ABSTRACT

Title of Document:	AN INVESTIGATION OF MECHANISMS OF AN INTRON-MEDIATED GENE SILENCING
	Dandan Li, Doctor of Philosophy, 2014
Directed By:	Professor Zhongchi Liu Department of Cell Biology and Molecular Genetics

Introns, non-coding regions between exons, exist in most eukaryotic genes. Many studies have shown that introns regulate gene expression through both transcriptional and posttranscriptional mechanisms. I revealed that in *Arabidopsis thaliana*, over-expression of the first and longest intron of *CAULIFLOWER (CAL)* gene led to the silence of the endogenous *CAL* gene through DNA methylation, which is mediated by the intron-derived 24 nt siRNAs.

I investigated mechanisms of this intron-mediated gene silencing phenomenon through several different approaches, including northern blot, qRT-PCR, small RNA sequencing, bisulfite sequencing, McrBC-PCR, and bioinformatics. The *CAL* first intron does not show evolutionary conversation among different species in Rosid family. A distinctly folded stable secondary structure was found in the *CAL* first intron but its relevance to the silencing remains to be determined. Further, the *CAL*

first intron likely possesses regulatory sequences demonstrated by the intron's ability to induce *GUS* reporter gene expression when fused upstream of a *TATA* box and the *GUS* gene. Antisense long non-coding RNAs (lncRNAs) from the intron was detected by qRT-PCR, which may pair with the over-expressed sense *CAL* first intron transcript from the transgene to from double stranded RNAs and subsequently generate 24 nt siRNAs. Therefore, this study provides a potentially novel and simple method to silence target genes by over-expressing *cis* regulatory elements either in introns or in promoters.

I investigated how the intron-mediated silencing is inherited and showed that the silencing of *CAL* occurs in seeds and the silencing efficiency is dependent on the length of seed storage time. The extent of methylation in the *CAL* first intron increases when the seed age increases. This work has important biotechnology implications.

Combined, my research not only describes a novel phenomenon of intron-mediated silencing but also sheds light in the mechanism of intron- or *cis* regulatory element-mediated gene silencing. Hence my research work will have broad implications both in basic research and in biomedical and agricultural applications.

AN INVESTIGATION OF MECHANISMS OF AN INTRON-MEDIATED GENE SILENCING

By

Dandan Li

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2014

Advisory Committee:
Dr. Zhongchi Liu, Chair
Dr. Antony Jose
Dr. Douglas Julin
Dr. Steve Mount
Dr. Shunyuan Xiao

Dedication

I dedicate this to my dear father, who cannot be with me at every important moment in my life.

Acknowledgements

It would not be possible for me to finish this doctoral dissertation without the support, encouragement, and guidance from my committee members, my family and friends.

Above all, I would like to express my deepest gratitude to my advisor, Dr. Zhongchi Liu, for her excellent guidance, caring, patience and understanding. I have been so fortunate to continue my graduate study in Dr. Liu's lab. She has spent much time in training me to work with plants, teaching me to do new experiments, guiding me to think independently, revising my proposal and thesis, helping me to practice every presentation, and cheering with me for every little progress I made. No matter when I encountered problems in my project, she always supported and encouraged me. Her recommendations and instructions have enabled me to assemble and complete this dissertation effectively. I thank Dr. Liu for everything she has done for me.

I would like to thank my committee members, Dr. Steve Mount, Dr. Shunyuan Xiao, Dr. Antony Jose and Dr. Douglas Julin, for their insightful comments, hard questions, brilliant ideas, and encouraging words, allowing me to reach this accomplishment. Special thanks go to Dr. Julin, who was willing to join my thesis committee at the last moment.

I would like to thank the previous and present members of Liu lab. They form a collaborative and friendly atmosphere. Specifically, I thank Yifan Wang and Jing Wang for the happy times we spent in the lab.

Finally and most importantly, I would like to thank my loving family standing by me through the good times and bad. I thank my mother for her unreserved dedication and for her faith in me. When I worked into the late nights, she always took good care of my daughter without any complaints. Thanks to my little baby, Emay, for being such a good and sweet girl. Thanks to my younger brother for the moral and emotional support.

Table of Contents

List of Tables

Chapter 2

Chapter 3

Chapter 4

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 1: Introduction

Introns

Introns are non-coding DNA regions between exons in eukaryotic genes. Previously they were considered as intervening "functionless" sequences. Introns are transcribed together with exons into primary messenger RNA (pre-mRNA) by RNA polymerase II, but subsequently removed by splicing machineries to generate a mature messenger RNA (mRNA) prior to translation. According to the prediction of intron structure and the type of RNA splicing reaction, introns are classified into five distinct categories: introns of nuclear protein-coding genes that are removed by spliceosomes (denoted as spliceosomal introns), introns of transfer RNA genes that are removed by specific tRNA splicing enzymes, and self-splicing group I, II and III introns (Clancy, 2008). Specifically, spliceosomal introns are present in most eukaryotic organisms and are most studied among the five categories of introns.

Intron splicing

Spliceosomal introns possess the *cis*-acting consensus sequences in the 5'-end splice site (5'SS), the 3'-end splice site (3'SS) and the middle-located branch point sequences (BPS) (Figure 1.1A). These specific sequences are required for proper and efficient splicing by the spliceosome (Wahl et al., 2009). Spliceosome is a large and highly dynamic complex, mainly constituted by five Uridine-rich (U-rich) small nuclear ribonucleoproteins (snRNPs)—the U1, U2, U4/U6, and U5 snRNPs. The assembly of spliceosome is a sequential process in which snRNPs and different non-

snRNP proteins recognize the *cis*-acting sequences and load onto the pre-mRNA. In the E complex, U1 snRNP binds to the GU sequence at 5'SS, and Splicing Factor 1 (SF1)/Branch Binding Protein (BBP) binds to BPS, whereas the U2 auxiliary factor (U2AF) subunits U2AF65 and U2AF35 bind to BPS-downstream-located polypyrimidine tract (py-tract) and 3'SS, respectively. Then U2AF proteins recruit the U2 snRNP to the BPS with ATP hydrolysis, thereby forming A complex. Afterwards, a pre-assembled U4/U6.U5 tri-snRNP is recruited to form the B complex, which is still catalytically inactive although all snRNPs are present. Subsequently, U1 and U4 snRNPs are destabilized or released, and eventually spliceosome converts to the active B* complex, which changes its composition and performs the first step of catalytic reaction to generate the C complex (Figure 1.2). During this first splicing, the 2'OH of a specific adenine at BPS attacks the 5'SS and cleaves the sugar-phosphate backbone. The exposed 5'-end of the intron covalently linked to the adenine at the BPS, creating a free exon I and an "intron-lariat-exon II" intermediate (Figure 1.1B). Then the spliceosome undergoes additional rearrangement to facilitate the second splicing step, in which the released 3'-OH of exon I reacts with the 5'-end of exon II, consequently resulting in ligated exons and a free intron lariat (Figure 1.1B). Following this second catalysis, the spliceosome dissociates and releases U2, U5 and U6 snRNPs for recycling (Figure 1.2). The machinery of splicing is mostly conserved between plant and animal cells. (Wahl et al., 2009; Wachtel and Manley, 2009)

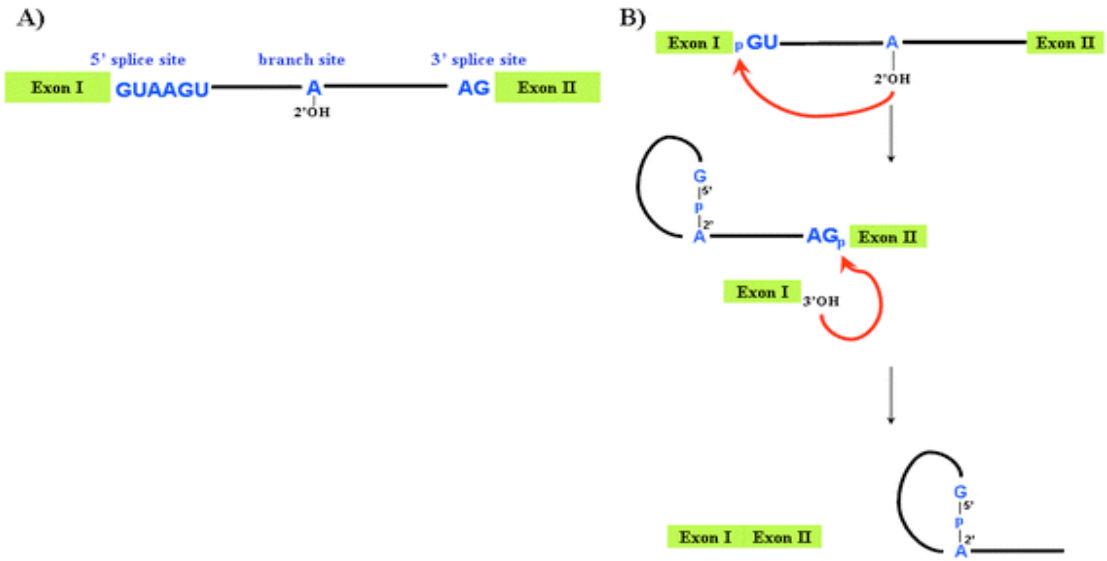


Figure 1.1. The *cis*-acting consensus sequences (A) and a depiction of the two-step catalytic reaction (B) involved in the intron-splicing process (Wachtel and Manley, 2009).

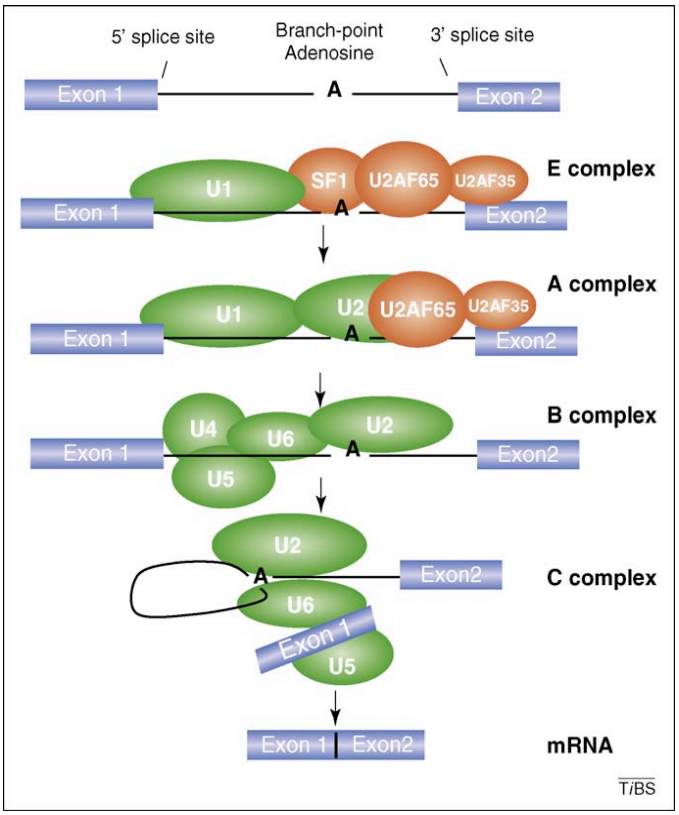


Figure 1.2. Spliceosome components cooperate to accomplish intron splicing (Warf and Berglund, 2010).

Indisputably, the replication and the transcription of introns, as the by-product of gene maintenance and expression, consume an immense amount of energy and metabolic resource. However, introns are still maintained within the eukaryotic genome throughout evolution, suggesting introns are indispensible to eukaryotes and may confer crucial functions.

Introns play important functions in eukaryotic gene expression and evolution

It has been acknowledged that introns facilitate rapid evolution through partial or entire loss or gain, and meiotic sequence crossing-over. Introns' position, length, density and content are considerably variable among different species, even between related lineages. In general, the introns of higher eukaryotic organisms are more complicated. For instance, the *Homo sapiens* (human) genome contains 139418 introns (8.4 introns/gene), whereas the *Encephalitozoon cuniculi* (unicellular fungus) genome has only 15 introns with 5'-end bias in the position within genes (0.0075 introns/gene) (Mourier and Jeffares, 2003). However, an essential catalyst of evolution as the main function of introns is still a conventional yet unproven view. To date, although the full function of introns remains mysterious, extensive researches have demonstrated that introns may play profound and once-neglected roles in gene function and evolution. One of the vital roles of introns is to regulate gene expression either negatively or positively.

Alternative splicing of introns is an important means of gene regulation

Alternative splicing (AS) is a process by which a single pre-mRNAs can be spliced in different arrangements to produce multiple mRNA and protein variants. This effectively increases the diversity of proteomes. In *Arabidopsis thaliana*, under normal growth conditions, the frequency of AS is higher than 61% of intron-containing genes, based on a recent extensive RNA-seq analysis (Syed et al., 2012). AS is regulated by a complex signaling network that responds to multifarious intracellular and extracellular signals, such as hormones and stresses. Thus, this AS frequency of 61% in *Arabidopsis* may change as plants in various environmental conditions or varied plant tissues at different developmental stages are analyzed (Syed et al., 2012).

Alternative splicing is also intimately tied to nonsense-mediated mRNA decay (NMD), a eukaryotic mRNA quality surveillance mechanism. Specifically, NMD recognizes mRNA transcripts containing premature termination codons (PTC) and directs these mRNA for degradation to avoid the production of truncated proteins (Lewis et al., 2003). The study in the *rice* (*Oryza sativa*) *waxy mutant* discovered that when possessing the identical PTC, the fully spliced RNA is degraded much quicker than the partially spliced RNA which retains the first intron, indicating the potential ability of introns to affect the efficiency of NMD in plants (Isshiki et al., 2001). Many researchers have demonstrated that like NMD pathway in mammalian cells and yeast, an intron located at downstream of PTC, as a *cis*-acting element, is required for NMD activation in plants (Maquat, 2004; Kertesz et al., 2006). Besides contributing to the

expansion of proteome diversity, AS also generates defective transcript isoforms that can be degraded by NMD. Therefore, under certain conditions, AS is able to switch the transcripts from normal translation to NMD pathway. In another word, coupling with NMD, AS greatly contributes to the regulation of gene expression in a tissue-specific or developmental stage-specific manner (Maquat, 2004; McGlincy and Smith, 2008).

Introns may possess regulatory elements or structures

In addition to intron splicing enhancers (ISE), which can serve as *cis*-acting auxiliary sequences to facilitate splicing, it also has been well documented that RNA structures derived from both introns and exons, inhibit or aid spliceosomal component to be loaded onto pre-mRNA, thereby directly regulating splicing (McManus and Graveley, 2011; Warf and Berglund, 2009).

Moreover, Large-scale bioinformatic analyses revealed that significantly longer first introns exist in various species, including mammals, insects, plants and fungi, suggesting a common feature in eukaryotic genes. In addition, through the study of 18217 human ref-sequence genes, a particular G-quadruplex (G4) structure within the first 100bp of these long first introns was discovered (Eddy and Maizels, 2008). The G4 structure possesses at least four runs of guanines, with at least three guanines per run, separated by regions of DNA, forming a G-tetrad stacking fold. A generally adopted pattern is $G_{[3-5]}$-N_{L1}-$G_{[3-5]}$-N_{L2} -$G_{[3-5]}$-N_{L3}-$G_{[3-5],}$ where $G_{[3-5]}$ means three to five guanines and $N_{L1/2/3}$ stands for the loops constituted by several any nucleotides between two G-tetrad layers. Usually, the length of loop sequence is variable from 1

nt to 7 nt (Todd et al., 2005). The G4 structure located in the first intron is often close to the promoter and translation start site (TSS) and thus provides a structural target for regulatory proteins of transcription or translation. (Barrett et al., 2012). For example, Telomeres are rich in Guanines and prone to form G4 structures. Recently, researches provided strong evidence that G4 structures are specifically bound and regulated by telomere-binding proteins throughout the cell cycle (Juranek and Paeschke, 2012). Also, G4 structures are required for the maintenance of telomere length (Watson et al., 2013).

Some introns harbor *cis* regulatory elements that are bound by trans-acting factors to facilitate transcription of genes. The best-studied example is the enhancer element in the second intron of *AGAMOUS (AG)* in *Arabidopsis thaliana*. *AG* is a C-class floral homeotic gene, and encodes a MADS box transcription factor required to specify the identity of stamens and carpels in the third and fourth whorls in developing flowers. In addition, *AG* is responsible for meristem determinacy (Yanofsky et al., 1990; Sieburth and Meyerowitz, 1997; Bowman et al., 1989). At the early flower developmental stage, *AG* RNA is only restricted to cells of floral meristem that will later give rise to stamens and carpels (Drews et al., 1991). In the mature floral organs, *AG* expression shows a distinct pattern, in which *AG* RNA is concentrated and restricted in the connective tissue of stamens and in the stigma and ovules of carpels (Drews et al., 1991; Bowman et al., 1991).

The second intron of *AG* is 3 kb in length and contains binding sites for several transcription factors including those of LEAFY (LFY) and WUSCHEL (WUS), as

well as CArG box and other functional motifs (Busch et al., 1999; Lohmann et al., 2001; Hong et al., 2003). Using phylogenetic footprinting and shadowing, at least 6 *cis*-regulatory elements have been identified in *AG* second intron and are conserved among multiple Brassicaceae species (Hong et al., 2003). Additionally, by beta-glucuronidase (*GUS*) reporter gene analyses, theses motifs were proven to be functionally important for *AG* expression (Hong et al., 2003; Sieburth and Meyerowitz, 1997). When the promoter of *AG* was fused to *GUS*, the construct failed to show floral organ-specific GUS staining (Sieburth and Meyerowitz, 1997). In contrast, a *GUS* fusion that includes both *AG* promoter and the second intron exhibited a floral organ (stamen and carpel)-specific staining pattern in flowers, same as the *AG* mRNA expression detected by *in situ* hybridization (Sieburth and Meyerowitz, 1997; Drews et al., 1991; Bowman et al., 1991). This suggested that the *AG* second intron is required for proper spatial, temporal and quantitative *AG* expression (Sieburth and Meyerowitz, 1997). In a more recent study, it was shown that *AG* second intron was able to drive stamen- and carpel-specific *GUS* reporter gene expression even in the absence of a minimal *TATA* box promoter (Singer et.al, 2010). Thus *AG* second intron alone is sufficient to provide all the necessary *cis* elements for stage- and organ-specific transcriptional initiation.

Some introns were found to possess cryptic alternative promoters, which can cooperate with AS to initiate distinct patterns of gene expression. For example, the *Arabidopsis SYN1* gene, which is essential for chromosome condensation and paring in meiosis, has an alternative promoter in its first intron; it can produce two distinct transcripts, a full-length BP2 transcript expressing at low levels in most tissues and a

truncated BP5 transcript expressed only in pre-bolting buds (Bai, 1999). Another study in human hepatoma cells demonstrated that the liver specific gene *α-fetoprotein* (*AFP*) harbors an alternative promoter in the first intron, which is active in the yolk sac and fetal liver, contributing to early expression of AFP (Scohy et al., 2000).

Intron-mediated enhancement (IME) is a phenomenon that promoter-proximal intron sequences enhance gene expression through the elevation of steady-stated mRNA accumulation (Rose, 2008, 2011). The DNA sequences exhibiting IME are preferentially located within 1 kb range of transcription start site and display a position-dependent feature, in which the IME ability disappears if the IME element is moved toward 3'-end of the gene (Rose, 2004). Compositional differences between promoter-proximal introns and other introns were noted and used to predict the IME ability by the computational tool called IMEter (Rose et al., 2008). The high IMEter score region often contains multiple and redundant TT NNGAT(c/t)TG(a/t, g/t) in *Arabidopsis* or T(c/a)GATc in *rice*. A recent report indicated that the IME is mediated through DNA not RNA (Ross, et al., 2011), so its distinction from traditional enhancer-containing introns becomes less obvious. While the specific mechanism underlying IME remains a puzzle, one likely explanation is that in the presence of IME sequences, RNA polymerase II maybe more easily associated with the rendered DNA (Rose 2011). IME has been observed in a wide range of organisms, such as fungi (Lugones et al., 1999), nematodes (Okkema et al., 1993), insects (Zieler and Huynh, 2002), plants (Vasil et al., 1989; Rose et al., 2008) and mammals (Palmiter et al., 1991). Hence IME may be a fundamental property of gene expression in all eukaryotes.

Introns and long non-coding RNAs

In *Arabidopsis thaliana*, less than 50% of the genome encodes proteins. However, thorough analysis of eukaryotic transcriptomes revealed that up to 90% of the genomes could be transcribed into protein-coding or non protein-coding RNAs (Kim and Sung, 2012). According to their diverse functions, the non-coding RNAs (ncRNAs) are divided into two major categories. One group is "housekeeping" ncRNAs, such as ribosomal RNAs, transfer RNAs, small nuclear RNAs and small nucleolar RNAs, all of which are required for maintenance of basic cellular functions. The other group is "regulatory" ncRNAs, including well-known microRNAs (miRNAs) and small interfering RNAs (siRNAs), which play important roles in transcriptional and post-transcriptional gene regulation. Recently, a large number of long non-coding RNAs (lncRNAs) have emerged and been recognized as potent regulators of gene expression (Kim and Sung, 2012).

LncRNAs, generally larger than 200 nt, are present in both animals and plants. They can derive from any regions throughout the whole genome, such as introns, intergenic locations, or others even overlapping with protein-coding sequences. Also, lncRNAs can be in either sense or antisense orientation. Some lncRNAs are precursors of small regulatory RNAs. For instance, *MIR* genes are transcribed by RNA Polymerase II (Pol II) to produce long transcripts that are further processed to mature miRNAs. In addition, plant-specific RNA Polymerase V (Pol V) generates a group of lncRNAs that are capped at 5'-end but do not possess poly-A tails at 3'-end. These lncRNAs serve as scaffolds for 24 nt siRNA-loaded ARGONAUTE4 (AGO4) proteins to

methylate target DNA loci in RNA-directed DNA Methylation (RdDM) pathway (Kim and Sung, 2012).

It has recently been discovered that non-coding RNA transcripts can originate from the sequences between distantly located enhancers and proximal promoters (Dobi and Winstn, 2007; Ho et.al, 2006; Masternak et al., 2003; Rogan et al., 2004; Tchurikov et al., 2009; Zhu et al., 2007; Singer et al., 2010). These non-coding RNAs are considered as the by-product of a facilitated tracking mechanism of enhancer-promoter communication, during which RNA polymerase II and other transcription factors move along the DNA until they encounter a competent promoter to initiate transcription. The intervening DNA sequence between the enhancer and the promoter may form a loop. During the tracking process, RNA pol II synthesizes a non-coding intragenic RNA from the intervening DNA sequence (Zhu et al., 2007). However, it has been not known if these non-coding RNAs possess any biological function.

In addition, some dormant promoters located in non-coding loci, are capable of generating lncRNAs in response to external stimuli. An interesting example is an intron-derived lncRNA that targets genes for repressive histone modifications and epigenetic silencing in the vernalization process of *Arabidopsis* (Heo and Sung, 2011; Kim and Sung, 2012). The potent floral repressor gene, *FLOWERING LOCUS C* (*FLC*), is epigenetically silenced, when plants undergo a prolonged winter cold, and hence in spring plants are competent to flower. An antisense non-coding RNA named *COOLAIR* from *FLC* was proposed to be involved in vernalization-mediated *FLC* repression (Swiezewski et al., 2009; Kim and Sung, 2012). However, the abolishment

of *COOLAIR* transcription failed to alleviate cold-induced *FLC* repression, indicating the *COOLAIR* lncRNAs are not critical for *FLC* silencing during vernalization (Helliwell et al., 2011; Kim and Sung, 2012). Nevertheless, a recent research found that during the cold exposure, the first intron of *FLC* is capable of producing an approximate 1.1 kb lncRNA transcript in the sense direction, which was termed *COLD ASSISTED INTRONIC NONCODING RNA* (*COLDAIR*). *COLDAIR* physically interacts with a component of POLYCOMB REPRESSIVE COMPLEX 2 (PRC2), a conserved repressive chromatin-remodeling complex, and as a scaffold, *COLDAIR* renders and directs PRC2 to *FLC*, resulting in histone H3 Lys27 methylation (H3K27me) and subsequent silencing of *FLC*. Furthermore, *COLDAIR* is required to enrich PCR2 at *FLC* chromatin and to maintain a stable repressive state of *FLC*. (Heo and Sung, 2011; Kim and Sung, 2012). In a similar case in human distal fibroblasts and metastatic breast cancers, a 2.2 kb lncRNA *HOTAIR* is expressed from *HOXC* locus. *HOTAIR* also acts as scaffold for multiple chromatin remodeling complexes, including PRC2, to promote repressive histone marks on its target chromatins (Tsai et al., 2010; Rinn et al., 2007).

In addition, the secondary structure formed by lncRNAs appears to be conserved among related species, such as “stem-and-loop” for recruiting PRC2, which is observed in *Xist* RNAs from the X chromosome in human and mouse (Jeon and Lee, 2011). Therefore, not only the sequences of lncRNAs, but also the compositional structures are implicated in their regulatory function.

Additionally, intron-derived small RNAs can act to regulate the expression of cognate

genes. High-throughput sequencing data in monocot *rice* (*Oryza sativa*) discovered certain special long hairpin introns (lhp intron). One example is the fifth intron of gene *LOC_Os07g01240.1,* which is able to produce a large quantity of siRNAs from their long stem regions, mainly in 21 nt, 22 nt or 24 nt (Chen et al., 2011). Especially, those lhp introns-derived 24 nt siRNAs can regulate their cognate host genes through RdDM in *cis*.(Chen et al., 2011). Additionally, *miR-7* of *C. elegans* (Stark et al., 2003) and *Mi-154* of mice (Seitz et al., 2003) are also encoded by introns, and involved in further gene regulation. Therefore, intron-derived regulatory small RNAs are ubiquitous, ranging from mammals, insects, nematodes to plants (Ying and Lin, 2004).

Intron and small RNAs in higher plants

Small RNA (sRNA) molecules of about 21-26 nucleotides (nt) encompass many different classes of non-coding RNAs. In higher plants, endogenous small RNAs are divided into two major classes--- siRNAs (small interfering RNAs) and miRNAs (microRNAs), both of which are able to exert regulatory function in both post-transcriptional gene silencing (PTGS) and transcriptional gene silencing (TGS). However, they are generated through distinct biogenesis pathways and function with specialized effector proteins.

miRNAs

The first miRNA, *lin-4*, was discovered in *Caenorhabditis elegans* (*C. elegans*) (Lee et al., 1993), and afterwards, a large number of miRNAs are found in both animals and plants (Chen, 2008). As illustrated in Figure 1.3, in plants, like protein-coding

genes, endogenous *MIR* loci are transcribed by RNA Polymerase II to generate the primary precursor pri-miRNAs with incomplete self-complementarity that can fold into a unique stem-loop structure (Pashkovskiy and Ryazansky, 2013; Zhang et al., 2006). Pri-miRNAs are first recognized and bound by TOUGH (TGH), SERRATE (SE), HYPONASTIC LEAVES1 (HYL1) and DAWDLE (DDL) proteins (Rogers and Chen, 2013; Pashkovskiy and Ryazansky, 2008). TGH binds single-stranded RNA (Ren et al., 2012), SE binds pri-miRNA (Machida et al., 2011), probably at junctions of single-stranded RNA and double-stranded RNA, HYL1 binds double-stranded RNA (Rasia et al., 2010; Yang et al., 2010) and DDL binds RNA and interacts with DICER LIKE 1 protein (DCL1), a RNase III-type enzyme (Yu et al., 2008; Pashkovskiy and Ryazansky, 2013). Subsequently, pri-miRNA is cleaved by DCL1 to be particular hairpin RNAs, known as pre-miRNAs, which then undergo the second-time processing by DCL1 and is further cleaved to be miRNA duplexes (Rogers and Chen, 2013). The RNA METHYLTRANSFERASE HUA-ENHANCER 1 (HEN1) stabilizes miRNA duplexes by depositing methyl group onto the 3'-terminal nucleotide (Yu et al., 2005). Unlike the biogenesis of miRNAs in animals, the methylation of plants miRNAs occurs after DCL proteins processing (Chen, 2008). After processed in the nucleus, the mature miRNAs are exported to cytoplasm by assistance of HASTY (HST) (Bollman et al., 2003; Park et al., 2005). The methylated miRNA guide strand associates with a specific ARGONAUTE (AGO) and other effector proteins to assemble RNA-induced silencing complex (RISC), triggering target mRNA destruction or translational inhibition. The complementary miRNA* strand is quickly degraded by SMALL RNA DEGRADING NUCLEASE

(SDN) (Carthew and Sontheimer, 2009; Rogers and Chen, 2013). In addition to the regulation of mRNA, some miRNAs also act to process the precursors of trans-acting siRNAs (tasiRNA) (Montgomery et al., 2008).

The biological roles of miRNAs are involved in the diverse developmental processes. They regulate leaf morphogenesis and patterning (Palatnik et al., 2003; Palatnik et al., 2007), leaf polarity establishment (Allen et al., 2007), vegetative developmental timing (Chen, 2004; Wu et al., 2009), floral organ identity (Aukerman and Sakai, 2003; Liu et al., 2010), and phytohormone signaling (Guo et al., 2005; Mallory et al., 2005). Recently, miRNAs emerged to act in the defense responses to pathogens (Naqvi et al., 2010) and in abiotic stress responses (Ding et al., 2013).

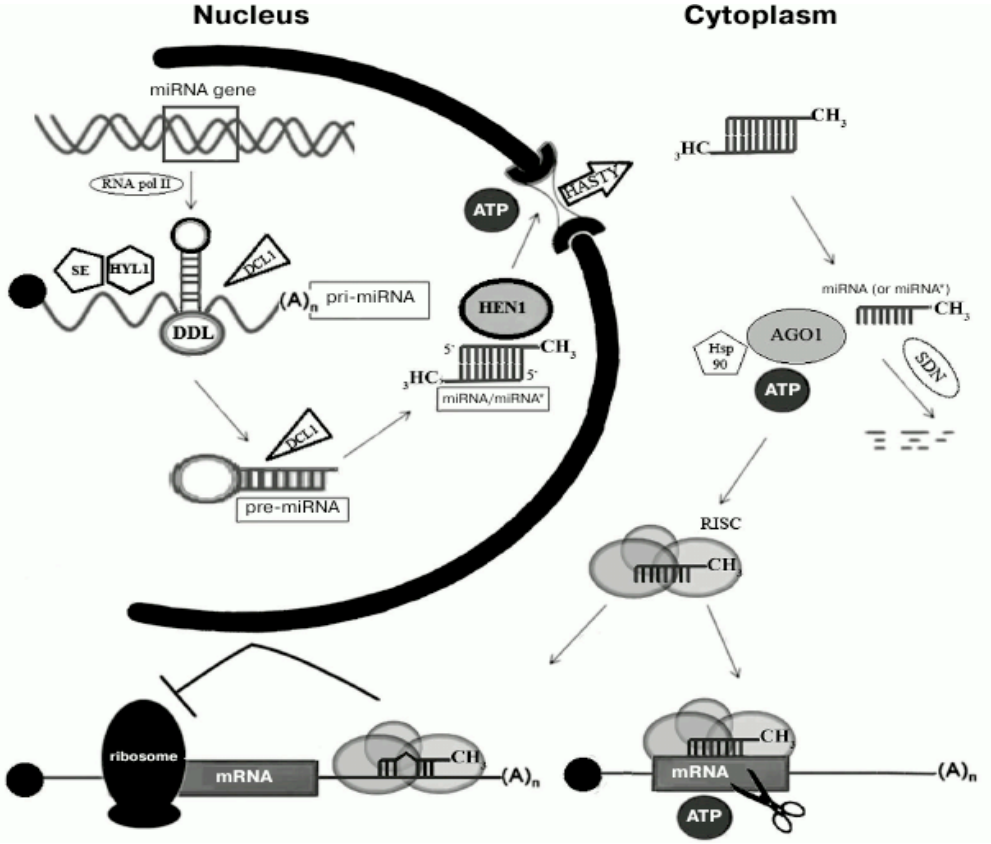


Figure 1.3. A model for miRNA biogenesis and function (Pashkovskiy and Ryazansky, 2013).

siRNAs

siRNAs are distinct from miRNAs in that they are derived from perfectly paired double stranded precursors (either from exogenous infectious viral dsRNAs, transgene, repeated sequences or endogenous transposons) (Molnar, 2011). According to the diversity of siRNA-generating loci and biogenesis pathways, endogenous siRNAs are divided into 3 categories (Figure 1.4). First, the siRNAs are from inverted and direct repeat sequences, including transposons, retroelements and inverted repeats. Second, the siRNAs originate from bidirectional transcripts. In plants, abiotic or biotic stresses can induce the production of antisense transcripts, which pair with the normally existing sense transcripts to form dsRNAs. These natural antisense transcript-derived siRNAs (natsiRNA) are involved in the biological response to internal and external stimuli. For the above two siRNA groups, The double-stranded RNA (dsRNA) precursor is further processed by DICER-LIKE proteins (DCL) into small dsRNAs around 20 bp. Unwinding into two single-stranded RNAs (ssRNA), the siRNA passenger strand is quickly degraded while the guide strand is incorporated by the AGO protein and loaded onto RISC. siRNAs direct RISC to corresponding perfectly complementary mRNA targets, which in turn are degraded. However, dissimilar to the other two groups, the third group is trans-acting siRNA (tasiRNA) arisen from specific TAS loci in plants. The primary transcript of TAS requires cleavage guided by a miRNA-loaded RISC, and then RNA DEPENDENT RNA POLYMERASE 6 (RDR6) is recruited to synthesize dsRNAs. The resulting dsRNAs are processed by DCL4 to generate 21 nt siRNAs. These tasiRNAs function via AGO1 or AGO7 to control leaf polarity and regulate hormone

responses in *Arabidopsis* (Montgomery et al., 2008; Chapman and Carrington, 2007).

Specifically, siRNAs in 21-22 nt length are the major players in the post-transcriptional gene silencing (PTGS) process (Schwab and Voinnet, 2010), while 24 nt siRNAs are mainly involved at transcriptional gene silencing (TGS) level (Zhang and Zhu, 2012). siRNAs can function in a remarkably wide range of biological processes, including chromosome rearrangement, genome reprogramming, antiviral defense and others (Finnegan and Matzke, 2003).

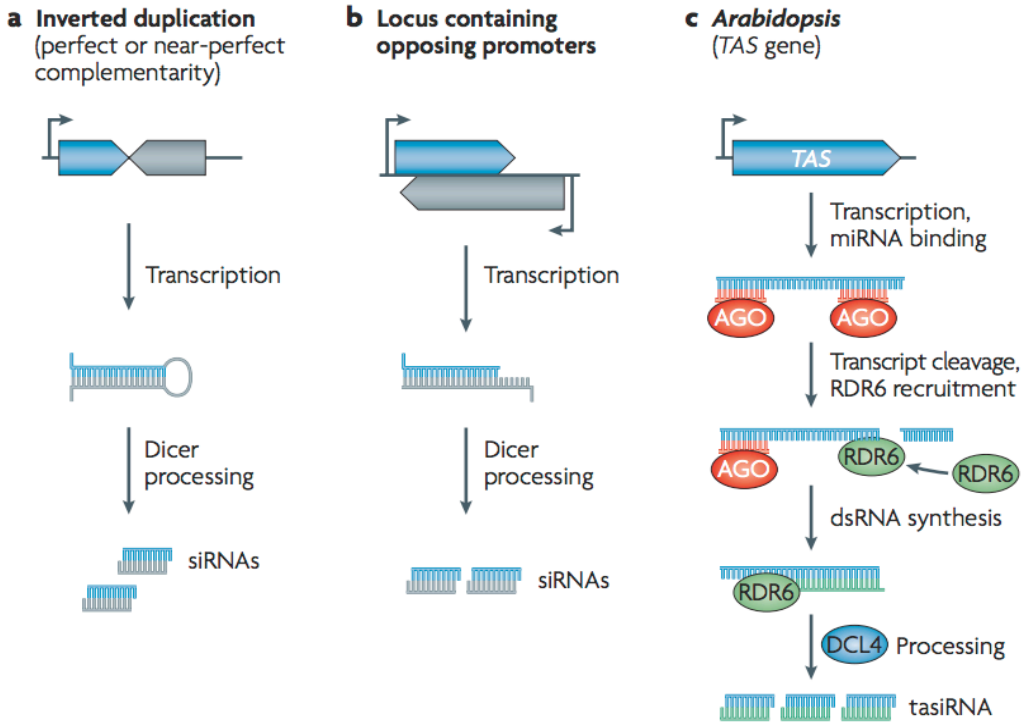


Figure 1.4. The biogenesis of siRNAs (Chapman and Carrington, 2007).

Transcriptional gene silencing and RdDM

Gene silencing can generally occur at two different levels: transcriptional gene silencing (TGS) and post-transcriptional gene silencing (PTGS). miRNA or siRNA-

mediated mRNA degradation and miRNA-mediated translational block represent PTGS. In the transcriptional level, 24 nt siRNAs are able to elicit de novo DNA methylation in cytosine at all three CG, CHG and CHH (H stands for A, T or C) contexts (Zhang and Zhu, 2011). DNA methylation can intervene gene transcription in two major ways. First, the methylated DNA itself could physically and directly obstruct the binding of transcription factors to the gene. Second, but more important, the methylation on DNA can recruit chromatin-modifying and remodeling proteins to methylate histone tails with repressive marks, thereby leading to a more compact and inaccessible heterochromatin-like structure. Strong evidence supports that DNA methylation can direct histone methylation and *vice versa* (Tariq et al., 2004). Strikingly, some introns are the target loci of DNA methylation and therefore affect gene expression negatively (Yegnasubramanian et al., 2011).

This above TGS phenomenon connecting small RNAs and epigenetic regulation is defined as RNA directed DNA methylation (RdDM), which is an important gene silencing mechanism through a small RNA-mediated epigenetic modification in both mammalian and plant cells (Simon and Meyers, 2011). The classical RdDM pathway can be divided into four major steps (Figure 1.5). First, in the biogenesis of 24nt siRNAs, RNA DEPENDENT RNA POLYMERASE 2 (RDR2) synthesizes the single stranded RNA transcribed by DNA DEPENDENT RNA POLYMERASE IV (Pol IV) into double-stranded precursors, which in turn are processed into 24 nt siRNAs by DCL3 (Zhang and Zhu, 2012; Simon and Meyers, 2011). Then HEN1 catalyzes 3'-end ribose methylation to stabilize the siRNAs. Second, both Pol V and Pol II have the ability to synthesize single-stranded scaffold RNAs from intergenic non-coding

regions, which are adjacent to the silenced loci (Zhang and Zhu, 2012; Simon and Meyers, 2011). Third, in the procedure of assembling the effector complex and recruiting methyltransferases, 24 nt siRNAs are loaded onto AGO4- or AGO6-containing RISC, and paired with nascent scaffold RNAs or complementary DNA targets, functioning as sequence-specific guide (Zhang and Zhu, 2012; Simon and Meyers, 2011). Through cooperation of all the specific effector proteins in RISC, DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) is recruited to initiate DNA methylation in target DNA loci. DRM2 is a plant-specific de novo DNA methyltransferase responsible for all symmetric and asymmetric sequence contexts and orthologous to mammalian DNA (cytosine-5) METHYLTRANSFERASE 3 (DNMT3) (Zhang and Zhu, 2012; Simon and Meyers, 2011). DNA methylation occurs prominently at the region of RNA-DNA matched sequences, supporting that the RNA-DNA base pairing acts as a substrate for RdDM. Fourth, in maintenance or reinforcement of DNA methylation pattern, DNA METHYLTRANSFERASE 1 (MET1), a plant ortholog of mammalian DNA (cytosine-5) METHYLTRANSFERASE 1 (DNMT1), appears to be responsible for sustaining CG methylation and CHROMOMETHYLASE 3 (CMT3), ubiquitous and unique in plants, is required for CHG methylation maintenance (Zhang and Zhu, 2012; Simon and Meyers, 2011). However, some evidence also support that MET1 and CMT3 can contribute to *de novo* methylation (Zhang and Zhu, 2012). Furthermore, distinct from CG and CHG methylation, CHH methylation cannot be retained by any maintenance methyltransferase and has to be reestablished by DRM2 during DNA replication. (Zhang and Zhu, 2012; Simon and Meyers, 2011)

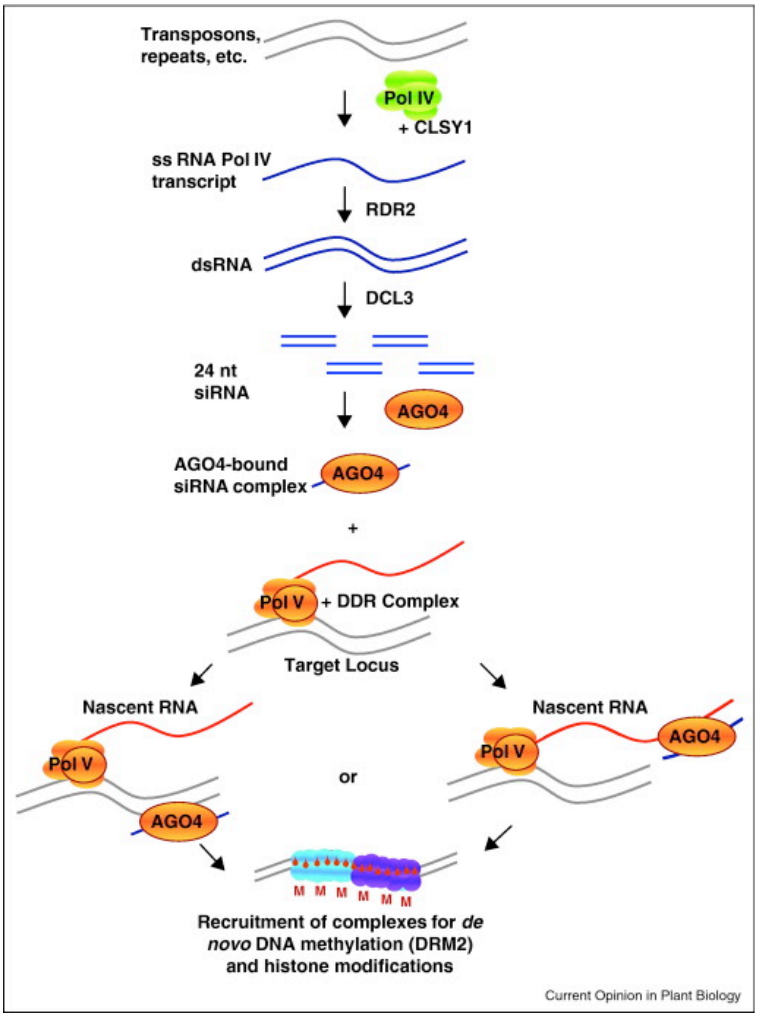


Figure 1.5. Summary of RdDM pathway (Simon and Meyers, 2011).

Epigenetics and Epigenetic Inheritance

Epigenetics is the functional modifications of genome without any alterations in the nucleotide sequence. Epigenetic marks consist of cytosine methylation, post-transcriptional modifications of histone tails and histone core variants, which control chromatin condensation and accessibility, and hence influence transposable elements and transcriptional activities of genes. Similar to genetic changes, epigenetic changes can be transmitted through mitosis and meiosis, as well as from one generation to the

next, which is referred to as epigenetic inheritance. The best-studied examples are paramutations and parental imprinting.

Paramutations were primarily observed in corn (Arteaga-Vazqueze et al., 2010), in which certain special homologous loci can communicate in trans to establish heritable expression states (Arteaga-Vazqueze and Chandler, 2010). In another word, the silenced allele can act in trans to paramutate the expression state of the non-silenced allele. Moreover, this epigenetic genotype can be stably inherited through generations. The phenomenon of paramutation was also described to occur between transgenes or between transgene and endogenous genes in multiple species (Khaitova et al., 2011; Suter and Martin, 2010). Although the mechanism behind paramutation is currently mysterious, it may involve an RNA-based mechanism for transferring regulatory information between alleles.

Another instance of epigenetic inheritance is parental imprinting, reported in both mammals and plants. In parental imprinting, a gene is expressed only when it was inherited from one parent. The copy derived from the other parent is inactive. Through the examination at the molecular level, the sole change found on the imprinted genes is the presence of extra methyl groups on DNA. The difference in the level of DNA methylation generally correlates with the difference in transcriptional activity of a gene.

DNA methylation takes place in three different sequence contexts, CG, CHG and CHH, where H represents A, T or C. DRM2 is a plant-specific *de novo* DNA methyltransferase responsible for methylation of all three sequence contexts and is

orthologous to mammalian DNMT3, which also functions in mammalian *de novo* DNA methylation. In mammals, CG methylation is maintained by DNMT1 and a cofactor named UBIQUITIN-LIKE CONTAINING PHD AND RING FINGER DOMAINS 1 (UHRF1), which recognizes hemimethylated DNA loci at replication fork. In *Arabidopsis*, MET1 and VARIATION IN METHYLATION (VIM) family proteins, orthologous to DNMT1 and UHRF1 respectively, are responsible for CG methylation (Feng et al., 2010). CHG methylation, ubiquitous and unique in plant genomes, requires a plant-specific DNA methyltransferase CMT3 and a histone methyltransferase KRYPTONITE (KYP) for maintenance (Feng et al., 2010). CHH methylation, also abundant in plants, cannot be retained by any maintenance methyltransferase and has to be reestablished by RdDM pathway-recruiting DRM2 after DNA replication. In mammals, CHG and CHH methylation mainly appear in stem cells, albeit in a relative low level, and can be maintained by PIWI-associated RNAs (piRNAs)-guided DNMT3 (Feng et al., 2010).

Besides methylation on DNA, modifications on histone tails also contribute to epigenetic regulation. In eukaryotic cells, nucleosome is a basic chromatin unit, in which DNA duplex wraps around a histone octamer core consisting of two of each H2A, H2B, H3 and H4. Linker DNA sequence, to which the linker histone H1 or H5 bind near the site of DNA entry and exit, connects nucleosomes (Henikoff et al., 2004). This combination of DNA and histones can modulate the condensation of local chromatin, thereby allowing more or less accessibility for transcriptional machineries. The principal post-transcriptional modifications of histone proteins, also called histone codes, take place on the N-terminal tails of H3 and H4 histones. The

modifications can be methylation, acetylation and ubiquitination on lysine, methylation of arginine, and phosphorylation on serine, resulting in positive or negative influence on gene expression (Rapp and Wendel, 2005). For instance, active transcription chromatin regions are commonly marked by H3K4me3 (histone H3 lysine 4 trimethylation), H3K36me3 and H3K79me2, whereas repressive chromatin is enriched for H3K9me3 or H3K27me3 (Zhou et al., 2011).

Some histone modifications can be maintained after mitosis and meiosis. For instance, *Arabidopsis* TRITHORAX-RELATED PROTEIN 5 (ATXR5) and ATXR6, the plant-specific histone H3K27 monomethyltransferaes, can interact with PROLIFERATING CELL NUCLEAR ANTIGEN (PCNA), which is a cofactor of DNA polymerase (Feng et al., 2010). During DNA replication, the methylation of H3K37 on the newly assembled nucleosomes could be established by ATXR5 and ATXR6. Nonetheless, the status of histone methylation is quite dynamic, as it is modulated by histone methyltransferases and demethylases, along with other histone-code reading proteins (Feng et al., 2010).

Furthermore, DNA methylation and histone modification are closely inter-related (Grewal and Moazed, 2003). Studies in *Drosophila* (Ahmad and Henikoff, 2002) and *Arabidopsis* (Soppe et al., 2002) revealed that CG methylation is a prerequisite for H3K9me. In addition, the *Arabidopsis* mutant of *Decrease in DNA methylation 1* (DDM1) encoding for a chromatin-remodeling factor not only exhibited a global reduction of DNA methylation in all cytosine sequence context (Singer et al., 2001), but also altered the pattern of H3K9me distribution (Gendrel et al., 2002; Tariq et al.,

2003). The above researches indicated that DNA methylation is able to direct histone methylation. In *Arabidopsis*, RNAi defective mutants were shown to lose *de novo* methylation mediated by DOMAINS REARRANGED METHYLTRANSFERASES (DRMs) (Chan et al., 2004). In fission yeast, mutations of genes essential for RNAi exhibited de-repression of transgene inserted at centromere and loss of typical H3K9me marks in centromere (Volpe et al., 2002). This indicated that siRNAs first direct DNA methylation, which in turn mediate histone marks H3K9me (Tariq and Paszkowski, 2004).

Additionally, studies in *Neurospora* revealed that the Dim-5 HMT, a H3K9 methyltransferase, is required for DNA methylation, suggesting that H3K9me3 directs DNA methylation (Tamaru et al., 2001; Tamaru et al., 2003). It was further supported by that the KFP, the first H3K9-specific methyltransferase found in plants, directly controls DNA methylation in the CHG contexts (Jackson et al., 2002). Therefore, epigenetic information also can flow from histone to DNA.

However, transgenerational epigenetic inheritance requires the passage of epigenetic modifications through germlines. In general, epigenetic marks that programmed during development must be restored in germ cells. For example, transposons and repetitive DNA elements must be silenced via methylation inheritance, in order to enforce the genome integrity. During genome reprogramming, a DNA glycosylase, DEMETER (DME) acts to erase cytosine methylation globally in the central cells of a mature female gametophyte and after fertilization, in the endosperm. Meanwhile, demethylation by DME may reactivate transposon expression, and subsequently transposon transcripts were tunneled into the RdDM pathway, producing 24 nt

siRNAs that mediate non-CG DNA methylation in egg cells and future embryos (Feng et al., 2010; Bourc'his and Voinnet, 2010).

For the male gametophyte pollen, in the vegetative nucleus, DME and its homolog REPRESSOR OF SILENCING 1 (ROS1) demethylate and reactivate the specific target transposons, leading to the accumulation of 24 nt siRNAs involved in RdDM machinery. In the sperm cell, methylation of symmetric CG and CHG is largely retained. CHH methylayion is initially lost in micropores and sperm cells, and then is restored in the embryo after fertilization, by *de novo* methyltransferase DRM2 guided by 24 nt siRNAs migrating prominently from the maternal origins (Slotkin et al., 2009; Calarco et al., 2012; Feng et al., 2010; Bourc'his and Voinnet, 2010). Therefore, small RNAs predominantly contribute to large-scale reprogramming and epigenetic heritance across generations.

Chapter 2: The discovery and characterization of an intron-mediated gene silencing phenomenon

Introduction

CAULIFLOWER (CAL) and *APETALA1 (AP1)*

APETALA1 (AP1) gene in *Arabidopsis thaliana* encodes a member of MADS-domain proteins, which are sequence-specific DNA-binding transcriptional activators. *AP1* gene is not only responsible for meristem identity, but also required for normal development of sepals and petals (Bowman et al., 1993; Gustafson-Brown et al., 1994; Irish et al., 1990). A distinguishing feature of *ap1* mutation is the loss of suppression of secondary floral meristems, causing extra flowers to form at the axils of sepals and sometimes even tertiary flowers at the axils of secondary flowers (Bowman et al., 1993; Gustafson-Brown et al., 1994; Irish et al., 1990). The severity decreases acropetally and normal flowers are eventually formed in apical positions, and the flowers are fertile (Figure 2.1).

CAULIFLOWER (*CAL*) gene in *Arabidopsis thaliana* also belongs to the MADS-box family and is closely related to *AP1* (Bowman et al., 1993; Kempin et al., 1995). *CAL* gene is a meristem identity gene, responsible for conferring floral characteristics to the lateral primordia produced by the shoot apical meristem (SAM) during the inflorescence (I) development (Ferrandiz et al., 2000). Genetic and molecular analyses revealed that the *CAL* is a paralog of *AP1* and partially redundant in function with *AP1* (Bowman et al., 1993; Kempin et al., 1995). Like *AP1*, *CAL* is expressed

strongly in the floral primordia (Bowman et al., 1993; Kempin et al., 1995). Mutations in *CAL* alone do not result in any phenotype. However, mutations in the *CAL* gene can enhance the *ap1* mutant phenotype. Plants homozygous for both *ap1* and *cal* loss of function mutations display the "cauliflower" phenotype (Figure 2.1), an indeterminate proliferation of inflorescence meristematic tissues, due to the inability to switch the meristems from inflorescences to flowers (Bowman et al., 1993; Kempin et al., 1995; Purugganan et al., 1998).

In this chapter, I documented in detail an intron-mediated gene silencing phenomenon using *CAL* first intron as my main experimental system.

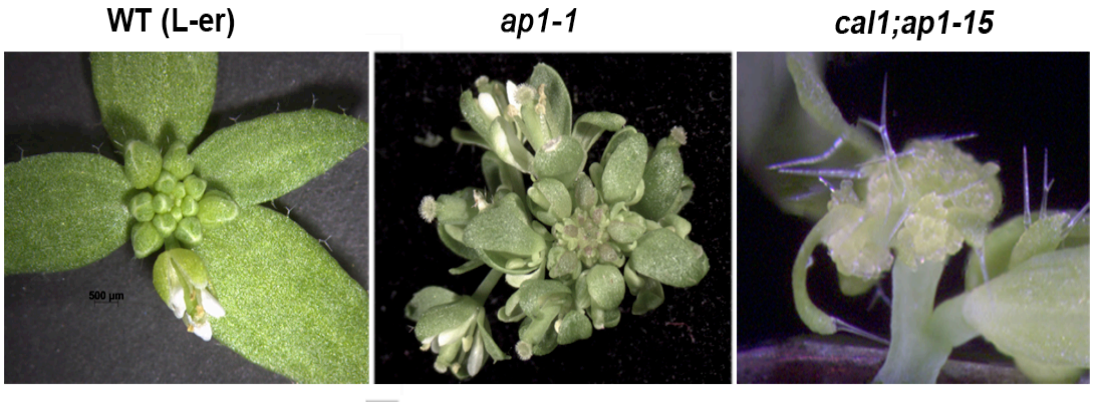


Figure 2.1. *Wild type* and *mutant* inflorescences.

Results

Over-expression of *CAL* first intron lead to silencing of endogenous *CAL*

Previously, our collaborator Dr. Zongrang Liu reported that a transgene containing the second and largest intron of *AG* in *Arabidopsis* could caused the silencing of endogenous *AG*, as these transgenic *Arabidopsis* plants exhibited floral phenotypes

similar to the *ag* loss-of-function mutants (Zongrang Liu, unpublished). To test whether this intron-mediated gene silencing phenomenon exists for other genes, four additional *Arabidopsis* genes were tested. These genes were chosen based on two criteria. First, the largest intron of the gene is more than 1 kb. Second, the respective mutant should exhibit a characteristic phenotype that is easily screened. These chosen genes are *STERILE APETALA (SAP), SHATTERPROOF1 (SHP1), APETALA1 (AP1)* and *CAULIFLOWER (CAL)*.

The largest intron of each gene was isolated and over-expressed under the control of *35S* promoter in either forward or reverse orientation. The constructs were transformed into *Arabidopsis* in appropriate genetic background, and the phenotypes of T1 transgenic plants were scored (Figure 2.2).

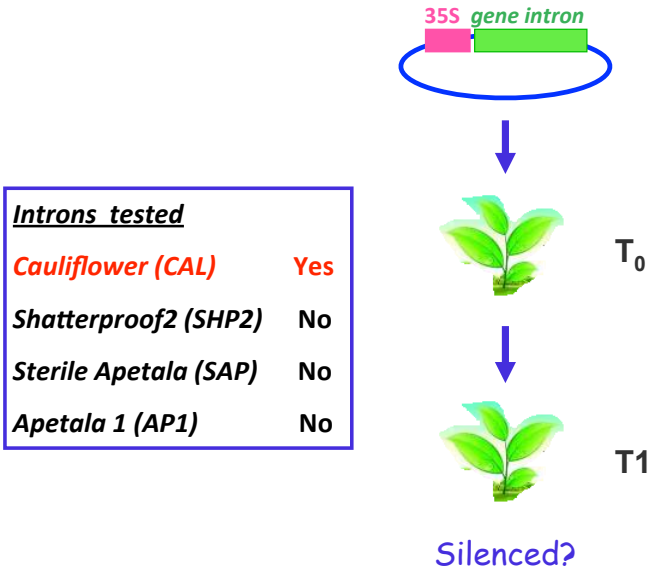


Figure 2.2. Scheme of over-expression of 4 individual introns to test the ability to induced gene silencing.

Gene *	Intron size	background	% Silence	# of T1	Binary vector
SHP1	1.3 kb	*shp2 (Ler)*	0	120	*pEARLEYGate 100*
SAP	2.8 kb	*Ler*	0	40	*pEARLEYGate 100*
AP1 (F)*	1.3 kb	*Ler*	0	406	*pEARLEYGate 100*
AP1 (R)*	1.3 kb	*Ler*	0	573	*pEARLEYGate 100*
CAL (F)*	1.0 kb	*ap1-1 (Ler)*	33.24	471	*pEARLEYGate 100*
CAL (R)*	1.0 kb	*ap1-1 (Ler)*	3.38	573	*pEARLEYGate 100*

Table 2.1. Summary of intron-mediated silencing in T1 transgenic *Arabidopsis* plants over-expressing the largest intron of respective genes. Only the first and largest intron of *CAL* gene can mediate endogenous gene silencing. F* or R* in parenthesis indicates the orientation of the intron. Experiments about introns of *SHP1* and *SAP* were done by Boyana Grigorova.

Among the introns tested, only the first intron of *CAL*, when introduced into *ap1-1* mutant background, was able to induce a typical loss-of-function-like phenotype. Precisely, 33.24% T1 generation transgenic plants carrying *35S::CAL Intron1* in forward orientation (*35S::CALI1.F*) exhibited cauliflower-like inflorescence (*cal-like* phenotype), resembling the typical *cal-1;ap1-15* double mutant (Table 2.1; Figure 2.3). The remaining 66.76% T1 transgenic lines were indistinguishable from the parent *ap1-1* plants and were thus referred to as *ap1-like* (Figure 2.3). Strikingly, the *cal-like* transgenic plants (at 33.24%) can be subdivided into "*cal-like*" and "*weak-cal-like*" subgroup at 25.20% and 8.04% respectively. The *weak-cal-like* plants showed loosely compacted inflorescences with more differentiated floral buds than *cal-like* inflorescences (Figure 2.3).

In contrast, T1 plants transformed with *35S::CALIntron1* in reverse orientation (*35S::CALI1.R*) presented a much lower frequency of *cal-like* phenotype (at 2.11% *cal-like* and 1.27% *weak-cal-like*, totally 3.38%). This general trend of a more potent

silencing effect of *35S::CALI1.F* than that of the *35S::CALI1.R* was also observed when the *CAL* first intron was cloned into a different binary vector (*pMDC32*) (Table 2.2).

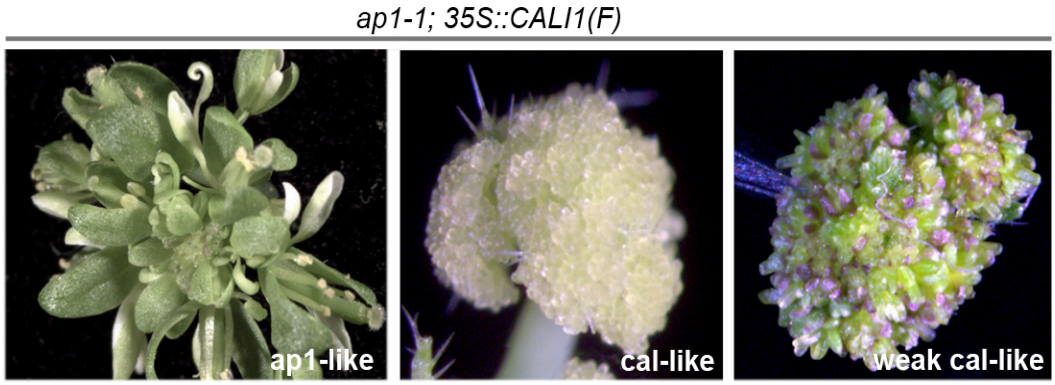


Figure 2.3. The phenotypes in T1 generation of *35S::CALI1* transgenic lines. From left to right, *ap1-like*, *cal-like* and *weak-cal-like* are displayed, respectively.

Intron-mediated silencing of *CAL* results from a reduction of *CAL* mRNA

To investigate the mechanism of *CAL* first intron-triggered silencing of endogenous *CAL*, we gauged *CAL* mRNA expression levels in inflorescences of *cal-like* plants and *cal;ap1* double mutant by northern blot (Figure 2.4) and qRT-PCR (Figure 2.4). *cal-1* is a missense mutation and hence, in *cal-1* mutants *CAL* mRNA level should not be affected. As *CAL* is predominantly expressed in floral meristem (Bowman et al., 1993; Kempin et al., 1995), mRNA from inflorescence tissues was examined. Since *cal;ap1* double mutants develop the equivalent density of young floral meristems as *cal-like* plants do, inflorescence of *cal;ap1* served as a more appropriate control than that of *wild type* or *ap1-1* plant. Northern blot, which is less quantitative due to over-exposure, showed about 3-fold decrease of *CAL* mRNA in *cal-like* plants when

compared with *cal-1; ap1-1* (Figure 2.4). Consistent to northern blot, qRT-PCR revealed about 5-fold reduction of *CAL* mRNA in *cal-like* plants (Figure 2.4). Our data suggested that transgenic *CAL* intron triggered endogenous gene silencing in either transcriptional or post-transcriptional level, leading to reduced *CAL* mRNA level.

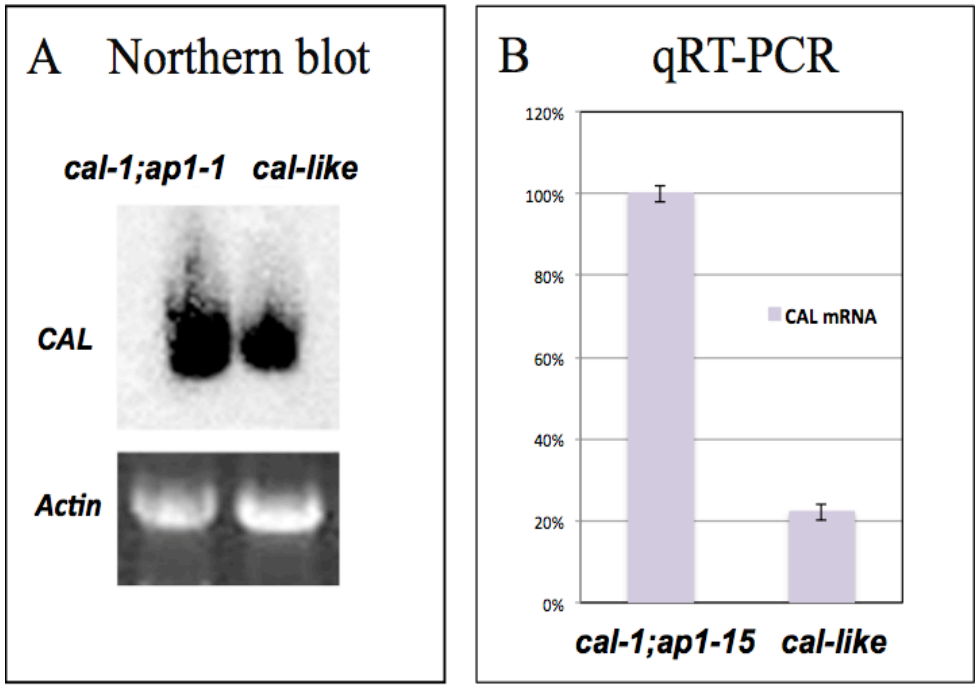


Figure 2.4. A reduction of *CAL* mRNA in *cal-like* inflorescences. (A) Northern blot showing *CAL* mRNA level (done by Boyana Grigorova). (B) qRT-PCR data showing *CAL* mRNA expression levels. Standard deviation was calculated from three biological replicates, each with three technical replicates.

Intron-mediated silencing depended on the production of intron-derived 24 nt siRNAs

Since eukaryotic gene silencing is predominantly conferred by small RNAs (sRNA) guidance, we tested whether silencing in *cal-like* plants could be directed by siRNAs derived from the *CAL* first intron. Northern blots of small RNAs were performed to

examine and compare siRNAs from inflorescences of *wild type Col, ap1-1*, transgenic *ap1-like,* and *cal-like* plants (Figure 2.5). The entire *CAL* first intron was radiolabeled and applied as a probe. An abundance of 24 nt siRNAs was detected in *cal-like* inflorescence tissues, but was absent from *wild type* and *ap1-1* inflorescences. A much lower amount of 24 nt siRNA was detected in *ap1-like* inflorescences, indicating that the 24 nt siRNAs had to reach a sufficient level to induce the significant silencing of endogenous *CAL*. Moreover, the quantity of 24 nt siRNAs was decreased in older *cal-like* inflorescence tissues, correlating with an alleviated phenotype in the older *cal-like* plants as these old inflorescences eventually differentiate into functional flowers. Additionally, no small RNAs were detected using the probes derived from *CAL* cDNA or *CAL* 5'-UTR (data not shown). The observation of 24 nt siRNA (instead of 21 nt RNA) strongly suggests that the intron-mediated silencing occurs at the transcriptional rather than post-transcriptional level. This is also consistent with the fact that intron-derived siRNAs will not be able to pair with mRNA.

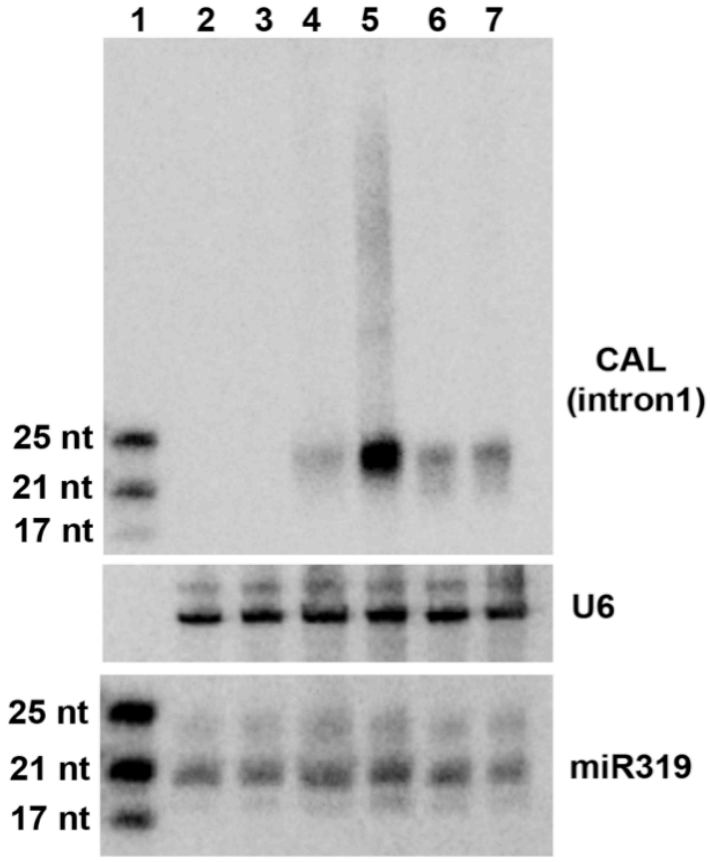


Figure 2.5. Small RNA northern blot showing the presence of 24 nt siRNAs. *cal-like* phenotype correlates with abundant intron-derived 24 nt small RNAs (done by Dr.Zongrang Liu's lab). Lane 1: RNA size ladder. Lanes 2-7 are inflorescence tissues of following genotypes. Lane 2: *wild type Col;* lane 3: *ap1-1*; lane 4: *ap1-like*, lane 5: *cal-like*; lane 6: old *ap1-like*; and lane 7: old *cal-like*.

While a positive correlation exists between an abundant intron-derived 24 nt siRNA and the corresponding silenced phenotype, we sought to determine if the 24 nt siRNAs are required for the silencing. RNA dependent RNA polymerase 2 (RDR2) is known to be required for the biogenesis and amplification of 24 nt siRNAs. The *rdr2-1* mutant was caused by a T-DNA insertion in the predicted exon 1 and thus a likely null allele (Xie et al., 2004). This *rdr2-1* allele was introgressed into the *Ler* background (obtained from Xuemei Chen), and we created the *rdr2-1; ap1-1* double

mutant line and then over-expressed *CAL* first intron in *rdr2-1; ap1-1* plants to test whether the blockade of 24 nt siRNA production can affect the silencing efficacy conferred by the *35S::CALI1*. The *35S::CALI1* construct in *pMDC32* vector caused silencing in *ap1-1* background, whereas it failed to induce silencing in the *rdr2-1; ap1-1* background (Table 2.2). The data indicated that the RDR2-dependent 24 nt siRNAs are required for the intron-mediated silencing.

	ap1-like	*cal-like*	Total
rdr2;ap1-1 35S::CALI1.R	147 (100%)	0 (0%)	147
rdr2;ap1-1 35S::CALI1.F	116 (100%)	0 (0%)	116
ap1-1 35S::CALI1.R	314 (99.68%)	1 (0.32%)	315
ap1-1 35S::CALI1.F	272 (90.37%)	29 (9.63%)	301

Table 2.2. RDR2 is required for the intron-mediated silencing. The *CAL* first intron was driven by *35S* promoter in the *pMDC*32 vector.

Abundant 24 nt siRNAs are derived from three specific regions of the *CAL* first intron

Although our northern blot can detect 24 nt siRNAs using the *CAL* first intron as a probe, it is not known if the 24 nt siRNAs originate from specific regions within the intron or throughout the entire intron. We thus isolated and sequenced small RNA populations from *cal-like* plants and *cal-1;ap1-15* control plants. Two biological replicates were conducted for each genotype. Two types of *cal-like* plants were utilized: *cal-like.*T1 plants were the first generation of transgenic line, while *cal-*

*like.*T4 plants were the fourth generation of transgenic line, in which the severity of silenced phenotype was moderate. Consistent with the previous northern blots, in *cal-like* plants, the majority of the siRNAs from the *CAL* first intron are 24 nt in length (Figure 2.6). Further, the 24 nt RNAs are mainly derived from three specific intronic clusters, at #83 to #106 (in antisense orientation), #105 to #128 (in sense orientation), and #605 to #628 (in antisense orientation), respectively (Figure 2.7). Conversely, 24 nt small RNAs were absent from *cal-1; ap1-15* plants.

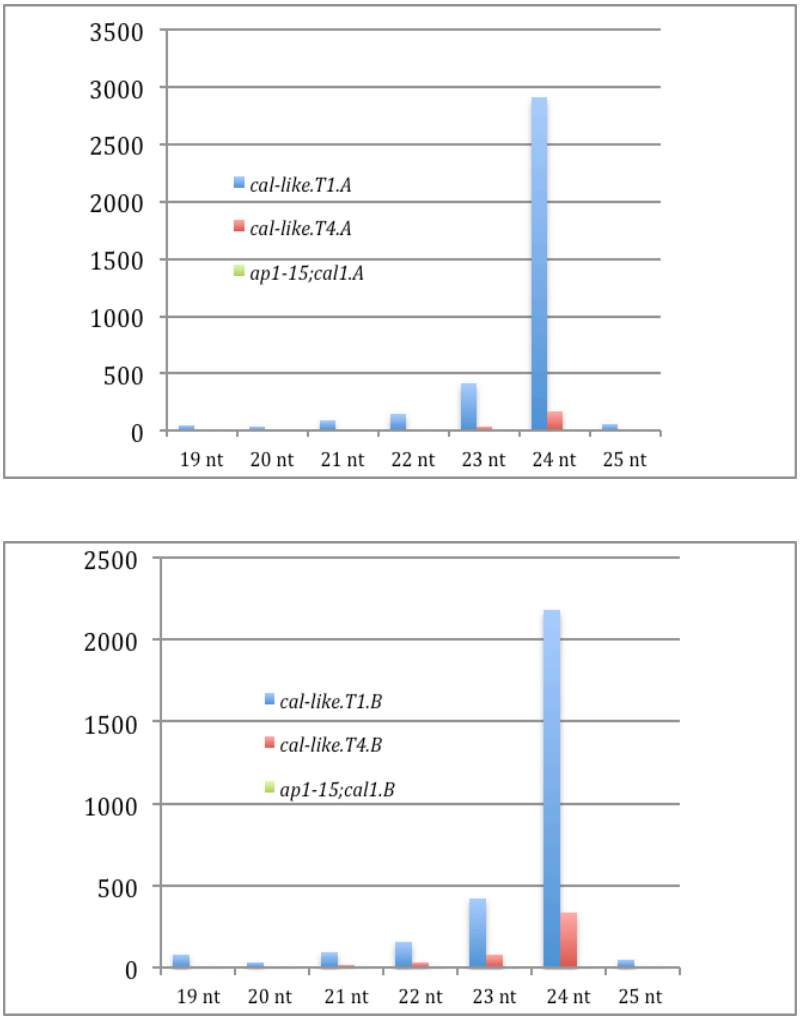


Figure 2.6. The normalized amount of different-sized siRNAs derived from *CAL* first intron in silenced transgenic lines and control plants. Y-axis indicates reads per 10 million (RPTM). Two biological replicates A and B were performed.

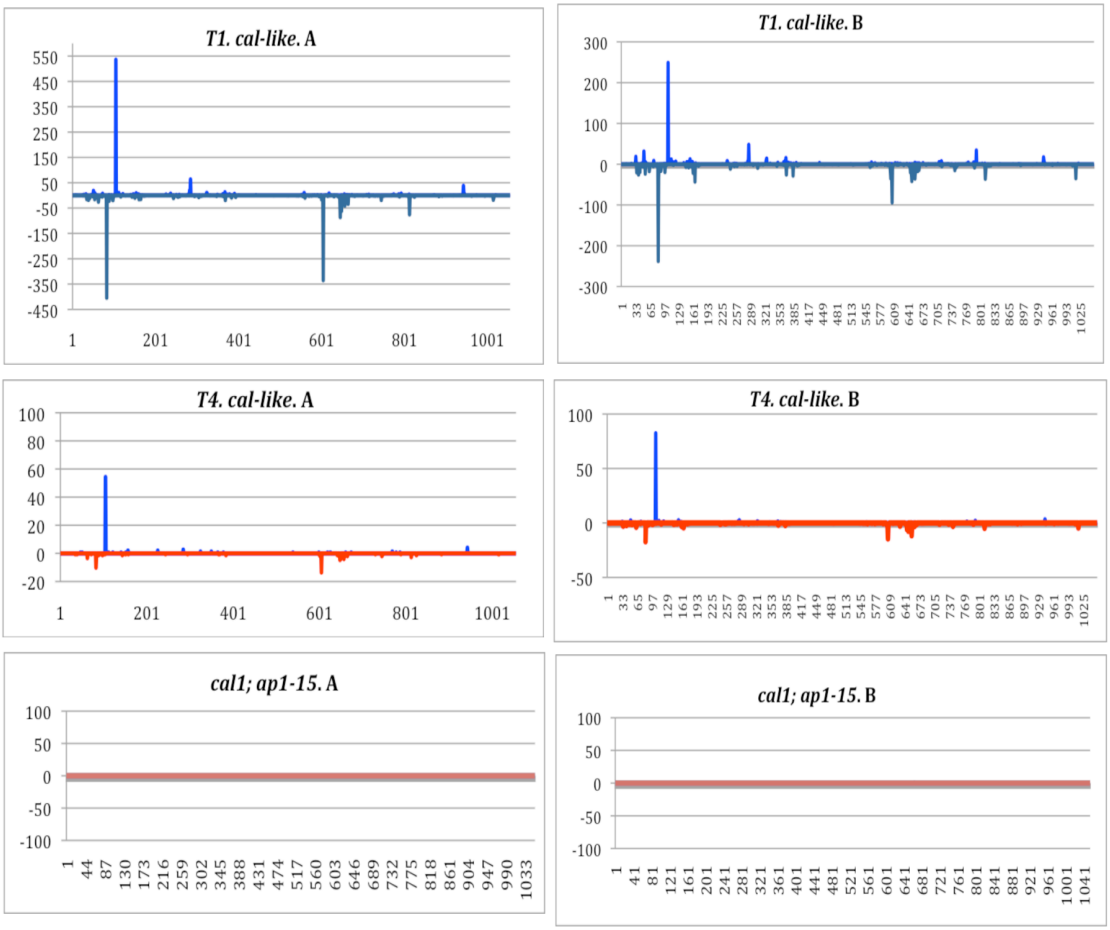


Figure 2.7. The profile of 24 nt siRNAs derived from specific regions of *CAL* first intron in silenced transgenic lines. Y-axis indicates reads per 10 million (RPTM). X-axis indicates nucleotide position starting from position 1 in the 5'-end of the intron. Two biological replicates A and B were performed.

Furthermore, *cal-like.*T4 plants exhibited a decreased level of 24 nt siRNAs in comparison with that of *cal-like.T1*, indicating that the 24 nt siRNAs contributing to endogenous *CAL* silencing are gradually diminished in subsequent generations. This is compatible with our observation that *cal-like.*T4 plants displayed a weak or intermediate *cauliflower* phenotype.

To determine if each of the three intronic regions, where the 24 nt siRNAs are derived, is essential for silencing, we over-expressed partially deleted *CAL* first introns in *ap1-1* mutants (Table 2.3). Our data demonstrated that removal of any one of the three intronic regions, from where 24 nt siRNA were produced, greatly diminished the proportion of plants with *cal-like* phenotype in T1 generation. Therefore, each intronic region contributes to the silencing ability of the intron.

Deleted regions	**Intron orientation**	**Number of T1 transgenic phenotype**		
		cal-like	***weak-cal-like***	***ap1-like***
Normal	Forward	163 (25.20%)	52 (8.04%)	432 (66.76%)
	Reverse	10 (2.11%)	6 (1.27%)	457 (96.62%)
#83-#104	Forward	19 (3.63%)	24 (4.60%)	479 (91.77%)
	Reverse	2 (0.32%)	2 (0.32%)	625 (99.36%)
#107-#128	Forward	31 (1.33%)	65 (2.80%)	2231 (95.87%)
	Reverse	10 (0.36%)	17 (0.61%)	2776 (99.03%)
#605-#628	Forward	33 (1.92%)	54 (3.15%)	1627 (94.93%)
	Reverse	2 (0.21%)	4 (0.42%)	948 (99.37%)

Table 2.3. Summary of silencing efficiency of *CAL* first intron with deleted regions. All introns were over-expressed from *35S* promoter in *pEarleyGate100*. Deleted region # is based on bp # in *CAL* first intron.

DNA methylation at the *CAL* first intron of silenced plants

The essential requirement of intron-derived 24 nt siRNAs in the silencing of *CAL* suggests that the *35S::CALI1* must have induced transcriptional gene silencing (TGS). Such TGS may be mediated by DNA methylation and chromatin compaction to repress the expression of *CAL* via the RNA directed DNA methylation (RdDM)

pathway. Thus, two independent experiments were conducted to examine methylation on the *CAL* first intron in *cal-like* plants.

Bisulfite converts unmethylated cytosines to uracils, but leaves methylated cytosines intact. By comparing the sequence of bisulfite-treated samples with reference, the methylated cytocines can be detected. Genomic DNA extracted from inflorescences of *ap1-1, ap1-like* and *cal-like* plants, exhibited noticeable DNA methylation at CG, CHG and CHH sites (H indicates either an A, T or C, but not G) in the *CAL* first intron of *cal-like* plants, but absent in the same regions of *ap1-1* and *ap1-like* plants (Figure 2.8). It implied that the over-expressed transgenic *CAL* first intron triggered RdDM at the endogenous *CAL* first intron locus, leading to switching off the *CAL* gene transcription.

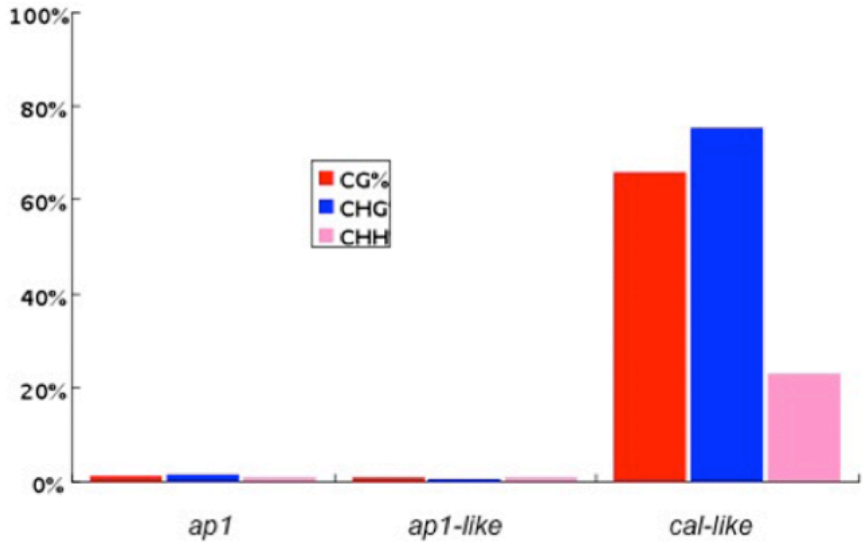


Figure 2.8. Bisulfite sequencing of *CAL* first intron indicated methylation in the silenced lines. X-axis shows different phenotype; Y-axis represents the percentage of methylated cytosines over total cytosines (done by Dr.Zongrang Liu's lab).

McrBC-PCR was the other method exploited to detect DNA methylation. McrBC enzyme specifically recognizes and cleaves methylated DNA (Gowher et al., 2000; Zhou et al., 2002). Hence, after McrBC digestion, methylated DNA template is cleaved and PCR product is reduced or absent. Comparing PCR products between McrBC-treated and untreated samples, *ap1-1* and *ap1-like* plants showed no obvious difference at the *CAL* locus and thus were unlikely methylated at the *CAL* (Figure 2.9). Oppositely, the two transgenic *cal-like* lines showed a reduction of PCR products specifically in the first PCR amplicon of *CAL* first intron (*Intron 1.a*; #35 to #278 bp at 5'-end of intron 1) (Figure 2.9), implying DNA methylation near the 5'-end of the intron in the silenced lines. However, this DNA methylation did not disperse along the gene, or into the promoter as indicated by unchanged PCR product amount in the promoter location or *Exon 2* through *Exon 3* region (Figure 2.9). Alternatively, McrBC digestion may not be sensitive enough to detect slight methylation spread.

The above McrBC-PCR data could not distinguish the methylation status of transgenic *CAL* intron from that of endogenous *CAL* intron. To examine if the endogenous *CAL* first intron was methylated, an *Exon 1* forward primer was paired with an *intron 1.a* reverse primer in the McrBC-PCR assay. Since *Exon 1* sequence is absent from the *35S::CALI1* transgene, *Exon1/Intron1.a* primer combination only can PCR-amplify the endogenous *CAL* gene region. This region showed decreased PCR products in *cal-like* plants, indicating that the endogenous *CAL* first intron (the *Intron1.a* region) was indeed methylated (Figure 2.9). Together, the McrBC-PCR

assay revealed that RdDM likely occurs at the first 278 bp at 5'-end of *CAL* first intron at the endogenous as well as transgenic *CAL* loci.

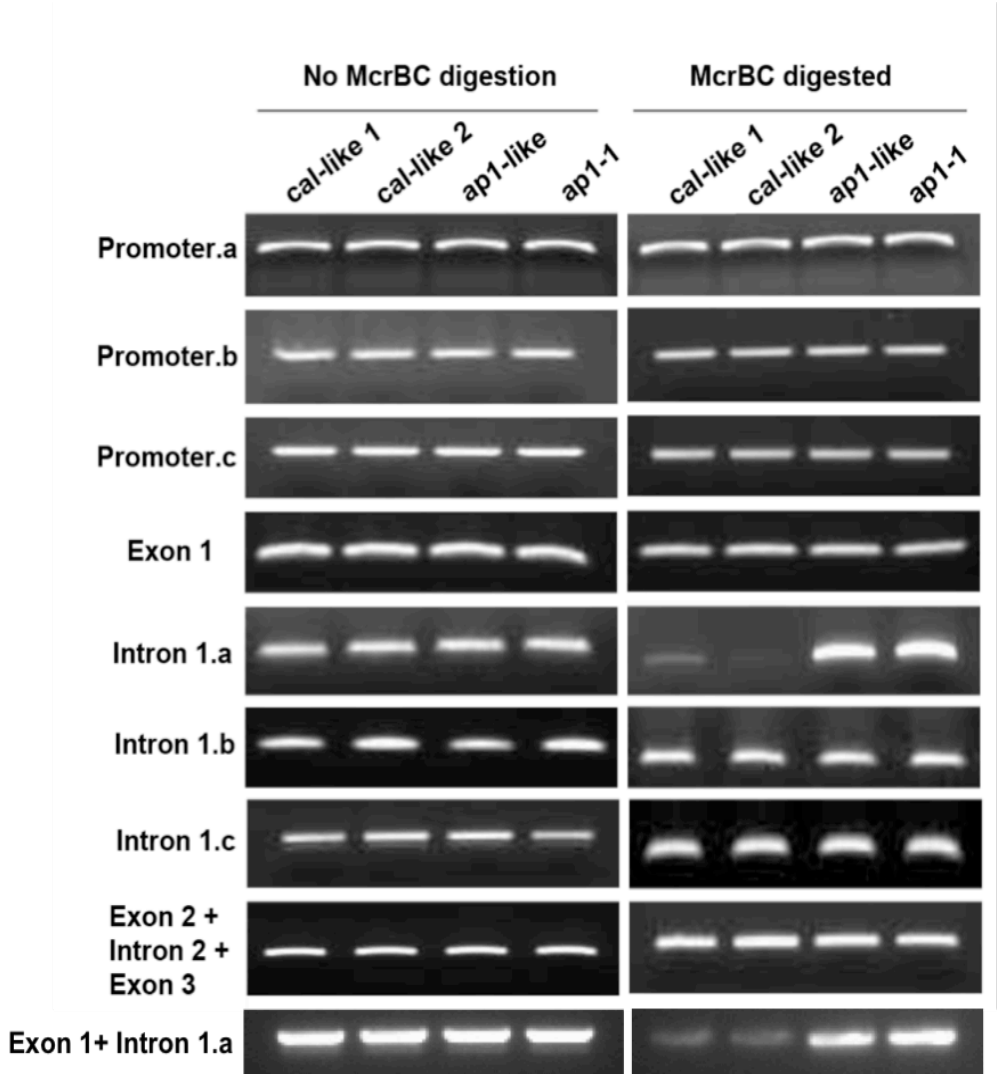


Figure 2.9. McrBC-PCR detected DNA methylation near the 5'-end of *CAL* first intron in the silenced lines. *cal-like* 1 and *cal-like* 2 are two different silenced T4 lines. The promoter and first intron (intron 1) are each sub-divided into three (a, b, c) PCR amplicons to ensure equal amplicon size. A reduction of PCR products in *CAL* intron 1.a was observed in these two silenced after McrBC-digestion. The last row examined methylation status of endogenous *CAL* first intron, confirming methylation in endogenous *Intron 1.a*.

Summary and discussion

In this chapter, I have shown that over-expression of *CAL* first intron was able to induce the silencing of *CAL* locus in *Arabidopsis*, and this intron-mediated gene silencing occurred on the transcriptional level. First, the over-expressed *CAL* first intron was able to elicit the production of 24 nt siRNAs mainly from three specific regions of *CAL* first intron. Second, the abundance of 24 nt siRNAs correlates with the silencing effect and is dependent on the amplification by RDR2. Third, DNA methylation was induced at *CAL* first intron in the silenced lines and may underlie the ability of the 24 nt siRNA to cause TGS. Taking together, the data led to a simple model that describes this epigenetic silencing process (Figure 2.10).

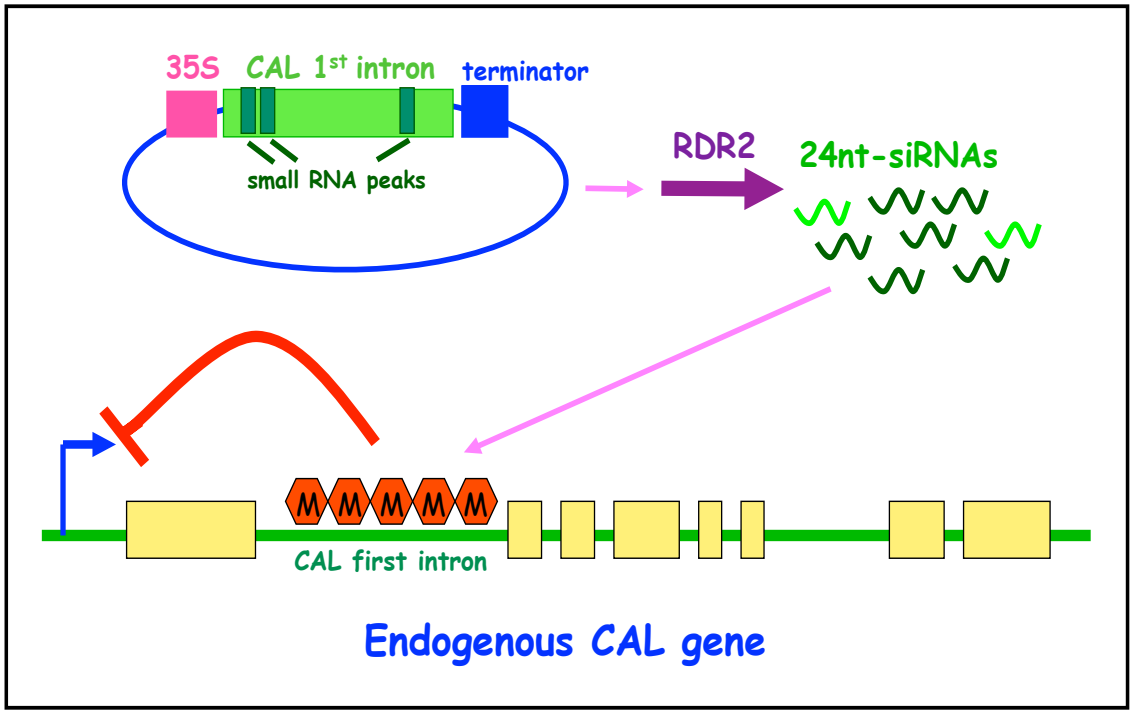


Figure 2.10. A model illustrating the phenomenon that over-expressed *CAL* first intron mediates the silencing of endogenous *CAL* gene.

ap1-like, *cal-like* and *weak-cal-like* phenotypes

After transformation, in T1 generation, three phenotypes were observed: *ap1-like*, *cal-like* and *weak-cal-like* phenotype. The varied phenotypes are likely due to the random loci, where the *35S::CALI1* transgene inserts in the genome. Some may be in highly active chromatin regions, resulting in high levels of *CAL* intron transcripts from the transgene and subsequent silencing of corresponding *CAL* gene. In this case, transgenic plants showed characteristic *cal-like* phenotype. Some transgenes were inserted into heterochromatin region with low transcription activity; accordingly, the amount of transgene transcript was low and not sufficient to elicit silencing of *CAL* (but enough to confer drug resistance for transgenic selection). In this case, plants displayed *ap1-like* phenotype. If the transcription of transgene acts intermediately, the plants may exhibit a *weak-cal-like* phenotype, as the endogenous *CAL* gene is not completely silenced. In some cases, the transgene may insert inside a gene, interrupting the function of the gene; however the probability of such occurrence is very low and the likelihood of such a high percentage of *cal-like* plants caused by the insertion into the same gene (*CAL*) is nonexistent.

Three peaks of 24 nt small RNAs originate from the *CAL* first intron

Another striking aspect is that only three peaks of 24 nt small RNAs were identified from *CAL* first intron. This is in contrast to what was expected that the 24 nt small RNAs could be generated from everywhere along the entire intron. However, why only distinct peaks were found remains elusive.

In a series of deletion-over-expression experiments, I showed that the removal of any of the intronic regions, where the 24 nt small RNAs peaks reside, abolished the ability of the *CAL* first intron to induce gene silencing. This indicates that each of the intronic regions is required for transgenic *CAL* first intron to achieve silencing. One explanation is that these three subintronic regions may possess certain specific features in nucleotide sequences or local secondary structures and hence they may favor or provide affinity for DCL3, the double-stranded RNA dicer. Alternatively, these 24nt siRNAs are particularly stable due to their special composition.

The silencing effect gradually diminishes with more generations

In my work with *cal-like* plants, the phenotype becomes weaker and weaker as the plants pass down in generations. This is also evident from our siRNA sequencing data (Figure 2.7) where the T4 (the fourth generation) *cal-like* plants showed the same three peaks of the 24 nt siRNAs but their reads are much lower than the T1 (the first generation) *cal-like* plants. Along the evolution, plants might have developed many strategies to protect integrity of their own genomes from external invasions. Therefore, after several generations, transgenes inserted into plant chromosome could be repressed, as well as the epigenetic marks on histone tails also can be erased. Consequently, the initial silence phenomenon becomes weaker. This weakening effect of gene silencing with increased generations was often observed in plants and reported in *C. elegans.* In *C. elegans*, H3K9me3, the repressing histone codon, once triggered by dsRNA, can only be maintained for three generations (Gu et al., 2011; Gu et al., 2012).

Orientation-preference of *CAL* first intron in silencing

I showed that over-expressing *CAL* first intron in sense (forward) orientation caused 10 folds higher percentage of *cal-like* plants than over-expression of *CAL* first intron in reverse (antisense) orientation. This orientation-preference for *CAL* first intron maybe explained by how the silencing initiates. As we know, production of siRNAs requires dsRNAs as precursor. Thus, the transcripts of over-expressed *CAL* first intron require to be converted into RNA duplex in perfect complement, either partially or entirely. According to the secondary structure of *CAL* first intron analyzed by mFOLD program, although the RNA of *CAL* first intron can fold back on itself to form relatively stable RNA duplexes in local areas, the length of the duplexes is less than 13 base pairs in sense-orientation, and 7 base pairs in anti-sense orientation (Figure 3.1). Apparently, 24 nt siRNAs are unlikely to originate from these regional dsRNAs. Recently, considerable intron-derived long non-coding RNAs (lncRNA) with various genetic functions have been revealed in both mammals and plants. Hence, we hypothesize that antisense lncRNA(s) may be transcribed from *CAL* first intron in endogenous *CAL* locus. Pairing between the sense-orientated transgenic intron transcript and the hypothesized antisense lncRNA from endogenous intron may result in dsRNA and trigger the 24 nt siRNA production.

Previous researches discovered that non-coding RNA transcripts could arise from the sequences between distant enhancers and promoters (Dobi and Winston, 2007; Ho et.al, 2006; Masternak et al., 2003; Rogan et al., 2004; Tchurikov et al., 2009; Zhu et al., 2007; Singer et al., 2010). Specifically, the regulatory enhancer elements possibly

associate transcription factors and RNA Pol II in *cis*. Then this transitive complex tracks along the sequence between the enhancer and the promoter, until encountering the promoter to activate transcription (Singer et al., 2011). Along the migrating path, the complex may synthesize by-product transcripts in the form of lncRNA (Figure 2.11A). Hence, promoter-upstream-located enhancers bring RNA Pol II and associated factors to slide along the intervening DNA sequence toward the cognate promoters, resulting in a non-coding RNA transcript in sense orientation (Figure 2.11A). Oppositely, in the case of *CAL* first intron, the hypothesized enhancer elements residing in the first intron may recruit RNA Pol II and other associated transcription factors to initiate transcription of lncRNA in reverse orientation from the intron region (Figure 2.11B). Therefore, in the next chapter, I examined this hypothesis by detecting whether the endogenous *CAL* first intron-derived antisense RNA is present.

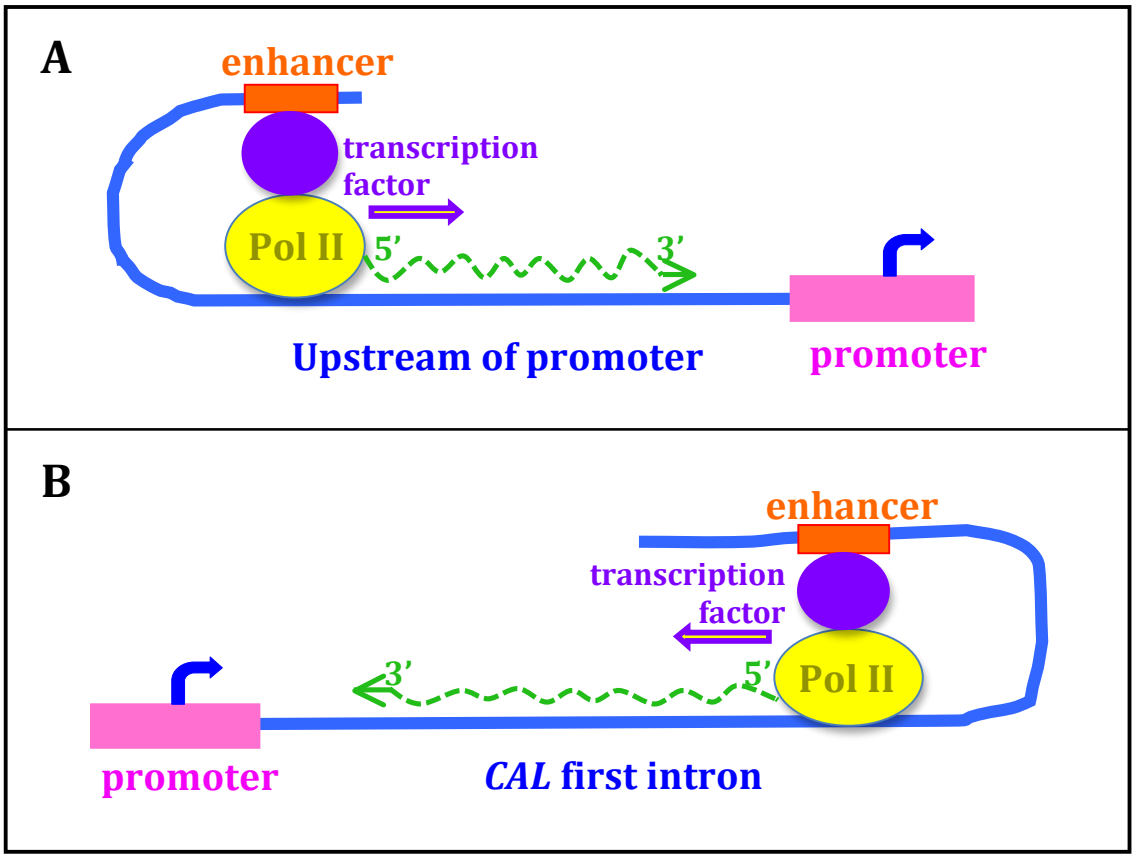


Figure 2.11. The model of enhancer-promoter interaction that initiates transcription of lncRNAs. During this facilitated tracking process, RNA pol II may generate a by-product transcript by using the intervening sequence between enhancer and promoter as a template. The orientation of the non-coding RNA transcript depends on whether respective enhancer is positioned upstream or downstream of the promoter.

Different vectors exhibit different efficiency in silencing

To investigate this transgenic intron-mediated silencing phenomenon, I utilized binary vectors *pEarleyGate100* and *pMDC32* to over-express *CAL* first intron in sense- and antisense- orientation in plants. Both *pEarleyGate100* and *pMDC32* express *CAL* intron via the *35S* promoter. However their transcription terminators and antibiotic selections in plants differ. *pEarleyGate100* employs *OCS* terminator and

confers Basta-resistance, while *pMDC32* uses *NOS* terminator and confers hygromycin-resistance. When both vectors carrying *CAL* first intron in sense-orientation were transformed into *ap1-1* plants, the T1 plants with *pEarleyGate100* resulted in 33.24% *cal-like* plants, whereas *pMDC32* led to only 9.63% *cal-like* individuals. In antisense orientation, T1 plants with *pEarleyGate100* showed 3.38% *cal-like* individuals, but *pMDC32* gave only 0.32% *cal-like* plants. Despite that different vectors induced silencing at different rates, the trend of silencing is similar. Both vectors are able to induce *cal-like* plants. Furthermore, both vectors caused higher number of *cal-like* plants in sense orientation than that in antisense orientation.

The reason that *pEarleyGate100* and *pMDC32* exhibit different silencing efficiency may be due to different drug selection. For *pEarleyGate100*, Basta was sprayed onto T1 seedlings, whereas for *pMDC32*, seeds were planted on hygromycin-containing 1/2 MS/agar. It is likely that plants with low transgenes expression can still survive during hygromycin selection while the same weak level of transgene expression cannot confer plants survival under BASTA spray. Thus, more transgenic plants that survive after hygromycin selection cannot exhibit silenced phenotype.

Discrepancy between McrBC-PCR and Bisulfite sequencing data

McrBC recognizes methylated cytosines preceded by a purine nucleotide (Adenine or Guanine), and cleaves the DNA sequence on one or both strands. The half-sites (G/A ^{met}C) can be separated by up to 3 kb, but the optimal distance is between 55 and 103 base pair (Gowher et al., 2000; Zhou et al., 2002). Most researchers employed McrBC to examine the methylation status of GC-rich region in animals or humans (Panne et

al., 1999). For genomic DNA of plants, the methylation pattern is more complicated, in which all three contexts CG, CHG and CHH (H represents A, T, and C) could be methylated. Especially, CHH contexts are prevalent in *CAL* first intron of *Arabidopsis thaliana* (Figure 2.12). Perhaps, McrBC is not sensitive to the methylation in CHG and CHH contexts. Therefore, our McrBC-PCR assay only detected methylation in the Intron 1.a fragment that contains three (G/A)CG sites separated by reasonable length.

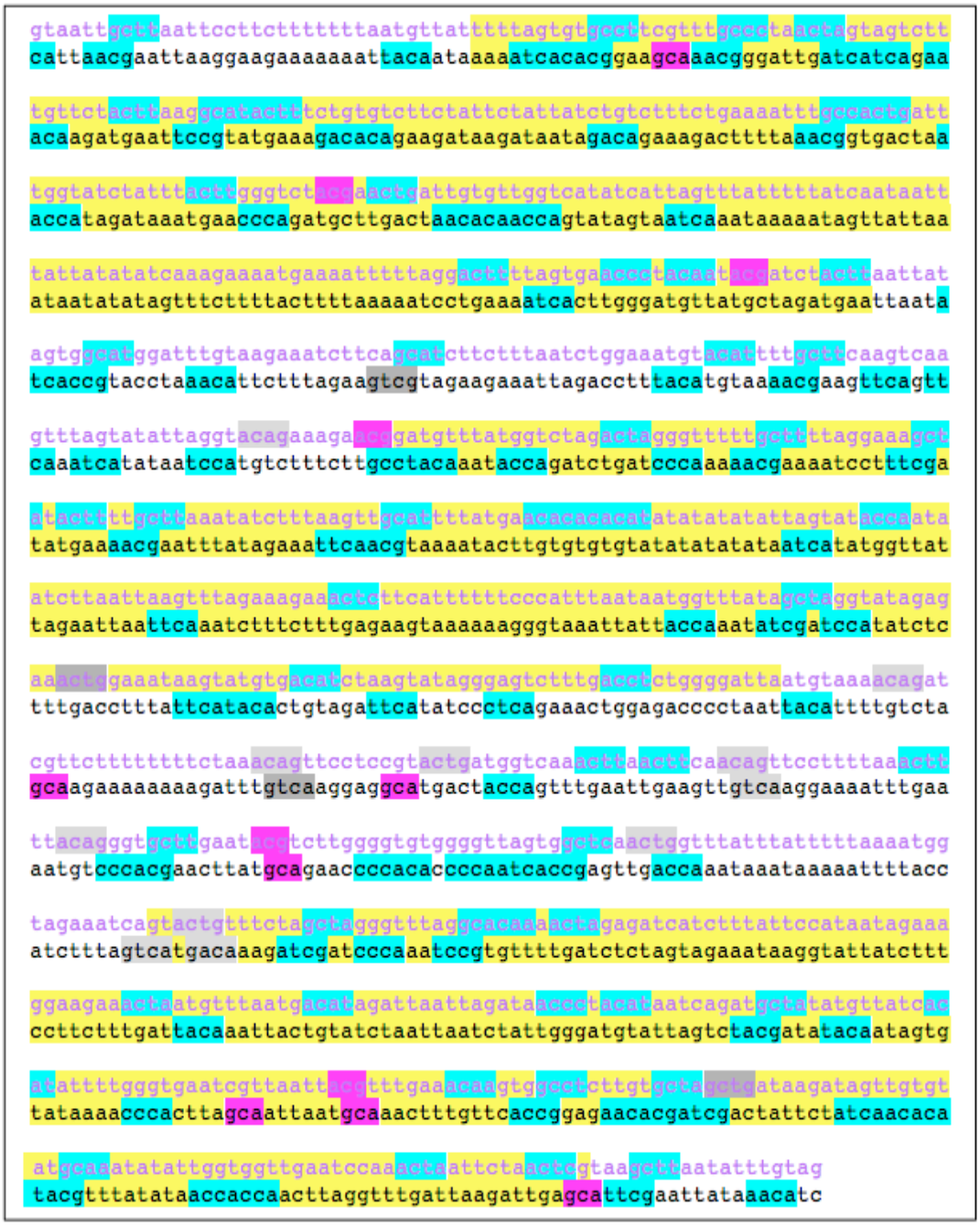


Figure 2.12. The profile of the cleavage sites of McrBC in both strands of *CAL* first intron. Blue shadow indicates G/A CHH, grey shadow indicates G/A CHH contexts, and pink shadow indicates CG contexts. Yellow shadow covers the regions amplified by McrBC-PCR, which correlate to Intron 1. a, b and c, respectively.

Material and Methods

Plant materials, growth conditions, transformation and selection of transgenic line

Seeds of *Arabidopsis thaliana* were stored at 4 ^{o}C in the dark for 3 days and sown directly onto Metromix soil (Griffin Greenhouse Supplies) under a 16-hour-light/8-hour-dark cycle at 22 ^{o}C in growth chambers. *ap1-1* single and *cal-1;ap1-1* double mutants as well as *rdr2-1;ap1-1* and the transgenic *35S:: CALI1; ap1-1* plants are in the *Landsberg erecta* (*Ler*) background, while *cal-1;ap1-15* is in the *Columbia* (*Col*) background. *cal-1;ap1-1* seeds were obtained from ABRC stock center (CS6161). *rdr2-1* (introgressed into *Ler*) (Xie et al., 2004) and *cal-1;ap1-15* seeds were bestowed by Dr. Xuemei Chen and Dr. Martin Yanofsky, respectively. Plant transformation was conducted via floral-dipping method. Shoots of young flowering plants were dipped into a solution of *Agrobacterium tumefaciens GV3101* carrying specific constructs for 5 minutes. After dipping, the T0 plants are returned to normal growth conditions until the seeds were collected (Mara et al., 2010). For constructs using *pMDC* binary vectors, the seeds of T0 plants were selected on 1/2 MS medium containing 50ug/ml of hygromycin. For constructs using the *pEarleyGate100* binary vector, the seeds of T0 plants were sown on soil and the T1 seedlings were sprayed by Basta for five times with one-day interval.

Construction of *35S::CALI1* as well as other intron-over-expression constructs

For the construction of *35S::CALI1, CAL* first intron was amplified by Phusion Taq (New England Biolabs) using the *CAL* first intron specific primers CALI1.F and CALI1.R. The amplified *CAL* first intron was cloned into *pCR8/GW/TOPO* using TA cloning kit (Invitrogen). Then the *CAL* first intron in either sense or antisense orientation was respectively introduced into *pEarleyGate100* and *pMDC32* binary vectors (Earley et al., 2006; Curtis and Grossniklaus, 2003) through the Gateway® technology (Invitrogen). The sense and antisense *35S::CALI1* in *pEarleyGate100* constructs were transformed into *Agrobacterium tumefaciens GV3101*, which was used to infect *ap1-1* plants. *35S::CALI1* in *pMDC32* was introduced into *ap1-1* as well as *ap1-1; rdr2-1* plants. For the *pEarleyGate100-CALI1.F* constructs, 647 *35S::CALI1.F; ap1-1* lines were generated. Detailed analysis was conducted on 11 *35S::CALI1.F;ap1-1* transgenic plants with *ap1-like* phenotype and 13 *35S::CALI1.F; ap1-1* transgenic lines with *cal-like* phenotype. The progeny of these lines were used for further analyses. For the *pMDC32-CALI1.F* constructs, 301 *35S::CALI1.F;ap1-1* transgenic lines were produced. Further detailed analysis was performed for the lines of 15 *ap1-like* plants and 20 *cal-like* plants. In addition, 116 *35S::CALI1.F* in *ap1-1; rdr2-1* background lines were obtained and scored for individual phenotype.

The largest introns of *AP1*, *SAP*, and *SHP1* were PCR amplified using cognate specific primers, which are AP1I.F and AP1I.R, SAPI.F and SAPI.R, SAPI.F and SAPI.R, respectively. Then these introns were subsequently cloned in sense as well as antisense orientations into *pEarleyGate100* vector using the same cloning method

described for *35S::CALI1*. Transformation and phenotype scoring were similarly conducted as described for *35S::CALI1*.

Genotyping to confirm transgenes

Genotyping of *35S::CALI1.F (pEarleygate100*) transgenic lines was performed on genomic DNA isolated with the DNeasy Plant Mini Kit (Qiagen, MD, USA). The primers pEG100.35S. Geno.F and CALI1.R, respectively, specific to the *35S* promoter of *pEearleyGate100* and to the 3'-end of the *CAL* intron, were used to identify the presence of the transgene. The Phire PCR kit (NEB) was utilized to genotype the *35S::CALI1.F* (*pMDC32*) transgenic lines according to manufacturer's instruction, with the primers pMDC32.35S.Geno.F specific to the promoter region of *pMDC32* and CALI1.R1 specific to the part of *CAL* first intron.

rdr2-1 is a T-DNA insertion SAIL_1277H08 line (Xie et al., 2004), which was introgressed into *Ler* and donated by Dr. Xuemei Chen. To genotype *rdr2-1*, we used the primers LP, RP and LB3 to determine homozygous lines. The *wild type RDR2* locus can be amplified by *RDR2* primers, RP and LP, yielding a PCR product in around 0.55 kb. The mutant *rdr2-1* allele can be detected using LB3 and RP that can generate an approximate 0.4 kb PCR product. *rdr2-1* was crossed into *ap1-1* to generate *rdr2-1; ap1-1* double mutants. Genotyping was used to confirm the double mutants.

RNA northern blot

For small RNA isolation, inflorescences of *wild type Col*, *ap1-1* single mutant, *35S::CALI.F; ap1-1 (ap1-like) and 35S::CALI1.F; ap1-1 (cal-like)* transgenic plants (both young and old) were subject to RNA extraction with TRI® Reagent (Sigma). A 15% acrylamide gel was employed to separate 30 ug of total RNA for each sample. Then RNA was transferred to Hybond-N nylon membrane (Amersham) by electroblotting, and cross-linked onto this membrane using EDC (N-(3-Dimethylamionopropyl)-N'-ethylcarbodiimide hydrochloride) (Sigma), according to a previously published protocol (Pall and Hamilton, 2008). *CAL* first intron DNA, excised from the *pCR8/GW/TOPO-CALI1* plasmid (described above) and purified from the gel, served as the probe labeled with alpha P^{32}-dCTP (Perkin-Elmer) with Ready-To-Go DNA Labeling Beads by following the manufacturer's instruction (Amersham). The small RNA blot was pre-hybridized with ULTRAhyb Ultrasensitive Hybridization Buffer (Ambion) for two hours, followed by hybridization with the probe at 42 ^{o}C overnight. Blots were washed twice at 50 ^{o}C with 2X SSC -2% SDS for 50 minutes, followed by two washes of 1XSSC-1% SDS for 50 minutes each.

For mRNA northern blot, the *CAL* cDNA fragment was excised using the *NotI* and *AscI* enzymes from a *CAL* cDNA clone (ABRC clone stock # CD3-736) and radiolabeled as described for small RNA blot. Total RNA was extracted from inflorescence of *cal-1;ap1-1* and *35S::CALI1;ap1-1 (cal-like)* plants using TRI® Reagent (Sigma). Then a 1% agarose gel was utilized to separate 15 ug of total RNA

for each sample. Subsequently, RNA was transferred onto a BrightStar-Plus membrane, hybridized and washed by using the Northern Max-Gly kit (Ambion).

qRT-PCR

Total RNA was isolated from inflorescences of *ap1-1; cal-15* and *35S::CALI1;ap1-1 (cal-like)* plants using RNeasy® Plant mini kit (Qiagen). First-strand cDNA was synthesized from 1 μg total RNA using QuantiSure™ First-strand cDNA kit (Accugen Biosciences). 1μl of 10x diluted cDNA was used as a template in real-time PCR analysis. iQ™ SYBR® Green Supermix (Bio-Rad Laboratories) was used to set up real-time PCR reactions, which were run and analyzed on CFX96™ Real-Time System (Bio-Rad Laboratories). Conditions for real-time PCR were as follows: 95°C for 3 min, followed by 40 cycles of 94°C for 15 s, 60°C for 20 s, 72°C for 30 s. Melting curve analysis was performed from 65°C to 95°C with increments of 0.5°C every 5 seconds. CALI1.qRT-PCR.F and CALI1.qRT-PCR.R, the primers specific to *CAL* cDNA were used, with efficiencies at 96%. The housekeeping gene *GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE C SUBUNIT 1* (*GAPC1*, *At3g04120*) was used as a reference. The specific primers for GAPC1 cDNA are GAPC1.qRT-PCR.F and GAPC1.qRT-PCR.R. The Pfaffl formula $2^{-\Delta\Delta Ct}$ was used to calculate relative gene expression differences. $\Delta C_t^{cal\text{-}like} = C_t^{cal\text{-}like} - C_t^{GAPC1}$. Similarly, $\Delta C_t^{cal\text{-}1;ap1\text{-}15} = C_t^{cal\text{-}1;ap1\text{-}15} - C_t^{GAPC1}$. $\Delta\Delta C_t$ was calculated as $\Delta C_t^{cal\text{-}1;ap1\text{-}15} - \Delta C_t^{cal\text{-}like}$. The ΔΔCT number is entered into the Pfaffl formula ($2^{-\Delta\Delta\ Ct}$) to yield "fold difference" between *cal-1;ap1-15* control and *cal-like*. Error bars represent standard deviation of three technical replicates each for 2 biological replicates.

Bisulfite sequencing

Genomic DNA was extracted from *ap1-1, 35S::CALI1;ap1-1(ap1-like)* and *35S::CALI1;ap1-1(cal-like)* inflorescence tissues by Qiagen DNeasy plant mini Kit (Qiagen). For the bisulfite CT conversion, 300ng of floral genomic DNA was treated by EZ DNA Methylation-Direct™ Kit (Zymo Research). JumpStart REDTaq DNA polymerase (Sigma) was used to amplify bisulfite-treated DNA. The primers to amplify *CAL* first intron region of bisulfite-treated DNA were MCalU1, MCalL1, MCalU2 and MCalL2. Then PCR products were cloned using the pGEM-T easy Kit (Promega). 15 colonies for each DNA type were sequenced. Methylation analysis was conducted using online analysis tool CYMATE (Hetzl et al., 2007).

McrBC assay for methylation status

Genomic DNA was isolated from inflorescences of *ap1-1, cal-1;ap1-15, cal-like* and *ap1-like* with the Qiagen DNeasy plant mini Kit (Qiagen). For McrBC treatment, 500 ng of genomic DNA was digested with 20 units of McrBC enzyme (New England Biolabs) in a 50 ul volume for 1 hour based on the manufacturer's instruction. 2 ul of McrBC-digested DNAs served as template for PCR. Equal quantities of non-digested genomic DNAs were used as template and served as control. The primers CALProm. F1 and CALProm. R1, CALProm. F2 and CALProm. R2, CALProm. F3 and CALProm. R3 were used to amplify CAL promoter region a, b and c, respectively. The primers CALE1.F and CALE1.R, CALE2+I2+E3.F and CALE2+I2+E3.R were used to amplify *CAL* first exon and the region covering *CAL* second exon, second intron and third exon, respectively. The primers CALI1.F1 and CALI1.R1, CALI1.F2

and CALI1.R2, CALI1.F3 and CALI1.R3 were used to amplify *CAL* first intron region a, b and c, respectively.

Deletion analysis of *CAL* first intron

To make a series of deletions of *CAL* first intron (*CALI1),* overlapping PCR was performed. Specifically, to delete *#83-#104* region of *CAL* first intron, an intron fragment from *#1* to *#82* was amplified by Phusion Taq polymerase with a primer pair of CALI1.F' and CALI1.82R. A second DNA fragment containing the deletion was amplified by a forward primer CALI1.(69-82)+(105-119).F with the sequences of *#69-#82* and *#105-#119*, and a reverse primer CALI1.R' from the 3'-end of *CAL* first intron. Thus, the 5'-end of this second PCR product overlaps with the 3'-end of the first PCR fragment. After gel extraction (Gel extraction kit, Qiagen), these two purified PCR products were mixed at 1:1 molar ratio and used as template for the third PCR with primers CALI1.F' and CALI1.R' that amplify the entire *CAL* first intron. The PCR program consists of a 2-step thermocycle at 98˚C for 30 s, followed by 5 cycles of 98˚C for 10 s, 50˚C for 30 s, 72˚C for 30 s, then 26 cycles of 98˚C for 10 s, 60˚C for 30 s, 72˚C for 30 s, and finally 72˚C for 5 min. After gel purification (Qiagen), the amplified *CAL* first intron with deletion was cloned into *pCR8/GW/TOPO* via TA cloning (Invitrogen) in either sense or antisense orientation, and subsequently into *pEarleyGate100* vector (Earley et al., 2006) through the Gateway® technology (Invitrogen).

To generate *35S::#107-#128-deleted CALI1* and *35S:: #605-#628-deleted CALI1* constructs, similar methods described above were utilized. Specifically, for *#107-#128* deletion construct, the primers amplifying the first fragment were CALI1.F' and CALI1.128R. The forward primer for the second PCR fragment, CALI1.(94-106)+(129-138).F, contains the intron sequence of *94#-106#* directly followed by *#129-#138*, and the same reverse primer CALI1.R' was used. For *#605-#628* deletion construct, the first fragment was amplified by the same forward primer CALI1.F' and a reverse primer CALI1.604R. The forward primer for the second DNA fragment, CALI.(593-604)+(629-641).F, contains sequence of *#593-#604* followed by *#629-#641* and the same reverse primer CALI1.R' was used. Transformation of these constructs into *ap1-1* plants, and selection as well as scoring of T1 plants are as the same as described for *35S::CALI1 in pEarleyGate100*.

Illumina sequencing of Small RNAs

Total RNAs were extracted from inflorescences of *cal-1;ap1-15* double mutant, *cal-like* T1 and T4 plants using microRNA extraction kit (Qiagen). Small RNA libraries were constructed by TrueSeq small RNA library prep and sequenced by Illumina HiSeq2000. Small RNA sequencing data was aligned to the sequence of *CAL* first intron and analyzed by Bowtie program.

Primer list

Primer name	Primer sequence
CALI1.F	5' TCC TCT GAA TCT TGG TAA TTG 3'
CALI1.R	5' TAG TAC CTT CTC CAT GCT AC 3'
AP1I.F	5' GTA ACT TCA ACT AAT TCT TTA CTT T 3'

AP1I.R	5’ CTG ATC ATT TCA CAG AAT TAG GG 3’
SAPI.F	5’ TCG GGT CAG TTC TAA CGT TC 3’
SAPI.R	5’ ACA GCG TGC CAC GTG GTG 3’
SHP1I.F	5’ ACA GGT ACG CTT CTC CTA C 3’
SHP1I.R	5’ CAC ACT AAT AAG TAA GAT CGC 3’
CALI1.F’	5’ GTA ATT GCT TAA TTC CTT CTT TT 3’
CALI1.R’	5’ CTA GTA CCT TCT CCA TGC TAC AA 3’
pEG100.35S.Geno.F	5’ GGAAGTTCATTTCATTTGG 3’
pMDC32.prom.Geno.F	5’ GAG AGG ACC TCG ACT CTA GAG GA 3’
LP	5’ ATG GTG TCA GAG ACG ACG ACG AAC CGA TCA AC 3’
RP	5’ ACA CAT TAG GAC TAA CAA ATT TAC C 3’
LB3	5’ TAG CAT CTG AAT TTC ATA ACC AAT CTC GAT ACA C 3’
CALI1.qRT-PCR.F	5’ CTC TGT TCT TTG TGA TGC CGA GGT T 3’
CALI1.qRT-PCR.R	5’ GCC TGC TAT ACT CCA TTG ACC AGT TC 3’
GAPC1.qRT-PCR.F	5’ CCA GTC ACT GTT TTC GGC ATC A 3’
GAPC1.qRT-PCR.F	5’AGC TGC AGC CTT GTC TTT GTC A 3’
MCalU1	5’ CCA AAA TTT CCC TTA TTR TCT TCT CCC AT 3’
MCalL1	5’ TTA TTA TTA AAT GGG AAA AAA TGA AGA GT 3’
MCalU2	5’ AAA ACT CTT CAT TTT TTC CCA TTT AAT A 3’
MCalL2	5’ GTG AGA GTT AGG TGY AAT TAG YTG T 3’
CALProm.F1	5’ CTT ACG TCG AGA TGT GTC TGC T 3’
CALProm.R1	5’ TGT GTC ATC CAA TCC AAT CGA G 3’
CALProm.F2	5’ TGG TCT GCT TAA AGT GAT CAT GGT GT 3’
CALProm.R2	5’ TGT TGA GTC AAT AGA TTT CAC GTT GGA 3’
CALProm.F3	5’ ACT GTT CTT ACC GCC GAG CAA T 3’
CALProm.R3	5’ GAC GTC TAG TGA CGT GAA AAT GGG T 3’
CALE1.F	5’ ATG GGA AGG GGT AGG GTT GA 3’
CALE1.R	5’ CAA GAT TCA GAG GAG TAC TCG 3’;
CALI1.F1	5’ TTT TAG TGT GCC TTC GTT TGC C 3’
CALI1.R1	5’ AAG TAG ATC GTA TTG TAG GGT TCA C 3’
CALI1.F2	5’ CGG ATG TTT ATG GTC TAG ACT AGG GT 3’
CALI1.R2	5’ TAA TCC CCA GAG GTC AAA GAC TCC CTA 3’
CALI1.F3	5’ GTA CTG TTT CTA GCT AGG GTT TAG GC 3’
CALI1.R3	5’ CGA GTT AGA ATT AGT TTG GAT TCA ACC 3’
CALE2+I2+E3.F	5’ CAT GGA GAA GGT ACT AGA ACG 3’
CALE2+I2+E3.R	5’ CTT TGG TTT CTC TCC AAA AGC 3’
CALI1.82R	5’ TAA GTA GAA CAA AGA CTA CTA GTT AG 3’
CALI1.(69-82)+(105-119).F	5’ CTT TGT TCT ACT TAA TTC TAT TAT CTG 3’
CALI1.128R	5’ ATA GAA GAC ACA GAA AGT ATG CC 3’
CALI1.(94-106)+(129-138).F	5’ CTG TGT CTT CTA TAT TTG CCA CTG 3’

CALI1.604R	5’ CCC TAT ACT TAG ATG TCA CAT ACT T 3’
CALI1.(593-604)+(629-641).F	5’ CTA AGT ATA GGG GTA AAA CAG ATC G 3’

Chapter 3: What are the special characteristics of *CAL* first intron that endow it the ability to induce gene silencing?

Introduction

The secondary structure of RNA is determined by specific base pairing interactions within the primary sequences. The precise formation of RNA secondary structure is required for its maturation, regulation and function. For example, ribosomal RNA (rRNA) must form intricate structure to interact with correct ribosomal subunit, allowing the proper formation of functional ribosome (Trappl and Polacek, 2011). Additionally a stem-loop structure with a stretch of poly-uracil formed by the RNA terminator sequence plays an important role in termination of transcription, (Mauger et al., 2013). Recent researches reported that instead of their primary sequence, the secondary structure of some lncRNAs is necessary for their function in regulating gene expression. For example, *COLDAIR* lncRNA in *Arabidopsis* and *HOTHAIR* lncRNA in human can fold into special "stem-and loop" structure as scaffold to render chromatin remodeling complex PRC2 to local chromatin, resulting in gene silencing (Sung, 2011). Recently, it was revealed that in *Arabidopsis* transcriptome, less structured mRNAs are significantly more abundant than those with high levels of folding (Li et al., 2012), implying that the mRNA secondary structure may have regulatory effects on its own expression.

In this chapter, I examined the secondary structure of *CAL* first intron RNA and test if there is any link between RNA secondary structure and its function in gene silencing.

In addition to secondary structure of intronic RNA, I also investigated the conservation of *cis*-element motifs in *CAL* first intron in closely related species as well as tested the hypothesis on the production of antisense lncRNAs from the *CAL* first intron. The results and their implications are discussed as well.

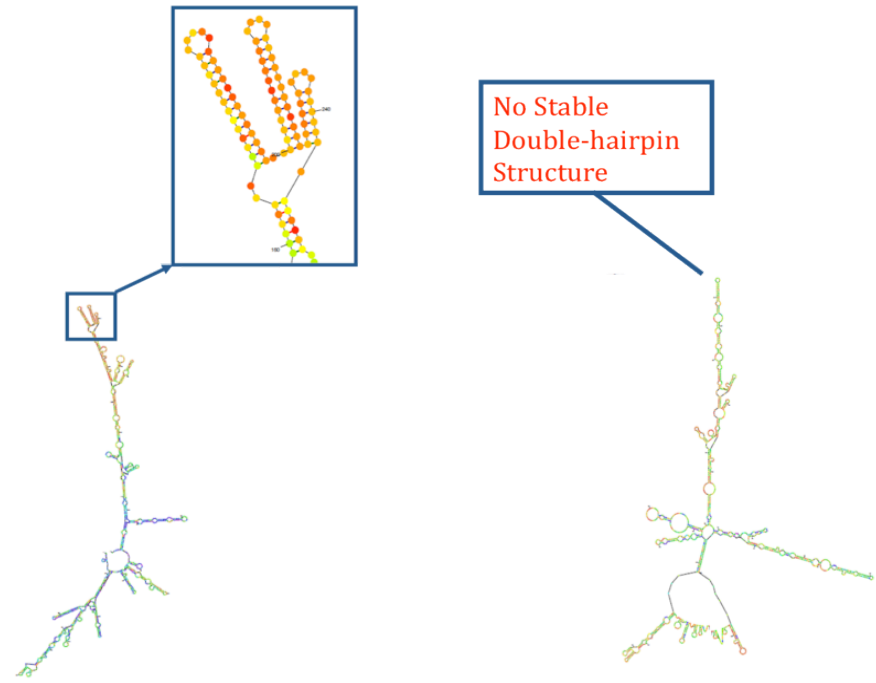
Results

Are there any stable secondary structures formed by *CAL* intronic RNA?

One immediate hypothesis is that *CAL* first intron RNA may possess special secondary structures or special primary sequence that can induce gene silencing. Since the largest intron of *AG* as well as of *CAL* are able to induce silencing, we investigated whether both *AG* second intron and *CAL* first intron share common structural features using mFOLD web server. In addition, the introns of *SHP* and *SAP*, which failed to induce gene silencing, served as negative comparisons. For each intron, both sense (forward orientation) and antisense (reverse orientation) strand were used as inputs individually, and among the output, the 5 top-ranking secondary structures (with lowest free energy and with P-num parameter screen) for each input were analyzed.

One promising and stable secondary structure was shared by the *AG* second intron sense- transcript and *CAL* first intron sense-transcript (Figure 3.1). The region between 167 bp and 243 bp in the sense RNA of *CAL* 1st intron formed a double-hairpin, a large stem-loop adjacent to a small stem-loop (Figure 3.1). This resembles the double-hairpin structure of *AG* second intron sense RNA at 2891 bp to 2980 bp and at 2401 bp to 2507 bp (Figure 3.1). However, no such stable double-hairpin

structure was observed in the antisense strand of the *CAL* and *AG* introns. Moreover, both sense and antisense RNA from the largest intronic sequences of *AP1, SHP* and *SAP* did not possess any such stable secondary structures.



CAL 1st intron. Forward

CAL 1st intron. Reverse

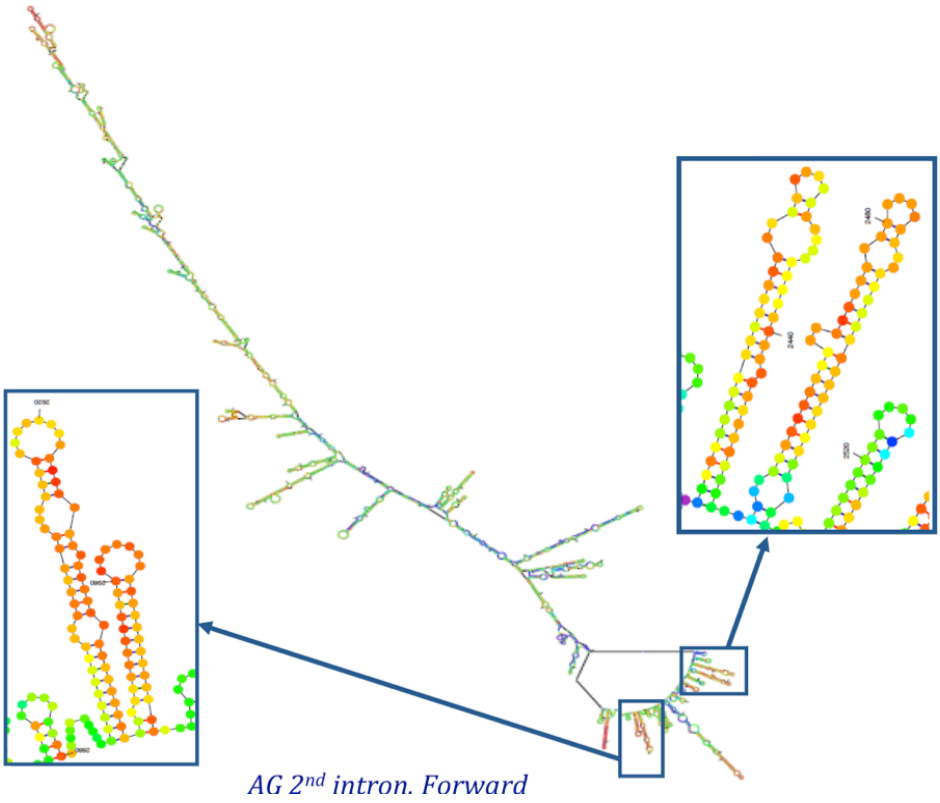


Figure 3.1. The special double-hairpin structures formed by sense RNA of *CAL* first intron and *AG* second intron.

This double-hairpin structure caught our attention for three reasons. First, the sense strand RNA of both *AG* second intron and *CAL* first intron can shape this structure, and they both have the capacity to induce TGS on corresponding gene, when over-expressed. Second, only the sense RNA of *CAL* first intron, but not antisense RNA, is able to form this structure and as discussed in Chapter 2, over-expression of *CAL* intronic sense RNAs yielded a much larger number of silenced plants than over-expressing *CAL* intronic antisense RNAs. Third, McrBC digestion demonstrated that the maximally methylated region, 35 bp to 278 bp of *CAL* first intron, overlaps with the region that forms double-hairpin at 167 nt to 243 nt of *CAL* first intron transcript.

To test whether the double-hairpin region is critical for gene silencing, we over-expressed *CAL* first intron with a deletion from 167 bp to 243 bp in *ap1-1* background using the *pEarleyGate100* vector. In T1 generation, the number of silenced plants was greatly decreased (Table 3.1) when compared with over-expression of intact CAL first intron in *pEarleyGate100* vector (Table 2.1), indicating that this double-hairpin forming region may be crucial in initiating silencing.

#167-#243 Deletion-Over-expression	*cal-like*	*weak-cal-like*	*ap1-like*
Forward	79 (2.18%)	75 (2.07%)	3476 (95.75%)
Reverse	59 (1.23%)	47 (0.98%)	4696 (97.79%)

Table 3.1. Summary of T1 transgenic plants after the deleted *CAL* first intron was over-expressed in *ap1-1* background. The *CAL* first intron contains a deletion from 167 bp to 243 bp. *pEarleyGate100* was utilized to overexpress transgenes.

Are there any evolutionary conserved sequence motifs or structures in *CAL* first intron?

If this double-hairpin region located at *CAL* first intron is functionally relevant, then the sequence or the structure should be evolutionarily conserved. Previously, The second intron of *AG* has been subjected to phylogenetic footprinting, and 12 conserved regulatory motifs among 29 Brassicaceae species were identified (Hong et al., 2003). Therefore in addition to structural conservation, conservation of primary sequences due to the presence of conserved *cis*-elements may also underlie the special ability of the *AG* and *CAL* introns.

In Phytozome database and based on syntenic relationship (Figure 3.2), we identified the *CAL* gene of five closely related species including *Arabidopsis thaliana (Ath), Arabidopsis lyrata (Aly), Capsella rubella (Cru), Brassica rapa (Bra)* and *Thellungiella halophila (Tha)*. *CAL* is apparently a recently evolved gene from duplication of *AP1* and is absent from more distantly related species (Bowman et al., 1993; Kempin et al., 1995). By sequence alignment of their largest introns, certain regions are more conserved than others (Figure 3.3). But the conserved regions (*#130-#166* bp; *#414-#445* bp, *#545-#623* bp) do not exactly overlap with the double hairpin nor the regions that covers 24 nt siRNA peaks (Figure 2.6). A limited number of conserved transcription factor binding sites were identified (Figure 3.3), this includes the binding site of SQUAMOSA-promoter binding protein-like (SPL), Leafy (LFY) and MADS box transcription factors (*CArG box*). Those *cis*-regulatory sequences may be functionally important. Whether these *cis*-elements are related to the ability of *CAL* intron to silence the endogenous *CAL* gene remains unresolved.

RNA structure analysis program mFOLD was also used to detect if the double hairpin structure is conserved by analyzing the largest intron transcript of *CAL* in *Aly, Cru, Bra* and *Tha*. However, no stable double-hairpin structure was found in the *CAL* of *Aly, Bra* and *Tha* (data not shown); only some stable double and single hairpins are found in the first intron of *CAL* in *Cru.*

Taken all together, the regulatory sequences rather than the secondary structure may have an evolutionarily conserved function, perhaps related to the competence of *CAL* intron to silence *CAL* gene.

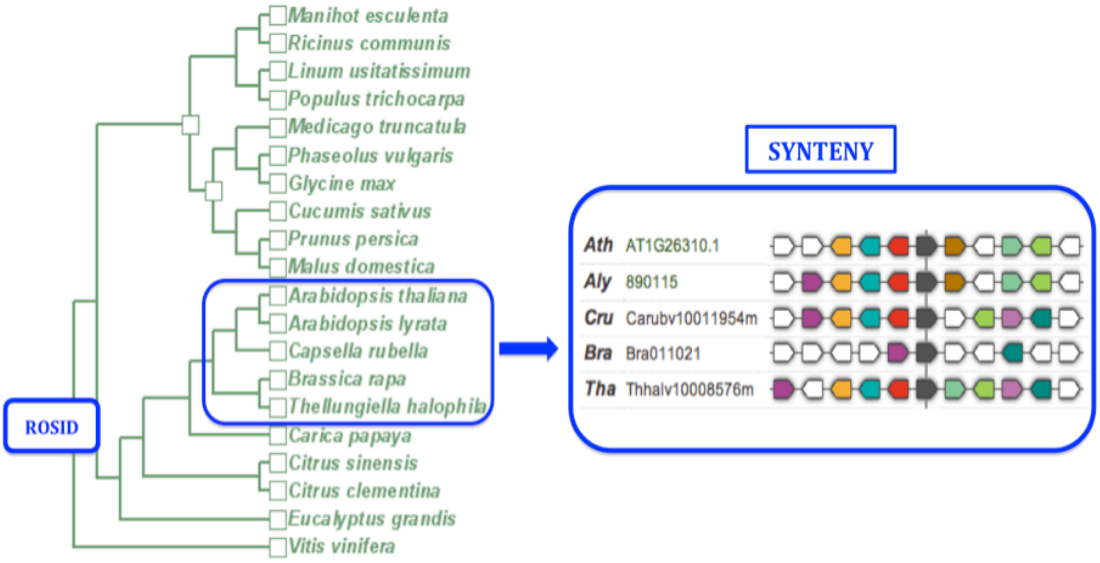


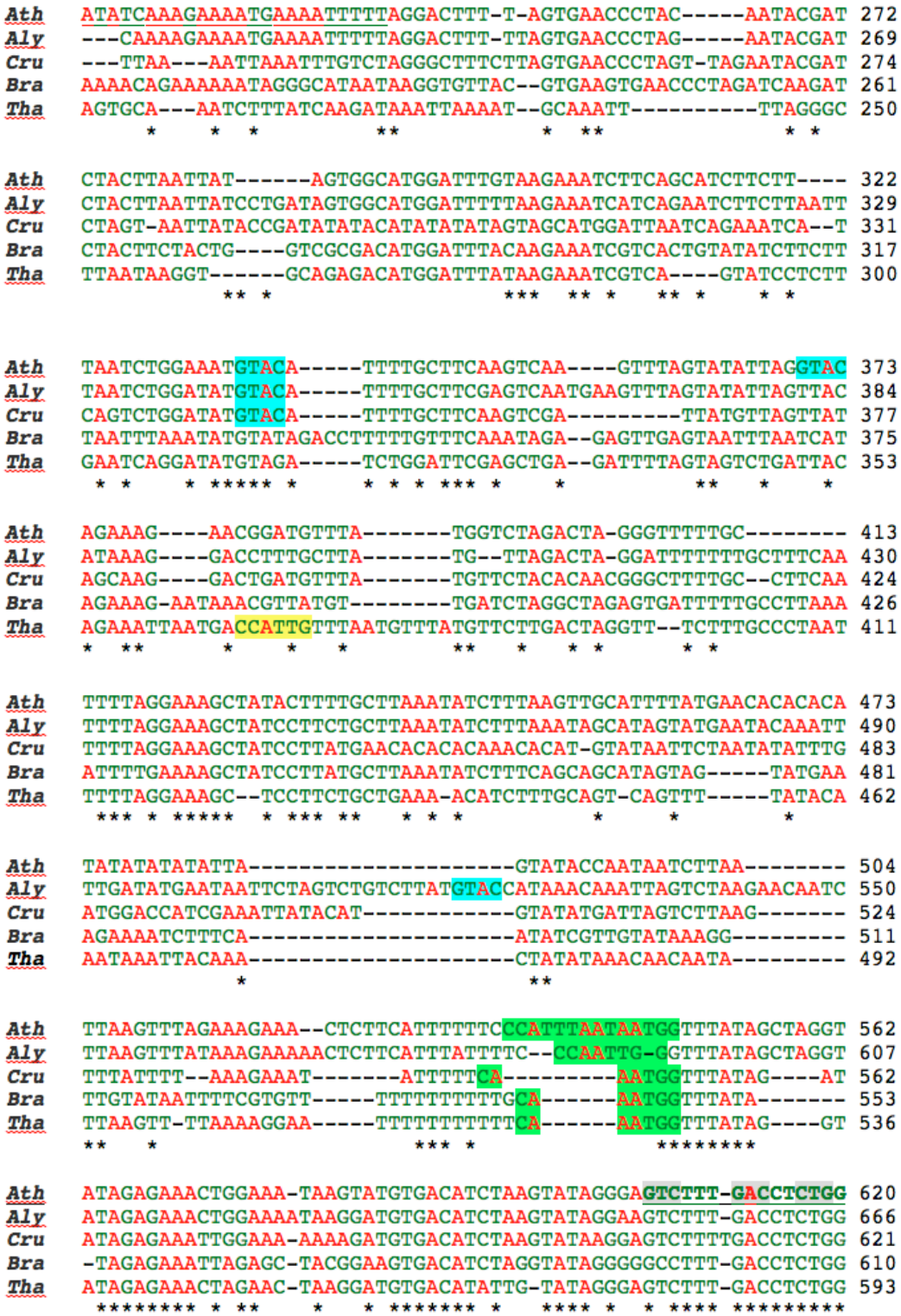
Figure 3.2. Phytozome analysis. Synteny was used in locating *CAL* gene in different spieces.

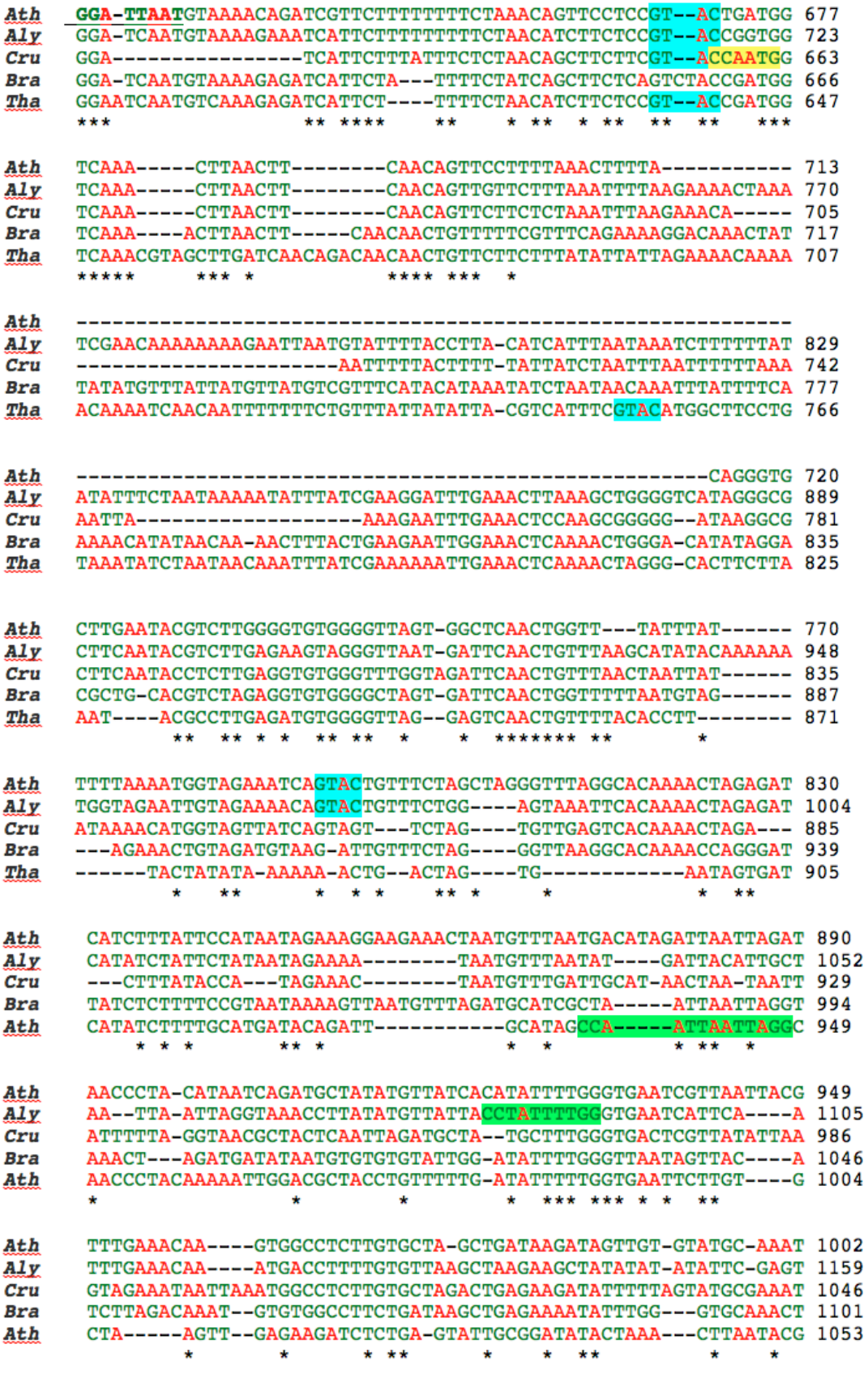
```
Ath   GTAATTGCTTAATTCCTTCTTT-TTTTAAT------GTTATTTT-TAGTGTGCCTTCGTT 52
Aly   GTAATTGCTTAATTCCTTTTTTGTTTTAAT------CTTATTTT-CAGTGTTCCTTTGTT 53
Cru   GTAATTGCATCATTCCCTCTTTTTATTTATTTATTAATTATATTGTAGTATGCCTTTGTT 60
Bra   GTAACTGCATAATTCCCCTTTTAA-------------TTGTTTT--AGTGTGCCTTTGTT 45
Tha   GTAATTATTTAATTCTTTCTATTT-------------TTGTCTT--TGTTTATTTCTTT- 44
      **** *   * ****    * *                ** * **   ** *   *   *

Ath   TGCCCTAACTAGTAGTCTTTGTTCTAC--TTAAGGCATACTTTCTGTGTCTTCTATTCTA 110
Aly   TGCCCTAACTAGTAGTTTTTGCGTTACAATTAAGGCCT--TTTCTGGATCTCCTCTTATC 111
Cru   TGCCCTAAAAAGTAGTTTTTGCTCTAC---TACTCTAC---TTCT-TGTCTCCTCTTCAT 113
Bra   TGCCCTAATAAATAGTTTTTGTTCTCCT-TTAGGCCTT---TTCTTGGTGTCTTCTTATG 101
Tha   ----CCAATGT-TCCTTTGTTTGCCCTA-ATTAGTAGT---TTCTTGCTCTACTCATCTG 95
          * **    *  * * *          *          ****   * *  *  *

Ath   TTATCTGTCTTTCTGAAAATTTGCCACTG----ATTTGGTATCTATTTACTTGGGTCTAC 166
Aly   TTCT----CTTCATGAAAATTTGTCACAG----ATTTGGTATCTGATTACTTGGATCTAT 163
Cru   CTGTCT--TTTTATGAAAATTTATCACAG----ATTTAGTATCTGATGACTTGGATCTAC 167
Bra   --------TTTTATGAAA-ATTCTCACAA----ATTTTGTAGTTAATTACTTGGATCTAC 148
Tha   --------CCTTATATTTTGTTGTCACAGCTATATTTTGCAATTGATTACTTGGATCTAC 147
                *  *      **  ***      **** * *  *  * ****** ****

Ath   GAACTGAT--TGT--GTTG--GTCATA-TCATTAGTTTATTTTTATCAATAATTTATTAT 219
Aly   GAACTGATATTGT--GTTG--GTCAAA-CCATTAGTTTGTTTTTATTATAAATTTAATAT 218
Cru   GAAATGAT--TGT--GTTA--GTCAAGCTCAAACCAATAGTTTTATTTTTATTTTTATTT 221
Bra   GAATTGATTTT----ACCAAAGTTAAA-CCATTATAGCATAATTGCTTATATCAGAAAAA 203
Tha   GAACTGGTTGTGTTGATTAATGTCAAA-CCATTAGCATATT-TTATCTTTATCAAAAGAA 205
      *** ** *  *          ** *    **           **      *
```


Ath ATATCAAAGAAAATGAAAATTTTTAGGACTTT-T-AGTGAACCCTAC-----AATACGAT 272
Aly ---CAAAAGAAAATGAAAATTTTTAGGACTTT-TTAGTGAACCCTAG-----AATACGAT 269
Cru ---TTAA---AATTAAATTTGTCTAGGGCTTTCTTAGTGAACCCTAGT-TAGAATACGAT 274
Bra AAAACAGAAAAAATAGGGCATAATAAGGTGTTAC--GTGAAGTGAACCCTAGATCAAGAT 261
Tha AGTGCA---AATCTTTATCAAGATAAATTAAAAT--GCAAATT----------TTAGGGC 250
Ath CTACTTAATTAT------AGTGGCATGGATTTGTAAGAAATCTTCAGCATCTTCTT---- 322
Aly CTACTTAATTATCCTGATAGTGGCATGGATTTTTAAGAAATCATCAGAATCTTCTTAATT 329
Cru CTAGT-AATTATACCGATATATACATATATATAGTAGCATGGATTAATCAGAAATCA--T 331
Bra CTACTTCTACTG----GTCGCGACATGGATTTACAAGAAATCGTCACTGTATATCTTCTT 317
Tha TTAATAAGGT------GCAGAGACATGGATTTATAAGAAATCGTCA----GTATCCTCTT 300
Ath TAATCTGGAAATGTACA-----TTTTGCTTCAAGTCAA----GTTTAGTATATTAGGTAC 373
Aly TAATCTGGATATGTACA-----TTTTGCTTCGAGTCAATGAAGTTTAGTATATTAGTTAC 384
Cru CAGTCTGGATATGTACA-----TTTTGCTTCAAGTCGA---------TTATGTTAGTTAT 377
Bra TAATTTAAATATGTATAGACCTTTTTGTTTCAAATAGA--GAGTTGAGTAATTTAATCAT 375
Tha GAATCAGGATATGTAGA-----TCTGGATTCGAGCTGA--GATTTTAGTAGTCTGATTAC 353
Ath AGAAAG----AACGGATGTTTA-------TGGTCTAGACTA-GGGTTTTTGC-------- 413
Aly ATAAAG----GACCTTTGCTTA-------TG--TTAGACTA-GGATTTTTTTGCTTTCAA 430
Cru AGCAAG----GACTGATGTTTA-------TGTTCTACACAACGGGCTTTTGC--CTTCAA 424
Bra AGAAAG-AATAAACGTTATGT--------TGATCTAGGCTAGAGTGATTTTTGCCTTAAA 426
Tha AGAAATTAATGACCATTGTTTAATGTTTATGTTCTTGACTAGGTT--TCTTTGCCCTAAT 411
Ath TTTTAGGAAAGCTATACTTTTGCTTAAATATCTTTAAGTTGCATTTTATGAACACACACA 473
Aly TTTTAGGAAAGCTATCCTTCTGCTTAAATATCTTTAAATAGCATAGTATGAATACAAATT 490
Cru TTTTAGGAAAGCTATCCTTATGAACACACACAAACACAT-GTATAATTCTAATATATTTG 483
Bra ATTTTGAAAAGCTATCCTTATGCTTAAATATCTTTCAGCAGCATAGTAG-----TATGAA 481
Tha TTTTAGGAAAGC--TCCTTCTGCTGAAA-ACATCTTTGCAGT-CAGTTT-----TATACA 462
Ath TATATATATATTA---------------------GTATACCAATAATCTTAA-------- 504
Aly TTGATATGAATAATTCTAGTCTGTCTTATGTACCATAAACAAATTAGTCTAAGAACAATC 550
Cru ATGGACCATCGAAATTATACAT------------GTATATGATTAGTCTTAAG------- 524
Bra AGAAAATCTTTCA---------------------ATATCGTTGTATAAAGG--------- 511
Tha AATAAATTACAAA---------------------CTATATAAACAACAATA--------- 492
Ath TTAAGTTTAGAAAGAAA--CTCTTCATTTTTTCCCATTTAATAATGGTTTATAGCTAGGT 562
Aly TTAAGTTTATAAAGAAAAACTCTTCATTTATTTTC--CCAATTG-GGTTTATAGCTAGGT 607
Cru TTTATTTT--AAAGAAAT-------ATTTTTCA---------AATGGTTTATAG----AT 562
Bra TTGTATAATTTTCGTGTT-----TTTTTTTTTTGCA------AATGGTTTATA------- 553
Tha TTAAGTT-TTAAAAGGAA-----TTTTTTTTTTTCA------AATGGTTTATAG----GT 536
Ath ATAGAGAAACTGGAAA-TAAGTATGTGACATCTAAGTATAGGGAGTCTTT-GACCTCTGG 620
Aly ATAGAGAAACTGGAAAATAAGGATGTGACATCTAAGTATAGGAAGTCTTT-GACCTCTGG 666
Cru ATAGAGAAATTGGAAA-AAAAGATGTGACATCTAAGTATAAGGAGTCTTTTGACCTCTGG 621
Bra -TAGAGAAATTAGAGC-TACGGAAGTGACATCTAGGTATAGGGGGCCTTT-GACCTCTGG 610
Tha ATAGAGAAACTAGAAC-TAAGGATGTGACATATTG-TATAGGGAGTCTTT-GACCTCTGG 593


Ath GGA-TTAATGTAAAACAGATCGTTCTTTTTTTTCTAAACAGTTCCTCCGT--ACTGATGG 677
Aly GGA-TCAATGTAAAAGAAATCATTCTTTTTTTTTCTAACATCTTCTCCGT--ACCGGTGG 723
Cru GGA----------TCATTCTTTATTTCTCTAACAGCTTCTTCGT--ACCAATGG 663
Bra GGA-TCAATGTAAAAGAGATCATTCTA---TTTTCTATCAGCTTCTCAGTCTACCGATGG 666
Tha GGAATCAATGTCAAAGAGATCATTCT----TTTTCTAACATCTTCTCCGT--ACCGATGG 647
TCAAA-----CTTAACTT--------CAACAGTTCCTTTTAAACTTTTA---------- 713
TCAAA-----CTTAACTT--------CAACAGTTGTTCTTTAAATTTTAAGAAAACTAAA 770
TCAAA-----CTTAACTT--------CAACAGTTCTTCTCTAAATTTAAGAAACA----- 705
TCAAA----ACTTAACTT-----CAACAACTGTTTTTCGTTTCAGAAAAGGACAAACTAT 717
TCAAACGTAGCTTGATCAACAGACAACAACTGTTCTTCTTTATATTATTAGAAAACAAAA 707
Ath ----------
Aly TCGAACAAAAAAAAGAATTAATGTATTTTACCTTA-CATCATTTAATAAATCTTTTTTAT 829
Cru ----------AATTTTTACTTTT-TATTATCTAATTTAATTTTTTAAA 742
Bra TATATGTTTATTATGTTATGTCGTTTCATACATAAATATCTAATAACAAATTTATTTTCA 777
Tha ACAAAATCAACAATTTTTTTCTGTTTATTATATTA-CGTCATTTCGTACATGGCTTCCTG 766
CAGGGTG 720
ATATTTCTAATAAAAATATTTATCGAAGGATTTGAAACTTAAAGCTGGGGTCATAGGGCG 889
AATTA----------AAAGAATTTGAAACTCCAAGCGGGGG--ATAAGGCG 781
AAAACATATAACAA-AACTTTACTGAAGAATTGGAAACTCAAAACTGGGA-CATATAGGA 835
TAAATATCTAATAACAAATTTATCGAAAAAATTGAAACTCAAAACTAGGG-CACTTCTTA 825
CTTGAATACGTCTTGGGGTGTGGGGTTAGT-GGCTCAACTGGTT---TATTTAT------ 770
CTTCAATACGTCTTGAGAAGTAGGGTTAAT-GATTCAACTGTTTAAGCATATACAAAAAA 948
CTTCAATACCTCTTGAGGTGTGGGTTTGGTAGATTCAACTGTTTAACTAATTAT------ 835
CGCTG-CACGTCTAGAGGTGTGGGGCTAGT-GATTCAACTGGTTTTTAATGTAG------ 887
AAT----ACGCCTTGAGATGTGGGGTTAG--GAGTCAACTGTTTTACACCTT-------- 871
TTTTAAAATGGTAGAAATCAGTACTGTTTCTAGCTAGGGTTTAGGCACAAAACTAGAGAT 830
TGGTAGAATTGTAGAAAACAGTACTGTTTCTGG----AGTAAATTCACAAAACTAGAGAT 1004
ATAAAACATGGTAGTTATCAGTAGT---TCTAG----TGTTGAGTCACAAAACTAGA--- 885
---AGAAACTGTAGATGTAAG-ATTGTTTCTAG----GGTTAAGGCACAAAACCAGGGAT 939
------TACTATATA-AAAAA-ACTG--ACTAG----TG-----------AATAGTGAT 905
CATCTTTATTCCATAATAGAAAGGAAGAAACTAATGTTTAATGACATAGATTAATTAGAT 890
CATATCTATTCTATAATAGAAAA--------TAATGTTTAATAT----GATTACATTGCT 1052
---CTTTATACCA---TAGAAAC--------TAATGTTTGATTGCAT-AACTAA-TAATT 929
TATCTCTTTTCCGTAATAAAAGTTAATGTTTAGATGCATCGCTA-----ATTAATTAGGT 994
CATATCTTTTGCATGATACAGATT----------GCATAGCCA-----ATTAATTAGGC 949
AACCCTA-CATAATCAGATGCTATATGTTATCACATATTTTGGGTGAATCGTTAATTACG 949
AA--TTA-ATTAGGTAAACCTTATATGTTATTACCTATTTTGGGTGAATCATTCA----A 1105
ATTTTTA-GGTAACGCTACTCAATTAGATGCTA--TGCTTTGGGTGACTCGTTATATTAA 986
AAACT---AGATGATATAATGTGTGTGTATTGG-ATATTTTGGGTTAATAGTTAC----A 1046
AACCCTACAAAAATTGGACGCTACCTGTTTTTG-ATATTTTTGGTGAATTCTTGT----G 1004
TTTGAAACAA----GTGGCCTCTTGTGCTA-GCTGATAAGATAGTTGT-GTATGC-AAAT 1002
TTTGAAACAA----ATGACCTTTTGTGTTAAGCTAAGAAGCTATATAT-ATATTC-GAGT 1159
GTAGAAATAATTAAATGGCCTCTTGTGCTAGACTGAGAAGATATTTTTAGTATGCGAAAT 1046
TCTTAGACAAAT--GTGTGGCCTTCTGATAAGCTGAGAAAATATTTGG---GTGCAAACT 1101
CTA-----AGTT--GAGAAGATCTCTGA-GTATTGCGGATATACTAAA---CTTAATACG 1053

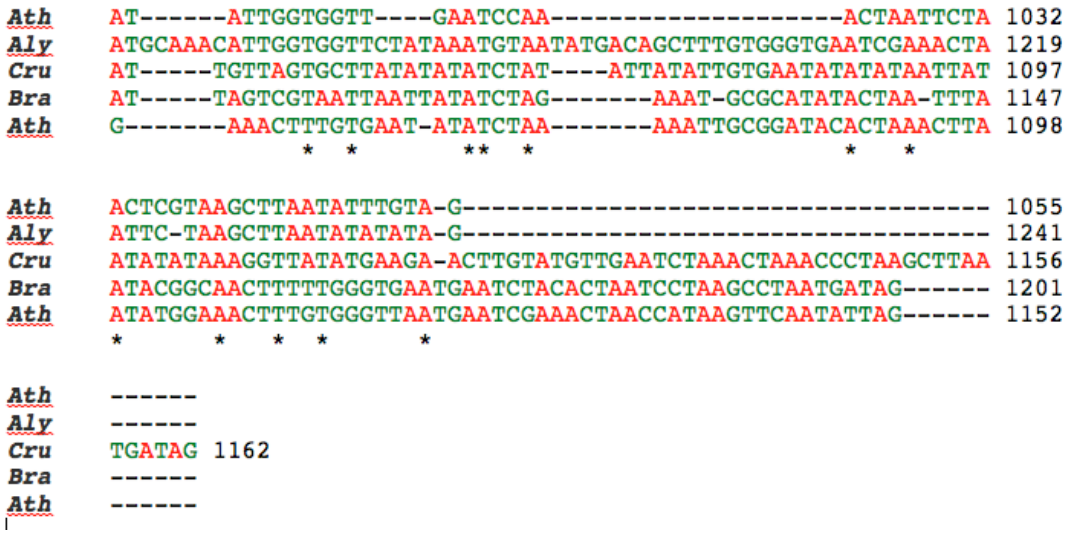


Figure 3.3. Alignment of *CAL* first intron of five relative spieces. Yellow color shadow indicates LFY binding sites, blue color shadow indicates SPL binding sites and green color shadow indicates CArG box.

Could the intronic regulatory elements direct *CAL* gene expression?

It is well known that *AG* second intron contains both activator and repressor binding sites and acts as an enhancer in transcription (Busch et al., 1999; Lohmann et al., 2001; Hong et al., 2003). To test if *CAL* first intron possesses *cis*-regulatory elements, *CAL* first intron was fused with the minimal promoter *TATA box*, which is placed upstream of beta-glucuronidase (*GUS)* reporter. If *CAL* first intron does act as an enhancer, the *GUS* expression would be observed in floral primordial tissues when stained with X-Gluc (5-bromo-4-chloro-3-indolyl-β-D-glucuronic acid), the substrate of *GUS* (Figure 3.4). Otherwise, the *TATA box* alone is insufficient to drive *GUS* expression. Our data support the idea that enhancer elements located in the *CAL* first intron may underlie the ability of *CAL* intron to initiate silencing.

Figure 3.4. *CAL* first intron sequence contains enhancer, which is able to assist *TATA box* to drive *GUS* expression in flowers.

Could promoter of any genes trigger silencing?

The proximal promoter regions upstream of most genes tend to contain primary regulatory sequences, such as enhancers. Therefore, if the functional regulatory elements are the critical key for *CAL* first intron to initiate silencing, then over-expression of a gene's upstream regulatory elements (including promoters) may also trigger gene silencing. Thus, we over-expressed the upstream promoter sequences of *CAL*, *AP1*, *LFY* (*LEAFY*) and *TFL1* (*TERMINAL FLOWER 1*) under 35S in the specific plant background. Currently, we are still waiting for the results from *AP1*, *LFY* and *TFL1* (Table 3.2).

Over-expression of promoters of gene	Background	Orientation	Silenced phenotype	Non-silenced phenotype
CAL	*ap1-1*	Forward	3 (0.57%)	524 (99.43%)
		Reverse	41 (6.21%)	619 (93.79%)
AP1 (+5'-UTR)	*Ler*	Forward	15 (8.47%)	162 (91.53%)
		Reverse	285 (66.13%)	146 (33.87%)
AP1	*Ler*	Forward		
		Reverse		
LFY	*Ler*	Forward		
		Reverse		
TFL1	*Ler*	Forward		
		Reverse		

Table 3.2. Summary of promoter-mediated gene silencing in T1 transgenic *Arabidopsis* plants. Forward or Reverse indicates the orientation of the promoter sequence. The *pEarleyGate100* vector was used.

Based on limited data on *CAL* promoter over-expression, it is interesting to note that over-expressing the *CAL* promoter sequence in the reverse orientation induces considerably higher percentage of silenced plants than the forward orientation. It is likely that the promoter region produces lncRNAs in the forward orientation (Figure 2.10), which more readily pair with transgenic transcripts in the reverse orientation.

Could 3'-terminator be involved in silencing?

One puzzling question is why the endogenous *CAL* first intron, presumably excised from the pre-mRNA splicing, did not evoke gene silencing? One possibility is that the spliced intron is extremely unstable and immediately degraded after splicing. This particularly considers that the spliced intron is in a lariat form and lacks poly-A tail. In contrast, the *35S::CALI1* has a terminator sequence resulting in a poly-A tail added to the *CAL* intron transcripts. In order to test this hypothesis, we used *35S* to drive *CAL* first intron, but meanwhile deleted terminator sequence in the binary vector. The

expressed intron transcripts should lack poly-A tails. However, the potency of the *35S*-driven *CAL* first intron in triggering silencing did not significantly change (Table 3.3). Therefore, the poly-A tails at 3'-end are not critical for the *CAL* intronic transcripts to silence *CAL* gene, rather the lariat form of excised endogenous intron might be extremely unstable or might possess this unique structure recalcitrant to form dsRNA.

	Orientation	*cal-like*	*ap1-like*
CAL1st intron.pEarleyGate100	Forward	33.24%	66.76%
	Reverse	3.38%	96.62%
CAL1st intron.pEarleyGate100 -OCS	Forward	34.72%	65.82%
	Reverse	2.66%	97.34%

Table 3.3. The 3'-end poly-A tail generated from the terminator of vector is not essential for *CAL* gene silencing.

Are there any antisense long non-coding RNAs derived from endogenous *CAL* first intron?

Previous studies have shown that non-coding RNAs are prevalent in the *Arabidopsis* genome. On the basis of the length and genomic locations, these non-coding RNAs are classified as small ncRNAs including miRNAs and siRNAs, natural antisense transcripts, long intronic non-coding RNAs and long intergenic non-coding RNAs. The last three categories are also referred to as long non-coding RNAs (lncRNAs) (Kim and Sung, 2012). In *Arabidopsis thaliana*, the lncRNAs exhibited a tissue-specific pattern of expression (Sung, 2011). Although a large number of the lncRNAs

in *Arabidopsis* have been shown to respond to abiotic stresses, the function of most of them still remain elusive (Matsui et al., 2008). Thus, I hypothesize, as discussed in chapter 2, that non-coding RNAs generated from the endogenous *CAL* first intron region may exist, perhaps in the antisense orientation. When the transcript derived from transgenic *CAL* first intron paired with the existing non-coding RNAs in antisense orientation to form double-stranded RNAs (dsRNA), the classical RDR2-DCL3-AGO4 RNAi pathway is triggered. To test this hypothesis, I examined RNA-seq data from available databases, such as PLncDB (Jin et al., 2013), focusing on the *CAL* first intron region. However, no lncRNA from the *CAL* first intron was found. Since most RNA-seq data was derived from poly-A RNA and lncRNAs may not necessarily possess poly-A tails, the existing databases may not detect these lncRNAs.

To experimentally test if there is non-coding RNA in *CAL* first intron region by qRT-PCR, total RNA was extracted from the inflorescences of *ap1-1* mutant (with *wild type CAL* locus), and first strand cDNA synthesis was performed using sense-directional primers specific to *CAL* first intron, which can only prime antisense intronic RNAs, but not primary mRNAs. Specifically, three primers in sense (forward) direction, 79F, 157F and 604F were designed according to the sequence proximal to the small RNA peak regions (Figure 3.5). As a negative control, cDNA synthesis was carried out in the absence of reverse transcriptase. As shown in Figure 3.6, each sense primer paired with an antisense RNA and yielded PCR products, implying the endogenous antisense RNA from *CAL* first intron does exist.

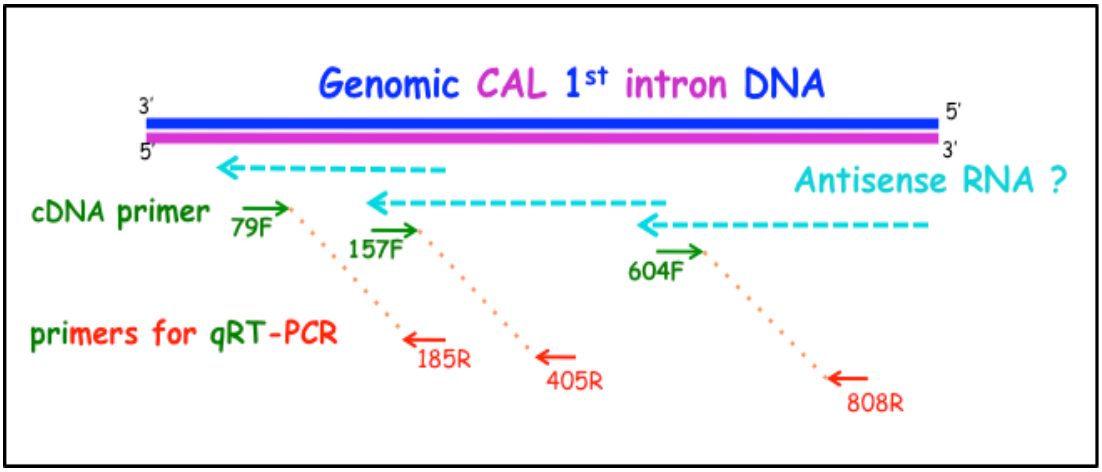


Figure 3.5. The location and direction of primers specific to examine the existence of antisense RNAs originated from endogenous *CAL* first intron.

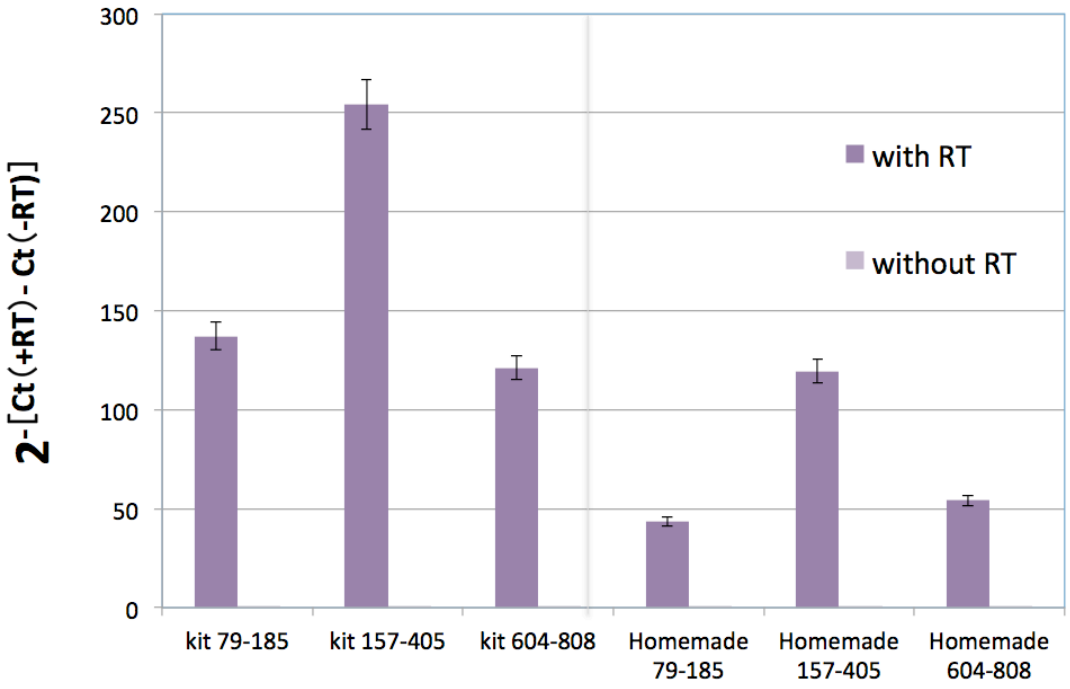


Figure 3.6. The antisense RNAs detected by three sense-directional primers specific to *CAL* first intron. The left and right panels are qRT-PCR results with RNA extracted with Qiagen RNeasy kit (left) and homemade protocol (right). Two biological replicates of *ap1-1* RNAs were utilized for each extraction method.

To determine whether these three qRT-PCR fragments *#79-#185*, *#157-#405* and *#604-#808*, are separate individual antisense RNAs, or different parts of the same long antisense RNA transcript, we performed regular PCR to amplify longer amplicons, *#79-#405*, *#79-#808* and *#157-#808,* and the PCR products of *#79-#405* and *#157-#808* were successfully obtained (Figure 3.7). This indicates that there are two independent antisense RNAs, 327 nt and 652 nt in length, respectively. These two antisense RNAs overlap in the *#157-#405* region, consistent with the qRT-PCR data, in which the amount of qRT-PCR product of *#157-#405* is 2-fold higher than that of *#79-#185* and *#604-#808* (Figure 3.6). However, the PCR for *#79-#808* failed to detect any PCR products, indicating a lack of antisense RNA from the beginning to the end of *CAL* first intron (Figure 3.7).

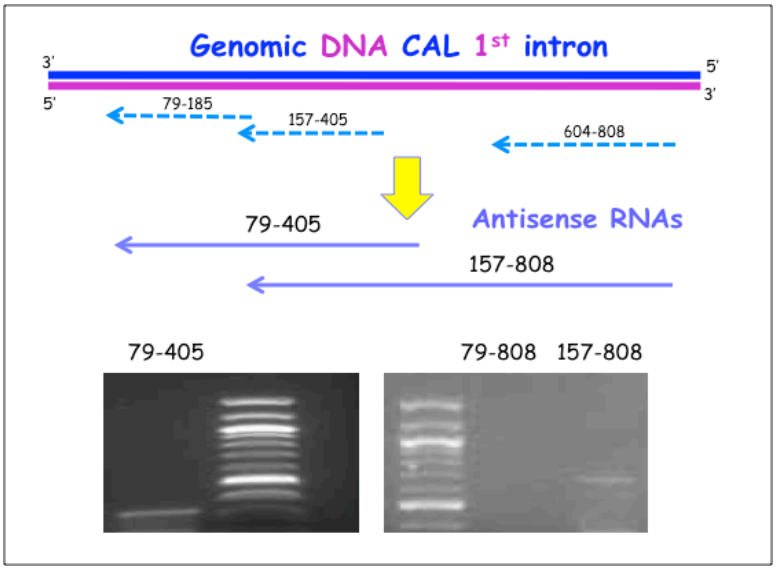


Figure 3.7. Two individual antisense lncRNAs generated from endogenous *CAL* first intron.

To test whether the expression of the lncRNAs derived from *CAL* first intron is tissue-specific, the total RNA extracted from leaves of *ap1-1* mutant and *wild type*

plants were examined by qRT-PCR. Using the primers specific to the region *#157-#405* of *CAL* first intron, we cannot detect any PCR products from leaf RNA, suggesting that the *CAL* first intron-derived lncRNAs is restricted to the inflorescence tissue.

Conclusion and Discussion

Combining all the above, we propose a model (Figure 3.8) that illustrates how this *CAL* intron-mediated gene silence is initiated. The enhancer elements located in the *CAL* first intron recruit Pol II and the other associated transcription factors. Then the Pol II complex migrates along the first intron toward the promoter to activate transcription. Along the tracking path, RNA Pol II synthesizes at least two antisense lncRNA by-products from *CAL* first intron. When the transcripts from *35S::CALI1* pair with these antisense lncRNAs, the resulting dsRNAs initiate the RdDM pathway. The dsRNA duplex is processed into 24 nt siRNAs by DCL3 and 24 nt siRNAs are amplified by RDR2-DCL3 amplification loop. Then, abundant 24 nt siRNAs are loaded onto AGO4 to guide methylation at the endogenous *CAL* first intron to repress *CAL* gene transcription (Figure 3.8).

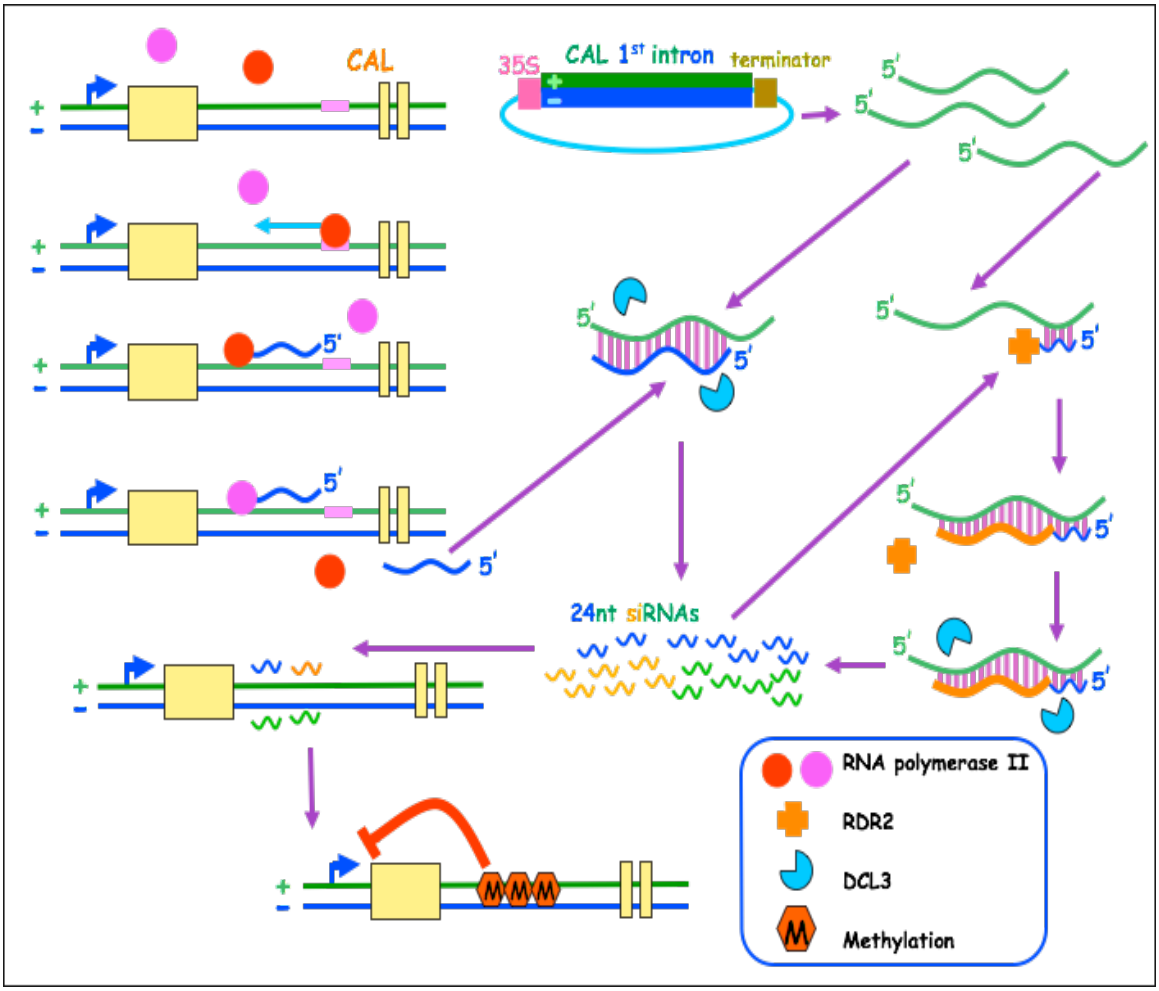


Figure 3.8. The proposed model illustrating the molecular basis of *CAL* first intron-mediated gene silencing.

In conclusion, our work indicates that over-expressing *CAL* first intron may induce gene silencing because of antisense lncRNAs produced from the *CAL* first intron. The production of lncRNAs from *CAL* first intron on the other hand may depend on the presence of enhancer elements in the intron. Therefore, one exciting potential application of our finding is the new way of silencing genes by over-expressing promoters or introns containing enhancer elements.

Methods

Phylogenetic analysis, alignment of sequences, and detection of transcription factor binding sites in *CAL* first intron

We used the Phytozome database (http://www.phytozome.net) to search for the *CAL* genes in other species. The *Arabidopsis thaliana (Ath*) CAL protein sequence was used as input and did BLAST against the Rosid plant family, which yielded 97 hypothetical Rosid genes including 5 in *Manihot esculenta (*Mes), 1 in *Ricinus communis (Rco*), 14 in *Linum usitatissimum (Lus),* 3 in *Medicago truncatula (Mtr*), 2 in *Phaseolus vulgaris (Pvu*), 28 in *Malus domestica (Mdo)*, 1 in *Arabidopsis lyrata (Aly),* 3 in *Capsella rubella (Cru),* 11 in *Brassica rapa (Bra),* 3 in *Thellungiella halophila (Th*a), 7 in *Carica papaya (Cpa),*3 in *Citrus sinensis (Csi),* 4 in *Citrus clementine (Ccl)* and 11 in *Eucalyptus grandis (Egr).* Subsequently, these genes were filtered by two criteria, similar synteny to *Ath CAL* gene and the presence of long first intron. 16 candidate genes from 9 Rosid species emerged, including *cassava4:1_031659m* gene in *Mes*, *LUS10034662, 10021140, 10007983, 10005081* genes in *Lus*, *Medtr5g049070.1* gene in *Mtr, Phvulv091015107m* and *091000547m* genes in *Pvu, #890115* gene in *Aly, Carubv10011954m* gene in *Cru, Bra012997, 014628, 011021* and *035952* genes in *Bra, Thhalv10006576m* gene in *Tha* and *orange1.1g046479m* gene in *Csi*. The first intron of the above listed genes were aligned against the first intron of *Ath CAL* by ClustalW2 (http://simgene.com/ClustalW), only the introns of 4 putative *CAL* genes exhibited some levels of conservation; they are *#890115* in *Aly, Carubv10011954m* in *Cru, Bra011201* in *Bra* and *Thhalv10006576m* in *Tha*. Well-defined LFY binding site

(CCANTG), SPL binding site (GTAC), and MADS box consensus sequence ($CC[A/T]_6GG$) were identified and highlighted in these introns.

mFold analysis

In the mfold Web Server, RNA Folding Form (Version 2.3 energies), was used to examine RNA secondary structures; the web link is http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form2.3 (Zuker, 2003). The *CAL* first intron sequence from the individual species, *Arabidopsis thaliana (Ath), Arabidopsis lyrata (Aly), Capsella rubella (Cru), Brassica rapa (Bra)* and *Thellungiella halophila (Tha),* as well as the sequence of *AGOUMAS (AG)* second intron in *Ath,* were used as input. The settings are linear RNA, 22˚C folding temperature and structure annotation in p-num. In the output, the top 5 folding structures with the least free energies were further analyzed. For each folding structure, the local regions in orange or red are the most stable, which are usually consistent among the inspected 5 folding structures. The green and blue color shadowed regions display varied local structures among the 5 folding data, indicating instability of these secondary structures.

Deletion of the hairpin-forming region of the intron

A similar method described earlier for *CAL* intron deletions was utilized to delete the hairpin-forming region *(#167-#243*) of the *CAL* first intron. The primers amplifying the first fragment were CALI1.F' and a reverse primer CALI1.166R starting from #166 of *CAL* first intron. For the the second PCR fragment, the forward primer,

CALI1.(152-166)+(244-259).F, consisting of *CAL* first intron sequence *#152-#166* followed by *#244-#259*, was paired with CALI1.R'. Hence, the second PCR product contains a deletion #167 to #243 but overlaps with the 3'-end of the first PCR fragment at the 5'-end. After gel purification, these two PCR products were mixed at 1:1 molar ratio and served as template for a third PCR reaction with primers CALI1.F' and CALI1.R' in a 2-step PCR with Phusion Taq: 98˚C for 30 s, followed by 5 cycles of 98˚C for 10 s, 50˚C for 30 s, 72˚C for 30 s, then by 26 cycles of 98˚C for 10 s, 60˚C for 30 s, 72˚C for 30 s, finally 72˚C for 5 min. After gel extraction, the PCR fragment was cloned into *pCR8/GW/TOPO* in either sense or antisense orientation, and then into *pEarleyGate100* vector (Earley et al., 2006). The transformation of the plasmid constructs into *ap1-1* plants, selection and scoring of T1 generation were as the same as described for *35S::CALI1* in *pEarleyGate100*.

Construction of *CALI1+TATA::GUS*

For the construction of *CALI1+TATA::GUS*, *CAL* first intron sequence was amplified by intron specific primer CALI1.F' and CALI1.R'. The minimal promoter *TATA box* sequence was amplified using *pMDC32* vector as template using primers CALI1F+TATA.F and TATA.R. The achieved *TATA box* PCR product comprises 10 bp at 5'-end that overlap with the 3'-end of *CAL* first intron fragment. After gel purification (Gel extraction kit, Qiagen), The *TATA box* PCR fragment was mixed with the *CALI1* PCR fragment at 1:1 molar ratio. Then the mixture was used as PCR template and CALI1.F' and TATA.R were used as primers to perform a 2-step Phusion PCR. Thermal conditions for PCR were as follows: 98˚C for 30 s, followed

by 5 cycles of 98˚C for 10 s, 50˚C for 30 s, 72˚C for 30 s, then by 26 cycles of 98˚C for 10 s, 60˚C for 30 s, 72˚C for 30 s, finally 72˚C for 5 min. Following gel extraction (Qiagen), the amplified *CAL* first intron with *TATA box* fragment at the 3'-end was cloned into *pCR8/GW/TOPO* by TA cloning kit in sense orientation, and then recombined into *pMDC162* vector through the Gateway® technology (Invitrogen). The *Agrobacterium tumefaciens GV3101* containing the construct was transformed into the *Ler* plants. About 15 transgenic lines were generated.

Analysis of *GUS* reporter gene expression

The inflorescence from *CALI1+TATA::GUS* transgenic plants were collected, placed in 90% cold acetone, and incubated at room temperature for 20 minutes. Acetone was replaced with staining buffer without X-Gluc (0.2% Triton X-100, 50mM NaHPO4 Buffer pH7.2, 2mM Potassium Ferrocyanide, 2mM Potassium Ferricyanide). Then the buffer without X-Gluc was replaced with the staining buffer with 2mM X-Gluc. The tissues in the staining buffer were placed under a vacuum until the tissues submerged beneath the surface of the solution. After incubation at 37°C overnight, the tissues were washed in successive ethanol series (20%, 35% and 50% ethanol) at room temperature for 30 minutes each time. FAA fixative (50% Ethanol, 5% Formaldehyde, 10% Acetic acid, rest water) was used to fix the tissues for 30 minutes at room temperature, and then replaced by 70% ethanol.

Directional cDNA synthesis and qRT-PCR

Total RNA was isolated from inflorescences of *ap1-1* and *wild type Ler* plants using RNeasy® Plant mini kit (Qiagen). First-strand cDNA was synthesized from 1 µg total RNA with *CAL* first intron specific sense-directional primer 79F, 157F or 604F, using first-strand cDNA synthesis kit (iScriptTM select cDNA synthesis kit, Bio-Rad Laboratories). 1µl of cDNA was used as a template in real-time PCR analysis. iQ™ Eva® Green Supermix (Bio-Rad Laboratories) was used to set up real-time PCR reactions, which were run and analyzed on CFX96™ Real-Time System (Bio-Rad Laboratories). Conditions for real-time PCR were as follows: 95˚C for 3 min, followed by 54 cycles of 94˚C for 15 s, 60˚C for 20 s, 72˚C for 30 s. Melting curve analysis was performed from 65˚C to 95˚C with increments of 0.5˚C every 5 seconds. The specific qRT-PCR primer pairs were 79F and 185R, 157F and 405R, as well as 604F and 808R. The efficiency of the above primers was above 95%. Error bars represent standard deviation of three technical replicates. Additionally, two biological replicates were performed.

Primer list

Primer name	Primer sequence
CALI1.F’	5’ GTA ATT GCT TAA TTC CTT CTT TT 3’
CALI1.166R	5’ GTA GAC CCA AGT AAA TAG ATA CC 3’
CALI1.(152-166)+(244-259).F	5’ TTT ACT TGG GTC TAC AGG ACT TTT AGT GAA C 3’
CALI1.R’	5’ CTA GTA CCT TCT CCA TGC TAC AA 3’
CALI1F+TATA.F	5’ ATA TTT GTA GCG CAA GAC CC 3’
TATA.R	5’ GGT CCT CTC CAA ATG AAA TG 3’
79F	5’ CTT AAG GCA TAC TTT CTG TGT CTT C 3’
157F	5’ TTG GGT CTA CGA ACT GAT TGT 3’
604F	5’ GAG TCT TTG ACC TCT GGG GAT T 3’
185R	5’ TGA CCA ACA CAA TCA GTT CG 3’
405R	5’ CCT AGT CTA GAC CAT AAA CAT CCG T 3’

808R	5’ CCT AGC TAG AAA CAG TAC TGA 3’
LFY promoter.F	5’ TGA ATT TTG AAA CAG TAA TAA ATA GCT GAA 3’
LFY promoter.R	5’ TCG CTA TTT TTG CAA TAA AGC ATT TAT GTG 3’
TFL1 promoter.F	5’ GGG TTA TGT TTA TAA CTT GGG TAA GCA G 3’
TFL1 promoter.R	5’ GGG GTT TTC CGT CTT AGA GAG AGA G 3’
AP1 promoter.F	5’ TGC TCA TGA TCT CCA TAT ACA T 3’
AP1 promoter+partial 5’UTR.R	5’ GAG CTC AGA CTT TGG TAT GAA C 3’
New AP1 promoter.F	5’ GCT CAT GAT CTC CAT ATA CA 3’
New AP1 promoter.R	5’ GAA AAG CTA AAG CTG GTT TCT C 3’
CAL promoter.F	5’ AAT TTC TTC TTT CTT ACG TCG AGA T 3’
CAL promoter.R	5’ GAA AGA GTG CTT CTT TCA CTC TTT A 3’

Chapter 4: Silencing of *CAL* is inherited in a time-dependent manner

Introduction

Epigenetic inheritance possesses unique features. During gamete development, epigenetic marks are programmed and erased, but can be completely restored in germlines (Bourc'his and Voinnet, 2010). Therefore, epigenetic modifications can be passed from one generation to the next. *De novo* methyltransferases and maintenance methyltransferases prominently contribute to re-establishment of DNA methylation in germ cells. Particularly, unlike CHG and CG methylation, CHH methylation, abundant in plants, cannot be retained by any maintenance methyltransferase and has to be established by RdDM pathway, which recruits *de novo* methyltransferase DRM2 after DNA replication (Feng et al., 2010). Thus, 24 nt small RNAs involved in RdDM play a crucial role in the genome reprogramming and epigenetic heritance across generations. Here, I report a time-dependent epigenetic inheritance phenomenon, existing in *CAL* intron-mediated silenced plants. I characterized this time-dependent trait in both silenced phenotype percentage and DNA methylation level.

Result

A time-dependent silencing phenomenon

As shown in earlier chapters that the silencing effect appeared to decrease as generations increase. T1 plants showed higher degree of *cal-like* phenotype than T4

plants. To observe how the silencing signal is inherited, we followed the *cal-like* and *ap1-like* transgenic plants in the T2 generation. When the seeds collected from T1 lines were sowed about one week after harvest, *ap1-like* and *cal-like* lines segregated in a ratio approximately 1:1. This implied that not all *35S::CALI1* transgenic plants exhibited a *cal-like* phenotype perhaps due to low levels of transgene expression. Surprisingly, we accidentally observed that seeds from the same *cal-like* T1 lines, when planted three weeks later, segregated above 95% *cal-like* progeny. To further investigate this phenomenon, we repeated the transformation and followed T2 phenotypes of five *cal-like* T1 lines (T1-A, B, C, D, E) as well as three *ap1-like* T1 lines (T1-1, 2, 3) by planting the seeds at an interval of one week for nine weeks. As shown in Table 4.1, close to 100% of the progeny were *cal-like* in all the five *cal-like* lines starting from week 3. This novel phenomenon indicates an unknown time-dependent silencing mechanism in the seeds during the first three weeks of seed maturation. Oppositely, the progeny of the *ap1-like* T1 lines remained roughly at 50% *cal-like* over the nine-week time course. This indicated that some intrinsic mechanistic differences might exist between *cal-like* and *ap1-like* lines in how the silenced phenotype of their progeny is inherited.

Week	1	2	3	4	5	6	7	8	9
*cal-like** (n)	*58% (115)*	*59% (213)*	**94% (105)**	**91% (125)**	**93% (116)**	**98% (159)**	**98% (126)**	**98% (192)**	**99% (153)**
*ap1-like** (n)	77% (79)	53% (106)	35% (43)	73% (56)	53% (53)	51% (67)	64% (96)	55% (74)	49% (84)

Table 4.1. Percentage segregation of *cal-like* plants in T2 changes with seed storage time. *: two different class T1 plants (*cal-like* and *ap1-like*) both segregated *cal-like* T2 plants. (n) represent total number of T2 progeny scored from 5 independent transgenic lines for *cal-like* row and 3 transgenic lines for *ap1-like* plants. (done by Boyana Grigorova)

One possibility may lie in the relatively low transgene expression in the T1 *ap1-like* lines, which is insufficient to trigger RdDM. However, in the T2 progeny of the *ap1-like* lines, the phenotype was determined by transgene segregation in which the doubling of transgene copies may boost the expression of the transgene and the resulting siRNA above the threshold levels.

To further verify the above time-dependent silencing phenomenon, we took a closer look at the seed development. Before seeds are fully matured, the siliques (seed pod) are green in color. When seeds are further matured, siliques become yellow. By the time the seeds are completely matured and dessicated, the siliques appear brown. As the color of the silique (seed pod) can serve as an indicator of seed maturation, a modified method was developed to confirm the time course of silencing related to seed age. We harvested the seeds from four siliques in green, yellow, and brown color, respectively, from 5 individual T2 *cal-like* plants. The seeds in green siliques

are youngest, and the seeds in brown ones are oldest. After 3-day air dry, we sowed the seeds at the same time, and documented the phenotype of the resulting plants. First, the seeds from different colored siliques germinated equally well. Second, plants derived from green siliques displayed 60% with *cal-like* phenotype and 17% with *weak-cal-like* phenotype. However, plants originated from brown siliques exhibited 94% *strong-cal-like* phenotype (Table 4.2). The number of plants from yellow siliques showed *cal-like* phenotype at a ratio between that of green and brown siliques.

Seeds from siliques in different color	*cal-like* phenotype	*weak-cal-like* phenotype	*ap1-like* phenotype
Green	144 (60.25%)	42 (17.57%)	53 (22.18%)
Yellow	241 (91.98%)	11 (4.20%)	10 (3.82%)
Brown	224 (94.12%)	0 (0%)	14 (5.88%)

Table 4.2. Plants originating from different aged seeds showed the different proportion of silenced phenotype.

Direct visualization of time-dependent silencing in seeds

To more precisely measure seed developmental time required for silencing and directly visualize the time-dependent silencing in seeds, we fused the entire genomic *CAL* locus from the START codon to right before the STOP codon in frame and upstream of the *GUS* reporter (*35S::gCAL-GUS*). The plants with transgenic *35S::gCAL-GUS* should express the GUS in most tissues including seeds. However, if the *gCAL* of *35S::gCAL-GUS* is silenced, the expression of *GUS* will be absent. Therefore, if the silencing increases during seed development, the number of *GUS*

positive seeds would decrease. Based on this rationale, *35S::gCAL-GUS* was transformed into *cal-like* plants that maintain the silencing signals. The seeds from five independent T1 transgenic lines (*35S::gCAL-GUS; cal-like*) were pooled and stained with X-Gluc (Figure 4.1). To rule out the possibility that pre-matured seeds are easier to stain than matured seeds, we fused *CAL cDNA* with *GUS* gene in frame (*35S::cCAL-GUS*), in which the transgene lacks *CAL* introns and shouldn't be subject to intron-mediated silencing. This *35S::cCAL-GUS* was also transformed into *cal-like* lines. Even though the silencing signals are still present in host plants, the *35S::cCAL-GUS* shouldn't be subject to silencing and the ratio of GUS-stained seeds should stay constant over the time course of seed maturation. The *35S::cCAL-GUS* transgenic plants are still growing at the moment and will be stained in a few weeks.

For *35S::gCAL-GUS;cal-like* transgenic plants, as the siliques pass from green to yellow and finally to brown, the GUS-staining positive (blue) seeds decreased from 15% to 1.2%, and ultimately to 0% (Figure 4.1). This indicated that silencing of the *35S::gCAL-GUS* increases with seeds' age. *35S::cCAL-GUS; cal-like* control lines are still being examined at this moment. In addition, our data demonstrated that once the *CAL* gene is silenced in seeds, this silenced status is maintained mitotically all the way from vegetative to reproductive phases.

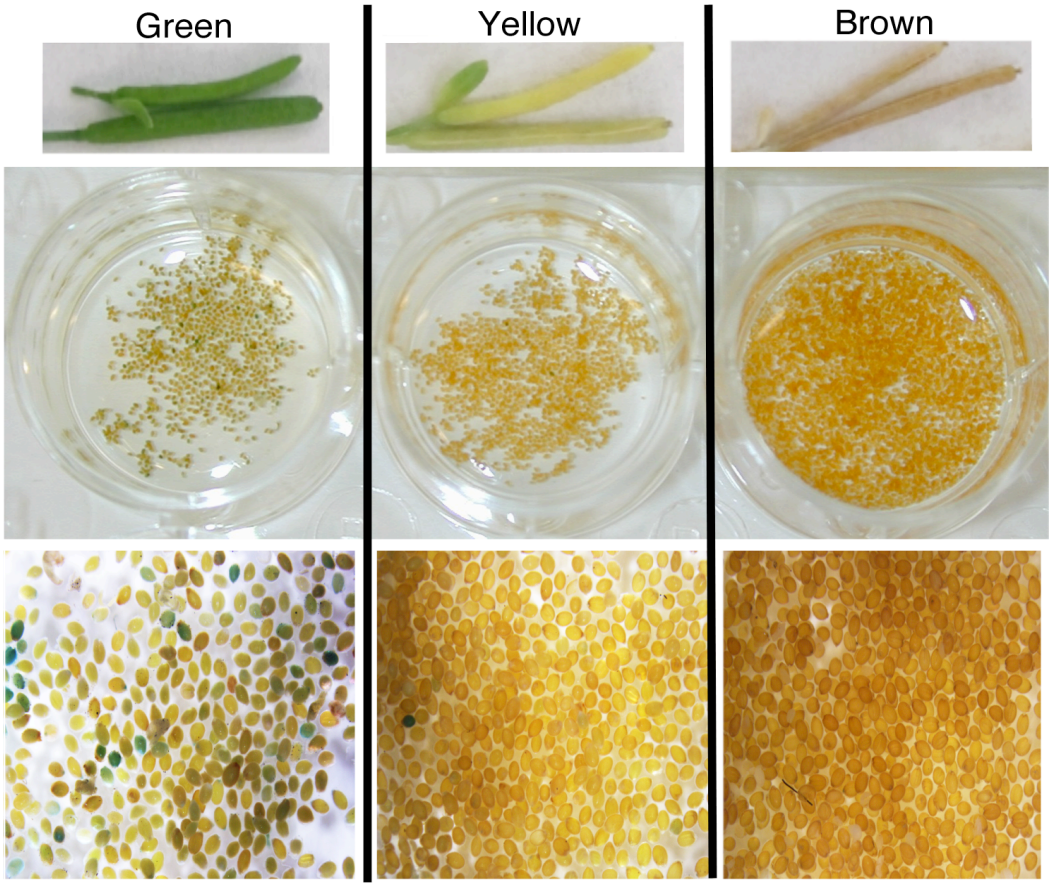


Figure 4.1. Time-dependent gene silencing in seeds. Seeds dissected from green, yellow, and brown silliques were stained with X-Gluc to illustrate the decreasing GUS-positive seeds as seeds age.

The time-dependent silencing may result from a period of time required to methylate newly synthesized DNA in embryos

To further investigate this time-dependent silencing phenomenon, we investigated DNA methylation state of embryos at three different stages using McrBC digestion. 100 embryos at heart stage and torpedo stage were dissected from fresh green seeds of 12 individual *cal-like* plants. Another 100 embryos at the bending cotyledon stage were dissected also from green seeds. Finally, 100 embryos at completely mature white stage were isolated from three-month old brown seeds of *cal-like* plants (Figure

4.3). The genomic DNA extracted from those embryos was treated by McrBC enzyme, and then subject to qPCR using the primers that specifically amplify the #35-#278 region of *CAL* first intron. This is the region that our previous McrBC digestion detected DNA methylation in *cal-like* plants (Figure 2.8). The experimental scheme is illustrated in Figure 4.2. Sensitivity to McrBC was detected by a significant reduction of PCR product in the #35-#278 nt region of *CAL* first intron, when compared with McrBC untreated controls in embryos from these *cal-like* plants (Figure 4.3). Moreover, the McrBC sensitivity increases when the embryos age increases as indicated by a decreased PCR products. This data supports that the time-dependent silencing may be due to the amount of time needed during embryo development to methylate DNA.

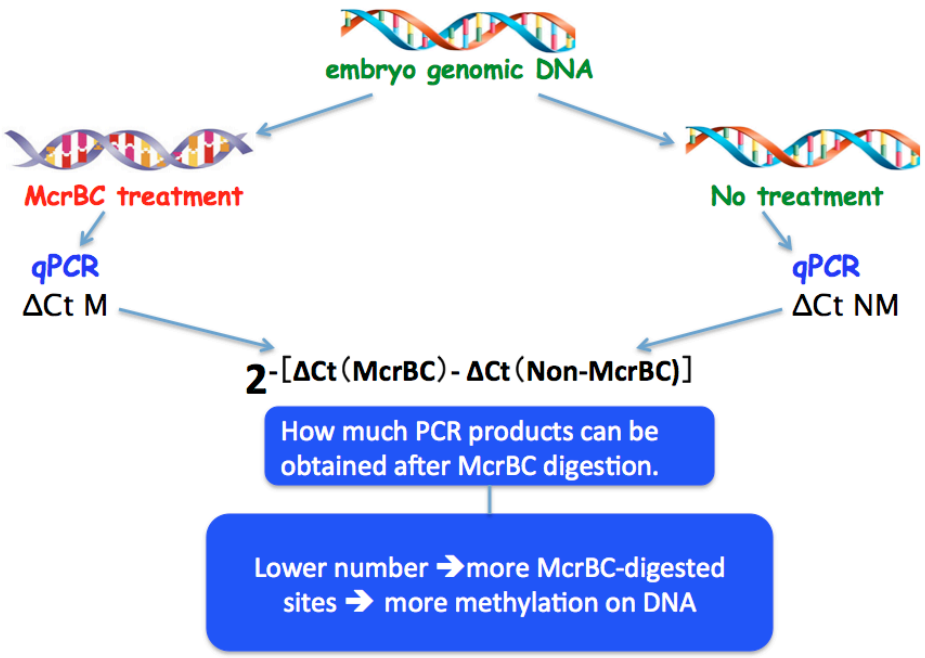


Figure 4.2. Scheme of the process of McrBC-qPCR for genomic DNA from embryos in different developmental stages.

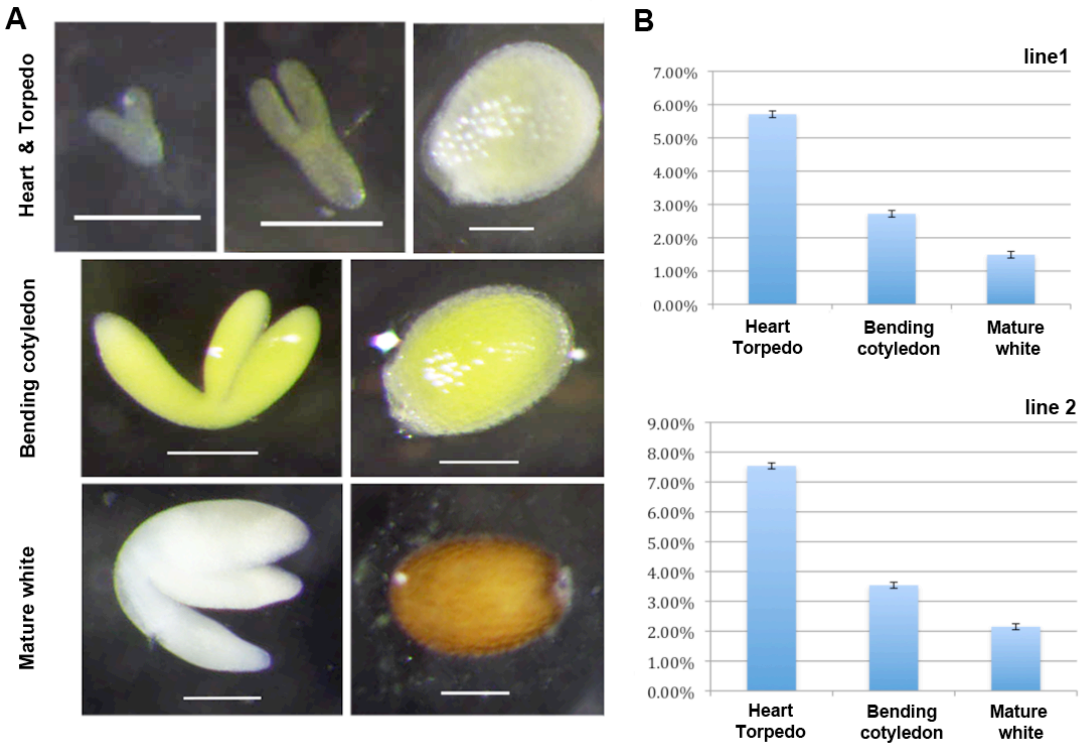


Figure 4.3. Developmental time-dependent DNA methylation in embryos. (A) Photograph of different stage embryos dissected from corresponding seeds. Three different stage embryo samples used in the study are heart and torpedo embryos, bending cotyledon embryos, and mature white embryos. (B) qPCR quantification of embryo DNA samples after McrBC treatment. Amplification cycles of embryo DNA after McrBC enzyme treatment was compared with those of embryo DNA before McrBC treatment. The DNA amount after McrBC treatment relative to DNA amount before McrBC treatment was expressed as % and plotted aginst the Y axis. The lower the %, the less DNA template, which would indicate more McrBC diggestion and more methylation. Two separate biological experiments using embryos collected from different transgenic lines are shown.

Inheritance of the silencing depends on the inheritance of the transgene

If most of the transgenic lines harbor a single copy of transgene, then about 75% of T2 progeny should possess the transgene while the other 26% of the T2 progeny should not inherit the transgene. However, our data revealed close to 100% *cal-like* phenotype in the T2 progeny of *cal-like* lines. One exciting possibility is that some T2

cal-like plants did not inherit the *35S::CALI1* transgene yet inherited the silencing signal, perhaps in the form of 24 nt siRNA. As a result these T2 plants showed close to 100% *cal-like* phenotype. To test this possibility, we extracted genomic DNA from 300 T2 plants, all of which exhibited the *cal-like* phenotype, from 9 independent T1 *cal-like* lines. Subsequently, we genotyped the presence of the *35S::CALI1* transgene in each of the 300 T2 plants. However, all the tested T2 *cal-like* plants harbored the transgene, indicating that *35S::CALI1* transgene is necessary to induce the silenced *cal-like* phenotype.

Discussion

Time-dependent methylation in embryo development may represent a common yet undiscovered phenomenon.

During early development of zygote and embryogenesis, the epigenetic marks inherited from parent generation are gradually established primarily through maintenance methyltransferases and 24 nt siRNAs-mediated RdDM pathway (Bourchis and Voinnet, 2010). Then the embryo has to undergo thousands of rounds of mitosis during development. As previously described in the introduction, once de novo methylation is established in *Arabidopsis* genome, symmetric CG and CHG methylation is prominently maintained by MET1 and CMT3, respectively (Zhang and Zhu, 2011). However, CHH methylation requires de novo methylation catalyzed by DRM2 that is recruited by RdDM machinery, after each round of DNA replication (Zhang and Zhu, 2011). In addition, a recent research on reprogramming of DNA methylation in pollen discovered that DRM2 was expressed at a very low level in the

sperm cells of matured pollens, implying that male gametes have a limited capability for de novo CHH methylation, and thereby maternal 24 nt siRNAs might be required to restore CHH methylation for the retrotransposons from sperm (Calarco et al., 2012). Although the expression level of DRM2 during the process of embryo maturation remains unknown, it cannot be ruled out the possibility that DRM2 expression may increase as time progresses, so that establishment of CHH methylation maybe dependent on gradual up-regulation of DRM2. Hence, when seeds are more mature, CHH methylation predominantly present in *CAL* first intron loci is more complete, and the offspring plants with a silenced phenotype are more prevalent.

Alternatively the time-dependent methylation during embryo maturation merely reflects the time required to maintain CG methylation by MET1 when rapid cell division in embryos couldn't be caught up by the maintenance methylation. The discovery of the time-dependent phenomenon may have important implications in agrobiotech. Transgenic seeds may need to be stored long enough to exhibit strongest effect by the silencing transgene.

Inheritance of transgene correlates with the inheritance of silenced phenotype

Ideally, transgenic lines only harbor a single copy of transgene. After self-fertilization, 25% of the T2 (second generation) plants should not inherit the transgene and thus may display *ap1-like* phenotype. However, in most of our *cal-like* lines, almost 100% T2 progeny exhibit *cal-like* phenotype. Extensive genotyping

showed that all of the T2 *cal-like* plants have inherited the transgene. One interpretation is that the efficiency of transformation was very high so that each transgenic T1 plant has obtained two or more independent transgenes. Hence, most of T2 plants contained at least one copy of the transgene and displayed the silenced phenotype.

Methods

McrBC-qPCR for embryo genomic DNA

Different aged siliques from 12 individual T3 plants that are decedents of, T1-7/T2-2 (T1-family #7 and T2 family #2) as well as 12 T3 plants from another transgenic line (T1-5, T2-26) were respectively collected. Both T2-2 and T2-26 are homozygous *cal-like* lines. Individual embryos were dissected from seeds and sorted into heart / torpedo, bending cotyledon and fully mature categories. For each category, approximately 200 embryos were acquired for genomic DNA extraction by DNeasy kit (Qiagen). 100 ng of embryo genomic DNA was digested with 20 units of McrBC (New England Biolabs) in 30 ul volume for 1 hour based on the manufacturer's instruction. 1 ul of McrBC-digested gDNA served as template for qPCR. The equal quantities of non-digested embryo gDNAs were used as control PCR templates. iQ™ Eva® Green Supermix (Bio-Rad Laboratories) was used in qPCR on CFX96™ Real-Time System (Bio-Rad Laboratories). qPCR Conditions were as follows: 95˚C for 3 min, followed by 40 cycles of 94˚C for 15 s, 60˚C for 20 s, 72˚C for 30 s. Melting curve analysis was performed from 65˚C to 95˚C with increments of 0.5˚C every 5 seconds. *CAL* first intron specific qPCR primers are CALI1.F1 and CALI1.R1, and

the corresponding primer efficiency was 97.2%. The housekeeping gene *HOBBIT* (*HBT, At2g20000*) was used as a reference gene. The specific primers for *HBT* are HBT.qPCR.F and HBT.qPCR.R, and the efficiency was 98.7%. The Pfaffl formula $2^{-\Delta\Delta Ct}$ was used to calculate relative gene expression differences. ΔC_t^{McrBC} equals $C_t^{McrBC} - C_t^{HBT}$. Correspondingly, $\Delta C_t^{non\text{-}McrBC} = C_t^{non\text{-}McrBC} - C_t^{HBT}$. $\Delta\Delta C_t$ was calculated as $\Delta C_t^{McrBC} - \Delta C_t^{Non\text{-}McrBC}$. The ΔΔCT number is entered into the Pfaffl formula ($2^{-\Delta\Delta Ct}$) to yield "fold difference" between non-McrBC-treated and McrBC-treated embryo gDNA. This indicates qPCR products obtained after McrBC digestion. The lower the value of $2^{-\Delta\Delta Ct}$, the more methylation of embryo gDNA. Error bars represent standard deviation of three technical replicates.

GUS reporter constructs and GUS staining

The *35S::gCAL-GUS* construct was generated as follows. Primers gCAL.F and gCAL.R were used to amplify the 3178 bp *CAL* genomic locus from the START codon ATG to just before the STOP codon, including all introns and exons. The LongAmp Taq PCR kit (NEB) was used in PCR. The PCR fragment was cloned into *pCR8/GW/TOPO* using TA cloning kit (Invitrogen). The *gCAL* was inserted into the gateway binary vector *pMDC140* (Curtis and Grossniklaus, 2003), fusing in frame *CAL* N-terminus to the *GUS* reporter. The *Agrobacteria tumefaciens GV3101* containing the construct was transformed into the *cal-like* plants. About 36 transgenic lines were generated. Seeds from 8 T1 individuals were collected at different ages and stained with X-Gluc according to a previous protocol (Mara et al., 2010).

CAL cDNA was amplified by Phusion Taq (NEB) using template cDNAs converted from total RNA of *wild type Ler* inflorescences. The primers cCAL.F and cCAL.R were used for amplifying *CAL* cDNA, which is 765 bp. *35S::cCAL-GUS* was similarly constructed and transformed into *cal-like* plants as was described for *35S::gCAL-GUS*.

Primer list

Primer name	Primer sequence
CALI1.F1	5’ TTT TAG TGT GCC TTC GTT TGC C 3’
CALI1.R1	5’ AAG TAG ATC GTA TTG TAG GGT TCA C 3’
HBT.qPCR.F	5’ CAG GCC ATT ACC TTC TTG GA 3’
HBT.qPCR.R	5’ CCT CAT ATG CAG CCC AAA GT 3’
gCAL.F	5’ ATG GGA AGG GGT AGG GTT GAA TTG AAG AGG A 3’
gCAL.R	5’ AGC GGC GTA ACA GCC AAG GTA ATT GTA AAT 3’
cCAL.F	5’ ATG GGA AGG GGT AGG GTT GAA TTG AAG AGG A 3’
cCAL.R	5’ AGC GGC GTA ACA GCC AAG GTA ATT GTA AAT 3’

Chapter 5: Conclusion, discussion, and future direction

My study showed that the over-expression of *CAL* first intron triggers silencing of the endogenous *CAL* gene. This is because the sense intron RNA transcribed from the transgene can pair with the antisense noncoding RNA from the endogenous *CAL* first intron. The resulting dsRNAs led to 24nt siRNAs that direct the methylation on the endogenous *CAL* first intron. This results in the chromatin closure and reduction of *CAL* gene expression.

I showed that the *CAL* first intron contains *cis*-regulatory elements. I proposed that the intronic *cis*-regulatory elements may interact with *CAL* promoter to generate intron-derived antisense lncRNAs. This may underlie the ability of *CAL* first intron in sense orientation, when over-expressed, to silence the endogenous *CAL* gene. In addition, I characterized the time–dependent silencing phenomenon, which concerns the inheritance of the silencing. Specifically, longer seed storage time enhanced the expression of the silenced phenotype. I further showed that increased seed storage time correlated with increased methylation of the *CAL* first intron.

Two striking observations are discussed here. The first concerns the *CAL* first intron derived-lncRNAs. Although numerous lncRNAs have been mapped in transcriptome data in eukaryotic organisms (Kim and Sung, 2012), the functional roles of most of them still remain elusive. The possible molecular mechanisms of lncRNA function are summarized in four major predictive models (Wang and Chang, 2011; Rinn and Chang, 2012). First, as decoys, lncRNAs could distract transcription factors or other DNA-binding proteins. Second, as scaffolds, lncRNAs could assemble relevant

molecular components to form a functional complex, or stabilize spatial nuclear structures and signaling complexes. Third, as guide, lncRNAs can directly recruit chromatin modification enzyme to genes, either in *cis* or *trans*, through RNA-DNA or RNA-DNA binding protein interactions. Fourth, as signal enhancer, lncRNAs could loop chromosome through an enhancer-like model, to exert gene regulation in space and time (Wang and Chang, 2011; Rinn and Chang, 2012). My detection of naturally existing antisense lncRNAs from *CAL* intron indicate that the intron-derived lncRNAs may either play an important role during normal flower development or simply be produced as a by-products. I favored the first possibility and proposed that the *cis*-elements (enhancer) in the *CAL* first intron may communicate with the *CAL* promoter, for example, by attracting RNA Polymerase II, which then glide toward the promoter in anti-sense orientation. The intron-derived antisense RNAs are the products during the gliding of Pol II. When *CAL* transcription is highly active, the antisense lncRNAs is also produced abundantly. In *wild type* plants, the intron sequence of the primary mRNA or intron lariat after splicing may base pair with these lncRNAs to form dsRNAs, and the produced 24 nt siRNAs can target and gradually build up methylation on the endogenous *CAL* intron region. Hence, *CAL* first intron and intron-derived lncRNAs may regulate *CAL* expression in a negative feedback way.

Thus, qRT-PCR should be carried out to examine and compare the *CAL* first intron-derived antisense lncRNAs in the total RNA extracted from young inflorescences, mature flowers and leaves. In addition, the bisulfite sequencing should be utilized to detect the methylation status and degrees of methylation of endogenous *CAL* first

intron DNA extracted from different aged flowers and vegetative tissues. This discovery could expand our understanding of the role of lncRNAs in the epigenetic regulation of gene expression.

The second striking finding is the time-dependent silencing phenomenon. As we discussed in Chapter 4, newly replicated DNAs may require time to build up methylation directed by 24 nt siRNAs. In plant germlines, transposable elements undergo reprogramming in which methylation is erased in gametes and restored in embryos through epigenetics mechanisms, such as RNA mediated DNA methylation and chromatin remodeling. Especially, maternal 24 nt siRNAs guide the restoration of CHH methylation in seeds (Slotkin et al, 2009; Calarco et al, 2012). In addition, the reprogramming of imprinted genes also requires 24 nt siRNAs in sperm cells (Slotkin et al, 2009; Calarco et al, 2012). Epialleles may also acquire 24 nt siRNAs originated from vegetative nucleus to establish methylation in embryos (Slotkin et al, 2009; Calarco et al, 2012). Therefore, seeds may need to be stored long enough to exhibit full silencing of transposons or full effect of imprinted genes and epialleles can be realized only after extended storage. To integrate this conclusion, I will do McrBC-qPCR on embryo genomic DNA by using transposon-specific primers to test the methylation extent in different aged embryos.

Reference

Ahmad, K., and Henikoff, S. (2002). The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. Molecular cell 9, 1191-1200.

Allen, R.S., Li, J., Stahle, M.I., Dubroue, A., Gubler, F., and Millar, A.A. (2007). Genetic analysis reveals functional redundancy and the major target genes of the Arabidopsis miR159 family. Proceedings of the National Academy of Sciences of the United States of America 104, 16371–16376.

Arteaga-Vazquez, M., Sidorenko, L., Rabanal, F.A., Shrivistava, R., Nobuta, K., Green, P.J., Meyers, B.C., and Chandler, V.L. (2010). RNA-mediated trans-communication can establish paramutation at the b1 locus in maize. Proceedings of the National Academy of Sciences of the United States of America 107, 12986-12991.

Arteaga-Vazquez, M.A., and Chandler, V.L. (2010). Paramutation in maize: RNA mediated trans-generational gene silencing. Current opinion in genetics & development 20, 156-163.

Aukerman, M.J., and Sakai, H. (2003). Regulation of flowering time and floral organ identity by a MicroRNA and itsAPETALA2-like target genes. Plant Cell 15, 2730–2741.

Bai, X., Peirson, B.N., Dong, F., Xue, C., and Makaroff, C.A. (1999). Isolation and characterization of SYN1, a RAD21-like gene essential for meiosis in Arabidopsis. The Plant cell 11, 417-430.

Barrett, L.W., Fletcher, S., and Wilton, S.D. (2012). Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. Cellular and molecular life sciences : CMLS 69, 3613-3634.

Becker, C., and Weigel, D. (2012). Epigenetic variation: origin and transgenerational inheritance. Current opinion in plant biology 15, 562-567.

Bollman, K.M., Aukerman, M.J., Park, M.Y., Hunter, C., Berardini, T.Z., and Poethig, R.S. (2003). HASTY, the Arabidopsis ortholog of exportin 5/MSN5, regulates phase change and morphogenesis. Development 130, 1493-1504.

Bourc'his, D., and Voinnet, O. (2010). A small-RNA perspective on gametogenesis, fertilization, and early zygotic development. Science 330, 617-622.

Bowman, J.L., Alvarez, J., Weigel, D., Meyerowitz, E.M., and Smyth. D.R. (1993). Control of flower development in *Arabidopsis thaliana* by APETALA1 and interacting genes. Development 119, 721-743.

Bowman, J.L., Drews, G.N., and Meyerowitz, E.M. (1991). Expression of the Arabidopsis floral homeotic gene AGAMOUS is restricted to specific cell types late in flower development. The Plant cell 3, 749-758.

Bowman, J.L., Smyth, D.R., and Meyerowitz, E.M. (1989). Genes directing flower development in Arabidopsis. The Plant cell 1, 37-52.

Burton, N.O., Burkhart, K.B., and Kennedy, S. (2011). Nuclear RNAi maintains heritable gene silencing in Caenorhabditis elegans. Proceedings of the National Academy of Sciences of the United States of America 108, 19683-19688.

Busch, M.A., Bomblies, K., and Weigel, D. (1999). Activation of a floral homeotic gene in Arabidopsis. Science 285, 585-587.

Calarco, J.P., Borges, F., Donoghue, M.T., Van Ex, F., Jullien, P.E., Lopes, T., Gardner, R., Berger, F., Feijo, J.A., Becker, J.D., et al. (2012). Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. Cell 151, 194-205.

Carthew, R.W., and Sontheimer, E.J. (2009). Origins and Mechanisms of miRNAs and siRNAs. Cell 136, 642-655.

Chapman, E.J., and Carrington, J.C. (2007). Specialization and evolution of endogenous small RNA pathways. Nature reviews Genetics 8, 884-896.

Chen, D., Meng, Y., Yuan, C., Bai, L., Huang, D., Lv, S., Wu, P., Chen, L.L., and Chen, M. (2011). Plant siRNAs from introns mediate DNA methylation of host genes. RNA 17, 1012-1024.

Chen, X. (2004). A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development. Science 303, 2022–2025.

Chen, X. (2008). MicroRNA metabolism in plants. Current topics in microbiology and immunology 320, 117-136.

Clancy, S. (2008). RNA splicing: introns, exons and spliceosome. Nature Education 1, 1.

Curtis, M.D., and Grossniklaus, U. (2003). A gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant physiology 133, 462-469.

Deyholos, M.K., and Sieburth, L.E. (2000). Separable whorl-specific expression and negative regulation by enhancer elements within the AGAMOUS second intron. The Plant cell 12, 1799-1810.

Ding, Y., Tao, Y., and Zhu, C. (2013). Emerging roles of microRNAs in the mediation of drought stress response in plants. Journal of Experimental Botany 64, 3077-3086.

Dobi, K.C., and Winston, F. (2007). Analysis of transcriptional activation at a distance in Saccharomyces cerevisiae. Molecular and cellular biology 27, 5575-5586.

Drews, G.N., Bowman, J.L., and Meyerowitz, E.M. (1991). Negative regulation of the Arabidopsis homeotic gene AGAMOUS by the APETALA2 product. Cell 65, 991-1002.

Earley, K.W., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K., and Pikaard, C.S. (2006). Gateway-compatible vectors for plant functional genomics and proteomics. The Plant journal : for cell and molecular biology 45, 616-629.

Eddy, J., and Maizels, N. (2008). Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. Nucleic acids research 36, 1321-1333.

Feng, S., Jacobsen, S.E., and Reik, W. (2010). Epigenetic reprogramming in plant and animal development. Science 330, 622-627.

Ferrandiz, C., Gu, Q., Martienssen, R., and Yanofsky, M.F. (2000). Redundant regulation of meristem identity and plant architecture by FRUITFULL, APETALA1 and CAULIFLOWER. Development 127, 725-734.

Finnegan, E.J., and Matzke, M.A. (2003). The small RNA world. Journal of Cell Science 116, 4689–4693.

Foerster, A.M., Hetzl, J., Mullner, C., and Mittelsten Scheid, O. (2010). Analysis of bisulfite sequencing data from plant DNA using CyMATE. Methods Mol Biol 631, 13-22.

Fuks, F. (2005). DNA methylation and histone modifications: teaming up to silence genes. Current opinion in genetics & development 15, 490-495.

Gendrel, A.V., Lippman, Z., Yordan, C., Colot, V., and Martienssen, R.A. (2002). Dependence of heterochromatic histone H3 methylation patterns on the Arabidopsis gene DDM1. Science 297, 1871-1873.

Gonzalez, M., and Li, F. (2012). DNA replication, RNAi and epigenetic inheritance. Epigenetics : official journal of the DNA Methylation Society 7, 14-19.

Gowher, H., Leismann, O., and Jeltsch, A. (2000). DNA of Drosophila melanogaster contains 5-methylcytosine. The EMBO journal 19, 6918-6923.

Gu, S.G., Pak, J., Guang, S., Maniar, J.M., Kennedy, S., and Fire, A. (2012). Amplification of siRNA in Caenorhabditis elegans generates a transgenerational sequence-targeted histone H3 lysine 9 methylation footprint. Nature genetics 44, 157-164.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell 130, 77-88.

Guil, S., and Esteller, M. (2012). *Cis*-acting noncoding RNAs: friends and foes. Nature structural & molecular biology 19, 1068-1075.

Guo, H.S., Xie, Q., Fei, J.F., and Chua, N.H. (2005). MicroRNA directs mRNA cleavage of the transcription factor *NAC1* to downregulate auxin signals for *Arabidopsis* lateral root development. Plant Cell 17, 1376–1386.

Gustafson-Brown, C., Savidge, B., and Yanofsky, M.F. (1994). Regulation of the arabidopsis floral homeotic gene APETALA1. Cell 76, 131-143.

Haag, J.R., Ream, T.S., Marasco, M., Nicora, C.D., Norbeck, A.D., Pasa-Tolic, L., and Pikaard, C.S. (2012). In vitro transcription activities of Pol IV, Pol V, and RDR2

reveal coupling of Pol IV and RDR2 for dsRNA synthesis in plant RNA silencing. Molecular cell 48, 811-818.

Hauser, M.T., Aufsatz, W., Jonak, C., and Luschnig, C. (2011). Transgenerational epigenetic inheritance in plants. Biochimica et biophysica acta 1809, 459-468.

Helliwell, C.A., Robertson, M., Finnegan, E.J., Buzas, D.M., and Dennis, E.S. (2011). Vernalization-repression of Arabidopsis FLC requires promoter sequences but not antisense transcripts. PloS one 6, e21513.

Henikoff, S., Furuyama, T., and Ahmad, K. (2004). Histone variants, nucleosome assembly and epigenetic inheritance. Trends in genetics : TIG 20, 320-326.

Heo, J.B., and Sung, S. (2011). Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. Science 331, 76-79.

Herr, A.J. (2005). Pathways through the small RNA world of plants. FEBS letters 579, 5879-5888.

Hetzl, J., Foerster, A.M., Raidl, G., and Mittelsten Scheid, O. (2007). CyMATE: a new tool for methylation analysis of plant genomic DNA after bisulphite sequencing. The Plant journal : for cell and molecular biology 51, 526-536.

Ho, Y., Elefant, F., Liebhaber, S.A., and Cooke, N.E. (2006). Locus control region transcription plays an active role in long-range gene activation. Molecular cell 23, 365-375.

Holeski, L.M., Jander, G., and Agrawal, A.A. (2012). Transgenerational defense induction and epigenetic inheritance in plants. Trends in ecology & evolution 27, 618-626.

Hong, R.L., Hamaguchi, L., Busch, M.A., and Weigel, D. (2003). Regulatory elements of the floral homeotic gene AGAMOUS identified by phylogenetic footprinting and shadowing. The Plant cell 15, 1296-1309.

Huettel, B., Kanno, T., Daxinger, L., Bucher, E., van der Winden, J., Matzke, A.J., and Matzke, M. (2007). RNA-directed DNA methylation mediated by DRD1 and Pol IVb: a versatile pathway for transcriptional gene silencing in plants. Biochimica et biophysica acta 1769, 358-374.

Irish, V.F., and Sussex, I.M. (1990). Function of the *apetala-1* gene during Arabidopsis floral development. The Plant cell 2, 741-753.

Isshiki, M., Yamamoto, Y., Satoh, H., and Shimamoto, K. (2001). Nonsense-mediated decay of mutant waxy mRNA in rice. Plant physiology 125, 1388-1395.

Jackson, J.P., Lindroth, A.M., Cao, X., and Jacobsen, S.E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature 416, 556-560.

Jeon, Y., and Lee, J.T. (2011). YY1 tethers Xist RNA to the inactive X nucleation center. Cell 146, 119-133.

Jin, J., Liu, J., Wang, H., Wong, L., and Chua, N.H. (2013). PLncDB: plant long non-coding RNA database. Bioinformatics 29, 1068-1071.

Johnson, L., Cao, X., and Jacobsen, S. (2002). Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. Current biology : CB 12, 1360-1367.

Juranek, S.A., and Paeschke, K. (2012). Cell cycle regulation of G-quadruplex DNA structures at telomeres. Current pharmaceutical design 18, 1867-1872.

Kalyna, M., Simpson, C.G., Syed, N.H., Lewandowska, D., Marquez, Y., Kusenda, B., Marshall, J., Fuller, J., Cardle, L., McNicol, J., et al. (2012). Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis. Nucleic acids research 40, 2454-2469.

Kempin, S.A., Savidge, B., and Yanofsky, M.F. (1995). Molecular basis of the cauliflower phenotype in Arabidopsis. Science 267, 522-525.

Kertesz, S., Kerenyi, Z., Merai, Z., Bartos, I., Palfy, T., Barta, E., and Silhavy, D. (2006). Both introns and long 3'-UTRs operate as *cis*-acting elements to trigger nonsense-mediated decay in plants. Nucleic acids research 34, 6147-6157.

Khaitova, L.C., Fojtova, M., Krizova, K., Lunerova, J., Fulnecek, J., Depicker, A., and Kovarik, A. (2011). Paramutation of tobacco transgenes by small RNA-mediated transcriptional gene silencing. Epigenetics : official journal of the DNA Methylation Society 6, 650-660.

Kim, D.H., and Sung, S. (2012a). Environmentally coordinated epigenetic silencing of FLC by protein and long noncoding RNA components. Current opinion in plant biology 15, 51-56.

Kim, E.D., and Sung, S. (2012b). Long noncoding RNA: unveiling hidden layer of gene regulatory networks. Trends in plant science 17, 16-21.

Lahmy, S., Bies-Etheve, N., and Lagrange, T. (2010). Plant-specific multisubunit RNA polymerase in gene silencing. Epigenetics : official journal of the DNA Methylation Society 5, 4-8.

Le, B.H., Cheng, C., Bui, A.Q., Wagmaister, J.A., Henry, K.F., Pelletier, J., Kwong, L., Belmonte, M., Kirkbride, R., Horvath, S., et al. (2010). Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. Proceedings of the National Academy of Sciences of the United States of America 107, 8063-8070.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to *lin-14*. Cell 75, 843-854.

Lewis, B.P., Green, R.E., and Brenner, S.E. (2003). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proceedings of the National Academy of Sciences of the United States of America 100, 189-192.

Liu, X., Huang, J., Wang, Y., Khanna, K., Xie, Z., Owen, H.A. and Zhao, D. (2010). The role of floral organs in carpels, an *Arabidopsis* loss-of-function mutation in MicroRNA160a, in organogenesis and the mechanism regulating its expression. The Plant Journal 62, 416–428.

Liu, Z., and Mara, C. (2010). Regulatory mechanisms for floral homeotic gene expression. Seminars in cell & developmental biology 21, 80-86.

Lohmann, J.U., Hong, R.L., Hobe, M., Busch, M.A., Parcy, F., Simon, R., and Weigel, D. (2001). A molecular link between stem cell regulation and floral patterning in Arabidopsis. Cell 105, 793-803.

Lugones, L.G., Scholtmeijer, K., Klootwijk, R., and Wessels, J.G. (1999). Introns are necessary for mRNA accumulation in Schizophyllum commune. Molecular microbiology 32, 681-689.

Machida, S., Chen, H.Y., and Adam, Y.Y. (2011). Molecular insights into miRNA processing by *Arabidopsis thaliana* SERRATE. Nucleic acids Research 39, 7828-7836.

Maenner, S., Blaud, M., Fouillen, L., Savoye, A., Marchand, V., Dubois, A., Sanglier-Cianferani, S., Van Dorsselaer, A., Clerc, P., Avner, P., et al. (2010). 2-D structure of the A region of Xist RNA and its implication for PRC2 association. PLoS biology 8, e1000276.

Mallory, A.C., Bartel, D.P., and Bartel, B. (2005). MicroRNA-directed regulation ofArabidopsis AUXIN RESPONSE FACTOR17 is essential for proper development and modulates expression of early auxin response genes. Plant Cell 17, 1360–1375.

Maquat, L.E. (2004). Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. Nature reviews Molecular cell biology 5, 89-99.

Masternak, K., Peyraud, N., Krawczyk, M., Barras, E., and Reith, W. (2003). Chromatin remodeling and extragenic transcription at the MHC class II locus control region. Nature immunology 4, 132-137.

Matsui, A., Ishida,J., Morosawa, T., Mochizuki, Y., Kaminuma, E., Endo, T.A., Okamoto, M., Nambara, E., Nakajima, M., Kawashima, M., Satou, M., Kim J., Kobayashi, N., Toyoda, T., Shinozaki, K., and Seki, K. (2008). Arabidopsis Transcriptome Analysis under Drought, Cold, High-Salinity and ABA Treatment Conditions using a Tiling Array. Plant and cell physiology 49, 1135-1149.

Matzke, M., Kanno, T., Huettel, B., Daxinger, L., and Matzke, A.J. (2007). Targets of RNA-directed DNA methylation. Current opinion in plant biology 10, 512-519.

Mauger, D.M., Siegfried, N.A., and Weeks, K.M. (2013). The genetic code as expressed through relationships between mRNA structure and protein function. FEBS letters 587, 1180-1188.

McGlincy, N.J., and Smith, C.W. (2008). Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? Trends in biochemical sciences 33, 385-393.

McManus, C.J., and Graveley, B.R. (2011). RNA structure and the mechanisms of alternative splicing. Current opinion in genetics & development 21, 373-379.

Molnar, A., Melnyk, C., and Baulcombe, D.C. (2011). Silencing signals in plants: a long journey for small RNAs. Genome biology 12, 215.

Molnar, A., Schwach, F., Studholme, D.J., Thuenemann, E.C., and Baulcombe, D.C. (2007). miRNAs control gene expression in the single-cell alga Chlamydomonas reinhardtii. Nature 447, 1126-1129.

Montgomery, T.A., Yoo, S.J., Fahlgren, N., Gilbert, S.D., Howell, M.D., Sullivan, C.M., Alexander, A., Nguyen, G., Allen, E., Ahn, J.H., et al. (2008). AGO1-miR173 complex initiates phased siRNA formation in plants. Proceedings of the National Academy of Sciences of the United States of America 105, 20055-20062.

Morello, L., Giani, S., Troina, F., and Breviario, D. (2011). Testing the IMEter on rice introns and other aspects of intron-mediated enhancement of gene expression. Journal of experimental botany 62, 533-544.

Mourier, T., and Jeffares, D.C. (2003). Eukaryotic intron loss. Science 300, 1393.

Naqvi, A.R., Haq, Q.M., and Mukherjee, S.K. (2010). MicroRNA profiling of Tomato leaf curl New Delhi virus (ToLCNDV) infected tomato leaves indicates that

miR159/319 and miR172 might be linked with leaf curl disease. Virology Journal 7, 281.

Newell-Price, J., Clark, A.J., and King, P. (2000). DNA methylation and silencing of gene expression. Trends in endocrinology and metabolism: TEM 11, 142-148.

Okkema, P.G., Harrison, S.W., Plunger, V., Aryana, A., and Fire, A. (1993). Sequence requirements for myosin gene expression and regulation in Caenorhabditis elegans. Genetics 135, 385-404.

Palmiter, R.D., Sandgren, E.P., Avarbock, M.R., Allen, D.D., and Brinster, R.L. (1991). Heterologous introns can enhance expression of transgenes in mice. Proceedings of the National Academy of Sciences of the United States of America 88, 478-482.

Panne, D., Raleigh, E.A., and Bickle, T.A. (1999). The McrBC endonuclease translocates DNA in a reaction dependent on GTP hydrolysis. Journal of molecular biology 290, 49-60.

Palatnik, J.F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J.C., Weigel, D. (2003). Control of leaf morphogenesis by microRNAs. Nature 425, 257–263.

Palatnik, J.F., Wollmann, H., Schommer, C., Schwab, R., Boisbouvier, J., Rodriguez, R., Warthmann, N., Allen, E., Dezulian, T., Huson, D., Carrington, J.C., Weigel, D. (2007). Sequence and expression differences underlie functional specialization of Arabidopsis microRNAs miR159 and miR319. Developmental Cell 13, 115–125.

Park, M.Y., Wu, G., Gonzalez-Sulser, A., Vaucheret, H., Poethig, R.S. (2005). Nuclear processing and export of microRNAs in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America 102, 3691–3696.

Parra, G., Bradnam, K., Rose, A.B., and Korf, I. (2011). Comparative and functional analysis of intron-mediated enhancement signals reveals conserved features among plants. Nucleic acids research 39, 5328-5337.

Pashkovskiy, P.P., and Ryazansky, S.S. (2013). Biogenesis, evolution, and functions of plant microRNAs. Biochemistry (Mosc) 78, 627-637.

Purugganan, M.D., and Suddith, J.I. (1998). Molecular population genetics of the Arabidopsis CAULIFLOWER regulatory gene: nonneutral evolution and naturally occurring variation in floral homeotic function. Proceedings of the National Academy of Sciences of the United States of America 95, 8130-8134.

Rapp, R.A., and Wendel, J.F. (2005). Epigenetics and plant evolution. The New phytologist 168, 81-91.

Rasia, R.M., Mateos, J., Bologna, N.G., Burdisso, P., Imbert, L., Palatnik, J.F., and Boisbouvier, J. (2010). Structure and RNA interactions of the plant MicroRNA processing-associated protein HYL1. Biochemistry 49, 8237-8239.

Ren, G., Xie, M., Dou, Y., Zhang, S., Zhang, C., and Yu, B. (2012). Regulation of miRNA abundance by RNA binding protein TOUGH in *Arabidopsis*. Proceedings of the National Academy of Sciences of the United States of America 109, 12817–12821.

Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. Annual Review of Biochemistry 81, 145-166.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., et al. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129, 1311-1323.

Rogan, D.F., Cousins, D.J., Santangelo, S., Ioannou, P.A., Antoniou, M., Lee, T.H., and Staynov, D.Z. (2004). Analysis of intergenic transcription in the human IL-4/IL-13 gene cluster. Proceedings of the National Academy of Sciences of the United States of America 101, 2446-2451.

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

Rose, A.B. (2004). The effect of intron location on intron-mediated enhancement of gene expression in Arabidopsis. The Plant journal : for cell and molecular biology 40, 744-751.

Rose, A.B. (2008). Intron-mediated regulation of gene expression. Current topics in microbiology and immunology 326, 277-290.

Rose, A.B., Elfersi, T., Parra, G., and Korf, I. (2008). Promoter-proximal introns in Arabidopsis thaliana are enriched in dispersed signals that elevate gene expression. The Plant cell 20, 543-551.

Rose, A.B., Emami, S., Bradnam, K., and Korf, I. (2011). Evidence for a DNA-Based Mechanism of Intron-Mediated Enhancement. Frontiers in plant science 2, 98.

Schwab, R., and Voinnet, O. (2010). RNA silencing amplification in plants: size matters. Proceedings of the National Academy of Sciences of the United States of America 107, 14945-14946.

Scohy, S., Gabant, P., Szpirer, C., and Szpirer, J. (2000). Identification of an enhancer and an alternative promoter in the first intron of the alpha-fetoprotein gene. Nucleic acids research 28, 3743-3751.

Seitz, H., Youngson, N., Lin, S.P., Dalbert, S., Paulsen, M., Bachellerie, J.P., Ferguson-Smith, A.C., and Cavaille, J. (2003). Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. Nature genetics 34, 261-262.

Sieburth, L.E., and Meyerowitz, E.M. (1997). Molecular dissection of the AGAMOUS control region shows that *cis* elements for spatial regulation are located intragenically. The Plant cell 9, 355-365.

Silverman, I.M., Li, F., and Gregory, B.D. (2013). Genomic era analyses of RNA secondary structure and RNA-binding proteins reveal their significance to post-transcriptional regulation in plants. Plant science : an international journal of experimental plant biology 205-206, 55-62.

Simon, S.A., and Meyers, B.C. (2011). Small RNA-mediated epigenetic modifications in plants. Current opinion in plant biology 14, 148-155.

Singer, S.D., Cox, K.D., and Liu, Z. (2010). Both the constitutive cauliflower mosaic virus 35S and tissue-specific AGAMOUS enhancers activate transcription autonomously in Arabidopsis thaliana. Plant molecular biology 74, 293-305.

Singer, S.D., Cox, K.D., and Liu, Z. (2011). Enhancer-promoter interference and its prevention in transgenic plants. Plant cell reports 30, 723-731.

Singer, T., Yordan, C., and Martienssen, R.A. (2001). Robertson's Mutator transposons in A. thaliana are regulated by the chromatin-remodeling gene Decrease in DNA Methylation (DDM1). Genes & development 15, 591-602.

Slotkin, R.K., Vaughn, M., Borges, F., Tanurdzic, M., Becker, J.D., Feijo, J.A., and Martienssen, R.A. (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell 136, 461-472.

Soppe, W.J., Jasencakova, Z., Houben, A., Kakutani, T., Meister, A., Huang, M.S., Jacobsen, S.E., Schubert, I., and Fransz, P.F. (2002). DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. The EMBO journal 21, 6549-6559.

Stark, A., Brennecke, J., Russell, R.B., and Cohen, S.M. (2003). Identification of Drosophila MicroRNA targets. PLoS biology 1, E60.

Suter, C.M., and Martin, D.I. (2010). Paramutation: the tip of an epigenetic iceberg? Trends in genetics : TIG 26, 9-14.

Swiezewski, S., Liu, F., Magusin, A., and Dean, C. (2009). Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. Nature 462, 799-802.

Syed, N.H., Kalyna, M., Marquez, Y., Barta, A., and Brown, J.W. (2012). Alternative splicing in plants--coming of age. Trends in plant science 17, 616-623.

Tamaru, H., and Selker, E.U. (2001). A histone H3 methyltransferase controls DNA methylation in Neurospora crassa. Nature 414, 277-283.

Tamaru, H., Zhang, X., McMillen, D., Singh, P.B., Nakayama, J., Grewal, S.I., Allis, C.D., Cheng, X., and Selker, E.U. (2003). Trimethylated lysine 9 of histone H3 is a mark for DNA methylation in Neurospora crassa. Nature genetics 34, 75-79.

Tariq, M., and Paszkowski, J. (2004). DNA and histone methylation in plants. Trends in genetics : TIG 20, 244-251.

Tariq, M., Saze, H., Probst, A.V., Lichota, J., Habu, Y., and Paszkowski, J. (2003). Erasure of CpG methylation in Arabidopsis alters patterns of histone H3 methylation in heterochromatin. Proceedings of the National Academy of Sciences of the United States of America 100, 8823-8827.

Tchurikov, N.A., Kretova, O.V., Moiseeva, E.D., and Sosin, D.V. (2009). Evidence for RNA synthesis in the intergenic region between enhancer and promoter and its inhibition by insulators in Drosophila melanogaster. Nucleic acids research 37, 111-122.

Todd, A.K., Johnston, M., and Neidle, S. (2005). Highly prevalent putative quadruplex sequence motifs in human DNA. Nucleic acids research 33, 2901-2907.

Tsai, M.C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. Science 329, 689-693.

Vaillant, I., and Paszkowski, J. (2007). Role of histone and DNA methylation in gene regulation. Current opinion in plant biology 10, 528-533.

Vaissiere, T., Sawan, C., and Herceg, Z. (2008). Epigenetic interplay between histone modifications and DNA methylation in gene silencing. Mutation research 659, 40-48.

Vasil, V., Clancy, M., Ferl, R.J., Vasil, I.K., and Hannah, L.C. (1989). Increased gene expression by the first intron of maize shrunken-1 locus in grass species. Plant physiology 91, 1575-1579.

Wachtel, C., and Manley, J.L. (2009). Splicing of mRNA precursors: the role of RNAs and proteins in catalysis. Molecular bioSystems 5, 311-316.

Wahl, M.C., Will, C.L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell 136, 701-718.

Wang, K.C., and Chang, H.Y. (2011). Molecular mechanisms of long noncoding RNAs. Molecular Cell 43, 904–914.

Warf, M.B., and Berglund, J.A. (2010). Role of RNA structure in regulating pre-mRNA splicing. Trends in biochemical sciences 35, 169-178.

Watson, L.A., Solomon, L.A., Li, J.R., Jiang, Y., Edwards, M., Shin-ya, K., Beier, F., and Berube, N.G. (2013). Atrx deficiency induces telomere dysfunction, endocrine defects, and reduced life span. The Journal of clinical investigation 123, 2049-2063.

Wu, G., Park, M.Y., Conway, S.R., Wang, J.W., Weigel, D., and Poethig, R.S. (2009). The sequential action of miR156 and miR172 regulates developmental timing in *Arabidopsis.* Cell 138, 750–759.

Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., and Carrington, J.C. (2004). Genetic and functional diversification of small RNA pathways in plants. PLoS Bioloy 2, 0642-0652.

Yang, S.W., Chen, H.Y., Yang, J., Machida, S., Chua, N.H., and Yuan, Y.A. (2010). Structure of *Arabidopsis* HYPONASTIC LEAVES1 and its molecular implications for miRNA processing. Structure 18, 594-605.

Yanofsky, M.F., Ma, H., Bowman, J.L., Drews, G.N., Feldmann, K.A., and Meyerowitz, E.M. (1990). The protein encoded by the Arabidopsis homeotic gene agamous resembles transcription factors. Nature 346, 35-39.

Yegnasubramanian, S., Wu, Z., Haffner, M.C., Esopi, D., Aryee, M.J., Badrinath, R., He, T.L., Morgan, J.D., Carvalho, B., Zheng, Q., et al. (2011). Chromosome-wide mapping of DNA methylation patterns in normal and malignant prostate cells reveals pervasive methylation of gene-associated and conserved intergenic sequences. BMC genomics 12, 313.

Ying, S.Y., and Lin, S.L. (2004). Intron-derived microRNAs--fine tuning of gene functions. Gene 342, 25-28.

You, W., Lorkovic, Z.J., Matzke, A.J., and Matzke, M. (2013). Interplay among RNA polymerases II, IV and V in RNA-directed DNA methylation at a low copy transgene locus in Arabidopsis thaliana. Plant molecular biology 82, 85-96.

Yu, B., Bi, L., Zheng, B., Ji, L., Chevalier, D., Agarwal, M., Ramachandran, V., Li, W., Lagrange, T., Walker, J.C., and Chen, X. (2008). The FHA domain proteins DAWDLE in *Arabidopsis* and SNIP1 in humans act in small RNA biogenesis. Proceedings of the National Academy of Sciences of the United States of America 105, 10073–10078.

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R., and Chen, X. (2005). Methylation as a crucial step in plant microRNA biogenesis. Science 307, 932–935.

Zhang, B., Pan, X., Cobb, G.P., and Anderson, T.A. (2006). Plant microRNA: a small regulatory molecule with big impact. Developmental biology 289, 3-16.

Zhang, H., and Zhu, J.K. (2011). RNA-directed DNA methylation. Current opinion in plant biology 14, 142-147.

Zhang, H., and Zhu, J.K. (2012). Seeing the forest for the trees: a wide perspective on RNA-directed DNA methylation. Genes & development 26, 1769-1773.

Zhang, M., Kimatu, J.N., Xu, K., and Liu, B. (2010). DNA cytosine methylation in plant development. Journal of genetics and genomics = Yi chuan xue bao 37, 1-12.

Zhang, R., Zhang, L., and Yu, W. (2012). Genome-wide expression of non-coding RNA and global chromatin modification. Acta biochimica et biophysica Sinica 44, 40-47.

Zheng, B., Wang, Z., Li, S., Yu, B., Liu, J.Y., and Chen, X. (2009). Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis. Genes & development 23, 2850-2860.

Zhou, Y., Bui, T., Auckland, L.D., and Williams, C.G. (2002). Undermethylated DNA as a source of microsatellites from a conifer genome. Genome / National Research Council Canada = Genome / Conseil national de recherches Canada 45, 91-99.

Zhu, X., Ling, J., Zhang, L., Pi, W., Wu, M., and Tuan, D. (2007). A facilitated tracking and transcription mechanism of long-range enhancer function. Nucleic acids research 35, 5532-5544.

Zieler, H., and Huynh, C.Q. (2002). Intron-dependent stimulation of marker gene expression in cultured insect cells. Insect molecular biology 11, 87-95.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic acids research 31, 3406-3415.