The Plant Cell, Vol. 21: 3063–3077, October 2009, www.plantcell.org © 2009 American Society of Plant Biologists

Endogenous, Tissue-Specific Short Interfering RNAs Silence the Chalcone Synthase Gene Family in *Glycine max* Seed Coats [W][OA]

Jigyasa H. Tuteja, Gracia Zabala, Kranthi Varala, Matthew Hudson, and Lila O. Vodkin[1]

Department of Crop Sciences, University of Illinois, Urbana, Illinois 61801S

Two dominant alleles of the *I* locus in *Glycine max* silence nine chalcone synthase (*CHS*) genes to inhibit function of the flavonoid pathway in the seed coat. We describe here the intricacies of this naturally occurring silencing mechanism based on results from small RNA gel blots and high-throughput sequencing of small RNA populations. The two dominant alleles of the *I* locus encompass a 27-kb region containing two perfectly repeated and inverted clusters of three chalcone synthase genes (*CHS1*, *CHS3*, and *CHS4*). This structure silences the expression of all *CHS* genes, including *CHS7* and *CHS8*, located on other chromosomes. The *CHS* short interfering RNAs (siRNAs) sequenced support a mechanism by which RNAs transcribed from the *CHS* inverted repeat form aberrant double-stranded RNAs that become substrates for dicer-like ribonuclease. The resulting primary siRNAs become guides that target the mRNAs of the nonlinked, highly expressed *CHS7* and *CHS8* genes, followed by subsequent amplification of *CHS7* and *CHS8* secondary siRNAs by RNA-dependent RNA polymerase. Most remarkably, this silencing mechanism occurs only in one tissue, the seed coat, as shown by the lack of *CHS* siRNAs in cotyledons and vegetative tissues. Thus, production of the trigger double-stranded RNA that initiates the process occurs in a specific tissue and represents an example of naturally occurring inhibition of a metabolic pathway by siRNAs in one tissue while allowing expression of the pathway and synthesis of valuable secondary metabolites in all other organs/tissues of the plant.

INTRODUCTION

Knowledge of the RNA silencing pathway in plants (also known as RNA interference) is now advanced (reviewed in Baulcombe, 2004; Matzke and Matzke, 2004; Zamore and Haley, 2005; Chapman and Carrington, 2007; Eamens et al., 2008; Ramachandran and Chen, 2008; Carthew and Sontheimer, 2009), but relatively few examples exist of regulation of a specific plant phenotype by naturally occurring variation in the pathway. The soybean (*Glycine max*) *I* (inhibitor) locus, an unusual cluster arrangement of chalcone synthase (*CHS*) genes that inhibits seed coat pigmentation, is one such example of a silencing locus (Todd and Vodkin, 1996; Tuteja et al., 2004) mediated through posttranscriptional RNA silencing that can be suppressed by a viral silencing suppressor protein (Senda et al., 2004). CHS is the first committed enzyme in the pathway to an extraordinarily diverse set of secondary products, including isoflavones in the seed cotyledons, defense compounds in the leaves, phenolic exudates of the roots, and anthocyanin pigments in the hypocotyls, trichomes, pods, and seed coats of certain genotypes. In this article, we report RNA analysis and high-throughput sequencing of small RNAs to detail that the biogenesis and accumulation of the *CHS* short interfering RNA (siRNA) silencing signal is limited to the seed coats of dominant *I* genotypes, thus explaining how the soybean plant can still express *CHS* transcripts required for the synthesis of secondary products in other tissues with *I* silencing genotypes.

In soybean, two dominant forms (*I* and i^i) of the *I* locus inhibit pigmentation of the seed coat in a spatial manner resulting in a colorless seed or light yellow on the entire seed coat (*I* allele) or yellow seed coat with pigmented hilum where the seed coat attaches to the pod (i^i allele). By contrast, the homozygous recessive *i* allele allows for pigment production and accumulation over the entire epidermal layer of the seed coat. Most cultivated soybean varieties have been selected for a yellow, nonpigmented seed coat (homozygous *I* or i^i alleles) to mitigate the undesirable effects of the black or brown anthocyanin pigments on protein and oil extractions during processing of soybean products (Palmer et al., 2004).

The *I* locus was initially identified as a region of duplicated and inverted *CHS* genes (*CHS1*, *CHS3*, and *CHS4*) (Todd and Vodkin, 1996) by analyzing a series of naturally occurring isogenic pairs that result from independently occurring mutations of the dominant silencing *I* allele to the recessive *i* allele (designated $I \rightarrow i$ mutations) or of the dominant silencing i^i allele to the recessive *i* allele (designated $i^i \rightarrow i$ mutations). Recently, in-depth BAC screening and sequence analyses revealed that five (*CHS1*, *CHS3*, *CHS4*, *CHS5*, and *CHS9*) of the nine nonidentical *CHS* gene family members are clustered in a 200- to 300-kb region (Clough et al., 2004; Tuteja and Vodkin, 2008) in the cultivar Williams containing the i^i allele. Three of these five genes, *CHS1*,

[1] Address correspondence to l-vodkin@illinois.edu.

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: Lila O. Vodkin (l-vodkin@illinois.edu).

[W]Online version contains Web-only data.

[OA]Open access articles can be viewed online without a subscription.

www.plantcell.org/cgi/doi/10.1105/tpc.109.069856

CHS3, and *CHS4*, were revealed to occur as two 10.91-kb perfect, inverted repeat clusters separated by 5.87 kb of intervening sequence that define the *I* locus based on deletions in this region that occur in recessive *i* mutations. Based on BLAST searches to the recently assembled 8X soybean genome sequence at the Department of Energy Joint Genome Institute (http://www.phytozome.net/soybean), the clustered *CHS* region of the *I* locus maps to chromosome Gm8, while four other *CHS* family members, *CHS2*, *CHS6*, *CHS7*, and *CHS8*, reside in different chromosomes, Gm5, Gm9, Gm1, and Gm11, respectively.

The six contiguous *CHS1-3-4* genes in the inverted repeat clusters lead to spontaneous deletions and truncations of *CHS* genes manifested as mutations of the *I* locus. Spontaneous mutations of the dominant, silencing *I* or i^i alleles to the recessive *i* alleles involve deletion of *CHS* promoter sequences from *CHS4* or *CHS1*, paradoxically resulting in increased *CHS7*/*CHS8* transcripts and pigmented soybean seed coats (Todd and Vodkin, 1996; Tuteja et al., 2004). Silencing by the naturally occurring *CHS* clusters parallels cosuppression, a phenomenon first described in plants transformed with extra copies of *CHS* genes (Napoli et al., 1990; Van der Krol et al., 1990). Using RNA gel blotting, the presence of small RNAs of ~21 nucleotides was found by Senda et al. (2004) in another yellow seed coated variety (Toyohomare) with a dominant *I* allele, and nuclear run-on experiments implicated a posttranscriptional mechanism mediated by siRNAs. Of the other studies reported thus far on RNA silencing involving endogenous alleles that are composed of multiple genes arranged in inverted repeat orientations (Kusaba et al., 2003; Della Vedova et al., 2005), the soybean system is unique in that it triggers tissue-specific gene silencing (Tuteja et al., 2004).

The involvement of gene silencing characterized by the production of the 20- to 30-nucleotide small RNAs in the regulation of plant development is now a well-established occurrence (Carrington and Ambros, 2003; Allen et al., 2004). Small RNAs, particularly microRNAs (miRNAs), have been identified and implicated in a variety of physiological and morphological processes through computational and cloning approaches (Llave et al., 2002; Bartel, 2004; Jones-Rhoades and Bartel, 2004; Sunkar and Zhu, 2004; Lauter et al., 2005; Borsani et al., 2005; Chuck et al., 2009). Further insights into the small RNA regulatory mechanisms are elucidated through the power of deep sequencing of small RNA populations in animals, plants, fungi, and protozoa (Lu et al., 2005; Nobuta et al., 2008).

Here, we present results from both small RNA gel blots and deep sequencing of small RNA populations from several genotypes of soybean and demonstrate that the *CHS* siRNAs accumulated only in the yellow seed coats having either the dominant *I* or i^i alleles and not in the pigmented seed coats with homozygous recessive *i* genotypes. However, the diagnostic *CHS* siRNAs did not accumulate in the cotyledons of genotypes with the dominant *I* or i^i alleles, thus demonstrating the novelty of an endogenous inverted repeat driving RNA silencing in trans of nonlinked *CHS* family members in a tissue-specific manner. This system demonstrates a naturally occurring feature of small RNA biogenesis and accumulation not well defined in other endogenous silencing examples.

Since CHS is the first committed enzyme of the flavonoid pathway, the endogenous tissue-specific silencing phenomenon of the *I* locus leads to selective downregulation of the flavonoid pathway and pigment inhibition only in the seed coats of silencing genotypes, whereas the cotyledons continue to accumulate high levels of isoflavones, other products of the flavonoid pathway that are characteristic of soybean seed (Dhaubhadel et al., 2007). In vegetative tissues, the roots use the flavonoid pathway to produce phenolic compounds involved in symbiosis with *Rhizobium* and the soybean leaves induce *CHS* transcripts upon pathogen challenge (Zabala et al., 2006). Thus, the silencing *I* and i^i alleles have economic value in that they inhibit the pigment in the seed coat, a desirable trait for soybean processing, yet they do not affect other essential functions of the flavonoid pathway in the cotyledons, leaves, and roots. The dominant alleles specifying yellow seed coat have been incorporated by breeders into the germplasm of all modern cultivated soybean varieties long before the mechanism of the locus was understood to be mediated by tissue-specific production of siRNAs.

RESULTS

CHS-Derived siRNAs Found in the Seed Coats of both *I* and i^i Dominant Allele Genotypes

The classically defined *I* locus (inhibitor) is characterized by its four alleles: *I*, i^i, i^k, and *i* (in order of dominant to recessive forms) that affect the production and accumulation of anthocyanins and proanthocyanidins in a spatial manner in the soybean seed coat (Todd and Vodkin, 1993; Wang et al., 1994). The dominant *I* allele inhibits pigmentation over the entire seed coat, resulting in a light or yellow color on mature harvested seeds, whereas the i^i and i^k alleles restrict pigmentation to the hilum and saddle shaped regions, respectively. The homozygous recessive *i* allele allows for pigment production and accumulation in the epidermal layer of the seed coat, thus imparting a buff, brown, or black coloration depending upon other anthocyanin pathway alleles present (Palmer et al., 2004).

We investigated the presence of *CHS*-related siRNA species in seed coats of the nonpigmented (Richland, *I*), and hilum-pigmented isoline (Williams, i^i) along with their corresponding mutant allele lines (T157, *i* and Williams 55, *i*) (Table 1) using RNA gel blotting. The siRNAs were visualized via RNA gel blots probed with an antisense, in vitro–transcribed *CHS7* probe. *CHS7* was chosen as the probe since the nearly identical *CHS7* and *CHS8* genes are downregulated by the silencing *I* locus (Senda et al., 2004; Tuteja et al., 2004). As shown in Figure 1, a strong hybridization signal between the 20- and 30-nucleotide RNA markers was detected in both Richland (*I*) and Williams (i^i) seed coat low molecular weight (LMW) RNA samples, while the RNA samples from the corresponding mutant isolines (T157, *i* and Williams 55, *i*) showed no evidence of *CHS* siRNAs. Thus, the presence of *CHS* siRNAs is limited to the yellow seed coat varieties with dominant *I* or i^i genotypes, which demonstrates that the mechanism of the dominant alleles is mediated by the siRNA silencing pathway. These results also agree with those of Senda et al. (2004), wherein small RNAs were visualized in RNA gel blots of seed coats from a different yellow seed coat cultivar, Toyohomare, which carries the *I* allele.

Table 1. Isogenic Lines, Alleles, and Tissues from Which the Sequenced Small RNA Populations Were Derived

Variety[a]	Allele[b]	Seed Coat Phenotype[c]	Source/Origin	Immature Seed Tissue Used for Small RNA[d]	Total Reads	Unique Signatures ≥5 Reads[e]
Richland	*I*	Yellow	Parent line, released 1926	Seed coat	2,885,864	32,870
T157	*i*	Pigmented	Mutant in Richland, 1938	Seed coat	Blots only	NA
Williams	i^i	Yellow, Ph	Parent line, released 1971	Seed coat	2,886,222	27,363
Williams	i^i	Yellow, Ph	Parent line, released 1971	Cotyledon	3,033,931	27,306
Williams 55	*i*	Pigmented	Mutant in Williams,1973	Seed coat	6,098,005	92,797

[a]Williams is sometimes referred to as Williams 43 or Williams 54, which are internal numbers used in the laboratory, as is Williams 55 to designate the isogenic mutant line. The official designation of the Williams 55 isoline in the USDA germplasm is L885-5495. The T number (T157) of the Richland mutant refers to the official line designation.

[b]All lines are homozygous for the *I* allele indicated. Dominance relationships are $I > i^i > i$.

[c]Ph, Pigmented hilum in the i^i genotype specifies pigment present in the hilum where the seed coat attaches to the pod with an otherwise yellow, nonpigmented seed coat.

[d]Seed coats and cotyledons samples are dissected from midmaturation, green seed at fresh weight range of 50 to 75 mg per seed. NA, not applicable, as no small RNA sequencing was conducted with this line.

[e]The number of unique signatures after adapter trimming that are represented by at least five reads.

Thus, both the dominant *I* allele and the dominant pattern form of the *I* locus, the i^i allele typical of the cultivar Williams, result in silencing mediated by *CHS* siRNA production. The two lines used in our study, Richland (*I*) and Williams (i^i), are the sources of the *I* and i^i alleles in many modern cultivated varieties. Williams is also the cultivar that has recently been sequenced by the Joint Genome Institute (http://www.phytozome.net/soybean).

CHS siRNAs Are Absent in the Cotyledons of Seeds with the Dominant *I* Genotype

We previously showed that the cytoplasmic *CHS* mRNA levels, while significantly lower in the seed coats of the yellow seeded varieties, did not show any reduction in the immature cotyledons dissected from the developing seed (Tuteja et al., 2004), thus predicting a tissue-specific silencing mechanism. Figure 2 shows that *CHS* siRNAs were again clearly detected in seed coats of Richland, the cultivar with the suppressive *I* allele, but not in the pigmented seed coats of T157 (*i*). More intriguingly, *CHS* siRNAs were not detected in cotyledons of either the yellow or the pigmented isolines. These results suggest that the *CHS* siRNA-mediated silencing of *CHS* expression in the immature soybean seeds is specific to the seed coat due to the absence of detectable *CHS* siRNAs in the cotyledons.

Highly Tissue-Specific Accumulation of *CHS* siRNA Conferred by the Dominant *I* and i^i Alleles

Our analysis of *CHS*-siRNAs was expanded to other tissues representing the vegetative parts of the plant. LMW RNA fractions from seed coats, cotyledons, roots, and leaves of the two isogenic pairs (Richland and T157 representing an $I \rightarrow i$ mutation and

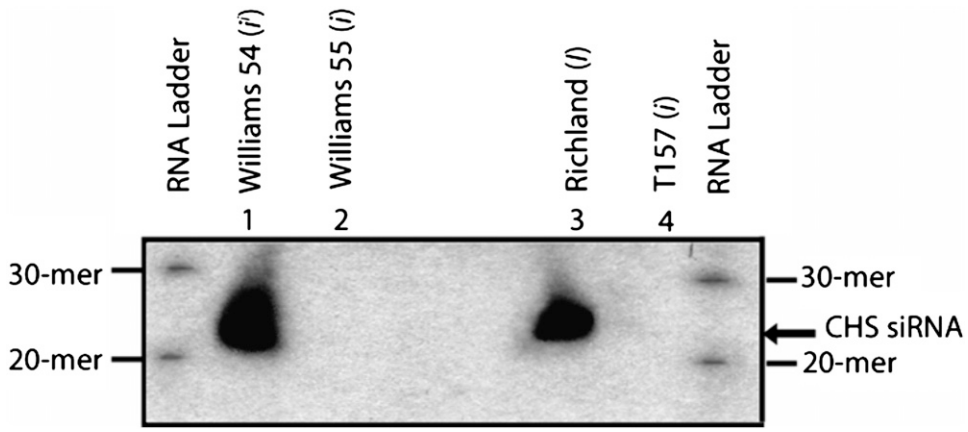

Figure 1. *CHS*-Derived siRNAs in Seed Coats of Soybeans with Silencing Genotypes, Williams (i^i) and Richland (*I*).

LMW RNA samples (75 μg) were fractionated in a 15% polyacrylamide gel and probed with an antisense *CHS7* riboprobe transcribed from a full-length *CHS7* cDNA. Radiolabeled LMW RNAs from both the yellow seed coat varieties Richland (*I*, yellow) and Williams (i^i, yellow seed coat with pigmented hilum) indicate the accumulation of *CHS* siRNA. By contrast, the LMW RNA fractions from the corresponding mutant isolines T157 (*i*) and Williams 55 (*i*) with pigmented seed coats lack *CHS* siRNA. Radiolabeled Decade markers (20 to 30 nucleotides) are shown at left and right.

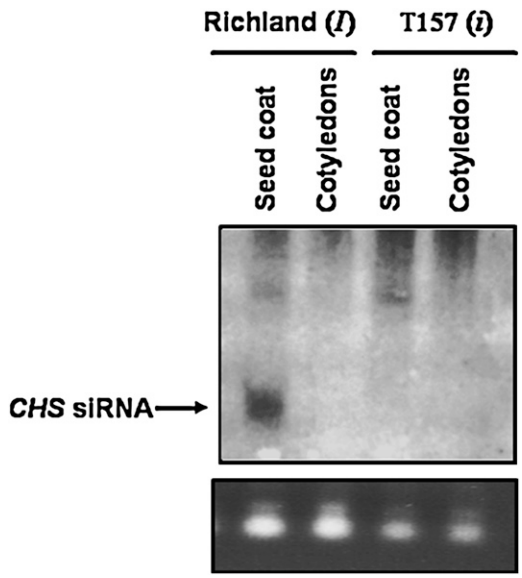

Figure 2. Cotyledons of Seeds with Silencing *I* and i^i Genotypes Do Not Accumulate *CHS* siRNA.

LMW RNA fractions (75 μg) from seed coats and cotyledons of the soybean isogenic lines Richland (*I*, yellow seed coat) and T157 (*i*, pigmented seed coat) were separated on 15% polyacrylamide gels, and the resulting RNA gel blots were probed with an antisense *CHS7* riboprobe transcribed from a full-length *CHS7* cDNA. *CHS* siRNAs were detected only in the seed coats of Richland (*I*), the cultivar with the yellow seed coats (top panel). The bottom panel shows small RNAs stained with ethidium bromide.

Williams and Williams 55 representing an $i^i \rightarrow i$ mutation) were separated on polyacrylamide gels and the RNA gel blots hybridized to the *CHS7* antisense probe as described before.

Figure 3 clearly shows that sense *CHS* siRNAs accumulated in the seed coats of both the nonpigmented Richland (*I*) and hilum-only pigmented Williams (i^i) cultivars. As shown before in Figure 1, no detectable hybridization to the *CHS* probe was observed in seed coats of their respective pigmented isolines, T157 (*i*) and Williams 55 (*i*). Intriguingly, no trace of *CHS* siRNAs was detected in the cotyledons, leaves, or roots of the yellow seeded cultivars Richland (*I*) and Williams (i^i). These results were in accordance with our previously published findings of exclusive tissue-specific effect on reduction of *CHS* mRNA transcript levels only in the seed coats of the yellow seeded cultivars and not in the cotyledons and vegetative tissues of the yellow seeded cultivars. Together, these data (Figures 1 to 3) demonstrate that there is a tissue-specific accumulation of *CHS* siRNAs only in the yellow seed coats and not in other tissues of the yellow seeded varieties.

High-Throughput Sequencing of Small RNA Populations from Seed Coats and Cotyledons

To ascertain and characterize the identity of the *CHS* siRNAs detected on gel blots, multiple small RNA libraries were generated and sequenced deeply by the Illumina high-throughput sequence by synthesis technology. Four small RNA libraries were sequenced (Table 1): seed coats from Richland (*I*, yellow seed coats), seed coats and cotyledons from Williams (i^i, yellow seed coats with black hilum), and seed coats from mutant line Williams 55 (*i*, pigmented seed coats).

A total of almost 15 million sequence reads (14,904,022) were obtained from the four libraries. As shown in Table 1, three libraries were sequenced at the same time to approximately three million reads. The fourth library from the pigmented genotype of the Williams 55 line carrying the *i* mutation produced twice the number of reads (six million) as it was sequenced at a later date when the yield from the Illumina flow cells had increased. The raw sequence reads were processed computationally to remove adapter sequences and from this pool of processed sequences, unique signatures representing at least five reads within each library were identified and selected for further analyses. The number of unique signatures ranged from 28,000 to 92,000 per library (Table 1). Normalization of the total counts of individual signatures was made based on three million raw reads.

BLASTn searches to the Sanger miRNA database (http://microrna.sanger.ac.uk/) found relatively few, <1000, with matches to currently known and curated miRNAs, indicating that many represent siRNAs or previously unknown miRNAs. In this report, we focus on the *CHS* siRNAs found in the different tissues and genotypes that are active physiologically to effect a change in plant pigment phenotype.

Multiple *CHS* siRNAs Accumulate in the Seed Coats but Not the Cotyledons of the i^i Yellow Seeded Genotype

To identify the *CHS*-derived siRNAs from these total small RNA populations, the unique sequence reads from each library were separately mapped to each of the five *CHS* containing BACs (77G7a, 56G2, 5A23, 28017, and 7C24). These BAC clones carry different members of the *CHS* multigene family and have been previously sequenced, annotated, and described in detail (Tuteja and Vodkin, 2008). Figure 4 and Supplemental Data Set 1 online demonstrate that out of a total of >500 kb of the soybean genome represented by these five BACs and spanning 91 predicted gene models, the small RNAs from the Williams seed coat versus cotyledon libraries mapped primarily to the coding regions of *CHS* with scattered matches to other reading frames and very few matches to the intergenic regions or introns. Excluding *CHS* genes, the highest numbers of siRNAs mapped to open reading frames with similarity to known retrotransposable elements. While BACs 5A23 and 28O17 have single copies of *CHS7* and *CHS8*, respectively, BAC77G7a has eight *CHS* family members, six of which form the 27-kb-long inverted repeat region consisting of two clusters of three genes (*CHS1-3-4* and *CHS4-3-1*) that define the i^i allele. Numerous small RNAs with homology to the different *CHS* genes were found, but they were present only in the seed coat library of the hilum-pigmented, yellow seeded Williams (i^i) cultivar and not in the cotyledon library constructed from the same Williams (i^i) cultivar.

Figure 4 illustrates the striking differences observed when both distribution and total counts of *CHS* siRNAs from the seed coat and cotyledon libraries of the hilum-pigmented yellow seeded Williams (i^i) cultivar were compared. Very high counts of ~25,000

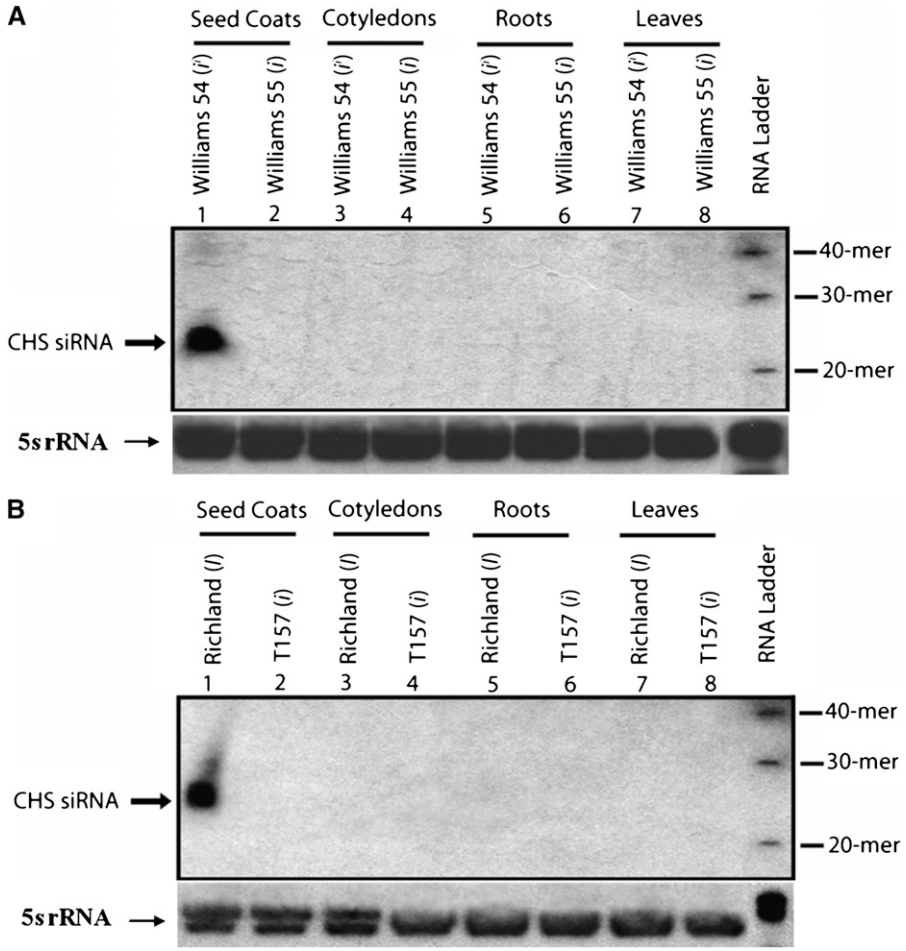

Figure 3. *CHS* siRNAs Accumulate in Seed Coats but Not in the Vegetative Tissues of Yellow Seeded Lines.

LMW RNA fractions (75 μg) were separated on 15% polyacrylamide gels and the RNA gel blots probed with an antisense *CHS7* riboprobe transcribed from a full-length *CHS7* cDNA. *CHS* siRNAs were detected only in the seed coats of the yellow seeded cultivars Williams (i^i; **[A]**) with the hilum pigmented yellow seed coats (lane 1, top panel) or**)** Richland (*I*; **[B]**) with yellow seed (lane 1, top panel) but not in cotyledons, leaves, and roots of either soybean line or their respective pigmented isolines Williams 55 (*i*; **[A]**) or T157 (*I*; **[B]**). Radiolabeled Decade markers (20 to 30 nucleotides) are shown at right. Lower panel shows hybridization of the same LMW RNA fractions to a 5S rRNA probe to show equal LMW RNA sample loading.

CHS siRNAs that align to the individual *CHS* genes were observed for the highly similar *CHS1-3-4* genes found in the inverted repeat *CHS* clusters of the silencing i^i allele. However, only eight or fewer occurrences of *CHS* siRNAs were found in the cotyledon sequences, thereby providing unequivocal evidence that *CHS* siRNAs were found uniquely in the yellow seed coats in a tissue-specific manner.

Tissue-Specific *CHS* siRNAs That Silence *CHS7* and *CHS8* in the Dominant i^i Genotype

We previously showed by analysis of genetic deletions that the origin of the silencing *I* locus is the inverted *CHS1-3-4* and *CHS4-3-1* cluster region, whereas the target genes are primarily the nonlinked *CHS7* and *CHS8* genes (Tuteja et al., 2004) since *CHS7* and *CHS8* are highly expressed in the developing seed coats of the pigmented isolines that carry the homozygous recessive *i* mutation but are downregulated in the yellow Williams seed coats with (i^i) genotype. As shown in Figure 4, ~39,000 total *CHS* siRNAs map to the *CHS7* and *CHS8* genes that are located on separate chromosomes from the *CHS1-3-4* and *CHS4-3-1* cluster regions. Thus, there are large numbers of *CHS* siRNAs available to downregulate the target *CHS7* and *CHS8* mRNAs in the developing seed coats of the Williams (i^i) yellow seeded cultivar, but none were detected in the cotyledons of the same i^i genotype.

The *CHS* multigene family has been divided into two subgroups on the basis of the degree of nucleotide identity in the open reading frames (Tuteja et al., 2004), and a phylogenetic tree has also been constructed previously (Matsumura et al., 2005). Supplemental Table 1 online summarizes the pairwise alignment of the nine *CHS* gene family members. *CHS* genes 1 through 6 grouped together, while *CHS7* and *CHS8* formed the second subgroup, with 82% similarity existing between the two groups.

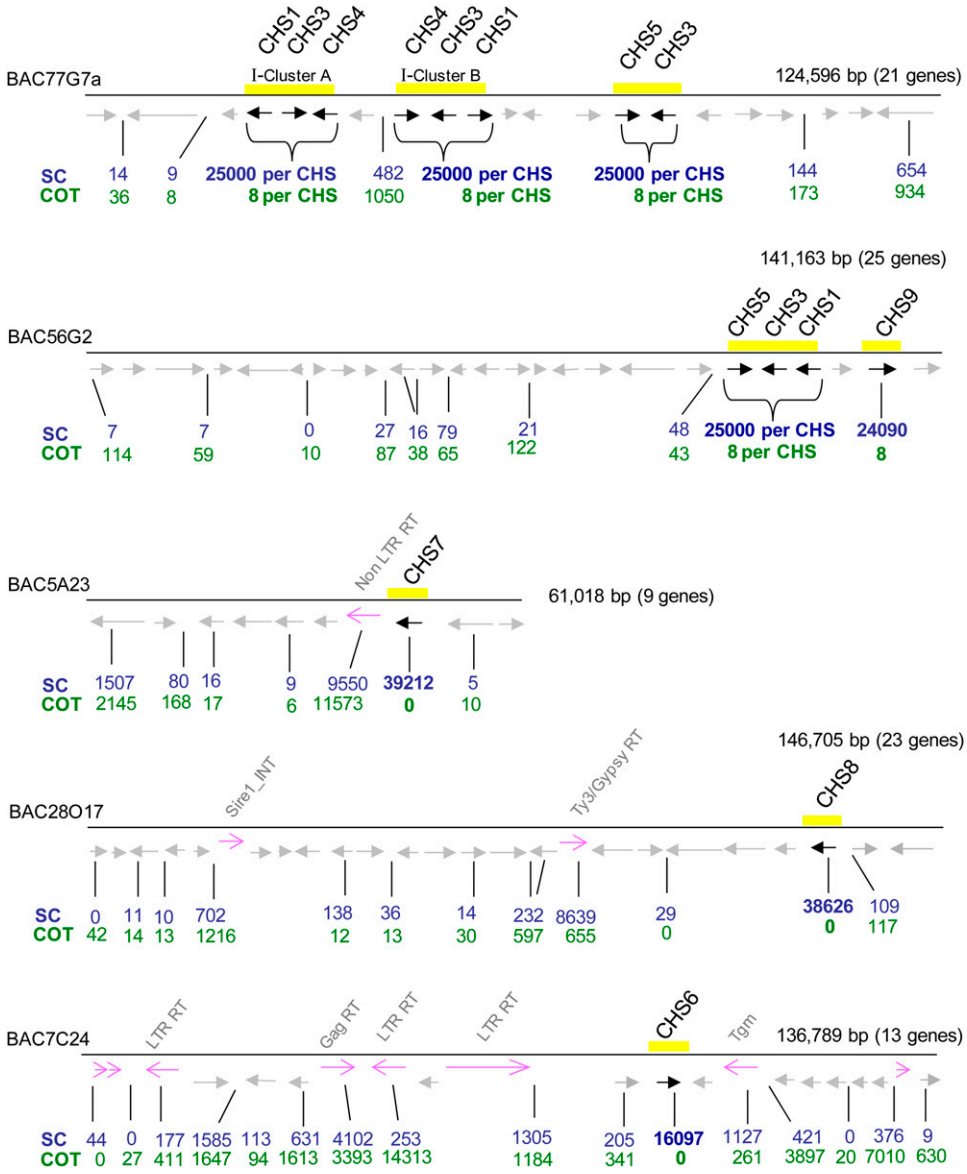

Figure 4. Schematic Diagram Mapping the Total Count of Small RNAs from the Seed Coat versus the Cotyledon Libraries Both Made from the Silencing Williams Genotype (*i*^{i}, Yellow Seeds) to Their Locations on Five BAC Clones Containing Members of the *CHS* Gene Gamily.

Total numbers of small RNA sequence reads related to the five BACs (77G7a, 56G2, 5A23, 28017, and C7C24) were obtained from the nearly three million sequence reads obtained by Illumina from seed coat (SC top line) or cotyledon (COT bottom line) libraries of Williams (*i*^{i}) yellow seeds. Closed arrows represent open reading frames in the indicated direction of transcription. Dark closed arrows indicate *CHS* genes, and light arrows represent other annotated genes as shown by Tuteja and Vodkin (2008). Annotations are shown only for *CHS* genes and for some of the transposon related open reading frames denoted by pink open arrows. The size of BACs in base pairs and the number of genes (excluding transposons) are given to the right of each BAC. See Methods for the BLAST criteria.

As much as 93 to 98% nucleotide sequence identity has been observed between *CHS* genes 1 through 6, with *CHS6* being the most divergent member of this subgroup. The two members of the second subgroup, *CHS7* and *CHS8*, are 97% identical. *CHS9*, a recently characterized member of this family exhibits greater homology to the first subgroup of *CHS* genes 1 through 6. Although very similar in sequence, multiple single or double nucleotide mutations distributed along the genes distinguish the family member genes, thus allowing their transcripts to be distinguished by quantitative real-time PCR (Tuteja et al., 2004).

Because the size of the target sequence influences the *e* value obtained from the BLAST algorithm and the BACs vary widely in size from 61,000 to >146,000 bases, we performed the BLAST analysis of each small RNA population to each of the nine

individual *CHS* genes to attain an accurate, comparative number for *CHS* siRNAs aligning to the individual *CHS* genes.

All *CHS* genes contain one intron at the same position, and excluding their introns, the *CHS* genes are nearly identical in size at 1167 bases (*CHS1-6* and *CHS9*) or 1170 bases (*CHS7* and *CHS8*) from the ATG to the stop codon. Supplemental Data Set 2 and Supplemental Tables 2 and 3 online present the results. The number of unique signatures with 100% identity to each *CHS* gene in a pairwise comparison (see Supplemental Table 2 online) indicates that while *CHS4* has 82% nucleotide similarity to both *CHS7* or *CHS8*, only ~15% of the *CHS* siRNAs have 100% identity to both *CHS4* and *CHS7* or *CHS4* and *CHS8* (see Supplemental Table 3 online). Thus, we chose *CHS7* and *CHS4* as representative genes of each of the two *CHS* subgroups.

Specifically, Figure 5A illustrates the alignment of *CHS* siRNAs from the Williams (*i*) seed coat with 100% identity to *CHS7*. Overall, the *CHS* siRNAs aligned through almost the entire length of the *CHS* gene exons and not at all to the introns. In contrast with the large number of *CHS* siRNA sequences that aligned with exon 2, only a few sequences aligned with exon 1. The majority of *CHS* siRNAs aligned with exon 2 to form a bell-shaped curve

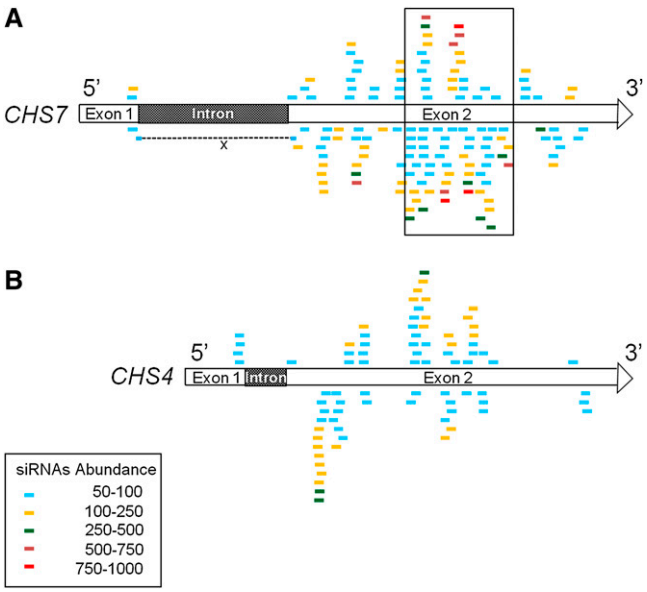

Figure 5. Diagram Representing Abundance and Alignments of *CHS* siRNAs with Sequence Signatures Identical to *CHS7* or *CHS4* Genes.

CHS siRNAs from the seed coat library of the yellow seed Williams (*i*) having 100% match to the *CHS7* (or *CHS4*) gene sequences were mapped based on their alignments to specific locations. The intron and two exons of *CHS7* and *CHS4* are depicted, and the orientation of transcription is indicated with an arrowhead. The colored segments represent the number of occurrences in the library and the location of their alignment with *CHS7* or *CHS4* sequences. The "x" denotes one siRNA signature spanning the intron. Those aligning to the sense strand are above the gene, and those aligning to the antisense strand are denoted below the gene. Number of occurrences: 50 to 100 (light blue), 100 to 250 (dark blue), 250 to 500 (green), 500 to 750 (orange), and 750 to 1000 (red). The boxed-in region highlights the most targeted portion of exon 2.

against both the sense and antisense strands. Figure 5 shows only the alignment results of the *CHS* siRNAs with more than 50 occurrences. As shown in Table 2, the majority (976) of the total (1118) unique signatures had very few occurrences (5 to 50), while the remaining 13% (141) were represented many times (50 to 1000). Only 38 *CHS* siRNA unique species, including only three siRNAs with more than 50 counts, aligned with exon 1 of *CHS7*. None aligned with the intron, although some did appear to span the border, indicating that they arose from processed transcripts.

Since the frequency of each small RNA signature in the library generally reflects its relative abundance in the sample, the sequence repeats provide a quantitative expression measurement. Strikingly, of the 1118 unique siRNA signatures with perfect matches to *CHS7* gene sequence, only 149 (13%) match perfectly to *CHS4* gene sequence (see Supplemental Table 2 online). This finding illustrates that many of the siRNAs matching 100% to *CHS7* originated from *CHS7* (or the similar *CHS8*) transcripts after intron splicing, most likely as a result of amplification by RNA-dependent RNA polymerase (RdRP), dicer-like (DCL), and argonaute (AGO)-like effector complex that synthesize and cleave aberrant double-stranded RNA (dsRNA) into phased 21- to 22-nucleotide secondary siRNAs.

CHS8 shows a very similar alignment of the *CHS* siRNAs (Table 2), as expected from the high sequence similarity between *CHS7* and *CHS8* (97% similar). The siRNAs that aligned uniquely to *CHS1*, *CHS3*, and *CHS4* are evidence that they originated from transcripts of the inverted repeat on chromosome Gm8 where those *CHS* genes reside. We propose that some of these siRNA signatures with perfect matches to genes in the *CHS1-3-4* and *CHS4-3-1* clusters represent the primary siRNA guides that trigger the silencing of all *CHS* genes.

The *CHS7* and *CHS8* sequence region that aligned with the largest number of siRNA signatures with very high counts must be the region most targeted by the primary siRNA-guided RNA-induced silencing complexes (RISC) (Figure 5A, framed region). This portion of the sequence comprises the central region 748 bp of exon 2 (975 to 1281 bp relative to the initiation codon). Likewise, the alignments of the Williams (*i*) seed coat library *CHS* siRNAs to *CHS1*, *CHS3*, and *CHS4* sequences of the inverted repeat also produced a similar alignment pattern (as shown in Figure 5B for the alignment of *CHS* siRNA with 100% identity to *CHS4*). This suggests that once silencing of all *CHS* genes is triggered by the *CHS1-3-4* primary siRNA guides, a multitude of *CHS* siRNAs originating from any of the expressed *CHS* genes become guides to advance the targeting and posttranscriptional suppression of the entire *CHS* gene family.

The 21-Nucleotide siRNAs Are the Predominant Size Class of siRNAs with 100% Match to Individual *CHS* Genes

Produced by endonucleolytic cleavage of dsRNA by different DCL-like orthologs and pathways, two major size classes of siRNAs, short (~21 nucleotides) and long (~24 nucleotides, have been detected in plants (Hamilton et al., 2002; Mallory et al., 2002; Tang et al., 2003). In our study, the sizes of the *CHS* siRNAs from the different libraries sequenced ranged primarily from 19 to 24 nucleotides. To determine the size class that dominated the different populations of *CHS* siRNAs with sequence identity to

Table 2. Number of Unique *CHS*-siRNAs from a Seed Coat Library (Williams i^i) Aligning with 100% Identity to Individual *CHS* Genomic Sequences and Their Frequencies of Occurrence

		Number of Unique *CHS*-siRNA Signatures and Their Frequencies[a]					
Frequencies		5 to 50	50 to 1000	5 to 1000	5 to 50	50 to 1000	5 to 1000
Gene	Strand	Exon 1		Intron	Exon 2		Total
CHS1	+	9	0	0	446	38	493
	–	36	4	0	382	36	458
CHS2	+	7	0	0	377	40	424
	–	17	1	0	273	25	316
CHS3	+	17	0	0	447	38	502
	–	31	4	0	382	39	456
CHS4	+	9	0	0	439	37	485
	–	36	4	0	394	39	473
CHS5	+	10	0	0	439	37	486
	–	36	4	0	394	39	473
CHS6	+	10	1	0	242	22	275
	–	39	3	0	165	12	219
CHS7	+	6	1	0	592	90	689
	–	29	2	0	349	49	429
CHS8	+	4	1	0	557	81	643
	–	28	2	0	345	49	424
CHS9	+	4	0	0	407	35	446
	_	15	1	0	354	32	402

[a]Data correlated from Supplemental Data Set 2 online for the exonic regions and from Supplemental Data Set 1 online for the introns.

each one of the nine *CHS* genes, *CHS* siRNAs from the Williams (i^i) seed coat library with 100% matches were categorized into size classes and plotted against the number of unique *CHS* siRNA signatures (Figure 6A) or total number of signature occurrences per *CHS* gene (Figure 6B). Interestingly, both graphs affirmed that the most abundant *CHS* siRNA size class is the small 21 nucleotides, with as many as 700 unique signatures totaling >30,000 counts for those matching 100% to the *CHS7* sequence. Based on the *Arabidopsis* model (Chapman and Carrington, 2007), these results suggest amplification by an RdRP6/DCL4 ortholog resulting in 21nucleotide secondary *CHS* siRNAs from *CHS7*/*CHS8* mRNAs. Significantly, as illustrated in Figure 6B, the higher number of signature occurrences for genes *CHS7* and *CHS8* is in accordance with our earlier gene expression results of the individual family members. We had shown that the dominant i^i allele executes its suppressive effect by inhibiting the accumulation of *CHS7* and *CHS8* transcripts. The increase in total *CHS* mRNA levels in the seed coats and consequential pigmentation of both the $i^i \rightarrow i$ and $I \rightarrow i$ mutations was attributed to a 7- to 25-fold increase in the *CHS7*/*CHS8* transcript levels (Tuteja et al., 2004).

The Dominant *I* Allele Also Produces Complex, Heterogeneous *CHS* siRNAs

Sequencing of the small RNA population from the immature seed coats of Richland (*I*, yellow) at the same stage of development as those of the Williams cultivar (i^i, yellow with pigmented hilum) also yielded close to three million raw sequence reads and >30,000 unique small RNAs. Large numbers of *CHS* siRNAs were found, agreeing with the blot data of Figures 1 to 3. The total number of *CHS* siRNAs found in the Richland population that map to each *CHS* gene is generally similar to that found for Williams (summarized in Table 3 from Supplemental Data Set 2 online). The *CHS* siRNAs from Richland also represented both strands and primarily mapped to exon 2. Some of the most abundant *CHS* siRNAs were the same tags as in Williams. For example, Table 4 shows some of the more abundant *CHS* siRNAs and the counts found in each library. Thus, both the dominant alleles (*I* and i^i) are effective in silencing the targeted *CHS* genes through production of a heterogeneous siRNA population that largely maps to both strands in the middle of exon 2 of the individual *CHS* gene family members.

The $i^i \rightarrow i$ Mutation Abolishes *CHS* siRNAs from the Small RNA Population and Restores Pigmented Seed Coats

Structurally, the recessive *i* locus mutation in Williams 55 line is represented by a deletion that includes the *CHS* cluster B and extends into the promoter of *CHS4* of cluster A (as illustrated in Figure 7). Examination of the number and distribution of *CHS* siRNAs in the seed coats of the two i^i and *i* isogenic lines revealed the presence of a considerably higher number of *CHS* siRNA reads in the hilum-pigmented yellow seeded cultivar (Williams, i^i) relative to the black seed coat of Williams 55 (*i*) (Table 3). In a Williams 55 seed coat library with >90,000 unique signatures and processed reads, only 16 different signatures totaling around 108 molecules per three million raw sequence reads mapped to any of the individual *CHS* gene family members. These few constitute only 0.03% of the number of *CHS* siRNAs found in the Williams (i^i, yellow with pigmented hilum) and 0.06% of the *CHS* siRNAs found in Richland (*I*, yellow). The presence of large numbers of *CHS*-specific siRNAs in the seed coats of the dominant *I* and i^i

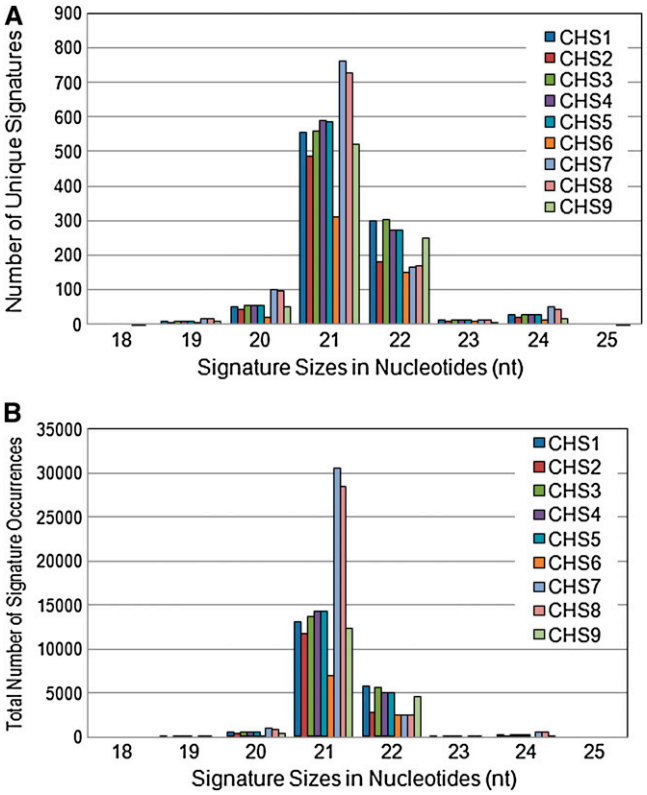

Figure 6. Size Distributions of *CHS* siRNAs for Each *CHS* Gene in a *CHS*-Silenced Seed Coat Library.

CHS siRNAs selected from the seed coat of yellow seed, Williams i^i, small RNA library were filtered to identify those with 100% identity to individual *CHS* genes. The nine *CHS*-siRNA subgroups were distributed according to the size of individual sequence signatures and plotted in two graphs.

(A) The number of unique signatures of a given size for each *CHS* gene.

(B) The same as in **(A)** but multiplied by the number of occurrences of each *CHS*-siRNA signature. The result for each *CHS* gene is color coded as indicated.

cultivars and not in the recessive *i* genotype is clear evidence of a suppressive effect of the inverted repeat *I* locus in soybean, which is mediated by a siRNA silencing pathway. Coupled with the small RNA gel blots of Figures 1 to 3, these results confirm that the naturally occurring deletions in the *CHS* genes at the *I* locus that accompany $i^i \rightarrow i$ and $I \rightarrow i$ mutations (Todd and Vodkin, 1996; Tuteja et al., 2004) serve to abolish the production of the small RNAs from the dominant forms of the *I* locus.

DISCUSSION

An Endogenous, Inverted Repeat, *CHS*-Coding Region Generates Heterogeneous *CHS* siRNAs

Sequencing small RNAs to a depth of three million reads from the seed coats of two silencing alleles of the *I* locus (*I*, yellow seed coats, and i^i, yellow seed coats with pigmented hilum) provided a wealth of data from which to determine the *CHS*-specific forms produced by these alleles. Based on alignments of these sequence signatures to the different *CHS* genomic sequences (Figure 4) and on data from RNA gel blots (Figures 1 to 3), we conclude that *CHS* small RNAs were generated and accumulated exclusively in seed coats of yellow seed soybeans with the dominant allele genotypes *I* and i^i and not in cotyledons. The *CHS* siRNAs found in the seed coats ranged in size from 19 to 24 nucleotides, although the small 21-nucleotide species was the most abundant size class. These sequences aligned to both the sense and antisense strands of exons 1 and 2 of all the *CHS* gene sequences (*CHS1* to *CHS9*). For all *CHS* genes, the numbers of signatures aligning to the sense strand were somewhat higher than to the antisense strand (Table 2). Many more siRNAs aligned to exon 2 than to exon 1. In exon 2, the larger numbers of unique signatures as well as the larger number of counts per signature were concentrated to the central sequence portion (Figure 5). Because there is sequence variation among the nine *CHS* genes, we were able to identify *CHS* siRNA signatures with 100% sequence identity to individual genes, thereby proving that the *CHS* siRNAs originated from transcripts of more than one *CHS* family member.

The silencing i^i allele contains six genes *CHS1-3-4* and *CHS4-3-1* as two perfectly repeated and inverted 10.91-kb clusters separated by a 5.87-kb intervening region as sequenced in two individual BACs (77G7-a and 104J7) (Clough et al., 2004; Tuteja and Vodkin, 2008). PCR indicates a similar clustered structure is also present in the Richland *I* allele (Tuteja et al., 2004). A number of studies have shown that inverted repeats delivered as transgenes to induce RNA interference are particularly potent silencers of gene expression (Muskens et al., 2000; Smith et al., 2000; Kusaba et al., 2003), and naturally occurring inverted gene duplications that may produce siRNAs are speculated to be the evolutionary progenitors of miRNAs (Allen et al., 2004). In animal

Table 3. Comparison of siRNAs Counts from Seed Coat and Cotyledon Libraries that Map to the Coding Regions of the Nine-Member *CHS* Gene Family

	Variety, Genotype, and Tissue of Library Origin			
CHS Gene	Richland (*I*) Seed Coat	Williams (i^i) Seed Coat	Williams (i^i) Cotyledon	Williams (*i*) Seed Coat
	CHS siRNA Counts per Three Million Sequence Reads			
CHS1	17,277	30,094	33	86
CHS2	16,891	28,547	38	76
CHS3	18,595	31,679	63	102
CHS4	18,368	31,841	43	112
CHS5	18,368	31,825	43	112
CHS6	14,066	24,006	83	187
CHS7	30,712	41,937	61	110
CHS8	30,382	42,045	77	137
CHS9	18,886	32,131	13	50

Alignments were made individually for each library to each *CHS* gene as described in Methods. Complete data are presented in Supplemental Data Set 2 online. The differences between the numbers in this table and those of Figure 4 (based on Supplemental Data Set 1 online) result from the variable lengths of the large BACs that influences the *e* value. The individual *CHS* genes are nearly identical in length.

Table 4. Some Abundant *CHS* siRNA Sequences Derived from the Yellow Seed Coats with Dominant Alleles from the Williams (I^i) or Richland (I) Cultivars and Their Alignments to *CHS* Genes

CHS siRNA Sequence	Williams Counts	Richland Counts	*CHS* Strand	*CHS* Gene Alignments[a]								
CAACTTGTGGAATTGGGTCAG	913	1100	–							7*	8*	
AAGCATTGTTTGGAGATGGAG	825	345	+	1	2	3	4	5	6	7*	8*	9
TTTGTATGAGCTTGTTTGGAC	803	799	+							7*	8*	
CACGGAATGTGACTGCAGTGA	723	325	–			3	4	5		7*	8*	9
TCAGAACCAACAATGACTGCA	551	629	–							7*	8*	
TCTCAACTTGTGGAATTGGGT	534	606	–							7*	8*	
TTACCCACTTGATCTTCTGCA	545	94	+							7*	8*	
TCAAAGAACATTGATAAGGCA	529	307	+							7*	8	
ACGGAATGTGACTGCAGTGAT	471	279	–			3	4	5		7*	8*	9
TCAAAGAACATTGACAAGGCA	466	331	+							7	8*	
CACCTTCGTTGGATGCAAGGCA	451	16	+	1*	2	3*	4*	5*	6			9*
CCTTCGTTGGATGCAAGGCAA	325	8	+	1*	2	3*	4*	5*	6			9*
CGCGGAATGTGACTGCGGTGA	298	123	–	1		3*	4*	5*		7	8	9*

Data are from Supplemental Data Set 2 online.

[a]Asterisk represents 100% match to exon 2 of the indicated *CHS* gene, and no asterisk denotes a single base mismatch of the *CHS* siRNA sequence to the indicated *CHS* gene. Blanks indicate more than one mismatch; + strand direction is the coding direction for all *CHS* genes.

systems, small RNA populations sequenced from mouse oocytes revealed the existence of endogenous primary siRNAs speculated to originate from naturally formed dsRNAs from pseudogene loci that contained inverted repeat structures (Tam et al., 2008; Watanabe et al., 2008). In the soybean system, the primary *CHS* siRNAs are derived from a cluster of functional *CHS* genes rather than pseudogenes. We have previously shown that *CHS* family members are transcribed in various tissues and in response to pathogen attack (Tuteja et al., 2004; Zabala et et al., 2006). These transcripts are likely to translate into functional proteins since six different CHS isomers have been identified in elicitor-treated soybean cell suspension cultures (Grab et al., 1985).

It is not difficult to envision that long transcripts initiated from one promoter within the two clusters of the I^i allele and that span very similar *CHS* genes in inverse orientation could fold and create aberrant dsRNAs that are then subjected to dicer-like enzyme complexes and resulting in 20- to 22-nucleotide primary *CHS* siRNAs. In *Arabidopsis* and other systems, different DCL ribonucleases orthologs (1 to 4) process the dsRNA into different siRNA size classes depending on their catalytic properties (Hamilton et al., 2002; Tang et al., 2003). DCL3 processes the dsRNA into 24-nucleotide siRNAs, while DCL4 processes dsRNAs into 21-nucleotide siRNAs. Even though not much is known about the DCL ribonucleases or the protein complexes that cleave these aberrant dsRNA structures in soybean, one can anticipate that based on the tight range of *CHS* siRNA sizes found in the seed coat siRNA population (Figure 6) that a DCL4 ortholog could be cleaving the aberrant dsRNAs in soybean.

Amplification of the Silencing Signal through the Action of an RNA-Dependent RNA Polymerase Is Deduced from the Specific *CHS* siRNA Sequences in the Seed Coat

It can be reasoned that the primary *CHS* siRNAs resulting from the cleavage of the *CHS* dsRNA (formed from the *CHS1-3-4* and *CHS4-3-1* cluster region) trigger sequence-specific degradation of *CHS7*/*CHS8* transcripts, explaining the observed 7- to 25-fold decrease in their expression levels in the yellow seed coats containing I and I^i alleles (Tuteja et al., 2004). These cleaved *CHS7* and *CHS8* mRNAs may in turn be substrates for further RdRP and DCL activity. After cleavage at the mRNA site targeted by the primary *CHS* siRNA guide, an RdRP could synthesize a complementary copy of the cleaved *CHS7* or *CHS8* mRNA, thus generating additional aberrant dsRNAs that are processed into secondary 21- to 22-nucleotide *CHS* siRNAs by dicer activity. These secondary *CHS* siRNAs would then fan out as multiple guides of AGO-RISC–like complexes that could target additional *CHS* mRNAs, amplifying the silencing response as well as spreading it over a larger region of the *CHS* mRNAs. The targeting of the *CHS* mRNAs by both the primary and secondary siRNA guides must also take place after intron splicing since no *CHS* siRNA signatures aligning to intron sequences were found in any of the small RNA populations examined (Table 2). A few of the *CHS* siRNA signatures are split by the intron when aligned to the *CHS* genomic sequences (Figure 5A), again confirming that they originate after intron splicing.

In summary, the pathway for *CHS* siRNA generation and accumulation must account for (1) the large number and distribution of unique siRNA signatures detected; (2) the range of siRNA sizes (20 to 22 nucleotides) with the small 21-nucleotide species being the most abundant, particularly for *CHS7* and *CHS8*; (3) the lack of siRNAs derived from intron sequences; (4) the existence of siRNA with 100% sequence identity to genes *CHS7* or *CHS8* and lower similarity to genes *CHS1*, *CHS3*, and *CHS4* and vice versa; (5) silencing of all nine nonidentical *CHS* gene coding regions, including those linked and those not linked to the long inverted repeat in Chromosome Gm8; and (6), the derivation of *CHS7*- and *CHS8*-specific siRNAs from both strands. The latter implies that such siRNAs are processed from dsRNA substrates produced by the pairing of sense transcripts with antisense copies derived from RdRP action. Figure 7 depicts the plausible succession of these steps schematically.

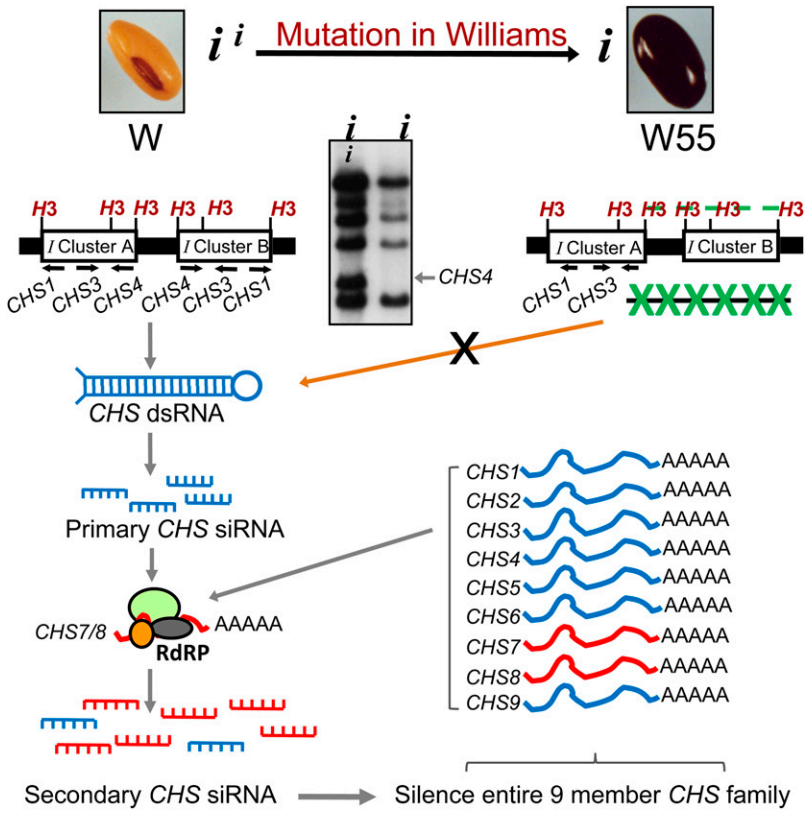

Figure 7. A Schematic Illustrating the Role of *CHS* Gene Clusters in Generation of *CHS* siRNAs in the Silencing i^i Allele and Its Comparison to the Recessive *i* Mutation.

Seed phenotypes are indicated for W = Williams (i^i, hilum-only pigmented seed coat) and the isogenic mutant line W55 = Williams 55 (*i*, black seed coat). The presence of an exact, base-by-base duplication of the 10.91-kb *CHS* clusters A and B at the *I* locus as revealed by BAC sequencing of the yellow genotype (i^i) is diagrammed, as is the deletion in the *i* mutation. Marked by green Xs, the deletion encompasses regions flanking *CHS* cluster B and extends into the promoter region, including the *Hin*dIII (*H3*) site of *CHS4* in cluster A. RFLP analysis also shows absence of the 2.3-kb *Hin*dIII fragment corresponding to *CHS4* genes in the pigmented genotype (*i*). (Summarized from Todd and Vodkin, 1996; Tuteja et al., 2004). The molecular events supported by the *CHS*-siRNA data presented in this report are diagrammed. A dsRNA generated from the inverted *CHS* repeats in the seed coat is cleaved into primary siRNAs representing both strands that are amplified by RdRP to generate secondary *CHS* siRNAs capable of downregulating all members of the *CHS* gene family, including the more distantly related *CHS7* and *CHS8* (denoted in red). These two genes are highly expressed in the pigmented seed coats in which *CHS* siRNA production has been abolished by the deletion in the mutant *i* allele (W55). Production of the primary *CHS* siRNAs is tissue specific, found only in the seed coats and not in other tissues of the yellow seeded (i^i) genotype.

As with the transacting-siRNAs of *Arabidopsis* (Yoshikawa et al., 2005; Chapman and Carrington, 2007), we found that there is a certain degree of phasing in the *CHS*-siRNAs (Figure 5), putatively as a result of periodic dicing of double-stranded *CHS* mRNA. We presume that the imprecise phasing observed in this case may be due to multiple initiation sites on the *CHS* mRNAs targeted by the primary *CHS*-siRNA guides originating at the *I* locus.

Tissue-Specific Biogenesis of the *CHS* siRNAs from the Inverted Repeat *I* Locus Clusters in Seed Coats Is More Plausible Than Lack of Signal Amplification in Other Tissues

More importantly, the results from this study present unequivocal evidence for the existence of an additional feature in siRNA regulation not described previously, a tissue specificity of endogenous siRNA generation from a cluster of genes that expresses normal mRNA transcripts in other tissue and organ systems. Several hypotheses can be put forward to explain the presence of *CHS* siRNAs in only one tissue, the seed coat. One possibility is that a cell or tissue-specific transcription factor in association with the structural peculiarities of the *I* locus could determine the seed coat–specific nature of *CHS* silencing. Previous expression studies of other genes in the anthocyanin pathway, such as flavonoid 3′ hydroxylase (*F3′H*), flavonone 3-hydroxylase (*F3H*), and flavonoid 3′,5′-hydroxylase (*F3′5′H*), have also shown tissue-specific expression in the seed coat for some of the family members (Zabala and Vodkin, 2003, 2005, 2007). Thus, a transcription factor (or a distantly located effector gene) could be regulating specific branches of the flavonoid pathway and possibly many other developmental pathways of the seed coat in a highly specific manner.

Conversely, the primary *CHS* siRNAs could potentially be generated from a dsRNA molecule produced in all tissues, but possibly they are not being amplified to detectable levels for lack of an RdRP enzyme in other tissues. RdRPs are involved in RNA amplification of primary siRNAs and generate more dsRNAs that are subsequently processed into the secondary siRNAs (Zamore and Haley, 2005; Chapman and Carrington, 2007). However, the lack of an RdRP function in so many different soybean tissues is implausible. As shown in Table 1, the cotyledon produces roughly the same number of 27,000 unique small RNAs as the seed coat libraries, although the cotyledon possesses only a handful of *CHS* siRNA molecules (Figure 4, Table 3). The distribution of non-*CHS* small RNAs that map to non-*CHS* coding regions is approximately the same in the Williams seed coat and the cotyledon. One of those signatures has over 11,000 occurrences that match to an long terminal repeat retrotransposon reverse transcriptase adjacent to *CHS7* on BAC5A23 (Figure 4). Additionally, another matches near the coding region for a gene with unknown function between the two *CHS4* inverted repeats of clusters A and B on BAC77G7a. Thus, the cotyledon is clearly capable of amplifying other non-*CHS* siRNAs. However, in the absence of *CHS* siRNAs, the soybean cotyledon continues to synthesize *CHS7* and *CHS8* mRNA transcripts in later stages of development, which result in accumulation of isoflavones and other flavonoid products in the soybean cotyledon. Thus, in contrast with the downregulation of the pathway in the seed coats by *CHS* siRNA-targeted destruction of *CHS7* and *CHS8* mRNAs in the yellow seed coats, the *CHS7* and *CHS8* transcripts continue to increase during cotyledon development, leading to the accumulation of large amounts of isoflavones in the mature soybean seed even in yellow seed coat varieties with the dominant *I* or i^i alleles. This system represents a targeted regulation of the flavonoid pathway in a specific tissue.

Likewise, we have sequenced libraries from other tissue and organ systems, including leaves and stems that also produce large numbers of small RNAs but only a handful of *CHS*-specific siRNAs similar to the very low percentages shown for the cotyledon library in Table 3. We have previously demonstrated that *CHS* transcripts in the leaves of Williams (i^i), including those for *CHS1*, *3*, *6*, *7*, and *8* in soybean leaves, are induced >1000-fold within 8 h after infection with the bacterial pathogen *Pseudomonas syringae* (Zabala et al., 2006). The induction of *CHS* transcripts would provide ample targets for RdRP amplification of a very low abundance *CHS*-siRNA silencing signal, should one exist in the pathogen challenged leaves of the Williams (i^i) genotype. However, posttranscriptional downregulation of *CHS* transcripts does not occur and the *CHS* mRNAs are highly expressed. These data reinforce that the tissue-specific nature of the *I* locus–mediated silencing effect is likely the tissue-specific biogenesis of the dsRNA and primary *CHS* siRNAs in the seed coats rather than failure to amplify secondary *CHS* siRNAs in other tissues.

The *CHS* siRNAs Are Not Transported from the Seed Coat to the Developing Cotyledons or Other Tissues

Systemic RNA silencing has been observed in plants, fungi, and in *Caenorhabditis elegans* (Voinnet et al., 1998; Winston et al., 2002; Mallory et al., 2003; Timmons et al., 2003). In plants, the cell-to-cell and systemic spread of some classes of small RNAs is considered to occur through plasmodesmata (Voinnet et al., 1998; Lucas et al., 2001; Himber et al., 2003; Lucas and Lee, 2004) and the phloem (Palauqui et al., 1997; Klahre et al., 2002; Mallory et al., 2003), respectively.

The soybean seed coat, derived from the maternal ovular integuments, encloses the filial tissues (the embryo and the cotyledons) and includes two vascular bundles (the phloem and xylem elements) at the hilum, the point of attachment to the pod (Thorne, 1981). The phloem conduit, comprising the sieve tube system, functions in the long-distance transport of nutrients by pressure-driven bulk flow of the translocation stream and thus provides for storage product accumulation in the cotyledons. The symplasmic discontinuity between the maternal and filial tissues in the soybean seeds necessitates an apoplasmic exchange localized to the maternal/filial interface (Thorne, 1981). In our system, there is currently no evidence for the active transfer of the *CHS* siRNAs generated in the immature seed coat to other tissues. This could be explained simply that the seed coat is an end point of phloem transport and is not likely able to transport siRNAs backward from the seed coat to other vegetative tissues. The seed coat obviously is a conduit for nutrients from the vegetative tissues of the plant to the developing seed cotyledon that it encloses; yet there is no evidence of transfer of the *CHS* siRNAs through the seed coat to the cotyledon underneath since they do not accumulate in the cotyledons.

Regulation of an Important Pathway by Tissue-Specific siRNA Biogenesis

To summarize, we have described an endogenous inverted repeat system in soybean that drives silencing of *CHS* genes in a tissue-specific manner, thereby inhibiting pigmentation of the seed coats. We present clear evidence that a large number of siRNAs with sequences identical to exons 1 and 2 of multiple members of the *CHS* gene family accumulated in the seed coats of soybean cultivars with dominant *I* or i^i alleles in a tissue-specific manner. The tissue-specific nature of the *CHS* siRNAs biogenesis adds another layer of complexity to the mechanisms of posttranscriptional regulation. Further study of this system should provide insight into the mechanism of tissue-specific gene silencing, which could be of practical use to target silencing to a restricted tissue or cell type.

While much emphasis has been placed to date on the evolutionarily ancient and highly conserved miRNAs, examples of siRNAs more uniquely tied to a particular species are likely to arise. As illustrated by the *CHS* siRNA system, expansion of duplicate genes can potentially spawn a unique regulatory system in a physiological process during natural selection and evolution or during domestication of a plant species. Thus, siRNA regulation could be an important addition to our knowledge of plant allelic diversity and short-term evolutionary mechanisms. Allen et al. (2004) have presented evidence that miRNAs have diverged from inverted gene duplications and represent older remnants of such events that once produced siRNAs.

The small RNA sequencing populations from the seed coat and cotyledons have revealed a vast number of additional small

RNAs (miRNAs or siRNAs) varying greatly in normalized sequence counts. Many have much higher occurrence than the *CHS* siRNAs characterized here and some also show tissue specificity. We have clearly shown that the *CHS* siRNAs are physiologically functional to downregulate a pathway and produce a visible trait difference, lack of seed coat pigmentation. Thus, we anticipate that continued investigation of the novel sequences revealed in these populations will lead to similar examples of regulation of other pathways in seed development as demonstrated here for the *CHS* siRNAs.

METHODS

Plant Materials and Genetic Nomenclature

The two isoline pairs of *Glycine max* used for this study were obtained from the USDA Soybean Germplasm Collections (Department of Crop Sciences, USDA/Agricultural Research Service University of Illinois, Urbana, IL). The genotypes of the four lines are described in Table 1. All lines are homozygous for the loci indicated, and only one of the alleles is shown for brevity in the tables and text.

Plants were grown in the greenhouse and tissues harvested from at least four plants of each isoline. Leaves and roots were harvested from 4-week-old plants and quick frozen in liquid nitrogen. Seed coats and cotyledons were dissected from seeds at varying stages of development based on the fresh weight of the entire seed: 10 to 25 mg, 25 to 50 mg, 50 to 75 mg, 75 to 100 mg, and 100 to 200 mg. Dissected seed coats and cotyledons from seeds of the 50 to 75 mg weight range were fast frozen in liquid nitrogen. All tissues were stored at −70°C till further use.

Small RNA Extraction and Gel Blot Analysis

LMW RNAs were isolated and probed as described previously (Hamilton and Baulcombe, 1999) with minor modifications. Total nucleic acids were extracted from the frozen seed coats, cotyledons, leaves, and roots of the two isogenic pairs using the standard phenol chloroform method (Todd and Vodkin, 1996) and precipitated with ethanol. Seed coats of the Williams 55 isoline produce procyanidins and were pretreated with proanthocyanidin binding buffer using the protocol of Wang et al. (1994), before extracting the total nucleic acids.

To the precipitate dissolved in water, polyethylene glycol (molecular weight 8000) and sodium chloride were added to a final concentration of 5% and 0.5 M, respectively, followed by incubation on ice for 30 min. High molecular weight nucleic acids were precipitated by centrifugation at 11,000 rpm for 20 min, while the LMW nucleic acids in the supernatant were recovered by ethanol precipitation at −20°C overnight. LMW RNA concentrations were measured on the NanoDrop ND1000 spectrophotometer (Nanodrop Technologies) and samples stored at −70°C until further use. For diagnostic purposes, the LMW RNA fractions were separated on a 1.2% agarose/3% formaldehyde gel and stained with ethidium bromide. The predominant stainable species of these gels was a band that runs at ~200 bp.

Seventy-five micrograms of LMW RNA concentrated in 16 μL 50% formamide was denatured at 70°C for 10 min. Denatured LMW RNAs were fractionated on 15% polyacrylamide 7 M urea denaturing gels, transferred to Hybond-NX membrane (Amersham) using a Bio-Rad Trans-Blot apparatus (Bio-Rad) at 100 V for 1 h. The membranes were equilibrated on 20× SSC saturated filters, air-dried, and UV cross-linked (Stratalinker; Stratagene). Prehybridization was performed in 50% formamide, 7% SDS, 0.05 M $NaHPO_4/NaH_2PO_4$, pH 7.0, 0.3 M NaCl, 5× Denhardt's solution, and 100 μg/mL sheared denatured salmon sperm DNA at 40°C for at least 2 h. Hybridization was performed in the same solution by adding the hydrolyzed [α-^{32}P]UTP-labeled riboprobe or the [γ-^{32}P]dATP-labeled oligoprobe at 40°C for 15 to 20 h. The filters were washed in 2× SSC and 0.2% SDS at 40°C for 15 min and exposed to Hyperfilm (Amersham).

For accurate sizing of the siRNA species, an RNA ladder (10 to 150 nucleotides) was used and radiolabeled with [γ-^{32}P]dATP following the protocol provided with the Decade Markers Kit from Ambion. In the case of the RNA gel blot shown in Figure 2, 50 pmoles of two sense DNA oligonucleotides, a 20-mer (CHS7RT-1F), and a 25-mer (CHS7RT-si25) corresponding to a region in the second exon of *CHS7* were also run on the same gel (data not shown).

The *CHS* antisense riboprobe used for LMW RNA analysis was transcribed in vitro from the T7 promoter of a *Bam*HI cleaved *CHS7* EST, AI437793, by means of the MAXIscript In Vitro Transcription Kit (Ambion). AI437793 contains the full-length *CHS7* open reading frame. Riboprobes were treated with RNase free DNase to remove the DNA template, and the 20 μL probe was hydrolyzed to an average size of 50 nucleotides with 300 μL of 0.2 M carbonate buffer (0.08 M $NaHCO_3$ and 0.120 M Na_2CO_3) by incubating at 60°C for 3 h. Subsequently, 20 μL of 3M NaOAc, pH 5.0, was added to the hydrolyzed probe before adding the probe to the hybridization solution.

The 5S rRNA oligoprobe was used as a loading control. A 27-mer oligo (5′-GGTGCATTAGTGCTGGTATGATCGCAC-3′) antisense to the soybean 5S rRNA encoding gene was γ-radiolabeled using the DNA 5′ End-Labeling System (Promega) according to the manufacturer's instructions. Unincorporated nucleotides were removed using BioSpin 6 chromatography columns (Bio-Rad).

Sequencing of Small RNA Libraries and Data Analysis

Gel purification, cloning, and sequencing of small RNAs from multiple tissue samples (seed coats and cotyledons of Williams [*i*], seed coats of Williams 55 [*i*], and seed coats of Richland [*I*]) were performed at Illumina using the SBS (sequencing by synthesis) technology. Briefly, 2.5 to 5 μg of the purified LMW RNA fraction of each of the four samples was provided to Illumina, which subsequent to quality checks, was separated on 15% polyacrylamide gels containing 7 M urea in TBE buffer (45 mM Tris-borate, pH 8.0, and 1.0 mM EDTA). A gel slice containing RNAs of 15 to 35 nucleotides was excised and eluted. Gel-purified small RNAs were ligated to the 3′ adapter (5′-TCGTATGCCGTCTTCTGCTTG-3′), and the small RNA libraries sequenced using the Illumina Genetic Analyzer. Sequence information was extracted from the image files with the Illumina Firecrest and Bustard applications.

A total of three to six million reads that were 33 bases long were obtained from the deep sequencing of the above-mentioned libraries. Adapter trimming was performed using the first occurrences of substring TCG as the unique identifier for the beginning of the adapter (5′-TCGTATGCCGTCTTCTGCTTG-3′). The sizes of the small RNAs after adapter trimming ranged from 14 to 33 nucleotides, with the majority in the range of 19 to 24 nucleotides. Adapter trimmed sequences were compared to obtain the number of unique sequences and occurrences of each. At this stage, all sequences present more than five times were carried forward for subsequent comparisons.

Alignments of these curated small RNAs to each individual BAC sequence were made using BLAST (Altschul et al., 1990) with minimum match length of 16 bases with no mismatches or 20 bases with one mismatch allowed. Also, alignments were made to individual *CHS* sequences with at least 14 bases with no mismatches or 18 bases with one mismatch allowed. For the alignments to individual *CHS* sequences, the variable length intron was omitted so that the CHS protein coding regions would be in maximum alignment throughout their 1167 bases (for *CHS1-6* and *CHS9*) and 1170 bases (for *CHS7* and *CHS8*). A total of 200 bases from the genomic sequence 5′ of the ATG start codon and 200 bases 3′ of the stop codon of each gene were taken to represent the flanking regions,

which brings the sequences to 1567 or 1570 nucleotides. The results from BLAST analyses were further characterized, cross-compared, and scrutinized with Excel tools. In some instances detailed alignments were performed with the *MultAline* program (http://bioinfo.genotoul.fr/multalin/multalin.html).

Accession Numbers

Sequence data used in this article can be found in the GenBank/EMBL databases under the following accession numbers: EF623854, EF623856, EF623857, EF623858, and EF623859, corresponding to the five *CHS* containing BACs, 77G7a, 56G2, 5A23, 28017, and 7C24, respectively (Tuteja and Vodkin, 2008). The sequences for the *CHS* family member genes were extracted from these BAC clone sequences, except for *CHS2*, which had accession number X65636. The accession number for the soybean 5S rRNA EST (Gm-c1015-7201) is X15199.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Table 1. Percentage of Genomic Sequence Similarity of Pairwise Alignments of the Nine Members of the *CHS* Gene Family.

Supplemental Table 2. Unique Small RNA Signatures from the Williams Seed Coat Library (i^i) with 100% Identity to *CHS* Genes in a Pairwise Comparison.

Supplemental Table 3. Percentage of Unique *CHS* siRNAs from the Williams (i^i) Seed Coat Library Aligning to *CHS* Sequences with 100% Identity That Are Shared between Different *CHS* Genes.

Supplemental Data Set 1. Small RNA Sequences from Seed Coat and Cotyledon Libraries That Align to Five BAC Sequences Containing *CHS* Genes.

Supplemental Data Set 2. Small RNA Sequences That Align to the Coding Regions of the Nine Individual *CHS* Genes.

ACKNOWLEDGMENTS

We thank Pam Long, Sean Bloomfield, and Martin Blistrabas for assistance with data analysis. This work was supported by grants from the University of Illinois Critical Research Initiative Program, the USDA, the Illinois Soybean Association, and the United Soybean Board.

Received July 10, 2009; revised September 3, 2009; accepted September 16, 2009; published October 9, 2009.

REFERENCES

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.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. **215:** 403–410.

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

Baulcombe, D. (2004). RNA silencing in plants. Nature **431:** 356–363.

Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R., and Zhu, J.K. (2005). Endogenous siRNAs derived from a pair of natural cis-antisense transcript regulate salt tolerance in Arabidopsis. Cell **123:** 1279–1291.

Carrington, J.C., and Ambros, V. (2003). Role of microRNAs in plant and animal development. Science **301:** 336–338.

Carthew, R.W., and Sontheimer, E.J. (2009). Origins and mechanisms of miRNAs and siRNAs. Cell **136:** 642–655.

Chapman, E., and Carrington, J.C. (2007). Specialization and evolution of endogenous small RNA pathways. Natl. Rev. **8:** 884–896.

Chuck, G., Candela, H., and Hake, S. (2009). Big impacts by small RNAs in plant development. Curr. Opin. Plant Biol. **12:** 81–86.

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 I locus. Genome **47:** 819–831.

Della Vedova, C.B., Lorbiecke, R., Kirsch, H., Schulte, M.B., Scheets, K., Borchert, L.M., Scheffler, B.E., Wienand, U., Cone, K.C., and Birchler, J.A. (2005). The dominant inhibitory chalcone synthase allele *C2-Idf* (*Inhibitor diffuse*) from *Zea mays* (L.) acts via an endogenous RNA silencing mechanism. Genetics **170:** 1989–2002.

Dhaubhadel, S., Gijzen, M., Moy, P., and Farhangkhoee, M. (2007). Transcriptome analysis reveals a critical role of *CHS7* and *CHS8* genes for isoflavonoid synthesis in soybean seeds. Plant Physiol. **143:** 326–338.

Eamens, A., Wang, M.-B., Smith, N.A., and Waterhouse, P.M. (2008). RNA silencing in plants: Yesterday, today, and tomorrow. Plant Physiol. **147:** 456–468.

Grab, D., Loyal, R., and Ebel, J. (1985). Elicitor-induced phytoalexin synthesis in soybean cells: Changes in the activity of chalcone synthase mRNA and the total population of translatable mRNA. Arch. Biochem. Biophys. **243:** 523–529.

Hamilton, A., Voinnet, O., Chappell, L., and Baulcombe, D. (2002). Two classes of short interfering RNA in RNA silencing. EMBO J. **21:** 4671–4679.

Hamilton, A.J., and Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science **286:** 950–952.

Himber, C., Dunoyer, P., Moissiard, G., Ritzenthaler, C., and Voinnet, O. (2003). Transitivity-dependent and -independent cell-to-cell movement of RNA silencing. EMBO J. **22:** 4523–4533.

Jones-Rhodes, M.W., and Bartel, D.P. (2004). Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Mol. Cell **14:** 787–799.

Klahre, U., Crete, P., Leuenberger, S.A., Iglesias, V.A., and Meins, F., Jr. (2002). High molecular weight RNAs and small interfering RNAs induce systematic posttranscriptional gene silencing in plants. Proc. Natl. Acad. Sci. USA **99:** 11981–11986.

Kusaba, M., Miyahara, K., Iida, S., Fukuoka, H., Takano, T., Sassa, H., Nishimura, M., and Nishio, T. (2003). Low glutelin content1: A dominant mutation that suppresses the glutelin multigene family via RNA silencing in rice. Plant Cell **15:** 1455–1467.

Lauter, N., Kampani, A., Carlson, S., Goebel, M., and Moose, S.P. (2005). MicroRNA172 down-regulates glossy15 to promote vegetative phase change in maize. Proc. Natl. Acad. Sci. USA **102:** 9412–9417.

Llave, C., Kasschau, K.D., Rector, M.A., and Carrington, J.C. (2002). Endogenous and silencing-associated small RNAs in plants. Plant Cell **14:** 1605–1619.

Lu, C., Tej, S.S., Luo, S., Haudenschild, C.D., Meyers, B.C., and Green, P.J. (2005). Elucidation of the small RNA component of the transcriptome. Science **309:** 1567–1569.

Lucas, W.J., and Lee, J.Y. (2004). Plasmodesmata as a supracellular control network in plants. Nat. Rev. Mol. Cell Biol. **5:** 712–726.

Lucas, W.J., Yoo, B.C., and Kragler, F. (2001). RNA as a long-distance information macromolecule in plants. Nat. Rev. Mol. Cell Biol. **2:** 849–857.

Mallory, A.C., Mlotshwa, S., Bowman, L.H., and Vance, V.B. (2003).

The capacity of transgenic tobacco to send a systemic RNA silencing signal depends on the nature of the inducing transgene locus. Plant J. **35:** 82–92.

Mallory, A.C., Reinhart, B.J., Bartel, D., Vance, V.B., and Bowman, L.H. (2002). A viral suppressor of RNA silencing differentially regulates the accumulation of short interfering RNAs and micro-RNAs in tobacco. Proc. Natl. Acad. Sci. USA **99:** 15228–15233.

Matsumura, H., Watanabe, S., Harada, K., Senda, M., Akada, S., Kawasaki, S., Dubouzet, E.G., Minaka, N., and Takahashi, R. (2005). Molecular linkage mapping and phylogeny of the chalcone synthase multigene family in soybean. Theor. Appl. Genet. **110:** 1203–1209.

Matzke, M.A., and Matzke, A.J.M. (2004). Planting the seeds of a new paradigm. PLoS Biol. **2:** 582–585.

Muskens, M.W., Vissers, A.P., Mol, J.N., and Kooter, J.M. (2000). Role of inverted DNA repeats in transcriptional and post-transcriptional gene silencing. Plant Mol. Biol. **43:** 243–260.

Napoli, C., Lemieux, C., and Jorgensen, R. (1990). Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. Plant Cell **2:** 279–289.

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.

Palauqui, J.C., Elmayan, T., Pollien, J.M., and Vaucheret, H. (1997). Systemic acquired silencing - transgene-specific post-transcriptional silencing is transmitted by grafting from silenced stocks to non-silenced scions. EMBO J. **16:** 4738–4745.

Palmer, R.G., Pfeiffer, T.W., Buss, G.R., and Kilen, T.C. (2004). Qualitative genetics. In Soybeans: Improvement, Production and Uses, 3rd ed, H.G. Boerma and J.E. Specht, eds (Madison, WI: American Society of Agronomy), pp. 137–233.

Ramachandran, V., and Chen, X. (2008). Small RNA metabolism in Arabidopsis. Trends Plant Sci. **13:** 368–374.

Senda, M., Masuta, C., Ohnishi, S., Goto, K., Kasai, A., Sano, T., Hong, J.-S., and MacFarlane, S. (2004). Patterning of virus-infected soybean seed coat is associated with suppression of endogenous silencing of chalcone synthase genes. Plant Cell **16:** 807–818.

Smith, N.A., Singh, S.P., Wang, M.B., Stoutjesdijk, P.A., Green, A.G., and Waterhouse, P.M. (2000). Totel silencing by intron-spliced hairpin RNAs. Nature **407:** 319–320.

Sunkar, R., and Zhu, J.K. (2004). Novel and stress-regulated micro-RNAs and other small RNAs from *Arabidopsis*. Plant Cell **16:** 2001–2019.

Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M., and Hannon, G.J. (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Science **453:** 534–538.

Tang, G., Reinhart, B.J., Bartel, D.P., and Zamore, P.D. (2003). A biochemical framework for RNA silencing in plants. Genes Dev. **17:** 49–63.

Thorne, J.H. (1981). Morphology and ultrastructure of maternal seed tissues of soybean in relation to the import of photosynthate. Plant Physiol. **67:** 1016–1025.

Timmons, L., Tabara, H., Mello, C.C., and Fire, A.Z. (2003). Inducible systemic RNA silencing in *Caenorhabditis elegans.* Mol. Biol. Cell **14:** 2972–2983.

Todd, J.J., and Vodkin, L.O. (1993). Pigmented soybean (*glycine-max*) seed coats accumulate proanthocyanidins during development. Plant Physiol. **102:** 663–670.

Todd, J.J., and Vodkin, L.O. (1996). Duplications that suppress and deletions that restore expression from a chalcone synthase multigene family. Plant Cell **8:** 687–699.

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., and Vodkin, L.O. (2008). Structural features of the endogenous *CHS* silencing and target loci in the soybean genome. Crop Sci. **48:** 49–69.

Van der Krol, A.R., Mur, L.A., Beld, M., Mol, J.N.M., and Stuitje, A.R. (1990). Flavonoid genes in petunia: Addition of a limited number of gene copies may lead to a suppression of gene expression. Plant Cell **2:** 291–299.

Voinnet, O., Vain, P., Angell, S., and Baulcombe, D.C. (1998). Systemic spread of sequence-specific transgene RNA degradation in plants is initiated by localized introduction of ectopic promoterless DNA. Cell **95:** 177–187.

Wang, C.S., Todd, J.J., and Vodkin, L.O. (1994). Chalcone synthase mRNA and activity are reduced in yellow soybean seed coats with dominant *I* alleles. Plant Physiol. **105:** 739–748.

Watanabe, T., et al. (2008). Endogenous siRNAs from naturally formed dsRNas regulate transcxripts in mouse oocytes. Science **453:** 539–543.

Winston, W.M., Molodowitch, C., and Hunter, C.P. (2002). Systematic RNAi in *C. elegans* requires the putative transmembrane protein SID-1. Science **295:** 2456–2459.

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

Zabala, G., and Vodkin, L.O. (2003). Cloning of the pleiotropic *T* locus in soybean and two recessive alleles that differentially affect structure and expression of the encoded flavonoid 3′ hydroxylase. Genetics **163:** 295–309.

Zabala, G., and Vodkin, L.O. (2005). The *wp* mutation of *Glycine max* carries a gene-fragment-rich transposon of the CACTA superfamily. Plant Cell **17:** 2619–2632.

Zabala, G., and Vodkin, L.O. (2007). A rearrangement resulting in small tandem repeats in the *F3′5′H* gene of white flower genotypes is associated with the soybean *W1* locus. Plant Genome **2:** S113–S124.

Zabala, G., Zou, J., Tuteja, J., Gonzalez, D.O., Clough, S.J., and Vodkin, L.O. (2006). Transcriptome changes in the phenylpropanoid pathway of *Glycine max* in response to *Pseudomonas syringae* infection. BMC Plant Biol. **6:** 26.

Zamore, P., and Haley, B. (2005). Ribo-gnome: The big world of small RNAs. Science **309:** 1519–1524.

The Plant Cell, Vol. 22: 1647, May 2010, www.plantcell.org

Correction

Jigyasa H. Tuteja, Gracia Zabala, Kranthi Varala, Matthew Hudson, and Lila O. Vodkin (2009) Endogenous, Tissue-Specific Short Interfering RNAs Silence the Chalcone Synthase Gene Family in *Glycine max* Seed Coats. Plant Cell **21:** 3063–3077.

Complete sequence data for the four libraries described in this article can be found in the Gene Expression Omnibus of the National Center for Biotechnology Information as series number GSE21825 and samples GSM543393, GSM543394, GSM543395, and GSM543396. Raw data is entered in the Sequence Read Archive as SRP002459.

www.plantcell.org/cgi/doi/10.1105/tpc.110.220560