected mRNA-seq data further showed that, for those IRs that are overlapping with 22-nt 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-nt siRNA loci have almost no detectable polyA signal (Figure 4B), as most 24-nt siRNAs are derived from Pol IV transcripts that do not have a polyA tail (Kuo et al., 2017). We also found that a smaller number of IRs mainly produced 21-nt siRNAs and are also transcribed by Pol II (Figure 4A, 4B).

A previous study showed that DCL4 prefers double-stranded RNA (dsRNA) substrates that are over 100 nt in length (Nagano et al., 2013). In addition, in a number of monocot species, 24-nt reproductive phasiRNAs 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-nt siRNA-generating IRs is much larger than that of 21-nt 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-nt siRNA-generating IRs were 254 nt and 1289 nt, 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 about 2800 nt and produced a large number of 22-nt 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 nt and produced a series of 21-nt 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 nt and shorter) are favored by DCL4 to produce 21-nt siRNAs.

DCL2-dependent 22-nt siRNAs regulate the seed coat color in soybean

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*, i^i, i^k, 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 i^i allele exhibits pigment on the hilum but not on the seed coat proper; the i^k allele

shows two-colored saddle pattern on seed coat; and the recessive *i* alleles feature fully pigmented seed coat (Tuteja et al., 2004; Cho et al., 2019). The i^{i} allele contains two similar inverted repeat clusters, and each cluster contains three *CHS* genes (*CHS1*, *CHS3*, and *CHS4*) (Tuteja et al., 2004; Tuteja et al., 2009) (as illustrated in Figure 5A). This locus is the source of siRNAs which target and silence other *CHS* genes in the genome in *trans*, and result in a yellow seed coat (Tuteja et al., 2004; Tuteja et al., 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 wild soybeans W05 showed that it also contains this large *CHS* cluster, suggesting 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 Wm82 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 (Shen et al., 2018; Xie et al., 2019; Liu et al., 2020) (as shown in Figure 5A). 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-nt siRNAs to silence other *CHS* genes (Xie et al., 2019). Recent studies found that AGO5 is involved in this process as a naturally occurring *ago5* mutant in soybean has altered color in the saddle region of the seed coat (Cho et al., 2017). DCL2 has not been implicated in this regulatory machinery as the majority of *CHS* siRNAs are 21-mers (Tuteja et al.,

2009). Yet, we inferred that 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).

To understand the role of DCL2 in regulating seed coat color, we *de novo* assembled the *I* locus in Tianlong1 (i^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 (expressed sequence tags) 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* inverted repeat region, connecting the last exon of the inserted subtilisin gene 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 nt (Figure 5A, 5B) in wild-type and disappeared in the Gm*dcl2a/2b* mutant (Figure 5A, 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-nt in length

and generated from both the sense and antisense strands (Figure 5B), indicating the possible involvement of an RNA-dependent RNA polymerase (RDR) in generating the dsRNA precursors. Interestingly, these 21-nt siRNAs also disappeared in the Gm*dcl2a/dcl2b* mutant (Figure 5C), suggesting that they are likely secondary siRNAs triggered by DCL2-dependent 22-nt *CHS* siRNAs (Figure 5D). This is consistent with a previous study that investigated sRNA profiles at ten different stages of seed coat development, and found that the 22-nt *CHS* siRNAs are more prevalent than the 21-nt siRNAs at the very early stages of seed coat development, including 4 to 24 days after flowering, whereas 21-nt *CHS* siRNAs become dominant at the later stages of seed coat development. The transition from 22-nt siRNAs at *CHS1/3* (*I* locus) to predominantly 21-nt secondary siRNAs at CHS7/*8* likely occurs at the 5-6 mg stages (Cho et al., 2013). In the *Gmdcl2a/dcl2b* mutant, the mRNA of *CHS2*, *CHS7*, and *CSH8* genes accumulate at much higher levels compared to the wild-type (Figure 5C), and the seed coat color changed from yellow to brown (Figure 1B). Our results demonstrated that the DCL2-dependent 22-nt siRNAs are required to maintain the silencing of the *CHS* gene family and regulate the color of seed coat in soybean (Figure 5D). In addition, by comparing mRNA-seq data of WT and *Gmdcl2a/2ab* 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 GmDCL2-dependent 22-nt siRNAs can regulate expression of genes besides the *CHS* family.

DISCUSSION

While our results showed that DCL2 can act on the transcripts containing long *CHS* inverted repeats from the *I* locus (i^i allele in Tianlong1 cultivar), 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 - a) 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 *I* locus, yet still have yellow seed coat (Liu et al., 2020); b) 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*, *AGO5*, or other genes involved in the silencing pathway, result in the inability to produce *CHS* siRNAs in W05 (Cho et al., 2017); c) the rearranged subtilisin promoter hypothesis cannot explain the two-color pattern observed in the i^i and i^k alleles, as both silencing (yellow) and non-silencing (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-nt siRNAs in soybean resembles the precursor of microRNAs in many ways: a single-stranded RNA transcribed by Pol II, forms a hairpin-like structure, and does not require the activity of RDR. Such types of IR precursors for siRNAs have also been observed for 24-nt reproductive phasiRNAs in some monocot species, including asparagus (*Asparagus officinalis*), lily (*Lilium maculatum*), and daylily (*Hemerocallis lilioasphodelus*) (Kakrana et al., 2018). Moreover, 22-nt siRNAs in soybean can also trigger the production of secondary 21-nt siRNAs, consistent with recent reports in Arabidopsis on the role of 22-nt siRNAs in amplifying silencing

signals (Wu et al., 2017; Wu et al., 2020). The parallels in the structure of precursor and the mode of function between the 22-nt miRNA and LIR-derived 22-nt siRNA imply a potential mechanism for the direct evolutionary descent of 22-nt miRNAs from 22-nt siRNAs. Our findings also support the previously proposed model that miRNA precursors can originate from local duplication that forms an inverted repeat structure (Allen et al., 2004; Cuperus et al., 2011). Therefore, the large number of LIR loci that generate 22-nt siRNAs in soybean could provide a rich foundation for miRNA origin.

While wild-type Arabidopsis generates few 22-nt siRNAs, we found that a large number of 22-nt siRNAs are produced mostly from LIR regions in soybean. These 22-mers are capable of triggering the production of secondary 21-nt 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-nt siRNAs at much higher levels compared to Arabidopsis (Lunardon et al., 2020). Our results here demonstrated that DCL2-dependent 22-nt 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-nt siRNAs in wild-type plants.

METHODS

Plant materials and growth conditions

For preparation of *Glycine max* leaf samples, the wild type Tianlong1 and CRISPR/Cas9-engineered mutant lines were grown under short-day conditions (10 h 300 μmol m^{-2} s^{-1} white light/14 h dark, 26 °C) in the greenhouse and the leaves of 10-day-old seedlings were

collected at ZT8 (8 hours after light exposure). For preparation of seed coat samples, the soybean plants were grown under long-day conditions (16 h 300 μmol m^{-2} s^{-1} white light/8 h 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 ZT11 (11 hours after light exposure). The RNA and DNA used in mRNA-seq, sRNA-seq and WGBS were isolated from tissues of wild-type (Tianlong1) and the *Gmdcl2a/2b-8* mutant. The seeds can be obtained by contacting the group of Dr. 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://cbi.hzau.edu.cn/crispr/) (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 X*ba*I and B*gl* II sites by the infusion technology (Clontech). The constructed GmDCL2-gRNA binary vector was introduced into the *Agrobacterium tumefaciens* strain EHA105 and transformed into WT 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 Supplemental Table 2.

Bi-allelic (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) are bi-allelic in both *GmDCL2a* and *GmDCL2b* loci, based on the genotyping of 23 T1 transgenic lines (all have brown seed coat) - ten lines were homozygous but in two kinds of genotypes (6 lines : 4 lines), 13 are still bi-allelic. Both of the T0 transgenic line #9 and line #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, and we named them *Gmdcl2a/2b-8*, *Gmdcl2a/2b-9*, and *Gmdcl2a/2b-16*, respectively.

RNA extraction and sequencing of mRNAs and sRNAs

The RNA samples of the wild-type (Tianlong1) and the *Gmdcl2a/2b-8* mutant were prepared by Spectrum Plant Total RNA Kit (Sigma, STRN50) according to the manufacturer's instructions. The strand-specific mRNA and sRNA libraries were prepared and sequenced at the BGI-Shenzhen.

Whole-genome bisulfite sequencing (WGBS) and analysis

Extraction of genomic DNA and the construction of whole-genome bisulfite sequencing 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 20-30 nt in size 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 to the mitochondrial and chloroplast genome (allowing two mismatches), the clean reads of each library were individually mapped to Tianlong1 genome by ShortStack (v3.8.5) (Axtell, 2013) using the parameters: -mismatches 0 --ranmax 5 –mincov 2rpm. And the result files were used to identify siRNA and miRNA loci as previously reported (Lunardon et al., 2020). The predicted RNA secondary structure of 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 pair-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 12000. The PCR duplication reads were removed by Picard (v2.18.22-SNAPSHOT). The uniquely mapped read count of genes, TEs, and IRs were extracted by featureCounts (v2.0.0) (Liao et al., 2014) with parameters -O -p. Due to the duplication of *CHS* genes, the read specifically mapped to a specific CHS gene, but not to other *CHS* genes were defined as read derived from this *CHS* gene. The FPKM (fragments per kilobase

of exon model per million mapped reads) values were calculated based on read count by edgeR packages (v3.22.5) (Robinson et al., 2010).

De novo assembly of the Tianlong1 genome

Genomic DNA was isolated and sequenced by BGI-Shenzhen. The 20kb library was sequenced on PacBio Sequel II System with 80X coverage. The 270-bp and 500-bp pair-end Illumina libraries were constructed by BGI and sequenced using HiSeq X Ten platform with 100X coverage. 0.5g leaves of 40-day-old plants were collected and used for *in situ* Hi-C library construction, following the procedure previously described with some modification (Moissiard et al., 2012). In brief, the homogenized tissues were fixed in 1% (vol/vol) formaldehyde and stopped with 0.125 M glycine. Then homogenate was filtered through two layers of miracloth. After centrifugation, the nuclei were digested and ligated *in situ* as previously described. After reverse-crosslink 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% SDS at 55℃ for 15min. Biotinylated DNA fragments were pulled down and amplified. The Hi-C library was sequenced on 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 re-sequencing data using Pilon(v1.23) (Walker et al., 2014). The

contigs had an N50 up to 4.34 Mb and were further assembled into chromosome with Hi-C data using 3D DNA (Dudchenko et al., 2017) and Juicebox Assembly Tools (Durand et al., 2016).

IR and TE identification of Tianlong1 genome

Inverted repeat (IR) identification was performed by running einverted (EMBOSS toolkit, v6.6.0) with parameter -threshold 150 -match 3 -mismatch -4 -gap 12 -maxrepeat 8500. And TEs were identified using RepeatMasker (A.F.A. Smit, R. Hubley & P. Green at http://repeatmasker.org) with the parameter "-nolow -no_is" and *Glycine max* specific library of repeat sequences.

Accession numbers

The genome assembly and annotation of Tianlong1 (v1.0) are available at NCBI with the accession number PRJNA645754. The sRNA-seq, mRNA-seq, and WGBS data generated in this study were deposited at NCBI under the accession number PRJNA644259 (https://dataview.ncbi.nlm.nih.gov/object/PRJNA644259?reviewer=4dqf4e0s98tsr4i017osb8784 9), and also hosted at http://ipf.sustech.edu.cn/priv/Tianlong1_public with Tianlong1 as reference. The sRNA data mapped to the Wm82 genome is also available at the MPSS plant sRNA website: https://mpss.danforthcenter.org/dbs/index.php?SITE=soy_sRNA_dcl2.

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-nt siRNAs in leaf.

Supplemental Table 1. The list of 22-nt siRNA loci with 22-nt siRNA TPM more than 5 and not decreased in *Gmdcl2a/dcl2b* mutant.

Supplemental Table 2. Primer sequences.

Supplemental Data Set 1. The list of genes showing differential mRNA expression between wild type and *Gmdcl2a/dcl2b-8* mutant in seed coat.

Supplemental Data Set 2. The list of genes showing differential mRNA expression between wild type and *Gmdcl2a/dcl2b-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 are grateful for the useful comments and edits suggested by the anonymous reviewers. Group of J.Z. is supported by the National Key R&D Program of China Grant (2019YFA0903903), the Program for Guangdong Introducing Innovative and Entrepreneurial Teams (2016ZT06S172), the Shenzhen Sci-Tech Fund (KYTDPT20181011104005), the Key Laboratory of Molecular Design for Plant Cell Factory of Guangdong Higher Education Institutes (2019KSYS006), and an NSFC to J.Z. (31871234),. Group of B.L. is supported by the Agricultural Science and Technology Innovation Program (ASTIP) of Chinese Academy of Agricultural Sciences, and the Central Public-Interest Scientific Institution Basal Research Fund.

Author Contributions

J.Zhai, B.L., J.J., and R.J. designed the research. J.J., R.J., Y.L., C.Q., D.L. and J.Zhan performed the experiments. J.J., Z.L., Y.Y., M.N. and L.F. analyzed the data. J.Zhai, B.L. and R.X. oversaw the study. J.J., R.J., J.Zhai and B.L. wrote the manuscript, and all authors revised the manuscript.

FIGURE LEGENDS

Figure 1. CRISPR/Cas9-engineered mutations in *GmDCL2a/2b* result in a brown seed coat.

(**A**) Deletion mutations at target sites in the *Gmdcl2a/2b* mutants. The gRNA target and PAM sequences are highlighted with bold in black and blue, respectively. The gene structures of *GmDCL2a* and *GmDCL2b* are shown above. The coding sequence (CDS) is shown as dark box, the untranslated region (UTR) is shown as grey box, and the intron region is shown as a line.

(**B**) Seed coat color phenotype of the *Gmdcl2a/2b* mutants. The scale bar represents 1 cm.

Figure 2 The accumulation levels of 22-nt siRNAs are sharply decreased in the *Gmdcl2a/2b* mutant.

(**A**) The length distribution of siRNAs in wild type (WT) and the *Gmdcl2a/2b* mutant. 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.

(**C**) The length distribution of sRNAs from loci predominantly generating 21-nt, 22-nt or 24-nt siRNAs.

(**D**) The comparison of sRNA accumulation levels of 22-nt siRNA loci in seed coat between wild type and the *Gmdcl2a/2b* mutant. The arrow and circle mark a miRNA-like locus shown in (**F**). The dash lines indicate a sRNA fold change of 10.

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

(**F**) A miRNA-like locus. The potential miRNA/miRNA* pairs and their expression levels in wild type and the *Gmdcl2a/2b* (*dcl2*) mutant were shown in the left panel. The depth of uniquely mapped mRNA reads and the accumulation levels of uniquely mapped sRNA reads in seed coat were shown in the right panel.

Figure 3 TE-derived 22-nt siRNAs in seed coat.

(**A**) TE enrichment analysis of 21-nt, 22-nt and 24-nt siRNA loci.

(**B**) Length distribution of sRNAs derived from different TE families. TPM: transcripts per million.

(**C**) Examples of 22-nt TE siRNA loci (siRNA loci overlapping with TE) and 21-nt TE siRNA loci. Only uniquely mapped RNA and sRNA reads are shown.

(**D**) Log2 fold change of sRNA and mRNA accumulation levels of TE siRNA loci in the *Gmdcl2a/2b* mutant compared with WT. The locus numbers (N) are indicated.

Figure 4 Inverted repeats derived 22-nt siRNAs.

(**A**) The IR enrichment analysis of 21-nt, 22-nt and 24-nt siRNA loci in seed coat samples.

(**B**-**C**) The RNA expression levels (FPKMs, fragments per kilobase of exon model per million mapped reads) (**B**) and repeat length (**C**) of the repeat region of IRs overlapped with 21-nt, 22-nt and 24-nt siRNA loci in seed coat samples. TPM: transcripts per million.

(**D**) The examples of 22-nt siRNA loci and 21-nt siRNA loci in seed coat samples. Only uniquely mapped mRNA and sRNA reads were shown.

Figure 5. 22-nt siRNAs suppress the expression of *CHS* genes which control the color of seed coat.

(**A**) The gene structures of *I* locus in wild soybean W05 genome as well as cultivated soybean ZH13 and Tianlong1 genome. The mRNA-seq and sRNA-seq results of Tianlong1 seed coat were showed below the gene structures. *: Due to the duplication of *CHS1* and *CHS3* gene, the read specifically mapped to *CHS1* or *CHS3*, but not to other *CHS* genes were also shown in "Uniquely mapped" panel. The junction read counts (n) are indicated. TPM: transcripts per million.

(**B**) The length distribution of *CHS* siRNAs in wild type.

(**C**) The sRNA and mRNA accumulation levels of *CHS* genes in the *Gmdcl2a/dcl2b* mutant and wild type. FPKM: fragments per kilobase of exon model per million mapped reads.

(**D**) A model of DCL2-dependent 22-nt siRNAs regulating the expression levels of *CHS* genes.

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.

Arikit, S., Xia, R., Kakrana, 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 **26,** 4584-4601.

Axtell, M.J. (2013). ShortStack: comprehensive annotation and quantification of small RNA genes. RNA **19,** 740-751.

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

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., 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 U S A. **104,** 3318-3323.

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 U S A. **107,** 15269-15274.

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 *Argonaute5* Illuminate Epistatic Interactions of the *K1* and *I* 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 CHSgenes comprising the Ilocus. Genome **47**, 819-831.

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

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.

Ding, S.-W., and Voinnet, O. (2007). Antiviral immunity directed by small RNAs. Cell **130,** 413-426.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics **29,** 15-21.

Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., and Aiden, E.L. (2017). 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., Rao, S.S.P., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). 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 Phaseoleae species by precursor-programmed uridylation. Proc Natl Acad Sci U S A. **115,** 8037-8042.

Gasciolli, 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., Mathioni, S.M., Huang, K., Hammond, R., Vandivier, L., Patel, P., Arikit, S., Shevchenko, O., Harkess, A.E., Kingham, B., Gregory, B.D., Leebens-Mack, J.H., and Meyers, B.C. (2018). Plant 24-nt reproductive phasiRNAs from intramolecular duplex mRNAs in diverse monocots. 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 U S A. **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., Du, H., Li, P., Shen, Y., Peng, H., Liu, S., Zhou, G.-A., Zhang, H., Liu, Z., Shi, M., Huang, X., Li, Y., Zhang, M., Wang, Z., Zhu, B., Han, B., Liang, C., and Tian, Z. (2020). Pan-Genome of Wild and Cultivated Soybeans. Cell **182**, 162-176.

Lunardon, A., Johnson, N.R., Hagerott, E., Phifer, T., Polydore, 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 U S A. **109,** 2461-2466.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal **17,** 10-12.

Moissiard, G., Cokus, S.J., Cary, J., Feng, S., Billi, A.C., Stroud, H., Husmann, D., Zhan, Y., Lajoie, B.R., McCord, R.P., Hale, C.J., Feng, W., Michaels, S.D., Frand, A.R., Pellegrini, M., Dekker, J., Kim, J.K., and Jacobsen, S.E. (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. (2013). Distinct substrate specificities of Arabidopsis DCL3 and DCL4. Nucleic acids Res. **42,** 1845-1856.

Nobuta, K., Lu, C., Shrivastava, R., Pillay, M., De Paoli, E., Accerbi, M., Arteaga-Vazquez, M., Sidorenko, L., Jeong, D.-H., Yen, Y., Green, P.J., Chandler, V.L., and Meyers, B.C. (2008). Distinct size distribution of endogeneous siRNAs in maize: Evidence from deep sequencing in the mop1-1 mutant. Proc Natl Acad Sci U S A. **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., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J., Cheng, J., Xu, D., Hellsten, U., May, G.D., Yu, Y., Sakurai, T., Umezawa, T., Bhattacharyya, M.K., Sandhu, D., Valliyodan, B., Lindquist, E., Peto, M., Grant, D., Shu, S., Goodstein, D., Barry, K., Futrell-Griggs, M., Abernathy, B., Du, J., Tian, Z., Zhu, L., Gill, N., Joshi, T., Libault, M., Sethuraman, A., Zhang, X.-C., Shinozaki, K., Nguyen, H.T., Wing, R.A., Cregan, P., Specht, J., Grimwood, J., Rokhsar, D., Stacey, G., Shoemaker, R.C., and Jackson, S.A. (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. Science China life sciences **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., Li, Y., Cao, X., and Qi, Y. (2019). MicroRNAs and Their Regulatory Roles in Plant-Environment Interactions. Annu Rev Plant Biol. **70,** 489-525.

Song, X., Li, P., Zhai, J., Zhou, M., Ma, L., Liu, B., Jeong, D.-H., Nakano, M., Cao, S., Liu, C., Chu, C., Wang, X.-J., Green, P.J., Meyers, B.C., and Cao, X. (2012). Roles of DCL4 and DCL3b in rice phased small RNA biogenesis. Plant J. **69,** 462-474.

Taochy, C., Gursanscky, N.R., Cao, J., Fletcher, S.J., Dressel, U., Mitter, N., Tucker, M.R., Koltunow, 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., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., and Earl, A.M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE **9,** e112963.

Wang, T., Deng, Z., Zhang, X., Wang, H., Wang, Y., Liu, X., Liu, S., Xu, F., Li, T., Fu, D., Zhu, B., Luo, Y., and Zhu, H. (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,** 62-14.

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., Li, B., Iwakawa, H.-O., Pan, Y., Tang, X., Ling-Hu, Q., Liu, Y., Sheng, S., Feng, L., Zhang, H., Zhang, X., Tang, Z., Xia, X., Zhai, J., and Guo, H. (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., 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.

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.

Xie, M., Chung, C.Y.-L., Li, M.-W., Wong, F.-L., Wang, X., Liu, A., Wang, Z., Leung, A.K.-Y., Wong, T.-H., Tong, S.-W., Xiao, Z., Fan, K., Ng, M.-S., Qi, X., Yang, L., Deng, T., He, L., Chen, L., Fu, A., Ding, Q., He, J., Chung, G., Isobe, S., Tanabata, T., Valliyodan, B., Nguyen, H.T., Cannon, S.B., Foyer, C.H., Chan, T.-F., and Lam, H.-M. (2019). A reference-grade wild soybean genome. Nat Commun. **10,** 1216-1212.

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 phasiRNAs in maize anthers. Proc Natl Acad Sci U S A. **112,** 3146-3151.

Zhai, J., Zhao, Y., Simon, S.A., Huang, S., Petsch, K., Arikit, S., Pillay, M., Ji, L., Xie, M., Cao, X., Yu, B., Timmermans, M., Yang, B., Chen, X., and Meyers, B.C. (2013). Plant microRNAs display differential 3' truncation and tailing modifications that are ARGONAUTE1 dependent and conserved across species. Plant Cell **25,** 2417-2428.

Zhai, J., Jeong, D.-H., De Paoli, E., Park, S., Rosen, B.D., Li, Y., González, A.J., Yan, Z., Kitto, S.L., Grusak, M.A., Jackson, S.A., Stacey, G., Cook, D.R., Green, P.J., Sherrier, D.J., and Meyers, B.C. (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.

Zhang, X., Zhu, Y., Liu, X., Hong, X., Xu, Y., Zhu, P., Shen, Y., Wu, H., Ji, Y., Wen, X., Zhang, C., Zhao, Q., Wang, Y., Lu, J., and Guo, H. (2015). Suppression of endogenous gene silencing by bidirectional cytoplasmic RNA decay in Arabidopsis. Science **348,** 120-123.

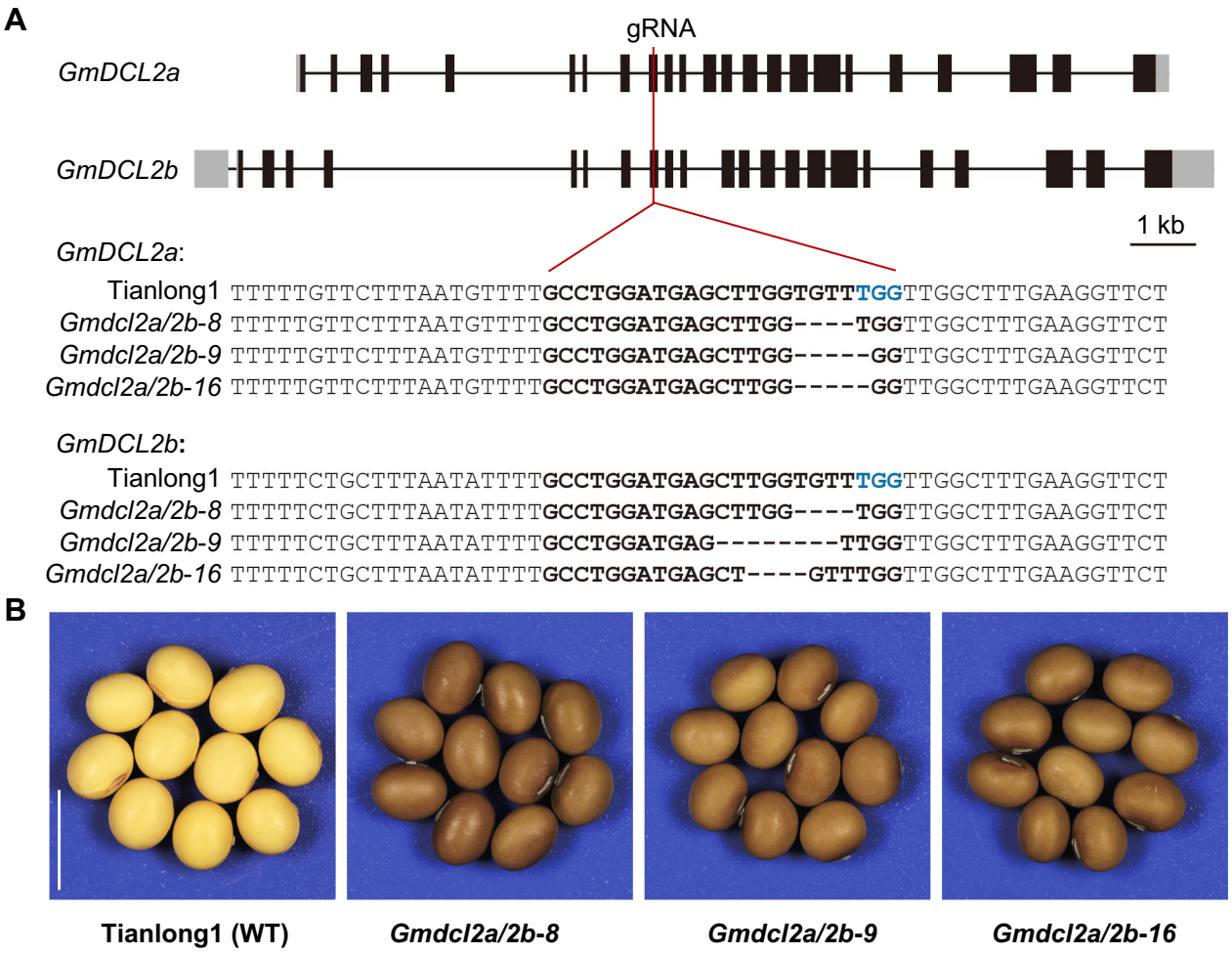

Figure 1. CRISPR/Cas9-engineered mutations in *GmDCL2a/2b* result in a brown seed coat.
(**A**) Deletion mutations at target sites in the *Gmdcl2a/2b* mutants. The gRNA target and PAM sequences are highlighted with bold in black and blue, respectively. The gene structures of *GmDCL2a* and *GmDCL2b* are shown above. The coding sequence (CDS) is shown as dark box, the untranslated region (UTR) is shown as grey box, and the intron region is shown as line.
(**B**) Seed coat color phenotype of the *Gmdcl2a/2b* mutants. The scale bar represents 1 cm.

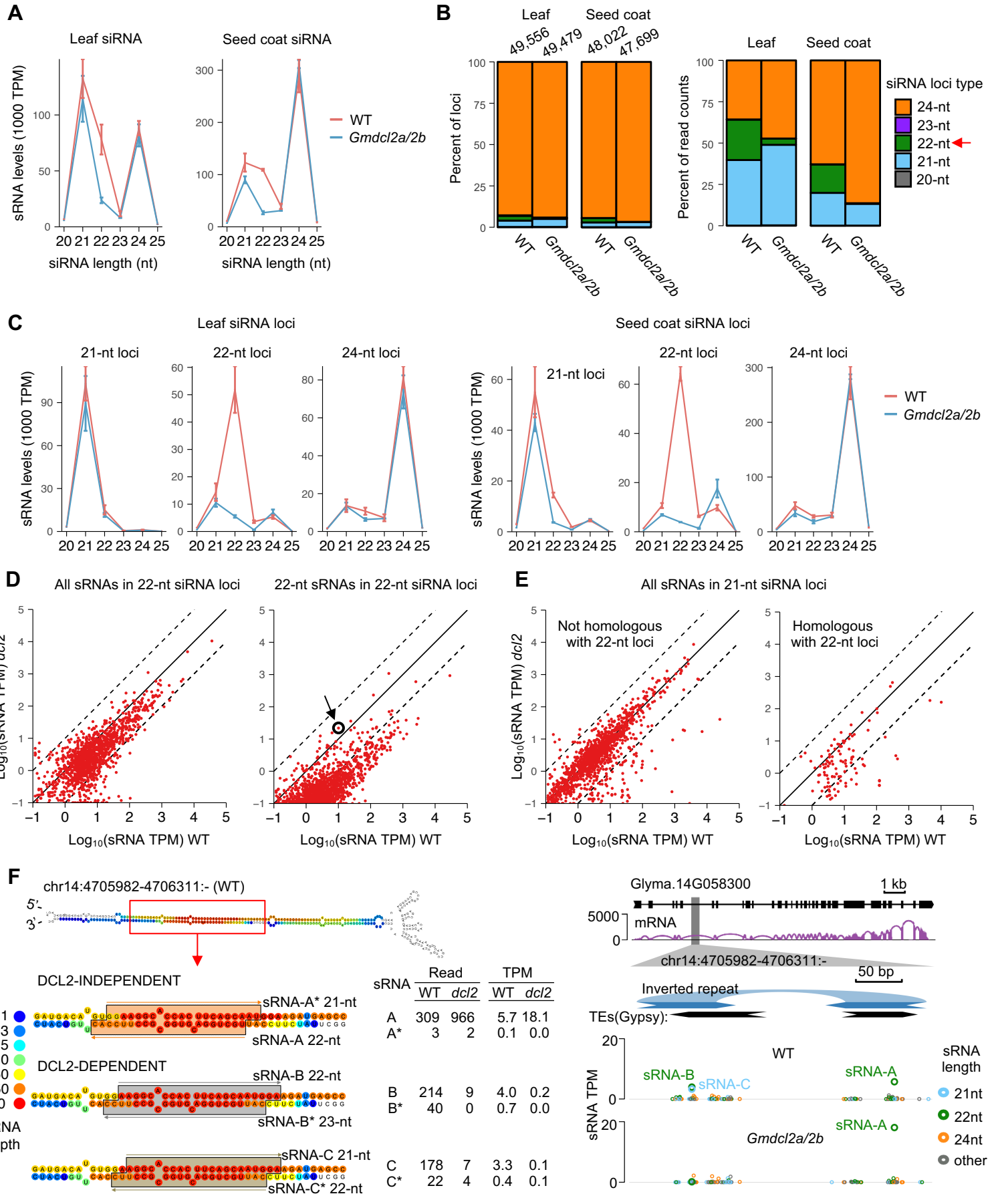
A
Leaf siRNA
Seed coat siRNA
sRNA levels (1000 TPM)
siRNA length (nt)
WT
Gmdcl2a/2b
B
Leaf
Seed coat
49,556
49,479
48,022
47,699
Percent of loci
Percent of read counts
siRNA loci type
24-nt
23-nt
22-nt
21-nt
20-nt
C
Leaf siRNA loci
Seed coat siRNA loci
21-nt loci
22-nt loci
24-nt loci
D
All sRNAs in 22-nt siRNA loci
22-nt sRNAs in 22-nt siRNA loci
Log10(sRNA TPM) dcl2
Log10(sRNA TPM) WT
E
All sRNAs in 21-nt siRNA loci
Not homologous with 22-nt loci
Homologous with 22-nt loci
F
chr14:4705982-4706311:- (WT)
DCL2-INDEPENDENT
DCL2-DEPENDENT
sRNA-A* 21-nt
sRNA-A 22-nt
sRNA-B 22-nt
sRNA-B* 23-nt
sRNA-C 21-nt
sRNA-C* 22-nt
sRNA depth
sRNA Read WT dcl2 TPM WT dcl2
A 309 966 5.7 18.1
A* 3 2 0.1 0.0
B 214 9 4.0 0.2
B* 40 0 0.7 0.0
C 178 7 3.3 0.1
C* 22 4 0.4 0.1
Glyma.14G058300
1 kb
mRNA
chr14:4705982-4706311:-
50 bp
Inverted repeat
TEs(Gypsy):
WT
Gmdcl2a/2b
sRNA TPM
sRNA length
21nt
22nt
24nt
other

Figure 2 The accumulation levels of 22-nt siRNAs are sharply decreased in the *Gmdcl2a/2b* mutant.
(A) The length distribution of siRNAs in wild type (WT) and the *Gmdcl2a/2b* mutant. 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.
(C) The length distribution of sRNAs from loci predominantly generating 21-nt, 22-nt or 24-nt siRNAs.
(D) The comparison of sRNA accumulation levels of 22-nt siRNA loci in seed coat between wild type and the *Gmdcl2a/2b* mutant. The arrow and circle mark a miRNA-like locus shown in (**F**). The dash lines indicate a sRNA fold change of 10.
(E) The comparison of sRNA accumulation levels of 21-nt siRNA loci in seed coat between wild type and the *Gmdcl2a/2b* mutant. The dash lines indicate a sRNA fold change of 10.
(F) A miRNA-like locus. The potential miRNA/miRNA* pairs and their expression levels in wild type and the *Gmdcl2a/2b* (*dcl2*) mutant were 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.

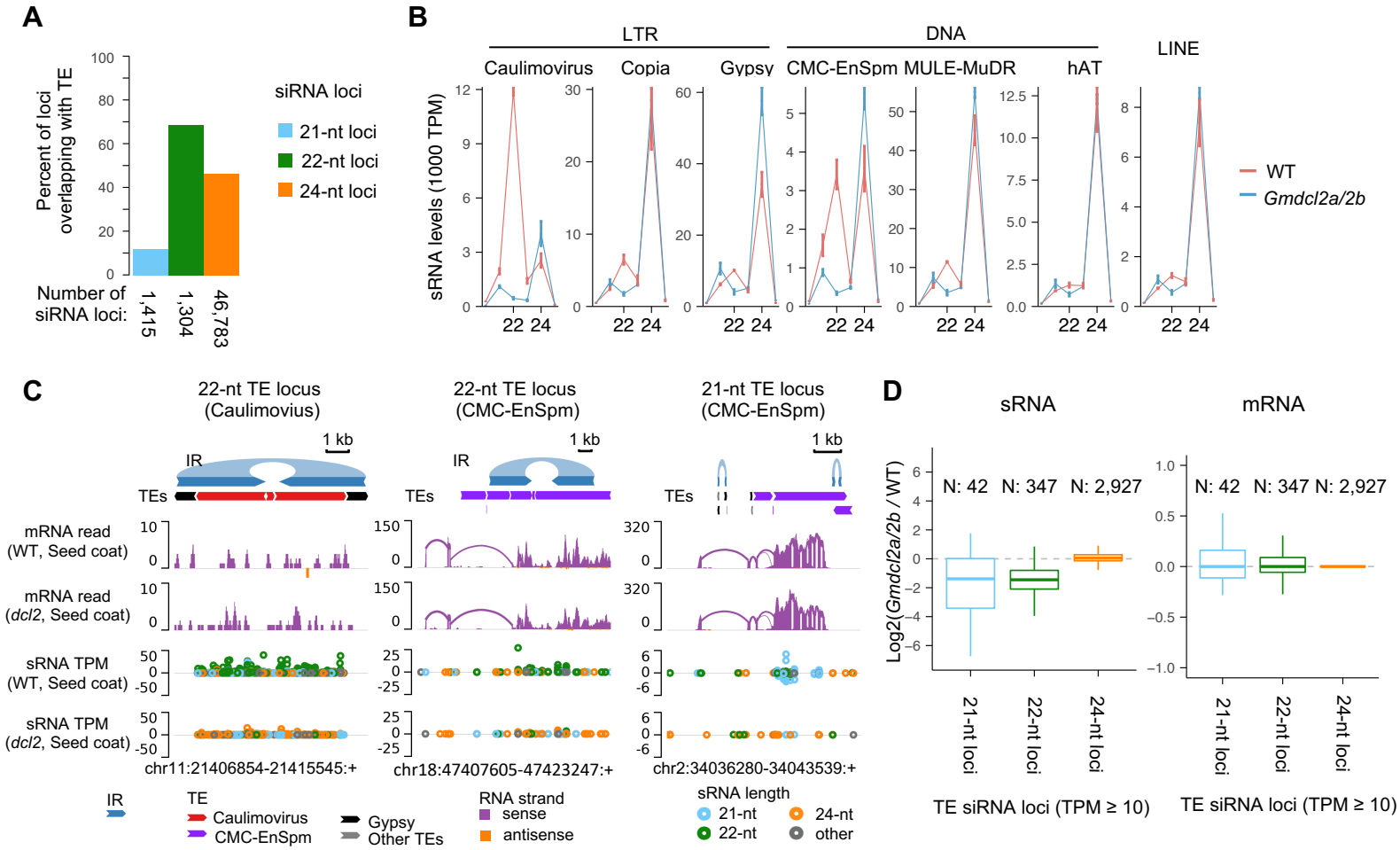

Figure 3 TE-derived 22-nt siRNAs in seed coat.

(**A**) TE enrichment analysis of 21-nt, 22-nt and 24-nt siRNA loci.

(**B**) Length distribution of sRNAs derived from different TE families. TPM: transcripts per million.

(**C**) Examples of 22-nt TE siRNA loci (siRNA loci overlapping with TE) and 21-nt TE siRNA loci. Only uniquely mapped RNA and sRNA reads were shown.

(**D**) Log2 fold change of sRNA and mRNA accumulation levels of TE siRNA loci in the *Gmdcl2a/2b* mutant compared with WT. The locus numbers (N) are indicated.

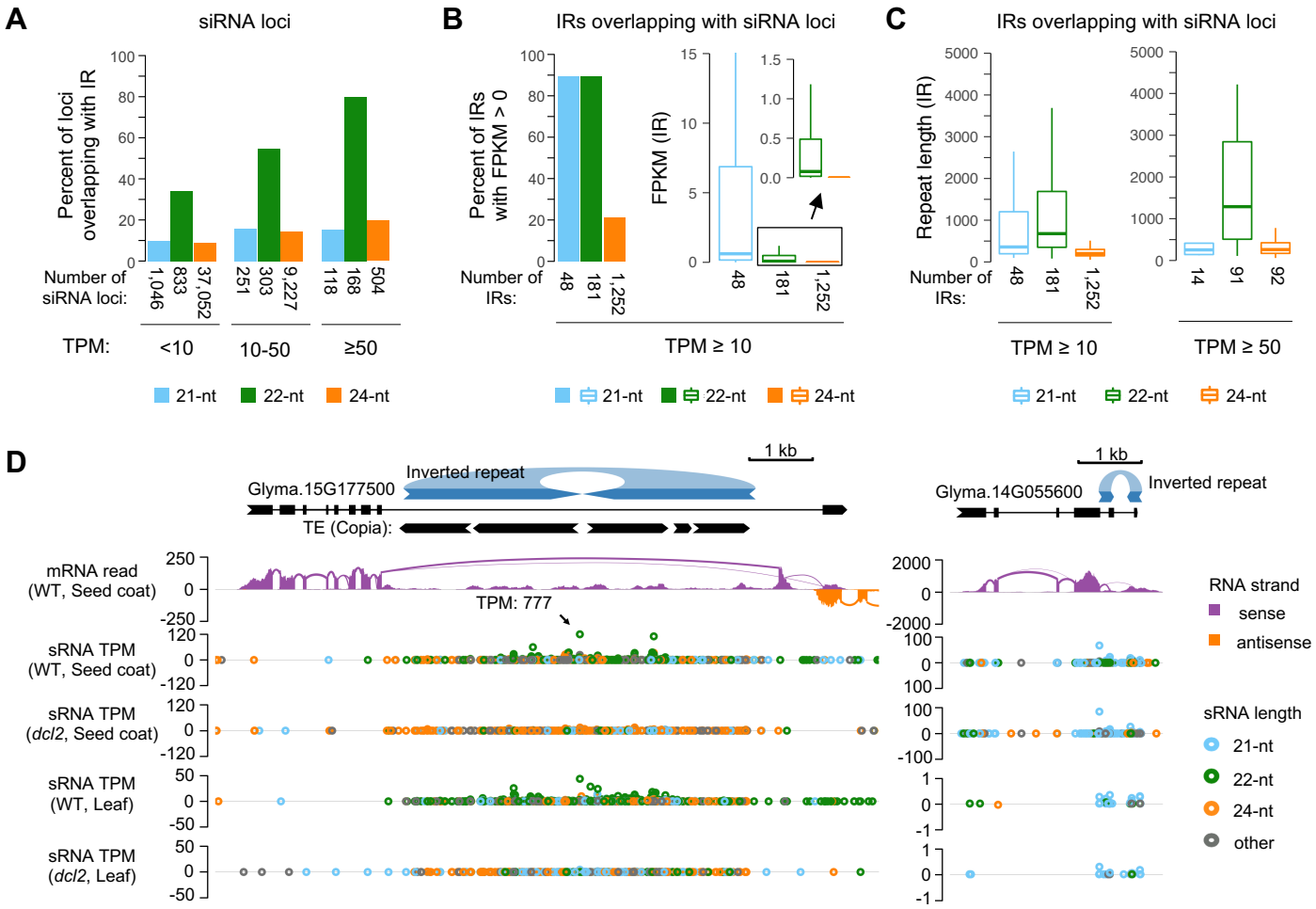

Figure 4 Inverted repeats derived 22-nt siRNAs.
(**A**) The IR enrichment analysis of 21-nt, 22-nt and 24-nt siRNA loci in seed coat samples.
(**B**-**C**) The RNA expression levels (FPKMs, fragments per kilobase of exon model per million mapped reads) (**B**) and repeat length (**C**) of the repeat region of IRs overlapped with 21-nt, 22-nt and 24-nt siRNA loci in seed coat samples. TPM: transcripts per million.
(**D**) The examples of 22-nt siRNA loci and 21-nt siRNA loci in seed coat samples. Only uniquely mapped mRNA and sRNA reads were shown.

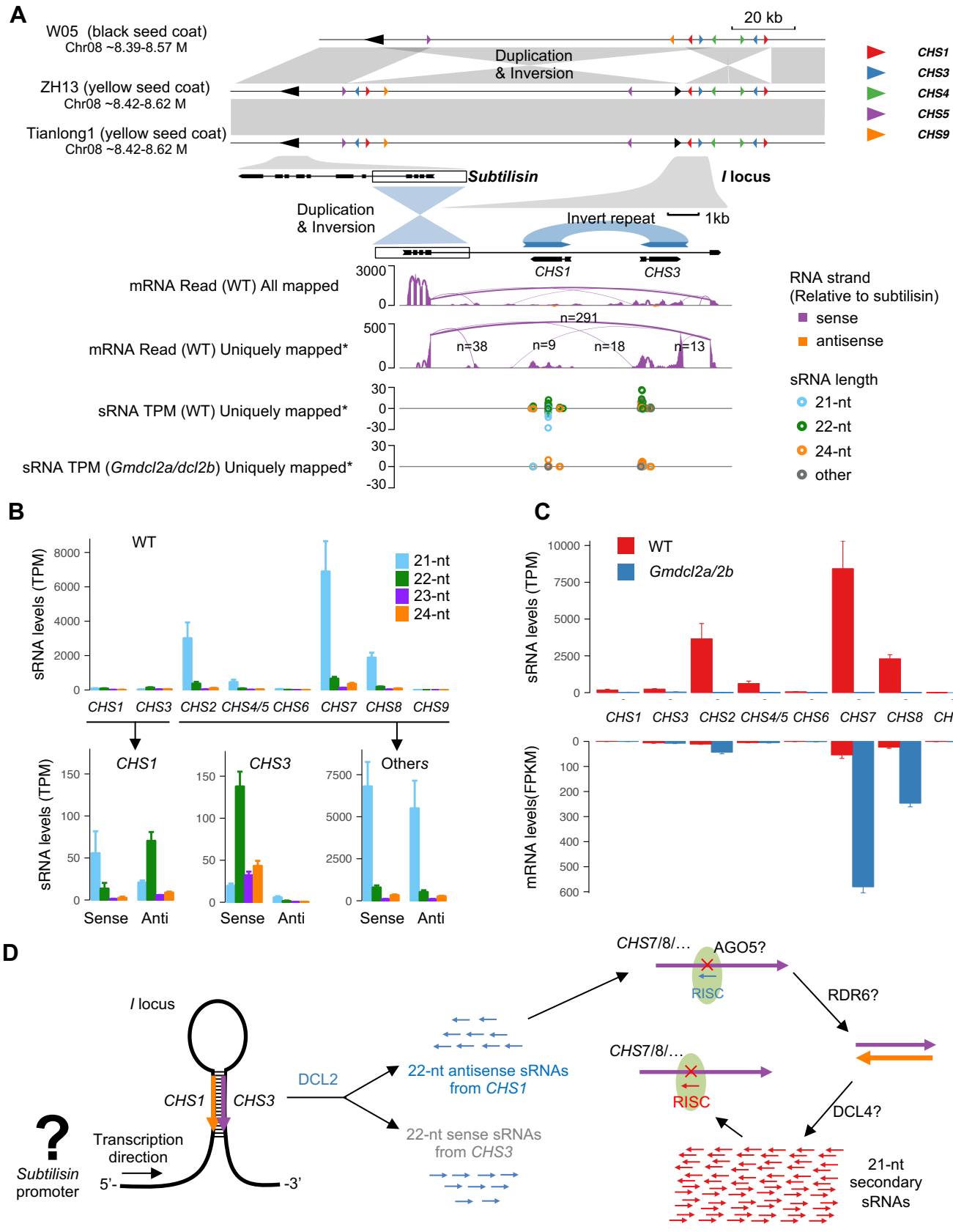

Figure 5. 22-nt siRNAs suppress the expression of *CHS* genes which control the color of seed coat.
(**A**) The gene structures of *I* locus in wild soybean W05 genome as well as cultivated soybean ZH13 and Tianlong1 genome. The mRNA-seq and sRNA-seq results of Tianlong1 seed coat were showed below the gene structures. *: Due to the duplication of *CHS1* and *CHS3* gene, the read specifically mapped to *CHS1* or *CHS3*, but not to other *CHS* genes were also shown in "Uniquely mapped" panel. The junction read counts (n) are indicated. TPM: transcripts per million.
(**B**) The length distribution of *CHS* siRNAs in wild type.
(**C**) The sRNA and mRNA accumulation levels of *CHS* genes in the *Gmdcl2a/dcl2b* mutant and wild type. FPKM: fragments per kilobase of exon model per million mapped reads.
(**D**) A model of DCL2-dependent 22-nt siRNAs regulating the expression levels of *CHS* genes.

Parsed Citations

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.
Google Scholar: Author Only Title Only Author and Title

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.
Google Scholar: Author Only Title Only Author and Title

Arikit, S., Xia, R., Kakrana, 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 26, 4584-4601.
Google Scholar: Author Only Title Only Author and Title

Axtell, M.J. (2013). ShortStack: comprehensive annotation and quantification of small RNA genes. RNA 19, 740-751.
Google Scholar: Author Only Title Only Author and Title

Axtell, M.J., Jan, C., Rajagopalan, R., and Bartel, D.P. (2006). A two-hit trigger for siRNA biogenesis in plants. Cell 127, 565-577.
Google Scholar: Author Only Title Only Author and Title

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.
Google Scholar: Author Only Title Only Author and Title

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 U S A. 104, 3318-3323.
Google Scholar: Author Only Title Only Author and Title

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 U S A. 107, 15269-15274.
Google Scholar: Author Only Title Only Author and Title

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.
Google Scholar: Author Only Title Only Author and Title

Cho, Y.B., Jones, S.I., and Vodkin, L.O. (2017). Mutations in Argonaute5 Illuminate Epistatic Interactions of the K1 and I Loci Leading to Saddle Seed Color Patterns in Glycine max. Plant Cell 29, 708-725.
Google Scholar: Author Only Title Only Author and Title

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.
Google Scholar: Author Only Title Only Author and Title

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 CHSgenes comprising the Ilocus. Genome 47, 819-831.
Google Scholar: Author Only Title Only Author and Title

Cuperus, J.T., Fahlgren, N., and Carrington, J.C. (2011). Evolution and functional diversification of MIRNA genes. Plant Cell 23, 431-442.
Google Scholar: Author Only Title Only Author and Title

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.
Google Scholar: Author Only Title Only Author and Title

Ding, S.-W., and Voinnet, O. (2007). Antiviral immunity directed by small RNAs. Cell 130, 413-426.
Google Scholar: Author Only Title Only Author and Title

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15-21.
Google Scholar: Author Only Title Only Author and Title

Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., and Aiden, E.L. (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science 356, 92-95.
Google Scholar: Author Only Title Only Author and Title

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst. 3, 95-98.
Google Scholar: Author Only Title Only Author and Title

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

Google Scholar: Author Only Title Only Author and Title

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 Phaseoleae species by precursor-programmed uridylation. Proc Natl Acad Sci U S A 115, 8037-8042.
Google Scholar: Author Only Title Only Author and Title

Gasciolli, 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.
Google Scholar: Author Only Title Only Author and Title

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.
Google Scholar: Author Only Title Only Author and Title

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.
Google Scholar: Author Only Title Only Author and Title

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.
Google Scholar: Author Only Title Only Author and Title

Kakrana, A., Mathioni, S.M., Huang, K., Hammond, R., Vandivier, L., Patel, P., Arikit, S., Shevchenko, O., Harkess, A.E., Kingham, B., Gregory, B.D., Leebens-Mack, J.H., and Meyers, B.C. (2018). Plant 24-nt reproductive phasiRNAs from intramolecular duplex mRNAs in diverse monocots. Genome Res. 28, 1333-1344.
Google Scholar: Author Only Title Only Author and Title

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.
Google Scholar: Author Only Title Only Author and Title

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.
Google Scholar: Author Only Title Only Author and Title

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.
Google Scholar: Author Only Title Only Author and Title

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.
Google Scholar: Author Only Title Only Author and Title

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.
Google Scholar: Author Only Title Only Author and Title

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 U S A 109, 1790-1795.
Google Scholar: Author Only Title Only Author and Title

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.
Google Scholar: Author Only Title Only Author and Title

Liu, Y., Du, H., Li, P., Shen, Y., Peng, H., Liu, S., Zhou, G.-A., Zhang, H., Liu, Z., Shi, M., Huang, X., Li, Y., Zhang, M., Wang, Z., Zhu, B., Han, B., Liang, C., and Tian, Z. (2020). Pan-Genome of Wild and Cultivated Soybeans. Cell 182, 162-176.
Google Scholar: Author Only Title Only Author and Title

Lunardon, A., Johnson, N.R., Hagerott, E., Phifer, T., Polydore, 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.
Google Scholar: Author Only Title Only Author and Title

Manavella, P.A., Koenig, D., and Weigel, D. (2012). Plant secondary siRNA production determined by microRNA-duplex structure. Proc Natl Acad Sci U S A 109, 2461-2466.
Google Scholar: Author Only Title Only Author and Title

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10-12.
Google Scholar: Author Only Title Only Author and Title

Moissiard, G., Cokus, S.J., Cary, J., Feng, S., Billi, A.C., Stroud, H., Husmann, D., Zhan, Y., Lajoie, B.R., McCord, R.P., Hale, C.J., Feng, W., Michaels, S.D., Frand, A.R., Pellegrini, M., Dekker, J., Kim, J.K., and Jacobsen, S.E. (2012). MORC family ATPases required for heterochromatin condensation and gene silencing. Science 336, 1448-1451.

Google Scholar: Author Only Title Only Author and Title

Nagano, H., Fukudome, A., Hiraguri, A., Moriyama, H., and Fukuhara, T. (2013). Distinct substrate specificities of Arabidopsis DCL3 and DCL4. Nucleic acids Res. 42, 1845-1856.
Google Scholar: Author Only Title Only Author and Title

Nobuta, K., Lu, C., Shrivastava, R., Pillay, M., De Paoli, E., Accerbi, M., Arteaga-Vazquez, M., Sidorenko, L., Jeong, D.-H., Yen, Y., Green, P.J., Chandler, V.L., and Meyers, B.C. (2008). Distinct size distribution of endogeneous siRNAs in maize: Evidence from deep sequencing in the mop1-1 mutant. Proc Natl Acad Sci U S A. 105, 14958-14963.
Google Scholar: Author Only Title Only Author and Title

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.
Google Scholar: Author Only Title Only Author and Title

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.
Google Scholar: Author Only Title Only Author and Title

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.
Google Scholar: Author Only Title Only Author and Title

Rogers, K., and Chen, X. (2013). Biogenesis, turnover, and mode of action of plant microRNAs. Plant Cell 25, 2383-2399.
Google Scholar: Author Only Title Only Author and Title

Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J., Cheng, J., Xu, D., Hellsten, U., May, G.D., Yu, Y., Sakurai, T., Umezawa, T., Bhattacharyya, M.K., Sandhu, D., Valliyodan, B., Lindquist, E., Peto, M., Grant, D., Shu, S., Goodstein, D., Barry, K., Futrell-Griggs, M., Abernathy, B., Du, J., Tian, Z., Zhu, L., Gill, N., Joshi, T., Libault, M., Sethuraman, A., Zhang, X.-C., Shinozaki, K., Nguyen, H.T., Wing, R.A., Cregan, P., Specht, J., Grimwood, J., Rokhsar, D., Stacey, G., Shoemaker, R.C., and Jackson, S.A. (2010). Genome sequence of the palaeopolyploid soybean. Nature 463, 178-183.
Google Scholar: Author Only Title Only Author and Title

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.
Google Scholar: Author Only Title Only Author and Title

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. Science China life sciences 61, 871-884.
Google Scholar: Author Only Title Only Author and Title

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.
Google Scholar: Author Only Title Only Author and Title

Song, X., Li, Y., Cao, X., and Qi, Y. (2019). MicroRNAs and Their Regulatory Roles in Plant-Environment Interactions. Annu Rev Plant Biol. 70, 489-525.
Google Scholar: Author Only Title Only Author and Title

Song, X., Li, P., Zhai, J., Zhou, M., Ma, L., Liu, B., Jeong, D.-H., Nakano, M., Cao, S., Liu, C., Chu, C., Wang, X.-J., Green, P.J., Meyers, B.C., and Cao, X. (2012). Roles of DCL4 and DCL3b in rice phased small RNA biogenesis. Plant J. 69, 462-474.
Google Scholar: Author Only Title Only Author and Title

Taochy, C., Gursanscky, N.R., Cao, J., Fletcher, S.J., Dressel, U., Mitter, N., Tucker, M.R., Koltunow, 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.
Google Scholar: Author Only Title Only Author and Title

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.
Google Scholar: Author Only Title Only Author and Title

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.
Google Scholar: Author Only Title Only Author and Title

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., and Earl, A.M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE 9, e112963.
Google Scholar: Author Only Title Only Author and Title

Wang, T., Deng, Z., Zhang, X., Wang, H., Wang, Y., Liu, X., Liu, S., Xu, F., Li, T., Fu, D., Zhu, B., Luo, Y., and Zhu, H. (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, 62-14.
Google Scholar: Author Only Title Only Author and Title

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.
Google Scholar: Author Only Title Only Author and Title

Wu, H., Li, B., Iwakawa, H.-O., Pan, Y., Tang, X., Ling-Hu, Q., Liu, Y., Sheng, S., Feng, L., Zhang, H., Zhang, X., Tang, Z., Xia, X., Zhai, J., and Guo, H. (2020). Plant 22-nt siRNAs mediate translational repression and stress adaptation. Nature 581, 89-93.
Google Scholar: Author Only Title Only Author and Title

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.
Google Scholar: Author Only Title Only Author and Title

Xi, Y., and Li, W. (2009). BSMAP: whole genome bisulfite sequence MAPping program. BMC bioinformatics 10, 232.
Google Scholar: Author Only Title Only Author and Title

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.
Google Scholar: Author Only Title Only Author and Title

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.
Google Scholar: Author Only Title Only Author and Title

Xie, M., Chung, C.Y.-L., Li, M.-W., Wong, F.-L., Wang, X., Liu, A., Wang, Z., Leung, A.K.-Y., Wong, T.-H., Tong, S.-W., Xiao, Z., Fan, K., Ng, M.-S., Qi, X., Yang, L., Deng, T., He, L., Chen, L., Fu, A., Ding, Q., He, J., Chung, G., Isobe, S., Tanabata, T., Valliyodan, B., Nguyen, H.T., Cannon, S.B., Foyer, C.H., Chan, T.-F., and Lam, H.-M. (2019). A reference-grade wild soybean genome. Nat Commun. 10, 1216-1212.
Google Scholar: Author Only Title Only Author and Title

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 phasiRNAs in maize anthers. Proc Natl Acad Sci U S A. 112, 3146-3151.
Google Scholar: Author Only Title Only Author and Title

Zhai, J., Zhao, Y., Simon, S.A., Huang, S., Petsch, K., Arikit, S., Pillay, M., Ji, L., Xie, M., Cao, X., Yu, B., Timmermans, M., Yang, B., Chen, X., and Meyers, B.C. (2013). Plant microRNAs display differential 3' truncation and tailing modifications that are ARGONAUTE1 dependent and conserved across species. Plant Cell 25, 2417-2428.
Google Scholar: Author Only Title Only Author and Title

Zhai, J., Jeong, D.-H., De Paoli, E., Park, S., Rosen, B.D., Li, Y., González, A.J., Yan, Z., Kitto, S.L., Grusak, M.A., Jackson, S.A., Stacey, G., Cook, D.R., Green, P.J., Sherrier, D.J., and Meyers, B.C. (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.
Google Scholar: Author Only Title Only Author and Title

Zhang, X., Zhu, Y., Liu, X., Hong, X., Xu, Y., Zhu, P., Shen, Y., Wu, H., Ji, Y., Wen, X., Zhang, C., Zhao, Q., Wang, Y., Lu, J., and Guo, H. (2015). Suppression of endogenous gene silencing by bidirectional cytoplasmic RNA decay in Arabidopsis. Science 348, 120-123.
Google Scholar: Author Only Title Only Author and Title