Methylation of Geminivirus Genomes: Investigating its role as a host defense and evaluating its efficacy as a model to study chromatin methylation in plants

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By
Priya Raja, M.Sc.
Graduate Program in Molecular, Cellular and Developmental Biology

The Ohio State University

2010

Dissertation Committee:
Dr. David M. Bisaro, Advisor
Dr. Biao Ding
Dr. Erich Grotewold
Dr. Deborah S. Parris
ABSTRACT

A major finding of this study is that plant hosts methylate geminivirus chromatin as an antiviral epigenetic defense.

The first part of this dissertation describes work that establishes methylation as a host defense. We evaluated the susceptibility of methylation-deficient mutant plants to geminiviruses of two distinct genera (Beet curly top virus, BCTV, a monopartite curtovirus, and Cabbage leaf curl virus, CaLCuV, a bipartite begomovirus). We found that Arabidopsis thaliana plants with mutations in genes encoding cytosine or histone H3 lysine 9 (H3K9) methyltransferases, novel plant polymerases Pol IV and Pol V, RNA-directed DNA methylation pathway components, or adenosine kinase, an enzyme associated with maintaining the methyl cycle, show enhanced symptoms in response to geminivirus infection. We carried out direct biochemical examination of viral chromatin from infected plants for evidence of methylation and it was found that cytosines in geminivirus DNA are methylated in the intergenic region, which contains the origin of replication and promoters for the replication gene and coat protein transcripts. Cytosine methylation in the intergenic region was reduced in the methylation mutants at specific cytosine residues. In addition, histones associated with geminivirus genomes were also found to carry methylation modifications on tail lysine residues that are characteristic of both active and repressed chromatin.
We proposed that methylation is a strategy employed by plant hosts to silence invading geminiviruses. Methylation is associated with transcriptional gene silencing as well as post-transcriptional gene silencing. The curtovirus L2 and begomovirus AL2 proteins are silencing suppressors that counter host-mediated silencing. It was previously found that L2 and AL2 can suppress silencing by multiple mechanisms, one of which involves interacting with and inhibiting adenosine kinase. Concurrent work in this lab has also shown that these proteins can reverse established transcriptional gene silencing and genomic methylation.

The second part of this dissertation establishes geminiviruses as a model to study chromatin methylation. In infected cells, geminivirus single-stranded DNA replicates through double-stranded intermediates that associate with histones to form minichromosomes. In our investigation, we sought to ascertain which members of the Arabidopsis dsRNA binding protein family (DRB2, 3, or 5) are associated with DCL3, the Dicer-like protein in the methylation arm of silencing. In infection studies with CaLCuV and BCTV, we found that drb3 mutants are uniquely hypersusceptible to geminivirus infection. Methylation analysis of the viral intergenic region showed that methylation is greatly reduced in the drb3 mutant. The drb3 mutant, like ago4 and dcl3, fails to recover from geminivirus disease. We also found that DRB3 interacts with DCL3 and AGO4 in distinct subnuclear locations using bimolecular fluorescence complementation analysis. While the production of 24 nt siRNAs was not affected in the drb3 mutant, the DRB3 protein clearly partners with DCL3 in the methylation pathway, likely by acting downstream of siRNA generation, possibly by facilitating RISC loading.
of the 24 nt siRNA species into AGO4-containing RISC complexes. This uncovers at least one dsRNA binding protein in Arabidopsis that is important for methylation. It also suggests that geminivirus genomes may be used as convenient probes to identify additional components and elucidate novel functions for known components in the chromatin methylation arm of RNA silencing.

The third part of this dissertation further explores methylation as a defense against geminiviruses, specifically, in the context of a phenomenon called ‘recovery’. Recovery occurs when axillary shoots that arise following an initial infection with BCTV L2 mutant virus show symptom remission and contain low levels of hypermethylated viral DNA. Work described in earlier parts of this thesis showed that recovery is dependent on DCL3, DRB3 and AGO4. In this study, we found that mutants defective in all non-CG methyltransferases (ddc) lost the ability to recover from infection, indicating that non-CG methylation is required for recovery. We also found that pol V mutants (nrpe; nrpd1b) were unable to recover, while pol IV mutants (nrpd; nrpd1a) showed delayed recovery. The dcl2/4 mutant, deficient in post-transcriptional gene silencing, showed mixed recovery, where some shoots recover, while others do not. Thus, both post-transcriptional and transcriptional gene silencing are associated with recovery. Interestingly, a dcl2/3/4 triple mutant was found to be methylation-competent, suggesting a possible new role for DCL1 in methylation. Recovery from disease is a keynote manifestation of defense and we report that recovery is mediated by Pol V-DCL3-AGO4-DRM2/CMT3. Further, this study identifies the host recovery assay as a sensitive and definitive tool for the identification of methylation pathway components.
To you, who must prove all that I have understood,

to be incomplete, or inadequate,

etirely, or in portions,

and change, modify, erase or extend,

and in doing so, give meaning to all that I have learned,

in the past six and more years of my life,

during the next four or five of your own.
ACKNOWLEDGMENTS

I thank my advisor and mentor, Dr. David M. Bisaro, to whom I owe a satisfying journey to this dissertation. I greatly value my experience as a student in his lab and am very deeply grateful for all the training, opportunities, patience, encouragement and support he provided throughout to enable this process.

I cherish with gratitude the interest and support shown by my committee members, Dr. Biao Ding, Dr. Erich Grotewold, and Dr. Deborah S. Parris. I thank them for serving on my committee, and sincerely appreciate their scientific critique and advice, and value the lessons I learned. I also want to acknowledge Dr. Dave Mackey and Dr. J.C. Jang for their interest and suggestions on my work. I owe thanks to Dr. Michael Boehm, Director of Plant Pathology who served as the Graduate Faculty Representative at my Final Oral Exam.

I owe special thanks to Dr. Kenneth J. Buckley for being a friend I could always count on since I started on my first rotation in graduate school.

I thank all my lab colleagues for scientific discussions, technical advice and friendship: Dr. Cody Buchmann, who started out as my colleague from zero-error
universe, but grew to be the most well-meaning labmate I could ask for, Jamie Wolf, for her cheerful and enthusiastic support with various projects in and out of lab, Gireesha Mohannath, for inspiring resilience through example and Sizhun Li, the most recent asset to our lab.

I am deeply appreciation of the help and cheer provided by Brad Sanville, Isaac Heard and Matt Balasco, undergraduate students that participated actively in experiments presented in this thesis.

I thank past and present members of the Bisaro lab: Xiaojuan Yang, Dr. Shaheen Asad, Natalie Villafranco, Dr. Yan Xie, Dr. Youn Lee, Allie Varner, Jeff Ostler, Veena Patil, Udit Singhal and especially Dr. Hui Wang, whose graduate work formed the basis for my thesis question.

I thank Litty and Poornima, Niloufer and Shalaka for their friendship, Isabel and Antje for hugs and support, and all the wonderful people at Rightmire Hall.

I am grateful to the staff of the Plant Biotechnology Center: Melinda Parker, Diane Furtney, Dave Long, Scott Hines, Joe Takayama, and MCDB Program Co-ordinator, Jan Zinaich for always being people I could rely on. I am also grateful to Jessie Siegman at the Molecular Genetics Department, Timothy Watson, Director of Graduation Services, and Sandy Shew at The Ohio State Graduate School for all their help with information and procedures during the dissertation process.

I extend my most sincere thanks to my family, especially my parents and grandparents, for their anchoring support, and to Ashish, for these years of togetherness, despite the distance.
VITA

May 3, 1978……………….  Born - Mumbai, India

1996 – 1998……………….  B.Sc., Life Sciences and Biochemistry, 
                      St. Xavier’s College, Mumbai, India

1998 – 2000……………….  M.Sc., Biochemistry, 
                      University of Mumbai, Mumbai, India

2003 – present…………….  Graduate Teaching and Research Associate, 
                            The Ohio State University, Columbus, Ohio, USA

2007………………………...I.E. Melhus Graduate Student Award by the American 
                            Phytopathological Society

2007………………………...Student Travel Grant by the American Society of Virology

2008………………………...Student Travel Grant by the American Society of Virology

2009………………………...Milt Zaitlin Student Award by the Plant Virology Club at 
                            the American Society of Virology

viii
PUBLICATIONS

PUBLISHED PAPERS


PUBLISHED ABSTRACTS

Wang, H., Buchmann, R.C., Yang, X., Raja, P., Buckley, K. J., and Bisaro, D.M. Suppression of RNA silencing by geminivirus proteins. 7th Annual Plant Molecular Biology & Biotechnology Research Symposium at The Ohio State University, Columbus, OH. April 15-16, 2005.


Raja, P., Buchmann, R.C., Buckley, K. J., and Bisaro, D.M. Geminivirus genome methylation as a host defense. 9th Annual Plant Molecular Biology & Biotechnology Research Symposium at The Ohio State University, Columbus, OH. March 30-31, 2007.

Raja, P., Buchmann, R.C., and Bisaro D.M. Methylation as a host defense against geminiviruses. 1st Annual Molecular Life Sciences Interdisciplinary Graduate Programs Symposium at The Ohio State University, Columbus, OH. May 12, 2007.
Raja, P., and Bisaro, D.M. Methylation as an antiviral defense against geminiviruses. 7th I. E. Melhus Graduate Student Symposium: Emerging and Changing Viral Pathogens – Biology and Molecular Mechanisms: Methylation as a host defense against geminiviruses, American Phytopathological Society Meeting, San Diego, CA. July 28-August 1, 2007

Buchmann, R.C.*, Raja, P.*, Asad, S.A., Sanville, B.C., and Bisaro D.M. Methylation as a host defense against geminiviruses. 10th Annual Plant Molecular Biology & Biotechnology Research Symposium at The Ohio State University, Columbus, OH. April 11-12, 2008. (*joint first authors)

Raja, P., Buchmann, R.C., Sanville, B.C., and Bisaro D.M. Viral genome methylation is an epigenetic defense against geminiviruses. 2nd Annual Molecular Life Sciences Interdisciplinary Graduate Programs Symposium at The Ohio State University, Columbus, OH. May 17, 2008.


OSU Comprehensive Cancer Center MCBG Retreat at The Hueston Woods Resort, Columbus, OH. Sept 15, 2009.


**FIELDS OF STUDY**

Major Field: Molecular, Cellular and Developmental Biology
Plant Virology and Epigenetics
# TABLE OF CONTENTS

Abstract ................................................................................................................................. ii

Dedication .............................................................................................................................. v

Acknowledgements ............................................................................................................... vi

Vita ........................................................................................................................................ viii

List of Tables ....................................................................................................................... xviii

List of Figures ....................................................................................................................... xix

Chapters:

CHAPTER 1 .......................................................................................................................... 1

INTRODUCTION ................................................................................................................ 1

1.1 GEMINIVIRUSES ........................................................................................................ 1

   1.1.1 Geminivirus virions, genome structure and organization................................. 2
   1.1.2 Geminivirus genes .............................................................................................. 4
   1.1.3 Geminivirus infection cycle .............................................................................. 8
   1.1.4 Geminivirus rolling circle replication – the role of Rep protein....................... 11
1.1.5 Geminivirus transcription ................................................................. 12

1.2 RNA SILENCING IN PLANTS .................................................................... 13

1.2.1 RNA silencing pathways .................................................................... 13
1.2.2 RNA silencing as an antiviral defense .................................................. 15
1.2.3 Post-transcriptional gene silencing (PTGS) ........................................... 16
1.2.4 Spread of the RNA silencing signal .................................................... 18
1.2.5 Viral suppressors of PTGS ................................................................. 20
1.2.6 RNA-directed DNA methylation and transcriptional gene silencing .... 21
1.2.7 Geminivirus silencing suppressors ....................................................... 26

1.3 CHROMATIN STRUCTURE AND EPIGENETICS ................................ 31

1.3.1 Nucleosome structure ......................................................................... 31
1.3.2 Histone tail modifications .................................................................... 32
1.3.3 Histone variants .................................................................................. 34
1.3.4 Nucleosome assembly ......................................................................... 37
1.3.5 Chromatin remodeling ......................................................................... 38
1.3.6 Can geminivirus chromatin be a model for eukaryotic chromatin? .... 39

CHAPTER 2 ..................................................................................................... 52

VIRAL GENOME METHYLATION AS AN EPIGENETIC DEFENSE AGAINST GEMINIVIRUSES ................................................................. 52

2.1 ABSTRACT ............................................................................................. 52
2.2 INTRODUCTION .............................................................................................................. 53

2.3 MATERIALS AND METHODS .......................................................................................... 56

2.3.1 Arabidopsis mutants ........................................................................................................ 56
2.3.2 Virus inoculation.............................................................................................................. 57
2.3.3 DNA isolation and Southern Blot analysis ..................................................................... 58
2.3.4 Bisulfite sequencing........................................................................................................ 59
2.3.5 Chromatin immunoprecipitation .................................................................................. 60

2.4 RESULTS ............................................................................................................................ 61

2.4.1 Methylation-deficient Arabidopsis mutants are hypersusceptible to
geminivirus infection. ............................................................................................................... 61
2.4.2 Plants deficient in DCL and RDR enzyme activities show moderately
increased susceptibility to geminiviruses .............................................................................. 64
2.4.3 Geminivirus genomes are methylated in vivo. ............................................................... 66
2.4.4 Analysis of cytosine methylation in the BCTV genome. ............................................... 68
2.4.5 ChIP reveals both active and repressive marks on geminivirus DNA. ...................... 70
2.4.6 Recovery of Arabidopsis plants from infection with BCTV L2 mutant virus
requires AGO4 and is associated with greatly increased cytosine methylation. ............ 71

2.5 DISCUSSION ...................................................................................................................... 72
CHAPTER 3 ................................................................................................................................. 88

ARABIDOPSIS DOUBLE-STRANDED RNA BINDING PROTEIN DRB3 PARTNERS WITH DICER-LIKE 3 AND ARGONAUTE 4 IN METHYLATION-MEDIATED DEFENSE AGAINST GEMINIVIRUSES ................................................................................................................................. 88

3.1 ABSTRACT ............................................................................................................................... 88

3.2 INTRODUCTION ....................................................................................................................... 90

3.3 MATERIALS AND METHODS .................................................................................................... 94

3.3.1 Arabidopsis mutants ............................................................................................................. 94

3.3.2 Virus inoculation .................................................................................................................. 94

3.3.3 Cloning for interaction studies .......................................................................................... 95

3.3.4 Yeast two-hybrid analysis .................................................................................................. 95

3.3.5 BiFC Analysis ...................................................................................................................... 96

Interactions between DCL and DRB proteins in plants were tested using BiFC (Bimolecular Fluorescence Complementation) (Hu et al., 2002). The construction of BiFC expression vectors using enhanced YFP has been described previously (Yang et al., 2007). BiFC expression vectors: ............................................................................................................. 96

3.4 RESULTS AND DISCUSSION ................................................................................................. 100

3.4.1 Arabidopsis drb3 mutants are hypersusceptible to geminivirus infection. 100

3.4.2 Arabidopsis dcl3 and drb3 mutants are unable to recover from geminivirus infection. ................................................................................................................................................................................... 102

3.4.3 Arabidopsis drb3 and dcl3 mutant plants cannot hypermethylate the viral genome.............................................................................................................................................................................. 104
3.4.4 DRB3 interacts with DCL3 and AGO4 in *Nicotiana benthamiana* cells. 106
3.4.5 DRB3 is not required for the biogenesis of 24 nt siRNAs by DCL3. 109

3.5 CONCLUSION .................................................................................................................. 110

CHAPTER 4 .......................................................................................................................... 118

IDENTIFICATION OF COMPONENTS REQUIRED FOR RECOVERY FROM GEMINIVIRUS DISEASE ........................................................................................................ 118

4.1 ABSTRACT ..................................................................................................................... 118

4.2 INTRODUCTION ........................................................................................................... 120

4.3 MATERIALS AND METHODS ...................................................................................... 123

4.3.1 *Arabidopsis* mutants .............................................................................................. 123
4.3.2 Virus inoculation ...................................................................................................... 125
4.3.3 5-azacytidine treatment .......................................................................................... 125
4.3.4 DNA isolation and Southern Blot analysis .............................................................. 126
4.3.5 Bisulfite sequencing ............................................................................................... 126

4.4 RESULTS .................................................................................................................... 127

4.4.1 Methylation is required for recovery. ...................................................................... 127
4.4.2 Non-CG methylation is necessary for recovery....................................................... 129
4.4.3 Pol V is necessary for recovery; Pol IV plays an important role............................... 131
4.4.4 DCL3 is the most important dicer-like enzyme for recovery, but DCL1, DCL2 and DCL4 can contribute ............................................................................................................ 132
5.1 Methylation as a defense against geminiviruses ........................................ 146
5.2 Using geminiviruses as a facile and efficient screen for new components of the
methylation pathway ....................................................................................... 151
5.3 Exploring chromatin biology using the geminivirus minichromosome ........ 154

BIBLIOGRAPHY .......................................................................................... 156
APPENDICES ......................................................................................... 175

APPENDIX A: VIRAL GENOME METHYLATION AS AN EPIGENETIC DEFENSE
AGAINST GEMINIVIRUSES - SUPPLEMENTAL DATA .............................. 176
APPENDIX B: ARABIDOPSIS DOUBLE-STRANDED RNA BINDING PROTEIN
DRB3 PARTNERS WITH DICER-LIKE 3 AND ARGONAUTE 4 IN
METHYLATION-MEDIATED DEFENSE AGAINST GEMINIVIRUSES ...... 213
APPENDIX C: DESIGNING PRIMERS FOR BISULFITE ANALYSIS .......... 229
LIST OF TABLES

Table 2.1 Methylation-deficient mutants are hypersusceptible to geminivirus infection...81
Table 4.1 Methylation-deficient mutants cannot recover from geminivirus disease......140
Table B.1 Interactions of DRB and DCL proteins in the Yeast Two-hybrid System.....225
LIST OF FIGURES

Figure 1.1 Geminivirus gene organization.................................................................42
Figure 1.2 Geminivirus replication origin.................................................................43
Figure 1.3 Geminivirus replication...........................................................................44
Figure 1.4 Antiviral RNA silencing pathways..........................................................46
Figure 1.5 The methyl cycle.....................................................................................48
Figure 1.6 Nucleosome structure............................................................................49
Figure 1.7 Mechanisms of chromatin remodeling.......................................................50
Figure 1.8 RNA silencing as a defense against geminiviruses.................................51
Figure 2.1 The methylation pathway in plants..........................................................79
Figure 2.2 Methylation-deficient mutants are hypersusceptible to geminiviruses.....81
Figure 2.3 The geminivirus intergenic region is methylated in vivo............................83
Figure 2.4 ChIP analysis of the CaLCuV intergenic region.......................................85
Figure 2.5 *Arabidopsis ago4* mutants do not recover after infection with BCTV L2' plants
........................................................................................................................................86

Figure 3.1 *Arabidopsis drb3* mutants show enhanced susceptibility to geminivirus infection........................................................................................................................................112

Figure 3.2 *Arabidopsis drb3* and *dcl3* mutants do not recover from infection with BCTV L2' mutant virus........................................................................................................................................112

Figure 3.3 *Arabidopsis drb3* and *dcl3* mutants fail to hypermethylate the viral genome........................................................................................................................................113

Figure 3.4 DRB3 independently interacts with DCL3 and AGO4 in distinct subnuclear bodies........................................................................................................................................114

Figure 3.5 DRB3 is not required for biogenesis of geminivirus-derived siRNAs........115

Figure 4.1 Methylation is required for recovery........................................................................................................................................135

Figure 4.2 Non-CG methylation is required for recovery........................................................................................................................................136

Figure 4.3 Recovery is correlated with hypermethylation of the viral genome and symptomatic tissue from *kyp2* and *cmt3* mutants does not show hypermethylation of the viral genome........................................................................................................................................138

Figure 4.4 The *pol V* mutant (*nrpd1b/nrpe*) does not recover, but the *pol IV* mutant (*nrpd1a/nrpd*) shows delayed recovery........................................................................................................................................139

Figure 4.5 The *dcl2/3/4* triple mutant is methylation-competent........................140

Figure 4.6 The *dcl3* mutant does not recover, but *dcl2/4* plants show mixed recovery..141

Figure B.1 *Arabidopsis drb3* and *dcl3* mutants do not recover from infection with BCTV L2' virus........................................................................................................................................209

Figure B.2 Cytosine methylation analysis of the BCTV L2' genome in wild-type, *ago4*, *drb3, drb4*, *dcl3* and *dcl4* mutant plants........................................................................................................................................210

Figure B.3 Interactions between selected DRB, DCL3 and AGO proteins using BiFC analysis in *N. benthamiana* epidermal cells........................................................................................................................................212
CHAPTER 1

INTRODUCTION


1.1 Geminiviruses

As early as 752 A.D., in the first known description of a plant virus disease, a Japanese poem described geminivirus symptoms (Saunders et al., 2003). More than twelve hundred years later, electronmicrographs of ultrathin sections of infected plants showed the first evidence of paired particles in phloem nuclei. Virions were later purified and characterized, confirming the presence of twin structures and lending this family of viruses the name Geminiviridae. The family of viruses was recognized in 1978 by the International Committee on the Taxonomy of Viruses. Geminivirus twin capsids encapsidate a single circular ssDNA molecule. Family members collectively have a wide host range within angiosperms, and infect important crop species such as tomato,
sugarbeet, beans, cassava, sweet potato, maize, pepper and cotton (Brown, 1996). They are classified into four genera, *Begomovirus, Curtovirus, Mastrevirus* and *Topocuvirus*, based on as their insect vector, genome organization, and host range (Fauquet et al., 2003). In basic research, geminiviruses have been used as models to study fundamental processes such as DNA replication, transcription, gene silencing and more recently, as described in this thesis, epigenetic regulation.

1.1.1 Geminivirus virions, genome structure and organization

While most plant viruses possess RNA genomes, geminiviruses, nanoviruses, (and pararetroviruses), are DNA viruses. Geminiviruses, which are by far the largest group of DNA viruses, contain circular ssDNA genomes that are about 2.5 to 3.0 kb in length. The ssDNA is encapsidated into a twin icosahedral particle whose structure is stabilized by protein–DNA interactions. The three-dimensional fine structure of virions has been resolved for *Maize streak virus*, type member of the genus *Mastrevirus* (Zhang et al., 2001) and *African cassava mosaic virus*, a begomovirus (Böttcher et al., 2004). Two joined incomplete icosahedra are twisted 20 degrees from each other to form a geminate particle. Each half of the virion is composed of 11 pentameric capsomers missing one capsomer to complete a T=1 structure. The missing faces are joined so that the interior of the geminate particle is a single compartment (Goodman, 1981).

Geminiviruses resemble animal polyomaviruses such as SV40 in their genome organization. Geminiviruses may be monopartite (including all mastre-, curto-, and
topocuviruses and some begomoviruses) or bipartite (as is common in all native begomoviruses e.g. New World viruses such as *Tomato golden mosaic virus* (TGMV), and *Cabbage leaf curl virus* (CaLCuV) and Old World viruses such as *African cassava mosaic virus* (ACMV)). Bipartite viruses have two components, called DNA A and DNA B, that share a sequence of about 200 - 250 bp within the intergenic region (IR) called the common region (CR). Both components are required for infectivity. The A component encodes information needed for replication and encapsidation, while the B component carries information for virus movement between cells and throughout the plant (Rogers et al., 1986; Sunter et al., 1987). The common region (CR) is conserved between two components of the same virus, and contains the origin of replication (ori) described in Section 1.1.2. Double-stranded replicative forms act as templates for transcription and replication. Both genes and proteins are numbered. Viral genes involved in replication and transcription are leftward oriented (designated L for left, or C for complementary), while the rightside typically contains the coat protein gene involved in packaging (R for right, or V for virion-sense) (Figure 1.1). This thesis uses the nomenclature L and R (Bisaro, 2006). In the A component, the leftward promoter is responsible for transcription of AL1, AL2 and AL3 genes, while the rightward promoter is responsible for coat protein (CP) or AR1 gene transcription.

Monopartite virus genome organization is similar to that of DNA A in bipartite viruses. While the replication genes (L1, L3) are highly conserved with bipartite begomoviruses, the rightward genes responsible for movement are not. Examples of a
bipartite begomovirus (*Tomato golden mosaic virus*, or TGMV) and a monopartite curtovirus (*Beet curly top virus*, or BCTV) are shown in Figure 1.1.

1.1.2 Geminivirus genes

Geminivirus DNA is transcribed bidirectionally from promoters located within the Intergenic Region (IR) by RNA Pol II. The intergenic region contains the origin (ori) which includes a highly conserved nine-nucleotide hairpin structure (TATAATT|AC) where rolling circle replication is initiated (Lazarowitz et al., 1992), as well as divergent promoters for early and late genes (Figure 1.2). Genes and proteins are numbered depending on whether they are transcribed leftward or rightward from the promoter. In a typical bipartite geminivirus, such as TGMV or CaLCuV, DNA A encodes four leftward genes – AL1, AL2, AL3, and AL4, and one rightward gene, AR1. DNA B contains two genes, BL1 and BR1. An additional promoter is located upstream of the AL2 and AL3 genes of begomoviruses, each of which are expressed from bicistronic transcripts (Shung et al., 2006; Shung et al., 2009).

Many viral proteins have multiple functions, and most genes and proteins have also been named according to their first described functions: Rep (Replication initiator protein, AL1), TrAP (Transcription Activator protein, AL2), REn (Replication Enhancer, AL3), CP (coat protein, AR1), MP (movement protein, BL1), and NSP (nuclear shuttle protein, BR1). The pre-coat protein (PCP) is present in mastre-, curto-, topocu-, and Old World monopartite begomoviruses, but is absent in New World bipartite begomoviruses.
The AL1 or Rep protein binds and inactivates plant retinoblastoma protein (RBR) relieving E2F-mediated repression of cellular genes such as proliferating cell nuclear antigen (PCNA) and conditioning cells to enter S-phase. This enables access to host replication machinery for virus replication (Hanley-Bowdoin et al., 2000; Egelkrout et al., 2001; Hanley-Bowdoin et al., 2004). AL1 also acts as an origin recognition protein and binds to the origin to recruit host replication machinery, including PCNA, replication protein A (RPA), and replication factor C (RFC) (Fontes et al., 1994; Luque et al., 2002; Castillo et al., 2003). It also serves as a replicative helicase (Choudhury et al., 2006; Clerot and Bernardi, 2006). AL1 also introduces a nick in the viral strand within the loop of the conserved hairpin, which provides the primer for initiation of DNA synthesis (Lazarowitz et al., 1992; Bisaro, 1996; Hanley-Bowdoin et al., 2004). These diverse activities of AL1 are modulated by its oligomeric state. AL1 also interacts with cellular kinases GRIK1 and GRIK2 that activate SnRK1 (Shen et al., 2006; Shen et al., 2009).

The AL2 or TrAP protein is a small 15 kD protein with multiple functions, including transcription activation, silencing suppression, and interaction with and inactivation of cellular proteins SnRK1 (SNF1) and ADK (Sunter and Bisaro, 1997; Hao et al., 2003; Wang et al., 2003; Wang et al, 2005). The AL2 protein has a basic region near the N-terminus that contains an NLS, a central region that contains a zinc finger-like motif (C-X₁-C-X₄-H-X₂-C), and an acidic C-terminal region that contains the minimal transcription activation domain. The central region is required for DNA and zinc binding, and contains invariant cysteine residues and a conserved histidine residue that is present in most begomovirus AL2 proteins (Hartitz et al., 1999; Dong et al., 2003). AL2 is
required for expression of late viral genes (Sunter and Bisaro, 1991; Sunter and Bisaro, 1992). AL2 is targeted to the CP promoter by protein-protein interactions. Transcription activation by AL2 involves both derepression (in the phloem) and transactivation of the CP promoter by interaction of AL2 with the cellular factor PEAPOD2 that binds the activation sequence (Sunter and Bisaro, 2003; Lacatus et al., 2009). AL2 dimerization, mediated in particular by the invariant cysteine residues, is required for transcriptional activation, although dimerization was not a prerequisite for nuclear localization (Yang et al., 2007).

The AL3 or REn protein is needed for optimal replication, although its mechanism of action remains unclear. TGMV AL3 mutants typically accumulate viral DNA in significantly reduced amounts (Sunter et al., 1990). In Tomato yellow leaf curl virus (TYLCV), AL3 has been found to dimerize, interact with cellular factors such as pRBR and PCNA, and interact with AL1 to enhance replication by protein-protein interactions (Settlage et al., 2001; Castillo et al., 2003; Settlage et al., 2005).

The AL4 or AC4 protein, in Old World begomoviruses, can act as a silencing suppressor by binding to single-stranded small interfering RNAs (Vanitharani et al., 2004; Vanitharani et al., 2005). The AR1 protein, or coat protein (CP), is a basic protein that binds non-specifically to nucleic acids and is necessary for cell-to-cell and systemic virus spread (in the monopartite geminiviruses) as well as plant-to-plant transmission (Gardiner et al., 1988; Liu et al., 1997; Qin et al., 1998).

In bipartite geminiviruses, the two genes encoded by DNA B play a role in cell-to-cell movement and systemic spread. BL1 is the movement protein (MP) and associates
with plasmodesmata, whereas BR1, the nuclear shuttle protein (NSP) binds ssDNA. BL1 and BR1 also interact with one another; possibly BR1 conveys viral genomes to BL1 for movement between cells (Rojas et al., 2005).

Monopartite viruses, such as the curtovirus BCTV, have similar functional organization as DNA A. The L1 and L3 genes are functional homologs of AL1 and AL3 (Stanley et al., 1992; Sunter et al., 1994; Choi and Stenger, 1995; Hormuzdi and Bisaro, 1995). However, unlike AL2, the L2 protein is not a transcription activator and it is not required for late viral gene expression. The L2 protein lacks the basic and acidic regions of AL2, but shares the central zinc finger-like motif with AL2. Nevertheless, L2 is a pathogenicity factor, and shares the silencing suppression function with AL2 (Wang et al., 2003; Wang et al., 2005). In the absence of a B component, the rightward genes for monopartite viruses carry out all movement functions. The R1 gene of BCTV encodes the coat protein, which along with R3, is essential for movement. R2 is essential for accumulation of ssDNA, which is the encapsidated form (Hormuzdi and Bisaro, 1993).

Allied genomic components such as Defective Interfering DNAs (DI DNAs) and satellite DNAs are associated with geminivirus genomes. DI DNAs may accumulate during an infection and, as their name indicates, these interfere with virus replication and attenuate symptom development. They almost always contain the viral ori (see Figure 1.2) and sometimes the left side of monopartite DNA, or DNA B. They are about 1,300 nt or smaller in size, and could be generated by nonhomologous recombination. Subgenomic or DI DNAs are often produced in a BCTV infection (Stenger et al., 1992). In addition to subgenomic DNAs that interfere with geminivirus replication, some
monopartite geminiviruses associate with and support a circular satellite DNA known as DNAβ. DNAβ cannot replicate on its own, and only shares homology with geminivirus DNA in the ori. DNAβ is typically half the size of a geminivirus genome component, though a very small satellite of only 682 nt has also been found (Dry et al. 1997; Li et al. 2007). DNAβ encodes a single gene called βC1, which is essential for symptom development in TYLCCNV infections, although the virus can independently replicate and move through plants (Cui et al., 2004; Saunders et al., 2004; Stanley et al., 2004; Saeed et al., 2005; Gopal et al. 2007). βC1 has also been identified as a silencing suppressor (Cui et al., 2005). Our recent work shows that DNAβ can assist a curtovirus such as BCTV and that its silencing suppression function may involve reversal of transcriptional gene silencing associated with methylation.

1.1.3 Geminivirus infection cycle

Geminiviruses belonging to the Mastrevirus and Curtovirus genera are transmitted by leafhoppers, while Topocuvirus members are transmitted by treehoppers. Begomoviruses are transmitted by the whiteflies, Bemisia tabaci (Bisaro, 2006; Jeske, 2009). The coat protein determines specificity between virus and insect (Briddon et al., 1990; Höhnle et al., 2001). Transmission is persistent and circulative. When these insects feed on the phloem of a plant, geminivirus capsids are taken up and pass through the midgut epithelium into the hemocoel and then enter salivary glands to be delivered to a new host. Geminiviruses are directly injected into phloem cells through the stylet. From
the sieve elements, geminiviruses must enter nucleated companion and phloem parenchyma cells for replication (Jeske, 2009).

Most geminiviruses are phloem-limited, but some (e.g. TGMV, CaLCuV) are found in leaf mesophyll tissues, and epidermis (Rushing et al., 1987). Geminiviruses may be restricted to the phloem due to lack of transport proteins, requirement for phloem-specific promoters, or better utilization of host replication machinery in the phloem (Morra and Petty, 2000). But transport proteins of phloem-limited AbMV and ICMV can efficiently move from cell to cell in the mesophyll when delivered biolistically. Most geminiviruses are not mechanically transmitted but the rate of mechanical transmission increases in transgenic plants containing AbMV DNA B that are able to constitutively express the transport proteins MP and NSP (Wege and Pohl, 2007).

Geminiviruses replicate in the nucleus via rolling circle replication. It is currently not known whether geminiviruses enter the nucleus as complete virions or as DNA, most likely through the nuclear pores. Geminiviruses preferentially release DNA at the top of virions, so it is also possible to conceive of virions docking at the nuclear membrane and releasing DNA within (Kittelmann and Jeske, 2008). After rolling circle replication, if DNA is concomittantly packaged into virions, there appears to be no defined origin of assembly. Virion assembly occurs within the nucleus (Goodman, 1981; Rushing et al., 1987). These virions must then find a way out of the nucleus in order to move long distance through the phloem, or be picked up by the insect vector. Cell-to-cell spread likely involves nucleic acid transport, not virions (Rojas et al., 1998).
Following injection of virions by the insect vector and disassembly of the capsid, circular single-stranded geminivirus DNA is replicated through circular double-stranded DNA intermediates which are wrapped around host nucleosomes to form viral minichromosomes. Geminiviruses rely on host polymerases for replication. How parental ssDNA is converted to the dsDNA replicative form intermediate is not known. However, mastrevirus virions package a small DNA-primer that binds to the small intergenic region (SIR) and probably serves as a primer for replication by host polymerases. Replication of all other geminiviruses likely relies on host machinery. Pol α:primase likely generates RNA primers from which second-strand synthesis is initiated by DNA polymerase ε. Following primer removal, the newly-synthesised complementary strand is ligated, twisted by topoisomerases and packaged into nucleosomes. The Rep gene conditions S phase in differentiated cells that the virus invades (Gutierrez, 2000a and b; Gutierrez et al., 2004; Pilartz and Jeske, 2003; Hanley-Bowdoin et al., 2004). Mastrevirus RepA protein contains the Rb binding LXCXE motif which binds plant retinoblastoma related (RBR) proteins to allow reentry into the S phase (Gutierrez, 2000a). The begomovirus AL1 protein contains different Helix 4 binding motifs for RBR interaction (Hanley-Bowdoin et al., 2004). Geminiviruses such as MSV and TGMV that are mutated in their RBR binding domains replicate less efficiently and are phloem-limited. This is because procambial cells of the vasculature are DNA replication-competent and RBR binding by Rep is only required to condition S-phase in terminally-differentiated mesophyll cells (Kong et al., 2000; McGivern et al., 2005; Shepherd et al., 2005; Jeske, 2009).
1.1.4 Geminivirus rolling circle replication – the role of Rep protein

Geminiviruses replicate by rolling circle replication similar to the ssDNA bacteriophage ΦX174. The single stranded (+) genome is first used as a template to synthesize a complementary negative strand (-). The invariant sequence in the geminivirus ori (TAATATT|AC) is nicked by the Rep protein (Figure 1.2). This provides the primer. Rep binds to the 5’ end to initiate rolling circle replication. It finally ligates the new 3’OH end to the 5’phosphate (Laufs et al., 1995). The double-stranded intermediate form of the genome is then used to synthesize (+) strand templated by the negative strand, forming a new genome. The resulting circular ssDNA may be immediately packaged into virions or template additional negative strands (Saunders et al., 1991; Stenger et al., 1991; Gutierrez et al., 1999; Hanley-Bowdoin et al., 2000; Jeske et al, 2001; Preiss and Jeske, 2003). The ligase activity is located in the N-terminal half of Rep (aa 1–180), that is responsible for origin recognition and transcriptional auto-repression. The C-terminal of Rep (aa 180—362) contains a helicase domain (Koonin and Ilyina, 1992; Choudhury et al., 2006; Clerot and Bernardi, 2006) that is believed to unwind templates in 3’ to 5’ direction. The diverse activities of AL1 appear to be regulated, at least in part, by the oligomeric state of the protein. The geminivirus replication cycle is shown in Figure 1.3.

Geminiviruses also use recombination-dependent replication to a lesser extent (RDR) and certain recombination hotspots have been identified, which correspond to nuclease-sensitive regions (Jeske, 2009).
1.1.5 Geminivirus transcription

Geminivirus genes are transcribed bidirectionally from the intergenic region by RNA polymerase II. The mRNAs are initiated downstream of a consensus TATA box (Figure 1.2). They are polyadenylated and capped. Most transcripts are unspliced, though the genome encodes overlapping transcripts, and transcripts containing multiple ORFs (open reading frames). Transcription occurs in a phased manner. AL1 and AL3 genes involved in replication are expressed early, AL2 is delayed-early, and is required for activation of late genes such as the coat protein (CP) and nuclear shuttle protein (NSP), which are needed for encapsidation and movement. The A component of bipartite begomoviruses contains three leftward (complementary sense) 3’ co-terminal transcripts for AL1, AL2 and AL3, which overlap the 3’ end of the rightward (viral sense) coat protein (CP) transcript. The promoter AL-62 directs transcription of an mRNA that spans AL1, AL2 and AL3 genes. This transcript includes AL4, contained within AL1, in a different reading frame. This mRNA is translated to yield AL1 and AL3 (and likely AL4) (Hanley-Bowdoin et al., 1989). AL2, however, is expressed under a downstream promoter, AL-1629. An additional downstream promoter, AL-1935, as well as AL-1629 can both direct expression of AL3 (Shung et al., 2006; Shung et al., 2009). AL1 represses its own transcription by binding the origin of replication, immediately upstream of AL-62 (Sunter et al., 1993; Eagle et al., 1994). Transcription from AL-1629 and AL-1935
promoters is dependent on AL1-mediated repression of AL-62 (Shung and Sunter, 2007). This results in delayed early expression of AL2 (Figure 1.1).

1.2 RNA silencing in plants

1.2.1 RNA silencing pathways

RNA silencing in Arabidopsis is triggered by double-stranded RNA which may result from endogenous loci with internal stem-loop, hairpin or inverted repeat structures within a transcript, overlapping transcripts, or as a result of dsRNA synthesis by RNA-dependent RNA polymerase (RDR) of an RNA virus. The dsRNA may also be synthesized from ssRNA by cellular RDRs (RDR1–6). Aberrant mRNAs or Pol IV transcripts (see Section 1.2.6, Figure 1.4) may be substrates for RDRs. The dsRNA is then diced into small RNA duplexes by Dicer-like (DCL1–4) proteins. Two classes of small RNAs are produced – micro RNAs (miRNAs) or small interfering RNAs (siRNAs). miRNAs are typically encoded by endogenous genes, and usually have effects on development. In plants, they are perfectly complementary to their targets, with up to four mismatches, usually at the 3’ end (Vaucheret, 2006). Specific dicers are involved in the biogenesis of different classes of small RNAs and there is specialization in their function. DCL1, DCL2, DCL3 and DCL4 produce 18–21 nt (miRNA), 22 nt, 24 nt and 21 nt long siRNAs, respectively. Arabidopsis dsRNA binding proteins (DRB1-5) associate with specific DCLs and play a role in processing of siRNAs and/or RISC loading (Chapter 3).
The small RNAs have 2 nt 3’ overhangs and are 2’-O-methylated at the 3’ end by HEN1 methylase which stabilizes the duplexes by preventing oligouridylation and degradation. RNA silencing mediated by siRNAs results in suppression of gene expression by means of post-transcriptional gene silencing (PTGS) or transcriptional gene silencing (TGS), based on whether target mRNA with sequence homology is cleaved post-transcriptionally, or whether transcription is repressed by promoter methylation. The small RNA duplexes may thus be exported to the cytosol for post-transcriptional gene silencing (PTGS) or retained in the nucleus to guide transcriptional gene silencing (TGS).

HUA-ENHANCER (HEN1) methylates the 3’ end of siRNAs preventing their uridylation and degradation. The guide strand (the strand with the weakest base-pairing at the 5’ end) of a small RNA is selectively incorporated into a RNA-induced silencing complex (RISC) that includes an ARGONAUTE (AGO1-10) protein and associates with complementary target RNA. DsRNA binding (DRB) proteins may serve to orient the siRNA duplex such that the DCL enzyme is bound to the most unstable end of the guide strand, while the DRB attaches to the opposite end to stabilize and orient the duplex for AGO loading. The DCL protein is switched for the appropriate AGO protein. AGO1, AGO4, AGO6, and AGO7 act in small RNA-directed silencing, and a slicer activity has been demonstrated for AGO1 and AGO4. The AGOs bind small RNAs through their PAZ domain and the PIWI domain is responsible for RNA slicing at the center of small RNA-target hybrids that have perfect complementarity. RISCs can also catalyze translational repression when there is imperfect complementarity, and this occurs more frequently in mammalian systems. Nuclear RISC complexes also direct DNA and/or
histone methylation (See reviews: Ding and Voinnet, 2007; Matzke et al., 2009; Ruiz-Ferrer and Voinnet, 2009).

1.2.2 RNA silencing as an antiviral defense

RNA silencing was first identified as an adaptive defense against viruses in plants (Lindbo et al., 1993; Ratcliff et al., 1997; Ruiz et al., 1998). In a first study, it was found that Tobacco etch virus (TEV)-CP transgenic plants that recovered from a systemic TEV infection were resistant to subsequent infection with TEV, but not another virus, such as Potato virus Y (Lindbo et al., 1993). This resistance was not graft-transmissible to plants that had never been infected with TEV, and was not inherited vertically by progeny plants. The observed resistance was found to be due to an inability of recovered plants to support replication of TEV. This first paper alluded to a “cytoplasmic event in which specific RNA sequences were targeted, functionally suppressed and eventually eliminated from the cell.” A “protein or nucleic acid factor” was the culprit. It was noted that cellular RDR levels were known to be elevated in virus and viroid infections, and a role was predicted for RDR in generating a dsRNA trigger that resulted in production and amplification of RNAs that mediated homology-dependent silencing (Lindbo et al., 1993).

That the resistance was mediated by RNA, and not a protein against the challenge virus, was reported by David Baulcombe in 1997. A similarity was established between transgene-induced gene silencing and natural phenomena such as recovery from virus
It was found that virus-induced gene silencing (VIGS) occurs if there is sequence similarity between the virus and an endogenous gene or an introduced transgene. RNA silencing was thus understood to be a natural antiviral defense (Ratcliff et al., 1997).

These early examples involved the study of RNA viruses, which are also the most common type of plant viruses, so RNA silencing was thought to be a cytoplasmic, post-transcriptional defense. This idea has since been expanded. It is now known that RNA silencing can be cytoplasmic and nuclear, post-transcriptional and transcriptional.

1.2.3 Post-transcriptional gene silencing (PTGS)

Inverted repeat structures in viral transcripts and double-stranded replicative intermediates act as triggers for the RNA silencing pathway (Molnar et al., 2005). Viral RNA-dependent RNA polymerases (RDRs) can also produce dsRNA. Antiviral post-transcriptional gene silencing (PTGS) is characterized by the appearance of 21 and 22 nt siRNAs. DCL4 generates 21 nt siRNAs that predominate, though DCL2 can also participate in antiviral PTGS through generation of 22 nt siRNAs. DCL2 acts redundantly when DCL4 is inactivated or silenced by a viral suppressor of silencing (Blevins et al., 2006; Bouche et al., 2006; Deleris et al., 2006; Diaz-Pendon et al., 2007; Qu et al., 2008). In dcl2/4 double mutants, PTGS is completely debilitated and this results in increased host susceptibility to virus with concomittant decrease in virus-derived siRNAs. DRB4 functions in association with DCL4 in PTGS (Adenot et al., 2006; Nakazawa et al., 2007;
Qu et al., 2008). DCL1 and DCL3 may have regulatory roles, but do not directly participate in defense against RNA viruses (Diaz-Pendon et al., 2007; Qu et al., 2008).

Cellular RDRs such as RDR6, and redundantly RDR1, generate dsRNA from virus-derived ssRNA such as aberrant transcripts that lack a 5’cap or a poly-A tail, or sliced transcripts. Suppressor of gene silencing 3 (SGS3) interacts with RDR6 and assists in this process (Mourrain et al., 2000). The dsRNAs are cleaved by DCL4 (and DCL2) to amplify siRNAs, which are loaded into AGO1/7 containing RISC complexes, depending on their structural features. AGO1 is the major cytoplasmic antiviral slicer, though AGO7 can redundantly perform this function (Baumberger et al., 2005; Zhang et al., 2006; Pantaleo et al., 2007; Qu et al., 2008).

PTGS can also target transcripts produced by DNA viruses, such as geminiviruses and pararetroviruses. RNA polymerase II transcribes a DNA virus and these transcripts can be converted to dsRNA by cellular RDRs. Secondary structure in transcripts, convergent transcripts and aberrant or sliced transcripts are also sources of dsRNAs (Chellappan et al., 2004; Akbergenov et al., 2006; Moissiard et al., 2006). DCL4 (and redundantly DCL2) are the main dicers involved in PTGS against DNA viruses. In addition, DRB4, a dsRNA binding protein which partners with DCL4, is targeted by the P6 protein, a silencing suppressor of the Cauliflower mosaic virus (CaMV), underscoring the importance of this pathway to DNA virus defense (Moissiard et al., 2006).
1.2.4 Spread of the RNA silencing signal

RNA silencing is non-cell autonomous. When an endogenous gene is targeted by PTGS, the effects of silencing may be perceived 10-15 cells away from the site of induction. This occurs in an RDR6-independent process and therefore, likely does not involve amplification of siRNAs. In the dcl4 mutants, however, this cell-to-cell spread is lost, leading to the notion that DCL4-generated 21 nt tasiRNAs may be responsible for cell-to-cell spread. The nature of the mobile silencing signal is still unresolved, and could be mature tasiRNA or long siRNA precursors (Voinnet, 2005; Dunoyer and Voinnet, 2009). Transgene sequences, unlike endogenous genes, are subject to transitivity, which is an extension of silencing to sequences outside the initial target. Transitivity is believed to be mediated by RDR6 and involves secondary amplification of siRNAs. DCL4, and redundantly DCL2, are required (Moissiard et al., 2007). High levels of transgene expression may play a role in this process. Cell-to-cell spread as a consequence of transgene silencing can occur in an rdr6 mutant. There is indirect evidence that the short-distance spread of silencing may occur through plasmodesmata. Guard cells of stomata that are symplastically isolated due to blockage of plasmodesmata escape silencing (Voinnet et al., 1998; Himber et al., 2003; Kalantidis et al., 2006). Besides DCL4, DRB4 and AGO1, Pol IV (NRPD) and RDR2 are additional factors implicated in spread. Pol IV and RDR2 may assist in the generation of dsRNA which is a substrate for DCL4 and important for the biogenesis of tasiRNAs (Smith et al., 2007). Alternatively, Pol IV and RDR2 may affect silencing downstream of DCL4, by
facilitating silencing spread by effecting the perception of the silencing signal in the recipient cells (Dunoyer et al., 2007). A SNF2-domain containing protein CLASSY1 (CLSY1) may play a role between Pol IV and RDR2 (Smith et al., 2007). DCL3, AGO4 and the 24 nt siRNA pathway are postulated to play no role in spread, or have a negative role. DCL1, surprisingly, was uncovered in a screen for components that enhance the spread of silencing, possibly by generating substrates for DCL4 (Dunoyer et al., 2007).

Besides cell-to-cell spread, RNA silencing can spread long-distance and cause systemic silencing. After induction at source, the systemic silencing signal travels through the vasculature and induces silencing in sink leaves (Palauqui et al., 1997; Voinnet et al., 1998). From grafting studies, it has been found that systemic silencing was abolished in nrpd1a and rdr2 mutant scions, and delayed in ago4 scions. None of these mutations had any effect in the rootstock, ruling out an involvement of this pathway in signal production (Brosnan et al., 2007). The Pol IV-RDR2-DCL3-AGO4 pathway is likely required upstream of RDR6–DCL4/DCL2 in systemic silencing, presumably for perception of the long-distance signal in the recipient cell (Dunoyer and Voinnet, 2008). The signal could traverse tissues that do not express homologous sequences to which the silencing was originally directed. Systemic signal can likely be generated against both transgenes and endogenous genes, but RDR6-mediated siRNA amplification is required to deliver the silencing signal outside of the phloem in a third kind of spread, called extensive local spread that occurs in sink leaf tissues.

Extensive local spread requires RDR6, DCL4 and the RNA helicase, SDE3 (SUPPRESSOR OF DEFECTIVE SILENCING 3). Secondary amplification of siRNAs
occurs. Intriguingly, extensive local spread has never been reported in source tissues, although substantial primary siRNAs are produced, and RDR6 as well as DCL4 are available in these tissues along with their substrates (Kalantidis et al., 2008).

1.2.5 Viral suppressors of PTGS

Viruses encode proteins that act as suppressors of host RNA silencing. They may attack multiple steps of the silencing pathway and suppress silencing either upstream of small RNA production, or by binding and sequestering siRNAs and preventing RISC loading, or by inhibiting key proteins of the silencing pathway (Diaz-Pendon et al., 2008; Roth et al., 2004; Donaire et al., 2009; Alvarado and Scholthof, 2009). Some suppressors employ multiple mechanisms. Helper Component Protease (HcPro) from the potyvirus, *Tobacco etch virus* was the first viral silencing suppressor identified. HcPro was shown to suppress transgene-induced silencing and VIGS (Anandalaksmi et al., 1998; Briugneti et al., 1998). Further, it could reverse established systemic silencing (Llave et al., 2000). HcPro acts by binding siRNAs and inhibiting RISC loading, and in addition, HcPro interacts with rgsCAM, an endogenous suppressor of silencing (Anandalakshmi et al., 2000). The protein p19 from the tombusvirus *Tomato bushy stunt virus* can suppress systemic spread and transgene-mediated gene silencing. p19 acts by binding and sequestering 21 nt siRNAs in a caliper-like fashion, thereby suppressing establishment of PTGS, but is unable to reverse established PTGS (Qu and Morris, 2002; Silhavy et al., 2002). It is often coexpressed and used to boost gene expression in transient
agroinfiltration based-systems (Voinnet et al., 2003). *Cucumber mosaic virus* protein 2b can suppress systemic silencing but not local silencing (Brigneti et al., 1998; Guo and Ding, 2002). CMV 2b has been shown to interact with AGO1 and interfere with slicer activity (Zhang et al., 2006). *Turnip crinkle virus* coat protein blocks local and systemic silencing by interacting with and suppressing the activity of DCL2 (Qu et al., 2003). In addition, TCV CP binds siRNAs in a size-independent fashion, and likely also sequesters dsRNA precursors for siRNA production (Meraï et al., 2006). Polerovirus P0 is an F-box containing protein that targets and mediates AGO1 degradation (Baumberger et al., 2007; Bortolamiol et al., 2007). Thus many silencing suppressors act by multiple mechanisms, inhibiting several discrete steps of the silencing pathway. In addition, there seem to be common strategies shared by multiple suppressors. For example, binding siRNAs is a general feature shared by several silencing suppressors including p19, HcPro, 2b, inhibiting HEN1 methylation and RISC loading.

1.2.6 RNA-directed DNA methylation and transcriptional gene silencing

RdDM was first discovered in viroid-infected plants. Viroid sequences were integrated into tobacco, and these transgenic plants were subsequently infected. It was found that replicating viroid RNA could direct methylation of homologous DNA sequences (Wassenegger et al., 1994). This was found to be mediated by dsRNA which was processed to small RNAs (siRNAs). When these siRNAs guide promoter methylation, this constitutes transcriptional gene silencing (TGS), while post-
transcriptional gene silencing may be associated with methylation in the gene body that impacts gene expression (Ingelbrecht et al., 1994; English et al., 1996; Jones et al., 1998; Hohn et al., 1996).

In *Arabidopsis*, the prevailing model for RNA directed DNA methylation involves a sequence of events mediated by enzymes Pol IV/V-RDR2-DCL3-AGO4. DsRNA, for example, convergent transcripts in a geminivirus infection, may be cleaved by DCL3 into 24 nt siRNAs, the guide strands of which are then incorporated into an AGO4 containing RISC complex. siRNAs that carry a 5’ adenosine are preferentially incorporated into AGO4 complexes (Mi S. et al., 2008). AGO4 physically interacts with a plant-specific polymerase, Pol V through conserved WG/GW motifs (El-Shami et al., 2007). Pol V is recruited to DNA by chromatin remodelers such as DEFECTIVE IN MERISTEM SILENCING (DMS3) or DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1 (DRD1) and generates a non-coding transcript, upon which AGO4 docks its siRNA cargo in a homology-dependent manner. SUPPRESSOR OF TY INSERTION 5-LIKE (SPT5L) or KTF1 is an RNA-binding adaptor protein that binds Pol V transcripts as well as AGO4 through WG/GW motifs. Thereafter, AGO4 has two independent activities: AGO4 has a catalytic slicer activity and AGO4-generated sliced transcripts serve as substrates for RDR2 and amplify secondary siRNAs that maintain methylation. AGO4, in addition, has a non-catalytic activity through its interaction with Pol V that is used to guide methylation of DNA and histones. It is not yet known how methyltransferases are then recruited to carry out de novo and maintenance methylation (Law and Jacobsen, 2010).
In fission yeasts that lack DNA methylation, heterochromatin can still be assembled, and can spread 100s of base pairs from initial site of siRNA-DNA homology. siRNAs pair with Pol II transcripts and this RNA-RNA interaction guides DNA methylation to regions where there is homology between siRNA and DNA (Irvine et al., 2006; Buhler and Moazed, 2007; Grewal and Elgin, 2007). But in most eukaryotic systems, including *Arabidopsis*, methylation is confined to regions of homology, and sequences as small as 30 nt can be targeted. This suggests that DNA methylation is the primary epigenetic mark that guides histone modifications (Huettel et al., 2007; Matzke et al., 2007).

Methylation in plants occurs at cytosines in all sequence contexts: CG, CNG and CHH (H=A,T,C). DOMAINS REARRANGED METHYLTRANSFERASE 1 and 2 (DRM1/2) are *Arabidopsis* homologs of mammalian Dnmt3 and direct *de novo* methylation in all sequence contexts. Methylation at CHH residues requires siRNA signalling for maintenance by DRM2, the protein expressed in *Arabidopsis* (Cao et al., 2003). CG methylation is carried out by METHYLTRANSFERASE 1 (MET1), previously called DECREASE IN DNA METHYLATION 2 (DDM2), the plant homolog of Dnmt1. CG methylation is symmetric and is maintained by MET1 and DECREASE IN DNA METHYLATION 1 (DDM1), a SWI2/SNF2 chromatin remodeler (Jeddeloh et al., 1999; Morel et al., 2000; Jones et al., 2001). DDM1 and MET1 were first identified in a screen for lines with decreased DNA methylation (Vongs et al., 1993). MET1 can carry out *de novo* as well as maintenance functions and, when provided with *in vitro* substrates, shows a preference for hemimethylated DNA (Cao et al., 2002). CG methylation is more
stable and can coordinate non-CG methylation (Mathieu et al., 2007). KRYPTONITE 2 (KYP2), also known as SU(VAR)3-9 HOMOLOG 4 (SUVH4) is the histone 3 lysine 9 (H3K9) methyltransferase. CNG methylation is maintained by DNA methyltransferase, CHROMOMETHYLASE 3 (CMT3) (Lindroth et al., 2001; Jackson et al., 2002; Malagnac et al., 2002; Jackson et al., 2004; Lindroth et al., 2004). The cmt3 mutant was identified in a reversion screen and was found to be impaired in CNG methylation (Lindroth et al., 2001) and CHH methylation (Bartee et al., 2001). The kyp2 mutant shows similar methylation deficiency as cmt3 mutants, but encodes H3K9 methylation which may have an indirect effect in maintaining DNA methylation. CMT3 can bind methylated histone peptides in vitro, and SUVH4 (through an amino terminal SRA domain) can bind methylated DNA in all sequence contexts, with a preference for CNG and CHH methylation (Johnson et al., 2002; Lindroth et al., 2004). Further, H3K9 methylation is associated with CNG methylation throughout the genome, especially at centromeric and pericentromeric heterochromatic regions, suggesting that CMT3 and SUVH4 reinforce each other (Bernatavichute et al., 2008). DDM1 is required for maintenance of CG methylation as well as H3K9 methylation (Gendrel et al., 2002; Johnson et al., 2002; Lippman et al., 2004).

In mammals, dnmt1 mutant mice are embryonic lethal, while in plants this is not the case because of considerable redundancy in the methylation pathway. DRM2 can methylate all sequence contexts, and DRM2 and CMT3 both function to establish and maintain non-CG methylation. AGO6 acts redundantly with AGO4 (Zheng et al., 2007).
SUVH4, SUVH5 and SUVH6 are H3K9 methyltransferases that can function in a redundant manner (Ebbs et al., 2005; Ebbs et al., 2006).

Genome-wide analysis of *Arabidopsis* DNA methylation using McrBC digestion, an enzyme that only cleaves methylated DNA, MeDIP (methylcytosine immunoprecipitation), bisulfite sequencing, or chromatin immunoprecipitation followed by DNA microarray (ChIP-on-chip) has revealed that DNA methylation predominates within the gene body, and not at the 5’ or 3’ ends. Most genes lack promoter methylation, but 20-30 per cent of genes show significant gene body methylation, which is almost exclusively found at CG sites (Zhang et al., 2006; Vaughn et al., 2007; Zilberman and Henikoff, 2007; Cokus et al., 2008; Lister et al., 2008). Transposable elements are methylated at both CG and non-CG sites, and gene body methylation may play a role in heterochromatin formation. Gene expression analysis of *metl* mutants showed that gene body methylation affects gene expression by preventing transcriptional elongation, with a striking effect on short genes. Interestingly, moderately expressed genes show the highest levels of gene body methylation and this could be directed by aberrant siRNAs produced from cryptic start sites exposed in the gene body. Highly expressed genes with frequent passage of RNA Pol II are less likely to produce these aberrant transcripts, while weakly expressing genes lie in condensed chromatin that prevents transcription initiation from cryptic start sites. Thus a feedback may exist between transcription and methylation-based silencing (Zilberman and Henikoff, 2007).
1.2.7 Geminivirus silencing suppressors

Plant DNA viruses encode silencing suppressors to counteract post-transcriptional (PTGS) and transcriptional gene silencing (TGS). Geminiviruses encode multiple silencing suppressors AL2/AC2/TrAP/C2, and AC4 proteins in bipartite geminiviruses; AV2 in monopartite geminiviruses. In addition, the βC1 protein encoded by satellite DNA associated with certain geminiviruses is also a silencing suppressor. *African cassava mosaic virus* AC2 was the first DNA virus suppressor identified, which could reverse established PTGS of a transgene (Voinnet et al., 1999).

As described in Section 1.1.2, the AL2 protein has an N-terminal basic region that contains an NLS, a central CCHC zinc finger-like motif and an acidic C-terminal that contains the minimal transcription activation domain. L2 shares the CCHC motif but lacks transcription activation function.

The requirements for suppressor function of AL2 have been studied in different viruses. The C2 protein of the monopartite *Tomato yellow leaf curl virus* requires the zinc finger domain to suppress silencing (van Wezel et al., 2002). The AC2 protein from *Mung bean yellow mosaic virus* (MBYMV) requires nuclear localization and transcriptional activity for suppression function (Trinks et al., 2005). In *Arabidopsis* protoplasts, it was found that AC2 expression leads to an increase in expression of about 30 host genes, one of which, Werner exonuclease-like 1 (WEL1), may act as a suppressor of PTGS (Trinks et al., 2005). WEL1 is a homolog of Werner exonuclease, WEX, which is required for PTGS against transgenes (Glazov et al., 2003), and upregulation of WEL1
might suppress WEX function (Trinks et al., 2005). AL2 does not bind siRNA or miRNA, and weakly binds ssDNA and dsDNA in a sequence-independent manner (Chellappan et al., 2005; Wang et al., 2005). Since AL2 requires an intact NLS, CCHC motif and C-terminal transcriptional activation domain in order to reverse PTGS and inhibit systemic spread, AL2 probably acts as a suppressor of PTGS either by activating transcription of host genes that are suppressors of silencing, or its activation domain mediates interaction with silencing pathway components in the nucleus (Bisaro, 2006). AL2 and its curtovirus counterpart L2, are both pathogenicity factors (Sunter et al., 2001; Hao et al., 2003; Wang et al., 2003; Baliji et al., 2007). L2 lacks a transcription activation domain and only shares the CCHC motif with AL2. Both L2 and AL2 interact with and inactivate SNF-related kinase, SnRK1 and adenosine kinase, ADK. Through interaction with SnRK1, these proteins inhibit the cellular stress response, a component of basal defense (Hao et al., 2003). Both AL2 and L2 can suppress silencing through interaction and inhibition of host enzyme ADK, which results in subsequent inhibition of the methyl cycle. Previous work in our lab showed that AL2 and L2 can inactivate ADK in vitro, in E.coli and in yeast. ADK activity was also found to be decreased in geminivirus-infected tissues and upon transgenic expression of AL2 and L2 (Wang et al., 2003). In plant cells, using bimolecular fluorescence complementation (BiFC) analysis, it was found that ADK interacts with AL2 and L2 mainly in the cytoplasm, while AL2 can interact with itself in the nucleus and activate transcription (Yang et al., 2007). ADK inactivation by AL2/L2 suppresses PTGS (Wang et al., 2005). ADK performs a variety of important functions in a cell, and is needed to maintain the methyl cycle (Lecoq et al., 2001; Weretilnyk et al., 2001).
2001; Moffat et al., 2002). The hog1 gene, identified in a screen for silencing suppressors, encodes methyl cycle enzyme S-adenosyl homocysteine hydrolase (SAHH) (Rocha et al., 2005). In our lab, it was found that knocking down SAHH in a transient agroinfiltration-based silencing suppression assay, resulted in suppression of GFP (Buchmann, unpublished).

It was recently shown that AL2 and L2 act as silencing suppressors that can also reverse TGS (Buchmann et al., 2009). A transgenic N. benthamiana line 16c containing a transcriptionally silenced 35S promoter-GFP transgene (16-TGS) was inoculated with a PVX vector designed to express AL2 and L2 proteins. Expression of AL2 and L2 restored GFP expression in these silenced lines, indicating that AL2 and L2 can reverse TGS. Since L2 and the transcriptionally inactive AL21-114 mutant could both reverse TGS, it was concluded that this reversal of silencing is not transcription-dependent. Knocking down SAHH and ADK expression in 16TGS plants also resulted in TGS reversal. Infecting 16-TGS plants with BCTV resulted in reversal of TGS in an L2-dependent manner. AL2 and L2 proteins were inducibly expressed under a dexamethasone responsive promoter in transgenic Arabidopsis lines, resulting in release of transcriptional silencing at endogenous loci such as AtSN1 (SINE element), Athila (LTR element), an F-box pseudogene and CACTA (DNA transposon). Silencing at all loci was reversed by AL2, but CACTA silencing was not reversed by L2, implying that a transcription-dependent mechanism was likely required to reverse CG methylation that is mainly responsible for CACTA silencing. All other loci were found to be expressed in mutants lacking non-CG methylation (Tran et al., 2005; Zhang et al., 2006). At the
AtSN1 locus, methylation was found to be reduced about 15% by AL2 expression. Genome-wide, methylation was reduced about 50% as assayed by a methylation-sensitive extension assay. Thus AL2 and L2 are suppressors of PTGS and TGS, and AL2 can act as a silencing suppressor in a transcription-dependent as well as a transcription-independent manner. AL2 and L2 can suppress PTGS as well as TGS in a transcription-independent manner (Wang et al., 2005; Buchmann et al., 2009).

In addition to AL2, AC4 has been identified as a silencing suppressor. AC4 has the least sequence conservation of all geminivirus proteins. Mutagenesis or transgenic overexpression of AC4 in some cases, results in no phenotype, while others produce phenotypes implicating it as a movement protein or a symptom determinant (Jupin et al., 1994; Latham et al., 1997; Krake et al., 1998). AC4 from Sri Lankan cassava mosaic virus and ACMV (Cameroon) can both suppress RNA silencing in an Agrobacterium-based silencing assay in N. benthamiana 16c plants (Vanitharani et al., 2004). These viruses are also associated with recovery-type symptoms in cassava. It was also found that AL2 and AC4 can act synergistically. For instance, mixed infection of cassava by ACMV (recovery-type, with a strong AC4 suppressor) and East African cassava mosaic virus (EACMV; non-recovery-type, with a strong AC2 suppressor) causes very severe disease. This suggests that AC2 and AC4 act at discrete steps in the silencing pathway (Vanitharani et al., 2005). AC4 from ACMV causes developmental defects when expressed as a transgene. The defects are associated with reduced accumulation of a specific miRNA, miR159 and over-accumulation of target mRNAs. ACMV AC4 also
binds ssRNA, suggesting that AC4 proteins act by binding miRNA and interfering with RISC loading (Chellappan et al., 2005).

The AV2 protein of monopartite begomoviruses such as TYLCV has been identified as a suppressor of PTGS through a mechanism that depends on its ability to interact with SISG3, the tomato homologue of Arabidopsis (SUPPRESSOR OF GENE SILENCING 3 (SGS3). SGS3 is a host factor that acts with RDR6 in the synthesis of dsRNA. Thus, V2 could interfere with siRNA amplification (Zrachya et al., 2007; Glick et al., 2008).

The most recently discovered geminivirus silencing suppressor is βC1 in β-satellites. The β-satellites are a non-autonomous small ssDNA family of approximately ~1360 nt that rely on a host geminivirus for replication and spread (Briddon et al., 2008). The β-satellites encode a single gene, βC1, which can act as a pathogenicity determinant and is required for the efficient infection of some hosts (Briddon et al., 2001; Saeed et al., 2005). βC1 has been identified as a suppressor of PTGS that localizes to the cell nucleus (Cui et al., 2005). When expressed using PVX or as a transgene, βC1 causes severe developmental defects, suggesting it may act in the miRNA pathway (Cui et al., 2004; Saunders et al., 2004; Saeed et al., 2005). Other studies have suggested that βC1 may also be involved in virus movement (Briddon et al., 2008). Recent evidence found in our lab suggests that βC1 acts as a suppressor of transcriptional gene silencing and likely acts by targeting a methyl cycle enzyme, SAHH and blocking methylation (Figure 1.5). In addition, preliminary evidence suggests that it may target methyl cycle components involved in PTGS.
1.3 Chromatin structure and epigenetics

1.3.1 Nucleosome structure

In 1974, Roger Kornberg described the nucleosome as a basic unit of chromatin. It consists of a histone core with two copies each of histones H2A, H2B, H3 and H4 forming an octamer. DNA (146 bp) is wrapped 1.75 times around this octamer. Histone H1 binds to a linker DNA between nucleosomes, at the point of entry and exit of DNA into the nucleosome. This unit, comprising a nucleosome plus linker DNA and histone H1, is called a chromatosome (Figure 1.6).

Evidence for nucleosome structure comes from electron micrographs that revealed a beaded pattern with uniform spacing, which has been confirmed by analysis of histone-histone interactions and studies of the digestion products of chromatin following DNase I or micrococcal nuclease treatment. Core nucleosome structure is conserved across all species, while the length of linker DNA varies (Cheah and Osborne, 1977). In plants, 22 bp of DNA binds histone H1. The linkers between nucleosomes vary from 0-90 bp in a species-specific and cell-type specific manner, resulting in 168 to 258 bp of DNA per chromatosome (Spiker, 1985). Histones are highly conserved among eukaryotes with little variation in amino acid sequence (Malik and Henikoff, 2003).

Histones are assembled as octamers, containing two H2A-H2B dimers, and an H3-H4 tetramer. The H2A-H2B dimer and H3-H4 tetramer are formed by hydrophobic interactions. Weak hydrogen bonds between histidine-lysine and histidine-tyrosine tether
the dimer to the tetramer (Eickbush and Moudrianakis, 1978). The H2A-H2B dimers associate with the H3-H4 tetramer in the presence of DNA. DNA binds to positively charged histones through electrostatic interactions and hydrogen bonds. The histones contain a conserved C-terminal histone fold domain that participates in interactions that assemble nucleosome structure, and unique N-terminal tails that are subject to various modifications as described in Section 1.3.2. Nucleosomes bind DNA in a non-sequence specific manner, but contact is preferentially through the AT-rich minor groove, i.e. once every 10 bp. The reason for this is that DNA bends easier in this orientation to be wrapped around nucleosomes (Ramakrishnan, 1997).

Chromatosomes coil into a 30 nm helical structure called 30 nm fiber, which opens into a 10 nm ‘beads-on-string’ fiber when RNA Pol II transcribes euchromatin.

1.3.2 Histone tail modifications

Histone tails can be post-translationally modified at their N-termini by acetylation, methylation, phosphorylation, ubiquitination, ADP-ribosylation, etc. These covalent histone modifications can extend the information contained in DNA (and the genetic code) leading to increased epigenetic informational and regulatory complexity. The histone code hypothesis was first proposed by Allis and Strahl in 2000, who suggested that “distinct histone modifications, on one or more tails, act sequentially or in combination to form a 'histone code' that is, read by other proteins to bring about distinct downstream events.”
These modifications have different outcomes on transcription and may act on their own or in combination with other histone marks. For example, H3K4 and H3K36 methylation are associated with activation of gene transcription, while H3K9, H3K27, H4K20 and H4R3 marks are associated with repression. Ubiquitination of H2B is an active mark. Acetylation of lysine and phosphorylation of serine and threonine residues is usually associated with activation (reviewed in Pfluger and Wagner, 2007).

All of these modifications alter access to DNA by chromatin remodelers and/or transcription factors. Histone lysine acetylation neutralizes the positive charge of the -N group, and introduces oxygen which is a H-bond acceptor, thus opening chromatin and recruiting other factors. Ubiquitination of lysine results in addition of a bulky moiety that increases the distance of histones from DNA. Lysine methylation increases the hydrophobicity and cationic nature of the -N group, contributing to closing of chromatin. Lysines can be mono- (me1), di- (me2), or tri-(me3)-methylated. From the modifications studied so far, it appears that methylation at a certain lysine residue, irrespective of whether it is mono-, di- or trimethylated, results in the same consequence i.e. it functions as an active or a repressive mark. For instance, H3K4me1, H3K4me2 and H3K4me3 are all associated with activation. Similarly, all H3K9 methylation has been associated with repression. But interestingly, while H3K9me1 and H3K9me2 marks occur most commonly in heterochromatin, H3K9me3 occurs in euchromatin. This is also the case with H3K27, where H3K27me1 and H3K27me2 occur in heterochromatin, while H3K27me3 occurs in euchromatin. H3K36 methylation is an activating modification and H3K36me1 and H3K36me3 occur in euchromatin, while H3K36me2 occurs in both
heterochromatin and euchromatin. H4K20 monomethylation is found in heterochromatin, while H4K20me2 and H4K20me3 are repressive marks that occur in euchromatin.

In *Arabidopsis*, the genomic distribution of H3 lysine methyl marks has been determined. H3K4 dimethylation (H3K4me2) is depleted in transposons (Lipmann et al., 2004). H3K9 dimethylation (H3K9me2), on the other hand, is almost exclusively found in transposons and other repeats (Lipmann et al., 2004; Turck et al., 2007). H3K27 trimethylation (H3K27me3) is found at a large number of silenced loci, and it is thought that it could be involved in maintaining tissue-specific gene expression. The function of H3K9 trimethylation has not yet been characterized in plants, but it is localizes to genes rather than repeat sequences and occurs at distinct loci from H3K4me2 and H3K27me3 marks (Zhang et al., 2007). H3K4me2 and H3K4me3 accumulate in promoters of 5’ regions of active genes, while H3K4me1 occurs within genes. H3K4me2 and H3K4me3 do not correspond to regions of DNA methylation, but H3K4me1 correlates with CG methylation within the gene body (Zhang et al., 2009).

1.3.3 Histone variants

Histones are expressed during the S phase of the cell cycle and package newly-synthesized DNA (Henikoff and Ahmad, 2005). Non-canonical histones, or histone variants of H2A, H2B and H3, on the other hand, are expressed during all phases of the cell cycle and are deposited in nucleosomes in differentiated cells in a replication-independent manner (Malik and Henikoff, 2003; Bernstein and Hake, 2006). Canonical
histones are usually encoded by multi-copy genes, while histone variants are encoded by single-copy genes. Genes encoding histone variants contain introns and transcripts are often polyadenylated. These features are important in post-transcriptional regulation and are absent in canonical histones. Canonical histones instead have a 3’ 25-26 nt hairpin structure that is bound by a stem-loop binding protein (Old and Woodland, 1984; Kamakaka and Biggins, 2005).

*Arabidopsis* has three isoforms of linker histone H1: two of these, H1-1 and H1-2 share extensive sequence similarity and probably arose from gene duplication, while the third, H1-3 is a divergent, drought-inducible variant (Wierzbicki and Jerzmanowski, 2004; Jerzmanowski et al., 2000). H1-1 and H1-2 variants localize mainly to heterochromatin, while H1-3 appears enriched in euchromatin. H1-3 is conserved among flowering plants and similar to H1\(^0\)/H5 variants in vertebrates is rich in arginine (Ascenzi and Gantt, 1999).

H2A variants differ from canonical H2A in sequence and length of their C-terminal tails, as well as their genome distribution. H2A.X and H2A.Z are constitutively expressed and localize throughout the genome, although H2A.Z is enriched in the intergenic regions. In general, H2A has the largest number of variants among core histones. H2A variants include H2A.Z, MacroH2A, H2A-Bbd, H2AvD and H2A.X, but some of these such as MacroH2A and H2A-Bbd are vertebrate-specific. H2A.Z is the most conserved among all organisms. The *Arabidopsis* genome contains 13 histone H2A genes, which are categorized into four major groups, plus one unique member. *Group I* consists of four histone H2As (H2A-1, -2, -10, and -13) that share over 92% amino acid
sequence identity. *Group II* consists of H2A-3 and H2A-5 (~76% amino acid sequence identity with H2A-1) and contain a SQEF motif in their C-terminal regions. This sequence is characteristic of H2AX, a histone H2A variant involved in double-strand DNA break repair (Rogakou et al., 1998). H2AX at sites of double-stranded breaks is phosphorylated at a C-terminal serine, and phosphorylation spreads to other H2AXs along the chromosome, recruiting INO80 chromatin remodelers. *Group III* has three histone H2As, H2A-6, -7, and -12 (~57% amino acid sequence identity with H2A-1) that have a plant-specific SPKK motif in their C-terminal regions. This motif is found in plant histone H2As that are cell cycle regulated (Huh et al., 1995). The SPKK motif can bind to the DNA minor groove in A/T-rich regions (Suzuki, 1991). *Group IV* (~52% sequence identity with H2A-1) includes H2A-8, -9, and -11. Group IV members show ~80% identity with histone H2A F/Z variants in other organisms. Histone H2A-4 is unique and has only 35% amino acid sequence identity with the other histone H2A family members and contains only 119 amino acids instead of 130-150 amino acids of other H2As. It is possible that H2A-4 maybe a pseudogene.

The only variants of H2B that have been identified are reported to have very specialized functions in transcription repression and chromatin packaging in sperm. There are none reported in plants.

The most prominent of H3 variants is CenH3, which is necessary for chromosome segregation during mitosis (Dalal et al., 2007). CenH3 contains a poorly conserved long extension of the N-terminal tail (Malik and Henikoff, 2003). Another H3 variant, H3.3 is similar in amino acid composition and length to H3 (also called H3.1), except at certain
sites [S31-A87-I89-G90 in H3.3; A31-S87-V89-M90 in H3 in Drosophila] (Waterborg and Robertson, 1996; Malik and Henikoff, 2003). In plants, positions 31, 87 and 90 are variant, and there is an additional difference at position 41 (Okada et al., 2005).

Incorporation of variants causes chromatin to become more open or compacted, facilitating or hindering transcription. In Drosophila, Xenopus, and mammals, H3.3 is associated with transcription (McKittrick et al., 2004; Loyola et al., 2006; Mito et al., 2007; Ng and Gourdon, 2008). Variant tails could be differentially modified. For instance, in Arabidopsis, the repressive H3K27 methyl mark appears enriched in histone H3, while the active H3K36 methyl mark is enriched in the variant histone H3.3 (Johnson et al., 2004). In general, H3.3 appears to condition a transcriptionally active state (Henikoff et al., 2004; Loyola and Almouzni, 2007).

There are no known sequence variants of histone H4.

1.3.4 Nucleosome assembly

CAF1 (Chromatin Assembly Factor 1) is involved in replication-dependent assembly of H3 and H4 (Verreault et al., 1996; Quivy et al., 2001; reviewed by Ingouff and Berger, 2010). CAF1 function is conserved from yeast to higher eukaryotes including Arabidopsis (van Nocker, 2003). In yeast, CAF1 is required for heterochromatin maintenance. The HIRA (HIR histone cell cycle regulation defective homolog A) complex is involved in H3.3/H4 deposition. CAF-1 and HIRA complexes contain CAC1, CAC2 and CAC3 proteins. The HIRA complex, in addition contains the HIRA protein. Both CAF1 and HIRA complexes also contain ASF1a and ASF1b histone chaperones.
(Tagami et al., 2004). In plants, the FASCIATA complex is the homolog of CAF1, whose components are FASCIATA1, FASCIATA2 and MULTICOPY SUPPRESSOR OF IRA1 (AtMSI) (Chen et al., 2008). An H3.3 chaperone homolog in plants, called HIRA1, has also been identified (Phelps-Durr et al., 2005).

1.3.5 Chromatin remodeling

Chromatin may exist as densely packed heterochromatin or open euchromatin that is accessible to transcription factors. Access to DNA wound in nucleosomes may be catalyzed by chromatin remodelers. ATP-dependent chromatin remodelers could function to regulate DNA access in one of three ways: (a) Transient DNA loops are created on the surface of nucleosomes by SWI/SNF proteins, exposing regulatory sites, (b) Nucleosome sliding is catalysed by Isw2 remodelers, that reposition nucleosomes laterally (c) Nucleosome eviction and replacement by RSC complex and histone chaperones, e.g. FACT (facilitates chromatin transcription), Asf1 (anti-silencing function 1), Chz1(H2A.Z-specific chaperone 1) (d) Replacement of histone subunits with variants, e.g. H2A is replaced by H2A.Z by SWR1 complex and H3 by H3.3 by CHD1 complex. (Clapier and Cairns, 2009) (Figure 1.7).

The Arabidopsis genome encodes about 40 putative SWI2/SNF2 chromatin remodelers. DRD1 is a plant-specific member of the SWI2/SNF2 family of ATP-dependent chromatin remodelers originally identified in a genetic screen that also yielded nrpe (Pol V) (Huettel et al., 2007). By comparing Pol V recruitment to three different
genomic loci (IGN5, IGN6 and solo LTR) in drd1 mutants relative to wild-type, it was found that DRD1 recruits Pol V to chromatin (Wierzbicki et al., 2008). DRD1 plays a role in de novo methylation in all sequence contexts, and also plays a role in erasure of methyl marks in the absence of the RNA trigger.

DDM1, another SWI/SNF chromatin remodeler in plants, has a homolog in mammals (Lsh) that has been shown to facilitate de novo methylation in mice. DDM1 assists MET1 in maintaining CG methylation, but it is not clear if it plays a role in de novo methylation in plants.

SUVH4/Kryptonite, is an H3K9 methyltransferase that helps maintain CNG methylation through interplay with CMT3. SUVH5 and SUVH6 act redundantly to maintain non-CG methylation at specific loci (Ebbs et al., 2005, Ebbs et al., 2006).

1.3.6 Can geminivirus chromatin be a model for eukaryotic chromatin?

Geminivirus dsDNA is packaged into cellular histones in vivo and is organized into episomal minichromosomes (Pilartz and Jeske, 1992). The only information available about plant virus minichromosome structure comes from what has been studied in the geminivirus, *Abutilon mosaic virus* (ABMV). Electron microscopic analysis revealed a ‘beads-on-a-string’ appearance that resembled eukaryotic chromatin. Analysis of nucleosome ladders, and an examination of topoisomers revealed that geminivirus minichromosomes consist of 12 nucleosomes per minichromosome with nucleosome-free gaps that correspond to the intergenic region and the downstream AL2 promoter on
DNA A or for the BC1 transcript on DNA B (Pilartz and Jeske, 2003). Interestingly, minichromosome populations can also exist in two alternate phases containing 11 or 13 nucleosomes each. Assuming an equal distribution of 200 bp DNA containing nucleosomes in a 2.6 kb genome, 13 nucleosomes would represent the closed or repressed form of a geminivirus minichromosome, while 11-12 would be the open or active form. Thus, it appears that geminivirus nucleosomes are not fixed in position, but that the minichromosome dynamically alternates between active and repressed forms (Pilartz and Jeske, 2003). Open nucleosome conformations may facilitate binding of chromatin remodelers and transcription factors.

Chapter 1 of this dissertation includes the first analysis of geminivirus associated histone-tail modifications, using Cabbage leaf curl virus in N. benthamiana and Arabidopsis. Using chromatin immunoprecipitation, evidence was found for both H3K9 dimethyl and H3 acetyl marks associated with the geminivirus genome, further supporting the idea that repressed and active populations of minichromosomes might exist in infected cells (Raja et al., 2008).

It is proposed in this dissertation that geminivirus minichromosomes can be used as a model to study various aspects of eukaryotic plant chromatin – the RNA silencing pathways leading to chromatin modifications, plant-specific polymerases, various histone tail modifications, chromatin remodelers, and histone-code readers. The work presented in this dissertation illustrates that geminiviruses may be used to screen for novel components, and chromatin architecture and modification mechanisms may be studied using geminiviruses (Figure 1.8).
Figure 1.1 Geminivirus gene organization

The genome structures of representative geminivirus genera are shown below. *Beet curly top virus* (BCTV, *Curtovirus*), *Tomato yellow leaf curl virus* (TYLCV, monopartite, *Begomovirus*), and *Tomato golden mosaic virus* (TGMV, A and B) and *Cabbage leaf curl virus* (CaLCuV, A and B) (bipartite, *Begomovirus*) are some examples of the genera represented. The solid arrows indicate the positions of viral genes, which are designated by number and the direction of transcription: leftward (L, complementary sense) or rightward (R, viral sense). Viral genes are also denoted by names based on their first described functions, such as Rep (replication initiator protein), TrAP (transcriptional activator protein), REn (replication enhancer), NSP (nuclear shuttle protein) and CP (coat or capsid protein). The position of the conserved hairpin is indicated by an asterisk within the intergenic region (IR), which is marked in color. The common region (CR), a sequence of ~230 bp that is nearly identical in the bipartite begomoviruses, is indicated by a colored box (Bisaro, 1996). Major transcripts (indicated by thin arrows) are shown using data from TGMV as a model for bipartite begomoviruses. Note that the bidirectional transcription units overlap at their 3' ends. In some cases, transcript names (AL-61, AL-1935, AL-1629) are provided.
Figure 1.1 Geminivirus gene organization.

Raja et al., 2010
Figure 1.2 Geminivirus replication origin. The replication origin of TGMV A is illustrated here. The origin includes Rep binding sites, the invariant sequence in the hairpin (TAATATTAC), and the site where replication is initiated. TATAA sequences, Rep and CP start sites, a putative G-Box element and a putative TrAP responsive element are indicated. The Rep binding site in inverse orientation has been identified by sequence analysis. Nucleotide coordinates are indicated below.
Figure 1.3 Geminivirus replication.

Geminivirus ssDNA is first converted into dsDNA, which is used as a template for transcription of viral genes such as Rep (AL1), the major replicative protein, and REn (AL3), a replication enhancer. The dsDNA acts as a template for rolling circle replication to produce new ssDNA molecules. Replication is facilitated by Rep (AL1, shown in blue) binding to Rb and inactivating it to condition S-phase. Rep also acts as an origin recognition protein by binding the origin of replication and recruiting host replication machinery. Rep additionally acts as a “nick-and-stick” protein to initiate and terminate replication. It nicks the viral strand at TAATATT/AC, creating a 3’-OH terminus to prime DNA synthesis, and then binds to the 5’-end of the nick site through a tyrosine residue. Replication proceeds 5’ to 3’ around the entire circular template and is carried out by host polymerases. When the origin is regenerated, Rep again nicks the nascent molecule releasing a unit length single strand which is subsequently ligated to circular form. Rep then associates with the 5’-terminus generated by the nick and the process continues. The mechanism by which replication enhancer protein (REn; AL3) stimulates replication is not yet known. The transcriptional activator protein (TrAP; AL2) is required for the expression of late viral genes including coat protein (CP) and MP. Newly-synthesised ssDNA (indicated in red) can be a template for transcription and replication or associate with the coat protein and be packaged into virions, or move outside the nucleus and to the neighboring cells with the help of the nuclear shuttle protein (NSP; BR1) and the movement protein (MP; BL1).
Figure 1.3 Geminivirus replication.
Figure 1.4 Antiviral RNA silencing pathways in plants. DNA viruses are targeted by TGS and their transcripts are subject to PTGS, while RNA viruses are targeted by PTGS. RNA silencing is elicited by convergent transcripts, or aberrant transcripts that are either overabundant, or lack 5’ caps or 3’ poly A tails. In the nucleus, DCL3 produces 24 nt siRNAs which can direct DNA or histone modification at homologous loci. This pathway is discussed in greater detail in Figure 1.5. In the cytoplasm, DCL4 is the preferred antiviral dicer involved in production of 21 nt siRNAs, though DCL2 can act redundantly. Viral RDRs generate dsRNA, while convergent and hairpin transcripts of DNA viruses naturally contain double-stranded features. This is cleaved by DCL4 in association with DRB4 to generate 21 nt siRNAs (or by DCL2 to produce 22 nt siRNAs), one strand of which is incorporated into an AGO1/7-containing RISC complex to guide degradation of homologous RNA. The resulting cleavage products are additional aberrant RNAs produced that can amplify the reaction by promoting further dsRNA production through the action of cellular RDR6 and SGS3.

Viruses encode silencing suppressors that block various steps in host silencing, and sometimes act in multiple ways. P19, HcPro and 2b can all bind siRNAs. 2b binds to and inhibits AGO1, and P0 targets the PAZ domain mediating degradation of transcripts by AGO1. V2 acts by inhibiting SGS3. AL2 acts as a silencing suppressor by two mechanisms – by inhibiting ADK and inhibiting viral genome methylation and TGS (and also PTGS), and also by transactivating endogenous silencing suppressors such as WEL1 and WEX.
Figure 1.4 Antiviral RNA silencing pathways in plants.
Figure 1.5 The methyl cycle.

SAM (S-adenosyl methionine) is the methyl donor for most transmethylation reactions. The resulting product, SAH (S-adenosyl homocysteine) is a methyltransferase inhibitor. SAH is converted to Hcy (homocysteine) and adenosine by SAHH (S-adenosyl homocysteine hydrolase). Phosphorylation of adenosine by ADK (adenosine kinase) is important because the SAHH reaction is reversible and the equilibrium lies in the direction of SAH synthesis. ADK activity serves to draw the SAHH-catalyzed reaction towards SAH hydrolysis resulting in the production of Hcy and adenosine. (THF, tetrahydrofolate.) Geminivirus proteins AL2 and L2 inhibit ADK, thereby inhibiting the methyl cycle, and this is one mechanism by which they act as silencing suppressors.
Figure 1.6 Nucleosome structure. Nucleosome comprising a histone octamer with two copies each of H2A, H2B, H3 and H4, around which 146 bp of DNA is wrapped. A chromatosome additionally includes the linker histone H1 and associated DNA.
Figure 1.7 Mechanisms of chromatin remodeling. Nucleosomes are depicted in blue and DNA in black. (a) Chromatin remodeling is required to compact open chromatin and to open compacted chromatin to reveal sites on DNA occluded by nucleosomes (shown in red) that can be accessed by DNA binding proteins. Chromatin remodeling is achieved by mechanisms depicted in (b) to (f). (b) Nucleosome sliding involves movement of nucleosomes laterally along the DNA strand revealing DNA binding sites. (c) DNA looping is achieved by a transient unwrapping of DNA from histone cores that allows access to DNA binding sites. (d) Nucleosome eviction is release of the nucleosome shielding the binding site in closed form. (e) Histone variants with variable affinity for DNA replace the H2A-H2B core dimer. (f) Histone variants are released from the nucleosome.
Figure 1.8 RNA silencing as a defense against geminiviruses. Geminiviruses are targeted by chromatin methylation. Geminivirus minichromosomes are transcribed by Pol II to generate coding transcripts that are used to make viral protein. But geminivirus minichromosomes are also transcribed by plant-specific polymerases Pol IV and Pol V. These non-coding RNAs are converted to dsRNA by RNA-dependent RNA polymerase (RDR2). The dsRNA generated serves as a template for DCL3 cleavage into 24 nt siRNAs. One strand of the 24 nt siRNA is selectively incorporated into a RISC complex containing AGO4. AGO4 docks on a scaffolding transcript generated by Pol V, and the siRNAs guide methylation of that region to which they bear homology. Methylation is carried out by de novo methyltransferase DRM2 and maintained by CMT3 at CNG and by MET1 at CG sites. Histone methylation is carried out by KYP2. DRD1 and DDM1 are chromatin remodelers likely involved in providing access to Pol V, Pol IV, and the methyltransferases.
CHAPTER 2

VIRAL GENOME METHYLATION AS AN EPIGENETIC DEFENSE AGAINST GEMINIVIRUSES

The following chapter has been previously published. Raja, P., Sanville, B. C., Buchmann, R. C., and Bisaro, D. M. (2008). Viral genome methylation as an epigenetic defense against geminiviruses. J. Virol. 82(18), 8997-9007. Brad Sanville contributed Figure 2.5A and assisted with Figure 2.5B, and R.C. Buchmann assisted with Figure 2.2.

2.1 Abstract

Geminiviruses encapsidate single-stranded DNA genomes that replicate in plant cell nuclei through double-stranded DNA intermediates that associate with cellular histone proteins to form minichromosomes. Like most plant viruses, geminiviruses are targeted by RNA silencing and encode suppressor proteins such as AL2 and L2 to counter this defense. These related proteins can suppress silencing by multiple mechanisms, one of which involves interacting with and inhibiting adenosine kinase (ADK), a cellular enzyme associated with the methyl cycle that generates S-adenosyl-
methionine, an essential methyltransferase cofactor. Thus, we hypothesized that the viral genome is targeted by small-RNA-directed methylation. Here, we show that *Arabidopsis* plants with mutations in genes encoding cytosine or histone H3 lysine 9 (H3K9) methyltransferases, RNA-directed methylation pathway components, or ADK are hypersensitive to geminivirus infection. We also demonstrate that viral DNA and associated histone H3 are methylated in infected plants and that cytosine methylation levels are significantly reduced in viral DNA isolated from methylation-deficient mutants. Finally, we demonstrate that *Beet curly top virus L2* mutant DNA present in tissues that have recovered from infection is hypermethylated and that host recovery requires AGO4, a component of the RNA-directed methylation pathway. We propose that plants use chromatin methylation as a defense against DNA viruses, which geminiviruses counter by inhibiting global methylation. In addition, our results establish that geminiviruses can be useful models for genome methylation in plants and suggest that there are redundant pathways leading to cytosine methylation.

2.2 Introduction

The RNA silencing machinery in higher plants is highly elaborated relative to other eukaryotes and is involved in a number of fundamental processes. These include the regulation of endogenous gene expression via the microRNA (miRNA), trans-acting small interfering RNA (siRNA), and natural antisense siRNA pathways, and repression of invasive endogenous DNAs, including repeated sequences and transposons, by siRNA-directed chromatin methylation (Baulcombe, 2004; Chan et al., 2005; Vaucheret, 2006).
Key players include multiple Dicer-like (DCL) ribonucleases, RNA-dependent RNA (RDR) polymerases, and Argonaute (AGO) proteins. The *Arabidopsis* genome encodes four DCL, six RDR, and 10 AGO proteins that functionally partner in specific ways to create distinct but partially overlapping pathways that are commonly triggered by double-stranded RNA (dsRNA) (Brodersen et al., 2004). In addition to modulating endogenous gene expression, mechanisms most likely related to the trans-acting siRNA and natural antisense siRNA pathways, collectively referred to as posttranscriptional gene silencing (PTGS), also serve as an important line of defense against invading plant viruses (Ratcliff et al., 1997; Voinnet, 2005; Ding and Voinnet, 2007). The importance of PTGS as an antiviral defense is evident from the observation that virtually all plant viruses encode proteins that suppress this process (Roth et al., 2004; Li and Ding, 2006).

While the vast majority of plant viruses have genomes composed of RNA, members of the *Geminiviridae* are true DNA viruses that replicate circular, single-stranded DNA genomes in the nucleus by a rolling-circle mechanism that employs host replication machinery (Saunders et al., 1991; Stenger et al., 1991; Hanley-Bowdoin et al., 1999; Hanley-Bowdoin et al., 2004). The double-stranded DNA (dsDNA) intermediates that mediate both viral replication and transcription associate with cellular histone proteins to form minichromosomes (Pilartz and Jeske, 1992; Pilartz and Jeske, 2003). Transcripts produced from these minichromosomes are subject to PTGS, and geminiviruses and their associated satellites have been shown to encode a variety of proteins that can suppress this defense (Voinnet, 1999; Dong et al., 2003; Vanitharani et al., 2004; Cui et al., 2005; Trinks et al., 2005; Wang et al., 2005; Zrachya et al., 2007). In
addition, given the role of RNA-directed methylation in silencing endogenous invasive DNAs, it is reasonable to propose that plants might also use methylation as a means to repress transcription and/or replication from a viral minichromosome (Bisaro, 2006; Ding and Voinnet, 2007). In support of this idea, we and others have shown that in vitro methylation of geminivirus DNA greatly reduces its ability to replicate in plant protoplasts (Brough et al., 1992; Ermak et al., 1993). We have also demonstrated that geminivirus AL2 (also known as AC2 or C2) and L2 proteins can act as silencing suppressors by interacting with and inhibiting adenosine kinase (ADK) (Wang et al., 2003; Wang et al., 2005; Yang et al., 2007). ADK is required for efficient production of the methyl group donor S-adenosyl methionine (SAM), and the primary defect of ADK-deficient yeast and plants is methylation deficiency (Lecoq et al., 2001; Moffatt et al., 2002). Thus, it is possible that one role of the AL2 and L2 proteins is to counter a methylation-based defense.

In this report, several lines of evidence are presented to support the hypothesis that plant hosts employ viral chromatin methylation as a defense against geminiviruses. First, we show that methylation-deficient mutant plants are hypersusceptible to geminivirus infection. Second, we show that cytosine residues in geminivirus DNA are methylated in infected plants and that hypersusceptibility is accompanied by reduced cytosine methylation. Third, we demonstrate that histone H3 methylated at lysine 9 (H3K9), a hallmark of repressed genes, is highly represented in viral chromatin. Finally, we show that host recovery from geminivirus infection, which occurs in plants infected
with $L2^-$ mutants, correlates with hypermethylation of viral DNA and that recovery requires the RNA-directed methylation pathway component AGO4.

2.3 Materials and Methods

2.3.1 Arabidopsis mutants

Mutants were obtained from the Arabidopsis Biological Research Center at The Ohio State University or from individuals. The following previously characterized seed stocks were used: wild-type Ler-0 (CS20) Landsberg erecta ecotype; $ago ~4-1$ $gll-1$ (CS6364/At2g27040) AGO4 (Zilberman et al., 2003); $kyp2 ~gll-1$ (CS6367/At5g13960) kryptonite/SuVH4 (Jackson et al., 2002); $cnt3-7$ (CS6365/At1g69770) chromomethylase (Lindroth et al., 2001); wild-type Ws-2 (CS22659) Wassilewskija ecotype; $drm1$ $drm2$ (CS6366/At1g28330 and At5g14620), domains rearranged methyltransferase (Cao and Jacobsen, 2002); $mmt$ (CS6398/At5g49810), methionine S-methyltransferase (Kocsis et al., 2003); wild-type Col-0 (CS60000) Columbia ecotype; $nrpd2a-1$ (SALK_095689/At3g23780), nuclear RNA polymerase D 2A (Herr et al., 2005); $ddml$ (SALK_000590/At5g66750), decreased DNA methylation 1 (Jeddeloh et al., 1998); $met1-7$ (SALK_076522/At5g49160), decreased DNA methylation 2/methyltransferase-1, hemizygous due to seed abortion phenotype (Jeddeloh et al., 1998); $hen1-5$ (SALK_049197/At4g20910), Hua enhancer (Vasquez et al., 2004); $hog1-1$ (CS1892/At4g13940), homology-dependent gene silencing, S-adenosyl-L-homocysteine hydrolase (Rocha et al., 2005); $adk1-1$ (SALK_040957/At3g09820.1), adenosine kinase 1
(Xie et al., 2004); adk 2-1 (SALK_000565/At5g03300), adenosine kinase 2 (Young et al., 2006); dcl2-1 (SALK_064627/At3g03300), Dicer-like 2 (Xie et al., 2004); dcl3-1 (SALK_005512/At3g43920), Dicer-like 3 (Xie et al., 2004); dcl4-2 (GABI_160G05/At5g20320), Dicer-like 4 (Xie et al., 2005); dcl2-1 dcl3-1 dcl4-2, generated from stocks above by J.C. Carrington; rdr2-2 (SALK_059661/At4g11130), RDR polymerase 2 (Herr et al., 2005); and rdr6-11 (CS24285/At3g49500), RDR polymerase 6, silencing defective 1, suppressor of gene silencing 2 (Peragine et al., 2004).

2.3.2 Virus inoculation

Agroinoculation of *Nicotiana benthamiana* plants with *Tomato golden mosaic virus* (TGMV), *Cabbage leaf curl virus* (CaLCuV), and *Beet curly top virus* (BCTV) was carried out as described previously (Sunter et al., 2001). Plants were inoculated 30 to 40 days after germination. An overnight culture of *Agrobacterium tumefaciens* containing tandem repeat copies of the appropriate virus genome was injected into the petiole of leaves using a Hamilton syringe. Three leaves were inoculated per plant, and 30 µl of inoculum was used per leaf. Infection of *Arabidopsis* plants with CaLCuV and BCTV was carried out mechanically as described for *N. benthamiana* and sugar beet (Hormuzdi and Bisaro, 1993). *Arabidopsis* plants were inoculated within 5 days of bolting. Bolts were cut where they emerged from the rosette, and inoculum was applied to the freshly cut stem, which was then punctured with an insect pin multiple times. CaLCuV and TGMV symptoms were observed, and plants were harvested 14 to 21 days
postinoculation. Due to an inherently longer latent period, BCTV symptoms were observed, and plants were harvested 21 to 30 days postinoculation. Symptomatic leaf tissue was harvested from *N. benthamiana* plants. Inflorescence tissue showing visible symptoms was harvested from *Arabidopsis* plants. For each sample, tissue was pooled from four infected plants. For BCTV recovery experiments, plants were agroinoculated with BCTV or *BCTV L2* mutant virus, and after the primary harvest, plants were allowed to continue growing under the same conditions. Observations were made about symptom development in secondary inflorescence tissue, which was then harvested as secondary tissue, 14 to 21 days after the harvesting of primary tissue.

2.3.3 DNA isolation and Southern Blot analysis

DNA was isolated from homogenized plant samples frozen in liquid nitrogen using DNeasy columns (Qiagen), and 500 to 600 ng of genomic DNA from infected plants was restricted overnight. After digestion, 1 to 1.2% agarose gel electrophoresis was carried out, followed by overnight transfer, using standard protocols, onto Nytran Super Charge membrane. The blot was UV cross-linked (1,700 µJ) and hybridized overnight using a $^{32}$P-labeled full-length virus probe at 48°C in Ultrahyb hybridization buffer (Ambion). Random primer-based $^{32}$P labeling was carried out using Strip-EZ DNA (Ambion). Signal intensity was quantitated using a Phosphorimager (Bio-Rad Molecular Imager FX).
2.3.4 Bisulfite sequencing

The bisulfite sequencing method is based on Frommer et al. (Frommer et al., 1992). DNA isolated from infected plant tissue was linearized overnight using appropriate restriction enzymes. Proteinase K digestion was subsequently carried out overnight, followed by bisulfite conversion using CT conversion reagent (EZ-DNA Methylation Gold; Zymo Research). Primers were designed against converted template, and the intergenic region of the viral genome was amplified by PCR. The PCR product was purified using Promega Wizard columns and TA cloned, and individual clones were sequenced at The Ohio State University Plant Microbe Genomics Facility. For conversion control, plasmids containing CaLCuV DNA or BCTV DNA were added to a vast excess of healthy plant DNA extract and treated with bisulfite reagent. The following forward and reverse primers were used to amplify CaLCuV and BCTV intergenic regions (IR) following bisulfite conversion: CaLCuV2556CF, GGGGATATGTAAAGAATATATTTGGG (forward), and CaLCuV359CR, TCCCCACCATAAAAACACCAC (reverse); BCTV2640CF, GGGATATGTAAAGAATATATG (forward), and BCTV147CR, TCTCCCCTTCTATTAACCAATCAAC (reverse).
2.3.5 Chromatin immunoprecipitation

The chromatin immunoprecipitation (ChIP) protocol was based on the method described by Johnson et al. (Johnson et al., 2002). Symptomatic Arabidopsis inflorescence or N. benthamiana leaf tissue (0.3 g) was cross-linked with formaldehyde for 20 min and then quenched with glycine for 10 min under vacuum. Tissue was then ground in liquid nitrogen and sonicated in lysis buffer under shearing conditions that resulted in fragments about 500 bp long (range, 250 to 1,000 bp). Protein A agarose beads and salmon sperm DNA were used for preclearing. Immunoprecipitation was carried out overnight at 4°C using commercially available antibodies (from Abcam or Upstate Biotechnologies) targeted to specific histone modifications. Cross-links were reversed at 65°C for 16 h, and then DNA was extracted using phenol-chloroform, followed by clean-up with Promega mini-prep columns. Purified DNA (2 to 4 µl) was used as a PCR template with primers specific for viral DNA and a control sequence. Primers used for CaLCuV DNA A amplified exactly the same region used in the bisulfite PCR but were designed to original virus template and not bisulfite-converted template. The following forward and reverse primers were used to amplify Tnt, Ta3, and actin controls: TntF, CATTGGTTCTAAAGGATGTGCGGC (forward), and TntR, GAAATCTCATCTTGGTGCGCG (reverse); Ta3F, GATTCTTACTGTAAGAAACATGGCATTGAGAGA (forward), and Ta3R, TCCAAATTTCCTGAGGTGCTT (reverse); ActinF,
CGTTTCGCTTTCTTAGTGTAGCT (forward), and ActinR, AGCGAACGGATCTAGAGACTCACCTTG (reverse).

2.4 Results

2.4.1 Methylation-deficient Arabidopsis mutants are hypersusceptible to geminivirus infection.

We hypothesized that plant hosts use methylation as a defense against geminiviruses. A prediction of this hypothesis is that methylation-deficient mutant plants should display enhanced susceptibility to geminivirus infection. We tested this using Arabidopsis-infecting geminiviruses of two distinct genera. CaLCuV (genus Begomovirus) is a bipartite virus that produces an AL2 protein, and BCTV (genus Curtovirus) is a monopartite virus that encodes the related L2 protein.

Arabidopsis mutants were selected to represent an array of defects in RNA-directed methylation pathway components, the methyl cycle, and downstream effectors (Fig. 2.1) (Vaucheret, 2006; Henderson and Jacobsen, 2007). The last group included the DNA methyltransferase mutants drm1 drm2, cmt3, and met1. The enzymes encoded by these genes carry out de novo and/or maintenance methylation of cytosine to 5-methyl cytosine in a variety of sequence contexts, including CG, CNG (where N is A, T, C, or G), or CHH (where H is A, T, or C). MET1 primarily maintains CG methylation, whereas the DRM1/2 and CMT3 methyltransferases are important for methylation at non-CG sites (Bender, 2004; Chan et al., 2005). CMT3 additionally maintains CNG methylation in
cooperation with H3K9 and H3K27 methyltransferases, indicating linkage between these epigenetic marks (Jackson et al., 2002; Malagnac et al., 2002; Lindroth et al., 2004). Consequently, we also tested a histone H3K9 methyltransferase mutant (kyp2) for geminivirus sensitivity. RNA-directed methylation pathway mutants included plants lacking a functional chromatin remodeling enzyme (ddm1) (Jeddeloh et al., 1998), the AGO4 protein (ago4) (Zilberman et al., 2003; Zilberman et al., 2004), and a subunit common to RNA polymerases IVa (Pol IV, NRPD) and IVb (Pol V, NRPE) (nrpd2a) (Herr et al., 2005; Onodera et al., 2005). Selected methyl cycle activities included ADK and methionine methyltransferase. *Arabidopsis* has two ADK genes, and the loss of all ADK function is lethal (Moffatt et al., 2002). However, it was possible to separately test *adk1* and *adk2* mutants, which are expected to have reduced levels of SAM and reduced global methylation. By contrast, *mmt* mutants show slightly elevated levels of SAM (Kocsis et al., 2003). We were unable to test mutants deficient for the methyl cycle enzyme S-adenosyl homocysteine hydrolase (*hogl*) (Rocha et al., 2005) and the siRNA/miRNA methyltransferase HEN1 (*henl*) (Park et al., 2002) due to severe morphological defects. Otherwise, the selected mutants did not exhibit phenotypes that complicated observation of symptom development. However, the timing of flowering is known to be altered in *Arabidopsis* methylation mutants, hastening it in some but delaying it in others, depending on the ecotype (Genger et al., 2003). To control for differences in developmental timing, wild-type and mutant plants were infected after they had both undergone the vegetative to floral transition. Because of this, rosette tissue was very much reduced by the time symptoms appeared, making floral tissue a better choice.
than leaves to assess disease development. In all cases, mutant responses to infection were compared with wild-type plants of the same ecotype. Importantly, to observe a range of symptom enhancement, plants were inoculated under conditions that normally produce relatively mild symptoms (Hormuzdi and Bisaro, 1993; Sunter et al., 2001).

As summarized in Table 2.1, we found that methylation-deficient plants were highly susceptible to geminivirus infection and typically developed severe symptoms compared to infected wild-type plants. Major differences were manifest in the marked stunting and enhanced deformation of flowers and siliques. Each mutant was tested with both viruses, and in all cases, both viruses yielded similar results. CaLCuV typically causes curling of the inflorescence and of siliques near the inflorescence. BCTV infection additionally results in shortened internodes and outgrowths on the inflorescence stem with increased anthocyanin pigmentation. These symptoms were far more pronounced in most of the mutants, and extensive stunting and deformation of inflorescence structures was observed (Fig. 2.2).

Most susceptible were the methyltransferase mutants drm1, drm2, cmt3, and kyp2; the methylation pathway mutants ddm1, ago4, and nrpd2a (Pol IV and Pol V); and the methyl cycle mutants adk1 and adk2 (Fig. 2.2 and Table 2.1). Symptom enhancement was less extreme in met1 mutant plants, suggesting that maintenance of CG methylation is less important than non-CG and H3K9 methylation in limiting infection. However, the plants used in these experiments were heterozygous for the met1-7 allele, so the symptom phenotype could have been moderated by remaining MET1 activity. Slight enhancement of the methyl cycle conditioned by the mmt mutation had no apparent effect on the
outcome of geminivirus infections. Importantly, no disease enhancement was noted when *Turnip crinkle virus*, an RNA virus, was used to infect *drm1 drm2* mutants (data not shown), indicating that the enhanced symptoms observed following geminivirus infection of mutant plants were not a general effect of reduced stress tolerance. Although Southern blot analysis of infected plant DNA obtained from floral tissues did not reveal increases in viral DNA levels beyond that observed in normal plant-to-plant variation (data not shown), the nature of the enhanced symptoms suggests that CaLCuV and BCTV are more meristem-invasive in methylation mutants than in wild-type plants. Further work is needed to confirm this.

These experiments provide strong genetic evidence that RNA-directed methylation of cytosine residues and H3K9 plays an important role in defense against geminiviruses. Further, the sensitivity of *adk* mutants is in accord with AL2- and L2-mediated reduction of ADK activity during geminivirus infection (Wang et al., 2005, Wang et al., 2003).

2.4.2 Plants deficient in DCL and RDR enzyme activities show moderately increased susceptibility to geminiviruses.

In an attempt to identify the DCL and RDR enzymes most relevant to geminivirus defense, we examined additional mutant plants for susceptibility to CaLCuV and BCTV. These included individual and multiple Dicer mutants (*dcl2, dcl3, dcl4,* and *dcl2 dcl3 dcl4*) and two RDR polymerase mutants (*rdr2* and *rdr6*). DCL3 generates the larger 24-nucleotide (nt) siRNAs associated with DNA methylation in plants (Xie et al., 2004).
RDR2 is also involved in methylation whereas RDR6 has been linked to PTGS and defense against meristem invasion by certain RNA viruses (Xie et al., 2004; Schwach et al., 2005). We were unable to study the miRNA-associated Dicer (dcl1) because morphological defects associated with this mutation complicated observation of viral disease symptoms.

As indicated in Table 2.1, little or no disease enhancement was seen in rdr2 and rdr6 mutants, suggesting either that de novo dsRNA synthesis does not play a significant role in geminivirus defense or that there is considerable functional redundancy between the several RDR proteins in Arabidopsis. Functional redundancy among the four Arabidopsis DCL enzymes in the biogenesis of endogenous siRNAs has been previously reported and was also apparent here in the response of dcl mutants to geminiviruses (Table 2.1) (Xie, Z. et al., 2004; Gasciolli et al., 2005; Henderson et al., 2006). The moderately enhanced susceptibility of dcl3 plants to BCTV and CaLCuV is consistent with the involvement of this enzyme in the methylation pathway but at the same time suggests that other DCL activities are able to partially fulfill its role. The sensitivity of dcl2 mutants is in keeping with the observation that geminivirus transcripts are targeted by PTGS and that this activity produces most of the PTGS-associated 22 nt viral siRNAs during geminivirus infection (Akbergenov et al., 2006). The dcl4 mutants showed no obvious symptom enhancement, and this enzyme may play little role in geminivirus defense when other DCL activities are present, at least in floral tissues. Interestingly, dcl2 dcl3 dcl4 triple mutants displayed only moderate symptom enhancement, suggesting a role for dcl1 when other Dicers are disabled. This is consistent with recent reports
suggesting that all four DCL enzymes are involved in the biogenesis of virus-specific siRNAs against DNA viruses and that, in the absence of the others, DCL1 can generate small RNAs in response to CaLCuV (Blevins et al., 2006, Moissiard and Voinnet, 2006).

2.4.3 Geminivirus genomes are methylated in vivo.

If methylation is a host defense, it should be possible to directly observe methylation of viral DNA in vivo. To assess methylation at high resolution, we carried out bisulfite sequencing of the CaLCuV genome, examining the IR that contains divergent viral promoters flanking the origin of replication. The experimental protocol involved treatment of DNA obtained from infected plants with bisulfite reagent to convert unmethylated cytosines to uracil, followed by PCR to amplify the viral strand of the IR. PCR products were cloned, and the sequences of 12 to 18 clones were compiled for each treatment.

Because any cytosine can potentially be methylated in plants, PCR primer design is critical. Primers that assume different amounts of cytosine conversion can differentially amplify relatively hypo- or hypermethylated templates. For CaLCuV we selected a primer set biased for templates with a relatively low density of methylation, which allowed us to examine the propensity for methylation at individual sites. The data generated showed evidence of methylation at some but not all cytosines in the IR (Fig. 2.3A) (see primary sequence data in Appendix A). Methylation patterns appeared somewhat stochastic in that a larger sampling tended to yield more sites, but distinct
patterns emerged. First, we observed that viral DNA methylation in wild-type Arabidopsis plants of different ecotypes (Ler-0, Ws-2, and Col-0) was variable in terms of the overall number of sites methylated (33 to 44% of the 79 cytosine residues surveyed) and that much of this difference could be attributed to lower levels of CG methylation in the Ws-2 ecotype (Fig. 2.3B). In addition, a larger proportion of non-CG relative to CG sites was methylated in all ecotypes. Similar methylation patterns were observed in CaLCuV IR DNA obtained from N. benthamiana plants although overall methylation was somewhat greater (~50%).

Interesting patterns could also be discerned at the level of individual cytosine residues (Fig. 2.3A). Methylation was detected at all but 12 of the 79 cytosines surveyed, and 10 were methylated in all three ecotypes. A region centered on the conserved hairpin showed a relatively high density of methylation in all ecotypes. In addition, a CNG site immediately adjacent to the directly repeated AL1 (replication initiator protein) binding sites was frequently methylated. A smaller region of high-density methylation adjacent to the coat protein start codon was also apparent. Importantly, several of the highly methylated cytosines were unmethylated or hypomethylated in DNA from the majority of the mutant plants tested (adk2, ago4, cmt3, drm1 drm2, and kyp2) (Fig. 2.3A).

Cytosine methylation levels in the IR of CaLCuV isolated from infected mutant plants are summarized in Fig. 2.3B. Reduced methylation in drm1 drm2 and cmt3 mutants confirms that these cytosine methyltransferases target viral DNA and is consistent with their known activities and partially overlapping functions (Cao et al., 2003). CHH methylation lacks a known maintenance mechanism and should be most
sensitive to the *drm1 drm2* mutations. Also as expected, viral DNA from *cmt3* mutants showed greater reduction at CNG sites than CHH sites. CaLCuV DNA from *kyp2* mutants displayed an equivalent reduction in CNG methylation, reinforcing the linkage between H3K9 methylation and CMT3-mediated maintenance of CNG methylation (Jackson et al., 2002; Malagnac et al., 2002; Lindroth et al., 2005). The methylation pathway mutant *ago4* showed significant reductions in both types of non-CG methylation, consistent with previous reports (Zilberman et al., 2003). The largest reduction in methylation at non-CG sites was observed in *adk2* mutants, which underscores the role of ADK in maintaining the methyl cycle and its value as a target for the AL2 and L2 proteins. By contrast, most of the mutations did not lead to significant reductions in CG methylation, suggesting that the CG maintenance pathway is relatively robust.

2.4.4 Analysis of cytosine methylation in the BCTV genome.

To further examine the correlation between cytosine methylation and enhanced susceptibility, studies were extended to include bisulfite sequencing of the BCTV IR from *N. benthamiana*, wild-type *Arabidopsis* Ler-0, and methylation-deficient mutants in this ecotype background (*ago4, kyp2*, and *cmt3*). In this case PCR primers used with bisulfite-treated DNA amplified a mixture of relatively densely and lightly methylated templates. More specifically, cytosine residues in individual clones tended to be either mostly unmethylated or mostly methylated, suggesting that populations of active and
repressed dsDNA replicative forms coexist in infected plants. As a consequence, all but two of the cytosines (total examined, 44) were methylated in at least one of the 7 to 10 clones examined for all treatments. Thus, data obtained for BCTV did not permit analysis of methylation at individual sites, but it was possible to discern changes in methylation density.

As illustrated in Fig. 2.3C, the overall methylation density in the BCTV IR was similar in *N. benthamiana* and wild-type *Arabidopsis* Ler-0 plants, and in each species 50 to 55% of the total cytosines represented in the 10 sample clones were methylated (see the Supplemental Material in Appendix A for primary sequence data). However, the amount of CG methylation was relatively greater, and CNG methylation was lower in *N. benthamiana* than in *Arabidopsis* plants. Similar to wild-type plants, virtually all cytosine residues in BCTV IR DNA obtained from *ago4, kyp2*, and *cmt3* mutant plants was methylated in at least 1 of the 10 clones examined for each treatment. But the proportion of relatively undermethylated clones was greater, and viral DNA from mutant plants thus displayed significant reductions in methylation density (~20 to 25%) in all sequence contexts (Fig. 2.3C).

Taken together, analysis of the CaLCuV and BCTV intergenic regions clearly indicated that hypersusceptibility of methylation-deficient *Arabidopsis* mutants to geminivirus infection was correlated with reduced cytosine methylation. Depending on the experimental conditions, either a reduced number of sites (Fig. 2.3B) or reduced methylation density (Fig. 2.3C) was observed.
2.4.5 ChIP reveals both active and repressive marks on geminivirus DNA.

Geminivirus dsDNA replicative forms associate with histones and exist as minichromosomes. We also observed that kyp2 mutants show enhanced susceptibility to geminivirus infection. Hence, we used ChIP to investigate modification of histone H3 associated with the CaLCuV IR in wild-type *N. benthamiana* and *Arabidopsis* plants. Viral DNA was sonicated to an average size of 500 bp (range, 250 bp to 1 kbp) and, following cross-linking and immunoprecipitation with antibodies specific for different histone H3 modifications, a 405 bp fragment spanning the 305 bp IR was amplified to determine whether viral DNA was associated with the modified histones.

H3K4 is associated with active genes, and H3K9 is associated with gene silencing (Jenuwein and Allis, 2001). Further, there is cross talk between histone and DNA methylation, as CMT3 interacts with histone H3 only when it is methylated at H3K9 and H3K27 (Lindroth et al., 2004). CMT3 interacts with LIKE HETEROCHROMATIN PROTEIN 1 (LHP1), a chromodomain containing protein that binds methylated histone tails. Using ChIP, we detected all three types of dimethylated histone H3 associated with CaLCuV DNA, and similar results were obtained with extracts from *N. benthamiana* and *Arabidopsis* (Fig. 2.4). Detection of H3K9 (and H3K27) is consistent with the enhanced susceptibility of kyp2 mutants and with the idea that these histone modifications are a component of defense against geminiviruses. In particular, H3K9 methylation appeared to be especially abundant in both host plants. On the other hand, H3K4 methylation and H3 acetylation are indicative of viral gene expression, which is to be expected during an infection. The presence of both active and repressive histone H3 marks within the viral
control region again suggests that populations of active and repressed genomes are present in infected plants.

2.4.6 Recovery of *Arabidopsis* plants from infection with *BCTV L2* mutant virus requires AGO4 and is associated with greatly increased cytosine methylation.

Begomovirus *AL2* mutants are not systemically infectious. By contrast, infections of *N. benthamiana* or *Arabidopsis* with BCTV or *BCTV L2* null mutants proceed similarly in primary infected tissue (Hormuzdi and Bisaro, 1995). However, after removal of the primary infected shoots, new growth (secondary infected tissue) is symptomatic in BCTV-infected plants but exhibits recovery (no symptoms) in *BCTV L2* mutant-infected plants. This recovery phenomenon is highly reproducible. In *Arabidopsis*, recovery rarely occurs in plants infected with wild-type BCTV but nearly always occurs in wild-type plants infected with the *BCTV L2* mutant virus. Viral DNA is present in recovered tissue but in much reduced amounts.

To determine if methylation plays a role in recovery, wild-type *Arabidopsis* (ecotype Ler-0) and plants lacking the methylation pathway component AGO4 were inoculated with wild-type BCTV and two different *BCTV L2* null mutants (*L2-1 – 70 a.a.* and *L2-2 – 72 a.a.*) (Hormuzdi and Bisaro, 1995). Primary infected tissue of wild-type and mutant plants showed symptoms of infection, and as before the symptoms on *ago4* mutant plants were considerably more severe (Fig. 2C). After removal of primary infected tissue, secondary shoots of wild-type and *ago4* plants infected with wild-type
BCTV again showed severe symptoms (10 to 16 plants per treatment) (data not shown). As expected, wild-type plants infected with *BCTV L2-1* or *BCTV L2-2* mutant virus showed recovery, with little or no evidence of symptoms. Remarkably, however, *ago4* mutant plants infected with these mutant viruses did not recover, and severe symptoms appeared in nearly all secondary shoots in all of the 16 plants inoculated with each mutant virus (Fig. 2.5A). The symptomatic secondary tissues in *ago4* plants contained much higher levels of *BCTV L2-1* and *BCTV L2-2* DNA than asymptomatic, recovered secondary tissue from wild-type plants (Fig. 2.5B). We concluded that AGO4 is required for recovery, linking L2 function and recovery with the methylation pathway.

The BCTV IR in viral DNA isolated from secondary infected tissues was examined for cytosine methylation by bisulfite sequencing (see Appendix A for primary sequence data). We found that *BCTV L2-1* and *BCTV L2-2* genomes obtained from recovered, secondary tissue of wild-type plants were hypermethylated. In the seven clones examined for each mutant virus, nearly all were highly methylated (about 80% of all cytosine residues examined) (Fig. 2.5C). Increased methylation was readily apparent in both non-CG and CG contexts.

### 2.5 Discussion

In this report, several lines of evidence are presented to support the hypothesis that plant hosts use small RNA-directed methylation as a defense against geminiviruses. We demonstrate that *Arabidopsis* plants harboring inactivating mutations in cytosine methyltransferases (*drm1 drm2, cmt3*, and *met1*), histone H3K9 methyltransferase (*kyp2*),
methylation pathway components (AGO4, DDM1, and NRPD2A), or methyl cycle enzymes (ADK1 and ADK2) are hypersusceptible to infection with the geminiviruses CaLCuV and BCTV. We show that cytosine residues in viral genomes are methylated in infected plants and that methylation is significantly reduced in methylation-deficient mutant plants that display enhanced susceptibility. We present evidence for populations of viral minichromosomes carrying marks associated with repressed chromatin (H3K9) or active chromatin (acetylated H3) in infected plants. Finally, we demonstrate that BCTV L2 mutant virus genomes are hypermethylated in tissue that has recovered from infection and that a methylation pathway component (AGO4) is required for Arabidopsis plants to undergo recovery. L2 is a silencing suppressor that interacts with and inhibits ADK (Wang et al., 2003). Thus, the absence of a viral pathogenicity factor that can oppose methylation is rendered redundant in plants containing a mutation that compromises methylation. In total, these observations indicate that plants employ an epigenetic mechanism, chromatin methylation, as a defense against DNA viruses.

Bisulfite sequencing of the CaLCuV IR, which was biased for relatively undermethylated clones, was revealing in several respects. First, it showed a greater proportion of non-CG methylation than CG methylation, as was previously observed within the C4 gene of Tomato leaf curl virus DNA from N. benthamiana plants (Bian et al., 2006). Further, cytosine residues in the vicinity of the conserved hairpin and AL1 binding sites were the most frequently methylated. The hairpin and AL1 binding sites lie within the divergent early and late gene promoters and in addition are integral components of the replication origin core (Hanley-Bowdoin et al., 2004). Thus, these
sequences are potentially high-value targets for defensive purposes. Increased methylation in this critical viral control region and reduced density in more susceptible mutants are consistent with a role for methylation in defense and with earlier findings that methylation negatively impacts viral transcription and replication (Brough et al., 1992; Ermak et al., 1993). However, how these sites are targeted remains unclear. The hairpin has a duplex stem consisting of only 11 bp, and the core AL1 binding site direct repeats are approximately 7 bp; thus, transcripts across this region would not be expected to possess a great deal of repetitive character or secondary structure.

It is also interesting that most of the mutants tested did not reduce CG methylation in the CaLCuV IR, yet all of them showed extreme sensitivity to geminiviruses. This suggests that in the context of a virus infection, methylation at non-CG sites is relatively more important for defense than maintenance methylation at CG sites. In addition, the relative consistency of CG methylation suggests that its maintenance is robust and less sensitive to perturbations in de novo methylation. Another striking observation is that while the methylation mutants displayed extreme sensitivity to geminiviruses, overall methylation levels in the CaLCuV IR were reduced only about 10% to 20% (Fig. 2.3B). Thus, geminiviruses are able to exploit relatively small changes in cytosine methylation. The relatively high levels of residual methylation also prompt us to conclude that there are redundant pathways leading to cytosine methylation, as suggested by others (Zilberman, D. et al., 2003; Henderson, I. R. et al., 2006). Finally, in mutant plants, methylation was usually reduced in a manner consistent with known activities of the inactivated genes. For example, cmt3 and kyp2 mutants showed the largest reductions in
CNG methylation, while CHH methylation was most affected in drm1 drm2 mutant plants (Fig. 2.3B). Thus, methylation of geminivirus chromatin appears to mirror host chromatin methylation in many respects, and it is likely that geminiviruses will prove to be useful models for the study of RNA-directed methylation pathways.

Bisulfite sequencing of the BCTV IR, which was biased toward more heavily methylated clones, was also revealing as it showed that viral genomes can be either mostly unmethylated or mostly methylated. This is in accord with the results of ChIP experiments, which showed that both active (acetylated H3) and repressive (H3K9) marks are associated with viral minichromosomes. Together, these results suggest that populations of active and repressed genomes are present in infected plants. Further, a larger proportion of relatively undermethylated clones was observed in DNA from mutant plants (ago4, kyp2, and cmt3), and overall cytosine methylation was reduced ~20 to 25%, confirming the correlation between enhanced susceptibility and reduced methylation. Conversely, most BCTV L2+ mutant clones obtained from asymptomatic, recovered tissue were nearly completely methylated (~80% of cytosines methylated), underscoring the role of methylation in inhibiting geminivirus replication.

It bears repeating that differences in CaLCuV and BCTV IR methylation reported here likely stem from the PCR primers chosen for amplification of bisulfite-treated DNA, with the former primer set being biased for relatively undermethylated templates. Indeed, experiments with a different CaLCuV IR primer set produced a mixture of mostly undermethylated and mostly methylated clones, similar to what was observed with BCTV (see Appendix A). A truer picture of cytosine methylation awaits analysis using unbiased
methods. However, given the differences observed between *Arabidopsis* ecotypes and *N. benthamiana* and the dramatic hypermethylation of viral DNA in recovered tissue, it is clear that viral cytosine methylation is dynamic. Factors that may influence methylation levels include the particular virus-host combination, tissue type, developmental age, environmental conditions, and the status of host defenses.

If plants employ both PTGS and methylation against geminiviruses, is it possible to evaluate the relative contributions of these pathways to antiviral defense? Earlier studies have shown that specific components of PTGS pathways, namely DCL4, RDR6, and SGS3, are required for silencing endogenous genes (virus-induced gene silencing; VIGS) using geminivirus vectors (Muangsan et al., 2004; Blevins et al., 2006). In particular, these enzymes are required for geminivirus VIGS in newly developed leaves, suggesting that systemic spread of the PTGS signal elicited by geminiviruses involves a pathway that utilizes these enzymes. In addition, slight symptom enhancement was observed in *dcl4*, *rdr6*, and *sgs3* mutant plants although viral DNA levels, at best, were only marginally enhanced. In contrast, we observed some enhancement of floral symptoms in *dcl2* and *dcl3* but not *dcl4* mutants, and there were no obvious differences between *rdr2* and *rdr6* mutants and wild-type plants (Table 2.1). Given the functional redundancy that exists among the DCL and RDR proteins and our incomplete knowledge of antiviral silencing systems, we find it difficult to make definitive statements about the relative importance of specific pathways at this time. However, one could speculate that, in *Arabidopsis*, both the methylation pathway (likely involving DCL3 and RDR2) and a pathway involving DCL4/RDR6/SGS3 are important for limiting the spread of
geminiviruses, with perhaps different roles in different tissues. The latter pathway may be more important for limiting geminivirus spread to shoot apical meristems, while the methylation pathway may play a greater role in preventing spread to secondary tissues (and conditioning recovery) and floral meristems. Further study is needed to answer these questions.

In summary, we conclude that geminivirus genomes are targeted for de novo methylation by both cytosine and histone methyltransferases and that geminivirus genomes will be useful models for examining RNA-directed methylation pathways. The antiviral nature of this methylation is clear from the hypersusceptibility of the methylation-deficient *Arabidopsis* mutants examined in this study and from the reductions in viral DNA methylation observed in the mutants. In addition, the reduced accumulation and hypermethylation of *BCTV L2-1* and *L2-2* mutant genomes in recovered tissue from wild-type plants and the absence of recovery in *ago4* mutant plants provide compelling evidence that host defense pathways leading to recovery require AGO4-mediated methylation, which can be opposed by the BCTV L2 protein. Taken together, our genetic and biochemical data strongly support the hypothesis that plants use cytosine and histone methylation as an epigenetic defense against invading DNA virus genomes, which most likely results in transcriptional gene silencing or direct inhibition of replication. This methylation-based defense occurs in addition to PTGS, which results in degradation of viral mRNAs. It follows that the geminivirus AL2 and L2 proteins, which are known to suppress PTGS and can inhibit methylation by inactivating ADK, may also
be capable of interfering with transcriptional gene silencing, and evidence has since been accumulated that this is indeed the case (Buchmann et al., 2009).
**Figure 2.1** The methylation pathway in plants. A putative pathway for RNA-directed DNA methylation in *Arabidopsis* is illustrated. Genomic and viral genome targets may be transcribed by an RNA polymerase IVa complex (Pol IVa; now called Pol IV; containing NRPD1A and NRPD2, a subunit common to Pol IV and Pol V). Resulting single-stranded RNA (ssRNA) is converted to dsRNA by complexes containing RDR2. The 24-nt siRNAs processed from dsRNA by DCL3 are loaded into complexes containing AGO4, which subsequently associates with Pol IVb (now called Pol V, containing NRPD1B and NRPD2, a subunit common to Pol IV and Pol V). The AGO4-associated siRNAs target the complex to homologous DNA sequences, where cytosine methyltransferases (e.g., DRM1/2) are recruited. Methylation also involves the SWI-SNF chromatin remodeling activities DRD1 and DDM1. Cytosine methyltransferases CMT3 and MET1 are primarily involved in methylation maintenance at CNG and CG sites, respectively. CNG methylation by CMT3 is also linked to H3K9 methylation carried out by KYP2.
Figure 2.1 The methylation pathway in plants.
Table 2.1. Methylation-deficient mutants are hypersusceptible to geminivirus infection

<table>
<thead>
<tr>
<th>Gene function and mutation(s)</th>
<th>Relative severity of disease$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methylation pathway components</td>
<td></td>
</tr>
<tr>
<td>nrpd2a (Pol IV and Pol V)</td>
<td>+++</td>
</tr>
<tr>
<td>ago4</td>
<td>+++</td>
</tr>
<tr>
<td>ddm1</td>
<td>+++</td>
</tr>
<tr>
<td>Cytosine methyltransferases</td>
<td></td>
</tr>
<tr>
<td>drm1 drm2</td>
<td>++++</td>
</tr>
<tr>
<td>cmt3</td>
<td>++++</td>
</tr>
<tr>
<td>met1</td>
<td>++</td>
</tr>
<tr>
<td>H3K9 methyltransferase, kyp2</td>
<td>+++</td>
</tr>
<tr>
<td>Methyl cycle enzymes</td>
<td></td>
</tr>
<tr>
<td>adk1</td>
<td>++++</td>
</tr>
<tr>
<td>adk2</td>
<td>+++</td>
</tr>
<tr>
<td>Mmt</td>
<td>+</td>
</tr>
<tr>
<td>DCL and RDR enzymes</td>
<td></td>
</tr>
<tr>
<td>dcl2</td>
<td>+++</td>
</tr>
<tr>
<td>dcl3</td>
<td>++</td>
</tr>
<tr>
<td>dcl4</td>
<td>+</td>
</tr>
<tr>
<td>dcl2 dcl3 dcl4</td>
<td>++</td>
</tr>
<tr>
<td>rdr2</td>
<td>+</td>
</tr>
<tr>
<td>rdr6</td>
<td>+</td>
</tr>
</tbody>
</table>

$^a$ Severity of disease symptoms in mutant Arabidopsis plants was based on the degree of stunting and floral deformation relative to wild-type plants of the appropriate ecotype (Col-0, Ler-0, or Ws-2) according to the following rating scale: +, typical, wild-type symptoms; ++, moderately enhanced; +++, severe; ++++, very severe with extensive stunting. Pol IV, polymerase IV and Pol V, polymerase V.
**Figure 2.2** Methylation-deficient mutant plants are hypersusceptible to geminivirus infection. Photographs are illustrative of geminivirus disease symptoms on wild-type and selected mutant plants. Plants were mock inoculated or inoculated with virus within 5 days of bolting, and symptoms were observed after 10 to 14 days for CaLCuV and 14 to 21 days for BCTV. Methylation-deficient mutants showed greater stunting and increased inflorescence deformation in response to geminivirus infection than wild-type (WT) plants of the same ecotype.

(A) *Arabidopsis* Ler-0 and cmt3 mutant plants inoculated with BCTV.

(B) Col-0 and adk1 mutant plants inoculated with CaLCuV.

(C) Typical inflorescence structures of BCTV-infected kyp2, cmt3, and ago4 mutants in the Ler-0 background.

(D) Inflorescence structures of CaLCuV-infected adk2, nrpd2a, and met1 mutants in the Col-0 background.

(E) Inflorescence structures of BCTV-infected ddm1, rdr2, rdr6, dcl2, dcl3, dcl4, and dcl2 dcl3 dcl4 mutants in the Col-0 background.

(F) Inflorescence structures of BCTV-infected drm1 drm2 and mmt mutants in the Ws-2 background.
Figure 2.2 Methylation-deficient mutant plants are hypersusceptible to geminivirus infection.
**Figure 2.3** The geminivirus intergenic region is methylated *in vivo*. The CaLCuV and BCTV IRs are methylated in infected plants, and methylation levels are reduced in methylation-deficient mutants. Viral DNA isolated from floral tissues of infected wild-type or mutant plants was treated with bisulfite, and the IR was amplified by PCR. Amplified fragments were cloned, and 12 to 18 (CaLCuV) or 7 to 10 (BCTV) independent clones representing the encapsidated, viral strand were sequenced for each treatment. (A) The sequence of the CaLCuV IR, summarizing methylation detected in viral DNA isolated from wild-type plants of the ecotypes Ler-0, Ws-2, and Col-0, is shown. Methylated cytosines are indicated in red. Cytosines that were unmethylated in all three ecotypes are indicated in green. Cytosines methylated in 5 to 10% of the clones are noted by a dot, and more frequently methylated cytosines (>10 to 20% of the clones) are marked by an asterisk. Sites that show reduced methylation (zero to one clone) in the mutants are underlined. The locations of the conserved geminivirus hairpin and the putative, imperfectly repeated AL1 binding sites (AGGGGAG/AGGAGAG) are shown. The locations of start codons for the divergent AL1 and coat protein genes are also indicated. (B) Histograms represent the percentages of cytosine sites methylated in different sequence contexts in CaLCuV IR DNA isolated from *N. benthamiana* and *Arabidopsis* wild-type (Ler-0, Ws-2, and Col-0) plants and the indicated mutant plants (*adk2, ago4, cmt3, drm1 drm2*, and *kyp2*). Mutants are compared to the appropriate wild-type background. (C) Histograms represent the percentages of total cytosine residues methylated in different sequence contexts in BCTV IR DNA isolated from *N. benthamiana* and *Arabidopsis* wild-type Ler-0 plants, *ago4, kyp2*, and *cmt3* mutants.
Figure 2.3 The IRs of CaLCuV and BCTV are methylated in infected plants, and methylation levels are reduced in methylation-deficient mutants.
Figure 2.4 ChIP analysis of the CaLCuV IR. Tissue from CaLCuV-infected *N. benthamiana* (A) or *Arabidopsis* (B) plants was cross-linked with formaldehyde and subjected to ChIP using antibodies specific for the indicated histone H3 modifications. PCR was performed to amplify a ~400-bp fragment spanning the ~300-bp IR. The H3 acetyl polyclonal antibody was raised against a peptide acetylated at lysines 9 and 14. Controls included actin and transposons Tnt and Ta3. Additional control experiments in which plasmid DNA or plasmid DNA containing the CaLCuV genome was added to healthy plant extracts yielded similar signals with actin primers, but no signals were detected with the CaLCuV IR primers (data not shown). AB, antibody.
**Figure 2.5** *Arabidopsis ago4* mutants do not recover after infection with *BCTV L2* mutant virus. (A) The photographs show secondary (2°) tissue of a wild-type (Ler-0) and *ago4* mutant plant infected with *BCTV L2-1*. Note recovery (absence of symptoms) in wild-type plants and severe disease symptoms in the *ago4* mutant. (B) Southern blot hybridization analysis of *BCTV L2-1* and L2-2 DNA in primary (1°) and secondary (2°) infected tissues of wild-type and *ago4* plants. DNA extracts were prepared from pools of at least six plants and digested with ScaI to linearize the circular viral genome and cleave host genomic 18S ribosomal DNA (rDNA) repeats. Blots were incubated with a BCTV-specific probe and later with a probe specific for 18S rDNA to provide a loading control. Viral single-stranded DNA (ss) and linear dsDNA (ds) are indicated. Note reduced *BCTV L2-1* and L2-2 DNA levels in recovered, wild-type secondary tissue. (C) Methylation of the BCTV and *BCTV L2-1* and L2-2 IRs. The histograms represent the percentages of total cytosine residues methylated in different sequence contexts in viral DNA obtained from wild-type and *ago4* mutants. Note the greatly increased methylation observed in *BCTV L2-1* and L2-2 DNA obtained from wild-type secondary tissue exhibiting recovery from infection. WT, wild type.
CHAPTER 3

ARABIDOPSIS DOUBLE-STRANDED RNA BINDING PROTEIN DRB3 PARTNERS WITH DICER-LIKE 3 AND ARGONAUTE 4 IN METHYLATION-MEDIATED DEFENSE AGAINST GEMINIVIRUSES

The following chapter has been submitted for publication as follows: Raja, P., Wolf, J., Heard, I. and Bisaro, D.M. Jamie Wolf contributed Fig. 3.5. Isaac Heard assisted with cloning of constructs, and participated in experiments presented in Fig. 3.4.

3.1 Abstract

*Arabidopsis* encodes five double-stranded RNA binding proteins (DRB), two of which are known to associate with specific DICER-LIKE (DCL) proteins in different
small RNA-mediated pathways. Despite considerable effort, discovering the roles of the remaining DRB proteins has proven difficult. In this investigation, we used a geminivirus model system to identify DRB proteins that function in the methylation pathway with DCL3 and ARGONAUTE 4 (AGO4). Plants methylate geminivirus minichromosomes to inhibit virus replication, and methylation-deficient mutants are hypersusceptible to geminivirus infection. Using a panel of drb mutants, we found that drb3 plants uniquely exhibit a similar hypersensitivity, and viral genome methylation is substantially reduced in drb3 compared to wild-type plants. In addition, like dcl3 and ago4 mutants, drb3 plants fail to recover from infection and cannot accomplish the viral genome hypermethylation that is typically observed in recovered tissues. Using bimolecular fluorescence complementation, we found that DRB3 associates with DCL3 but not DCL4. DRB3:DCL3 complexes accumulate throughout the nucleus and in subnuclear bodies frequently associated with nucleoli. DRB3 also exhibits nuclear interaction with AGO4, and DRB3:AGO4 complexes are concentrated in distinct, small spots that are not associated with nucleoli. Analysis of geminivirus-derived small RNAs showed that their size and abundance was similar in drb3 and wild-type plants, indicating that DRB3 acts downstream of siRNA biogenesis, possibly by facilitating RISC loading. These studies demonstrate that DRB3 partners with DCL3 and AGO4 in the methylation pathway that conditions an important defense against DNA viruses and cellular transposons.
3.2 Introduction

Plants use a sophisticated RNA silencing machinery to fine tune the expression of selected genes and repress resident transposons and repeated sequences. Silencing can occur post-transcriptionally via the microRNA (miRNA) pathway and various pathways mediated by small interfering RNA (siRNA), or transcriptionally by siRNA-directed chromatin methylation (Baulcombe, 2004; Broderson and Voinnet, 2006; Vaucheret, 2006; Henderson and Jacobsen, 2007; Matzke et al., 2009). In addition, RNA silencing acts as an effective defense against exogenous nucleic acids, including those of RNA and DNA viruses (Ding and Voinnet, 2007; Mlotshwa et al., 2008; Ruiz-Ferrer and Voinnet, 2009). The small RNAs characteristic of individual silencing pathways are generated by distinct Dicer-like (DCL) ribonucleases that process larger double-stranded RNA (dsRNA) and hairpin precursors. In Arabidopsis, DCL1 generates 21 nucleotide (nt) miRNAs, whereas DCL2, DCL3, and DCL4 produce siRNAs that are typically 22, 24, and 21 nt in length, respectively. These small RNAs program ARGONAUTE (AGO)-containing effector complexes known as RISC (RNA-induced silencing complex) that mediate sequence-specific translational inhibition, transcript cleavage, DNA methylation, or some combination of these, depending on the small RNAs and AGO proteins they contain.

In addition to RNase III domains, Dicer proteins have dsRNA binding motifs (dsRBMs) that mediate sequence non-specific dsRNA binding as well as protein-protein interactions (Saunders and Barber, 2003; Chang and Ramos, 2005; Hiraguri et al., 2005). An important group of Dicer-interacting factors includes proteins related to C. elegans
RDE-4 (Tabara et al., 2002). These proteins also have dsRBMs but lack an obvious catalytic domain, and those characterized to date cooperate with specific Dicers in small RNA biogenesis and/or RISC loading. For example, the *Drosophila* R2D2 protein is not involved in siRNA biogenesis, but a Dicer-2:R2D2 complex binds and loads siRNA into RISC (Liu et al., 2003). Another *Drosophila* dsRBM-containing protein, R3D1/Loquacious, enhances miRNA production by interacting with Dicer-1 and increasing its affinity for dsRNA substrates, and also likely participates in RISC loading (Jiang et al., 2005). Other structurally and functionally related proteins include TRBP and PACT in mammals, and in plants comprise a family known simply as dsRNA binding proteins (DRBs). The DRB family in *Arabidopsis* consists of five members (Hiraguri et al., 2005), only two of which have so far been assigned to distinct silencing pathways. DRB1/HYL1 is required for efficient miRNA accumulation, and interacts and functions exclusively with DCL1 (Han et al., 2004; Vasquez et al., 2004; Hiraguri et al., 2005; Curtin et al., 2008). More recent work indicates that DRB1 also is involved in selecting the guide strand that is loaded into RISC (Eamens et al., 2009). The DRB4 protein interacts and functions with DCL4 in trans-acting siRNA (ta-siRNA) biogenesis and in siRNA-mediated defense against RNA viruses (Adenot et al., 2006; Nakazawa et al., 2007; Curtin et al., 2008; Qu et al., 2008). In addition, the *Cauliflower mosaic virus* (CaMV) P6 silencing suppressor interacts with DRB4, suggesting that a DRB4:DCL4 complex also targets DNA virus transcripts (Haas et al., 2008). DRB4 is not required for the siRNA synthesis, however, the accumulation of 21 nt ta-siRNAs and virus-derived siRNAs is reduced in *drb4* plants. Determining the primary roles and DCL partners of
DRB2, DRB3, and DRB5 has proved difficult, possibly because of functional redundancy. What can be said at present is that none of the DRBs is required to generate the 22 and 24 nt siRNAs characteristic of DCL2 and DCL3, and none appears to be essential for the pathways in which these siRNAs are believed to function (Curtin et al., 2008).

Geminiviruses have small (2.5 to 3.0 kb) genomes of circular, single-stranded DNA that replicate in the nuclei of infected cells by a rolling circle mechanism that utilizes double-stranded DNA replicative form (dsDNA RF) intermediates (Hanley-Bowdoin et al., 2004; Rojas et al., 2005; Jeske, 2009). Geminivirus genomes typically specify four to seven proteins, none of which has polymerase activity. Instead, these viruses rely on host machinery for replication and transcription, both of which are believed to occur on viral chromatin templates composed of dsDNA RF associated with cellular histones organized as typical nucleosomes. Transcripts produced from viral minichromosomes are subject to post-transcriptional gene silencing (PTGS), and as a counterdefense several geminivirus proteins are known to suppress this aspect of silencing (Vanitharani et al., 2005; Bisaro, 2006). In addition, RNA-directed methylation of viral chromatin leading to transcriptional gene silencing (TGS) is also employed as an antiviral defense, and geminivirus proteins have also been shown to suppress methylation and TGS (Raja et al., 2008; Buchmann et al., 2009; Raja et al., 2010).

Genetic analysis has revealed a similarity between the pathways used to methylate geminivirus chromatin and to maintain methylation of resident transposons (Henderson and Jacobsen, 2007; Lisch, 2009). For example, *Arabidopsis* plants with mutations that
inactivate RNA polymerases IV and V, AGO4, or DDM1 show extreme sensitivity to geminiviruses, and ago4 mutants are unable to recover from infection. Transposon silencing and geminivirus defense also share multiple downstream effectors, as mutant plants deficient for the cytosine methyltransferases MET1, CMT3, and DRM1/2 (which promote methylation primarily at CG, CNG, and CHH sites, respectively), and the histone 3 lysine 9 (H3K9) methyltransferase KYP2/SUVH4, are also hypersusceptible to geminiviruses (Raja et al., 2008). Further, enhanced susceptibility is correlated with reduced viral genome methylation. Compared to wild-type plants, total methylation in the viral intergenic region (IR), which contains divergent early and late gene promoters flanking the origin of replication, was reduced approximately 10 to 25%, depending on the mutant, with most differences occurring at non-CG sites (Raja et al., 2008). That larger reductions were not observed is testimony to the functional redundancy of silencing pathway components. Nevertheless, these findings highlight the importance of viral genome methylation as a host defense, since even relatively modest reductions lead to a severe infection phenotype. From a practical perspective, these findings also suggest that geminiviruses could serve as sensitive probes for the identification of methylation pathway components and effectors. Here, we take advantage of this to show that DRB3 partners with DCL3 and AGO4 in the methylation pathway that mediates defense against geminiviruses.
3.3 Materials and Methods

3.3.1 Arabidopsis mutants

The dcl mutants (Xie et al., 2004) were obtained from Dr. J.C. Carrington, and all others from the Arabidopsis Biological Research Center at The Ohio State University. The following seed stocks were used: wild-type Col-0 (CS60000) Columbia ecotype, drb2 (SALK_012017 and CS849395/At2g28380), drb3 (SALK_022644/At3g26932), drb4 (SALK_000736/At3g62800) (Adenot et al., 2006), drb5 (SALK_031307C/At5g41071), dcl2-1 (SALK_064627/At3g03300), dcl3-1 (SALK_005512/At3g43920), dcl4-2 (GABI_160G05/At5g20320), wild-type Ler-0 (CS20) Landsberg erecta ecotype, ago 4-1 gl1-1 (CS6364/At2g27040) (Zilberman et al., 2003). Plants were reared in growth rooms at 22°C with 12 hour light/dark cycles.

3.3.2 Virus inoculation

Agroinoculation of Arabidopsis plants with CaLCuV or BCTV was carried out as previously described (Raja et al., 2008). Arabidopsis plants were inoculated within 5 days of bolting. Bolts were cut where they emerged from the rosette and inoculum was applied to the freshly cut stem, which was then punctured with an insect pin multiple times. CaLCuV symptoms were observed and plants harvested 14-21 days post-inoculation; BCTV symptoms were observed and plants harvested 21-30 days post-inoculation. Inflorescence tissue showing visible symptoms was harvested from Arabidopsis. For each sample, tissue was pooled from four infected plants. For BCTV
recovery experiments, plants were agroinoculated with BCTV or BCTV L2\(^{-}\) virus and after the primary harvest, plants were allowed to continue growing under the same conditions. The BCTV L2-I null mutant has been previously described (Hormuzdi and Bisaro, 1995). Observations were made about symptom development in secondary inflorescence tissue, which was then harvested as secondary tissue 14-21 days after harvesting of primary tissue.

3.3.3 Cloning for interaction studies

cDNA clones for DRB3 (U66272, 1080 bp), DRB4 (C104842, 1099 bp) and AGO4 (U09147, 2806 bp) were obtained from the Arabidopsis Biological Resource Center. DCL3 and DCL4 cDNAs were obtained from Dr. J.C. Carrington. PCR primers were used to amplify these genes with the introduction of a PacI site at the 5' end and an AscI site at the 3' end. The PCR products were subsequently digested with PacI and AscI and ligated into similarly digested yeast vectors pAS2 and pACT2, and BiFC vectors p2YN, p2YC, pYN and pYC.

3.3.4 Yeast two-hybrid analysis

Protein interactions involving DCL and DRB proteins were tested using the yeast two-hybrid system in the strain PJ69-4a (Strain a (\textit{trp1-901 leu2-3, 112 ura 3-52 his3-200 gal4A gal80A GAL2-ADE2 LYS2::GAL1-HIS3 met2::GAL7-lac7}) (9, 10). Genes were cloned into the bait plasmid pAS2 (Trp\(^{+}\)) and prey plasmid pACT2 (Leu\(^{+}\)) containing the
DNA binding domain and activation domain, respectively. Positive interactions between proteins were identified by growth of yeast cells transformed with pAS2 and pACT2 vectors containing the test proteins on media lacking adenine and histidine. Media also lacked leucine and tryptophan to ensure maintenance of the expression plasmids. (Data presented in Appendix B; Table B.1).

3.3.5 BiFC Analysis

Interactions between DCL and DRB proteins in plants were tested using BiFC (Bimolecular Fluorescence Complementation) (Hu et al., 2002). The construction of BiFC expression vectors using enhanced YFP has been described previously (Yang et al., 2007).

BiFC expression vectors:

p2YN: YFP codons 1 to 158 fused 3’ of gene.

p2YC: YFP codons 159 to 238 fused 3’ of gene.

pYN: YFP codons 1 to 158 fused 5’ of gene.

pYC: YFP codons 159 to 238 fused 5’ of gene.

The DRB and DCL genes were amplified by PCR with PacI and AscI sites at the 5’ and 3’ ends. The PCR products were cloned into BiFC vectors between PacI and AscI sites. Cloned plasmids were transformed into Agrobacterium tumefaciens strain GV3111, and cultures were used to infiltrate 3-4 week old N. benthamiana plants as previously described (Wang et al., 2005). Briefly, agrobacterium cells grown in liquid culture
(OD$_{600} = 1$) were sedimented and incubated for 3 hours in infiltration medium containing MES, acetylsyringone and MgCl$_2$. Cultures containing p2YN- and p2YC-based plasmids were mixed 1:1 immediately prior to infiltration of young leaves using a blunt syringe. RFP-histone 2B and RFP-fibrillarin (Chakrabarty et al., 2007) were used as markers for the nucleus and nucleolus respectively. Leaf tissue was analyzed by microscopy approximately 36 hours post-infiltration using a Nikon PCM 2000 confocal laser scanning microscope equipped with argon and green helium neon lasers with excitation wavelengths of 488 nm and 544 nm, respectively. To record YFP fluorescence, a band-pass emission filter (EM515/30HQ) with a 450- to 490 nm excitation wavelength and 515 nm emission wavelength was used. To record RFP fluorescence, a 565 nm long-pass filter (E565LP) was employed. Images were captured using Simple PCI Software and compiled with Adobe Photoshop.

3.3.6 Bisulfite Sequencing

Bisulfite sequencing was performed as described (Raja et al., 2008, Frommer et al., 1992). DNA isolated from infected plant tissue was linearized overnight using appropriate restriction enzymes (ScaI for BCTV, XmnI for CaLCuV). Proteinase K digestion was subsequently carried out overnight, followed by bisulfite conversion using CT conversion reagent (EZ-DNA Methylation Gold; Zymo Research). Primers were designed against converted template, and the intergenic region of the viral genome was amplified by PCR. The PCR product was purified using Promega Wizard columns and TA cloned, and individual clones were sequenced at The Ohio State University Plant
Microbe Genomics Facility. For conversion control, plasmids containing CaLCuV DNA or BCTV DNA were added to a vast excess of healthy plant DNA extract and treated with bisulfite reagent. The following forward and reverse primers were used to amplify CaLCuV and BCTV intergenic regions (IR) following bisulfite conversion:

CaLCuV2556CF, GGGGATATGTTAAGAATATTTTGGG (forward)
CaLCuV359CR, TCCCCACCATAAAACACCAC (reverse)
BCTV2640CF, GGGATATGTAAGAAATATG (forward)
BCTV147CR, TCTCCCCTTCTATTAACCAATCAAC (reverse).

The bisulfite reaction was carried out in a thermocycler under the following conditions:
1. 98°C for 10 minutes
2. 64°C for 30 minutes
3. 60°C for 6 minutes
4. 53°C for 30 minutes
5. Go to 3, 8 times
6. 4°C, forever

Data was analyzed and dot plots prepared using Kismeth (Gruntman et al., 2008).

3.3.7 Small RNA Analysis

RNA was isolated from 0.2 g of symptomatic floral tissue using Ribozol (Amresco) or Trizol (Invitrogen) reagent. Each sample was pooled from 3-4 plants. Small RNAs were analyzed as described (Qu et al., 2008). Total RNA (5-8 µg) was loaded onto a 16% denaturing polyacrylamide gel containing 8 M urea in 0.1X TBE. The
gel was run until 10 minutes after the bromophenol blue dye ran off the gel. The gel was electroblotted onto Hybond nylon membrane in 0.5X TBE, 4°C, at 40 V, for 1 hour. RNA was crosslinked to the membrane using a Stratagene UV crosslinker at 1700 kJ. The membrane was prehybridized in Ambion Ultrahyb oligo for 1 hour at 40-42°C. A mixture of oligonucleotides (20 µM) was used as probe. Oligonucleotides were labeled with Fermentas T4 polynucleotide kinase at 37°C for 30-60 minutes. Labeled probes were denatured and hybridization was carried out overnight at 40-42°C. The membrane was washed 3 times for 20 minutes each with 2xSSC, 0.5% SDS at the same temperature. Ethidium bromide stained rRNA was used as a loading control.

The following oligonucleotides were used as size markers:

BCTVCPF57: 5’-TCAAGGTGCGTGCAGCCGAAGAA-3’ (21 nt)

BCTVIRsR_2: 5’-TCTTCAGGAAGTTTCCGCTCA-3’ (22 nt)

BCTVIRqRTR: 5’-ACGGATGGCCCTTTATGAGTTGT-3’ (24 nt)

A mixture of the following sense and antisense oligos to the common region and coat protein were used as probes:

revcomBCTVCPF57: 5’-TTCTTCGGCCAGCGACTTTGA-3’ (21 nt)

revcomBCTVIRsR_2: 5’-TGAGCGGAAAACCTTCTGAAGA-3’ (22 nt)

revcomBCTVIRqRTR: 5’-ACAACTTCATAAGGCCATCCGT-3’ (24 nt)

BCTVCPF1122: 5’-TTCCCCGATAACGTCAGGCTAT-3’ (22 nt)

BCTVCPR1273: 5’-TGGCATTGAAGGAGCCTTGT-3’ (20 nt)

BCTVIRsF_1: 5’-GACACGTTAGGGAAGGGTACTGTAGAA-3’ (24 nt)

BCTVIRsF_2: 5’-TAAGATTTGTTGACTGGTCAATAGAAGG-3’ (28 nt)
BCTVIRsR\(_1\): 5’-GT TTGATTGTTGCCCATT TTTCG-3’ (27 nt)

BCTVIRqRTF: 5’-ATTGGACTCCGATGACGTGCTTA-3’ (24 nt)

3.4 Results and Discussion

3.4.1 *Arabidopsis drb3* mutants are hypersusceptible to geminivirus infection.

Methylation-deficient *Arabidopsis* mutants are hypersusceptible to geminiviruses (Raja et al., 2008). To assess whether *Arabidopsis* DRB proteins might be involved in methylation-mediated defense, *drb2*, *drb3*, *drb4*, and *drb5* mutant plants were challenged by geminivirus infection. DRB1 was not included in this study because it is known to participate in the miRNA pathway. Initial experiments were performed with *Cabbage leaf curl virus* (CaLCuV) and *Beet curly top virus* (BCTV) to evaluate responses to distinct types of geminiviruses. CaLCuV belongs to the genus *Begomovirus* and has a two component genome, while BCTV has a monopartite genome and is the type member of the genus *Curtovirus*. To observe a range of symptom enhancement, plants were inoculated under conditions that normally elicit mild symptoms.

Geminivirus-infected *drb2*, *drb4*, and *drb5* mutant plants showed mild to moderate disease enhancement relative to wild-type plants. By contrast, disease symptoms in *drb3* plants inoculated with either CaLCuV or BCTV were greatly enhanced and accompanied by considerable stunting (Fig. 3.1). A similar hypersusceptibility was previously observed with methylation-deficient mutants and was invariably correlated with reduced viral genome methylation, particularly at non-CG sites (Raja et al., 2008).
Bisulfite sequencing was carried out to assess the methylation status of the intergenic region (IR) of CaLCuV DNA A genomes from infected *drb3* plants. The sequence analyzed contains 79 cytosines in different contexts, including 19 CG, 13 CNG, and 47 CHH sites. Following bisulfite treatment to convert unmethylated cytosines to uracil, the viral strand of the IR was amplified by PCR. The PCR products were cloned, and sequences of 12 to 18 clones were analyzed per treatment. As shown in Fig. 1C, methylation of CaLCuV genomes in *drb3* mutants was substantially reduced, primarily at non-CG sites, relative to those from wild-type plants (Col-0 ecotype). CG methylation was essentially unchanged, while reductions of approximately 25% and 30% were observed in CHH and CNG contexts, respectively. We concluded that the disease phenotype of geminivirus-infected *drb3* mutant plants, and the reduced methylation of viral genomes in these mutant plants, was consistent with a defect in the methylation pathway.

An earlier study determined that methylation levels of several cellular transposons were unaltered in *drb3* plants (Curtin et al, 2008). We speculate that the increased demand imposed on the methylation pathway by geminivirus infection allowed us to detect significant differences in viral DNA methylation between wild-type and *drb3* plants.
3.4.2 *Arabidopsis dcl3* and *drb3* mutants are unable to recover from geminivirus infection.

Perhaps the most compelling argument for methylation as an antiviral defense comes from studies that have associated the methylation pathway with host recovery from geminivirus infection. Recovery occurs when tissues arising after the establishment of a systemic infection exhibit symptom remission and contain very little virus. Wild-type *Nicotiana benthamiana* and *Arabidopsis* plants rarely recover from infection with wild-type BCTV, but nearly always recover from infection with BCTV L2\(^{-}\) mutant virus, which lacks a pathogenicity factor that can suppress both PTGS and TGS and non-specifically inhibits DNA methylation (Hormuzdi and Bisaro, 1995; Wang et al., 2003; Wang et al., 2005; Raja et al., 2008; Buchmann et al., 2008). We and others have demonstrated that the reduced amount of geminivirus DNA in recovered tissue is hypermethylated (Raja et al., 2008; Hagen et al., 2008; Rodriguez-Negrete et al., 2009), and we further showed that *ago4* plants are unable to recover from BCTV L2\(^{-}\) infection, confirming that the siRNA-directed methylation pathway is required for recovery (Raja et al., 2008). We hypothesized that an analysis of mutant plants for recovery could provide a definitive and sensitive means of identifying methylation pathway components.

To test the hypothesis, *dcl2*, *dcl3*, and *dcl4* plants were examined. Considerable functional redundancy exists among *Arabidopsis* DCL enzymes, although DCL3 and the 24 nt siRNAs it generates have clearly been associated with chromatin methylation (Xie et al., 2004; Gasciolli et al., 2005; Henderson et al., 2006). However, no doubt because
of functional redundancy, only mild to moderate symptom enhancement was observed following BCTV or CaLCuV infection of *Arabidopsis dcl2, dcl3,* and *dcl4* plants in earlier studies, precluding unequivocal identification of an individual DCL protein as the key player in methylation-mediated defense (Blevins et al., 2006; Raja et al., 2008). Similar results were again obtained following inoculation of these *dcl* mutants with BCTV *L2* virus. However, after the establishment of systemic infections and removal of primary infected tissue, only *dcl3* mutant plants were unable to recover from infection. Severe symptoms appeared in new secondary, axillary shoots in all of the 32 *dcl3* plants tested (Fig. 3.2). By contrast, symptom remission was observed in new shoots of all wild-type, *dcl2,* and *dcl4* plants (32 each, Appendix B: Fig. S1). Thus, the host recovery assay could discriminate between functionally redundant DCL activities and identify DCL3 as the enzyme most closely associated with methylation.

To confirm the role of DRB3 in methylation-mediated defense, *drb2, drb3, drb4,* and *drb5* mutant plants were inoculated with BCTV *L2* and, as observed with wild-type BCTV, symptoms were much more severe in *drb3* plants than in other *drb* mutant or wild-type plants. As expected, new secondary shoots of wild-type plants infected with BCTV *L2* recovered from infection and showed little evidence of symptoms, as did most new shoots of the *drb2, drb4,* and *drb5* mutant plants (32 plants each, Appendix B: Fig. S1). However, while new tissues of *drb4* plants did not show floral deformation typical of BCTV symptoms, the plants were occasionally somewhat stunted compared to wild-type plants (Fig. 3.2). In contrast, *drb3* mutants did not recover, and severe symptoms appeared in all new shoots of the 32 *drb3* plants inoculated with BCTV *L2* (Fig 3.2).
Taken together, these and previous studies indicate that the methylation pathway conditioned by DCL3 and AGO4 is required for host recovery from geminivirus infection, and strongly link DRB3 to this pathway. In contrast, DRB2, DRB4, and DRB5 are not required for recovery, although supporting roles cannot be ruled out.

3.4.3 *Arabidopsis drb3* and *dcl3* mutant plants cannot hypermethylate the viral genome.

We previously showed that the relatively small amount of BCTV L2' DNA present in recovered tissue is hypermethylated, while the same virus from comparable but non-recovered tissue of *ago4* mutants is not (Raja et al., 2008). Thus, we compared the methylation status of BCTV L2' DNA obtained from recovered tissue with DNA from non-recovered tissue of *drb3*, *dcl3*, and *ago4* mutants. Mutant *drb4* and *dcl4* plants were included in this analysis because a DRB4:DCL4 complex plays a major role in PTGS-mediated defense against RNA viruses and DNA virus transcripts (Blevins et al., 2006; Bouche et al., 2006; Qu et al., 2008; Haas et al., 2008). Bisulfite sequencing was carried out as before, except that primers designed to amplify the viral strand of the BCTV IR were employed. PCR products were cloned and the sequences of 12 clones were determined per treatment. The BCTV IR region analyzed contains 44 cytosines in the following contexts: 10 CG, 8 CNG, and 26 CHH. Bisulfite sequencing data for all wild-type and mutant plants examined is illustrated graphically in Fig. 3.3A. Cytosine methylation profiles representing individual BCTV L2' IR clones obtained from wild-
type (Col-0), drb3, and dcl3 (Fig. 3.3B, 3.3C, and 3.3D) plants are also shown. Profiles for all wild-type and mutant plants are presented in Appendix B: Fig. S2.

As observed previously, viral genomes obtained from recovered tissue of wild-type plants was hypermethylated in all sequence contexts, with about 63% (Col-0) to 77% (Ler-0) of total cytosines methylated (Fig. 3A). Methylation levels were reduced in all sequence contexts in viral genomes obtained from drb3, drb4, dcl3, and ago4 mutants. However, reductions were considerably greater in viral genomes obtained from non-recovered drb3 mutant plants compared to recovered drb4 plants (34% and 54% total cytosines methylated, respectively). Likewise, methylation was reduced to a greater extent in viral genomes from non-recovered dcl3 plants than from recovered dcl4 plants (48% and 63% of total cytosines methylated, respectively). Methylation levels in dcl4 plants were similar to those seen in wild-type Col-0. The greatest reduction relative to comparable wild-type (Ler-0) was noted in genomes from ago4 mutant plants (77% compared to 38% total cytosines methylated) (Fig. 3.3A). These results confirm that geminivirus genomes obtained from recovered tissues (here from wild-type, drb4, and dcl4 plants) are more densely methylated than genomes from non-recovered drb3, dcl3, ago4 plants, and further associate DRB3 with the methylation pathway that is conditioned by DCL3 and AGO4 and required for recovery. The smaller reduction in methylation observed in genomes from drb4 plants suggests that this protein might also act in the methylation pathway to a limited extent.

We previously found that histone H3 associated with geminivirus DNA carries modifications characteristic of both active (acetylated H3) and repressed (H3K9me2)
chromatin, suggesting that populations of active and repressed genomes co-exist in infected plants (Raja et al., 2008). Cytosine methylation profiles support this idea, as they reveal the presence of both hypermethylated and hypomethylated (or unmethylated) genomes in infected plants. Thus, viral DNA populations present in recovered tissue from wild-type plants are skewed toward densely methylated genomes (Fig. 3.3B, Appendix B, Fig. S2), with a smaller proportion of lightly methylated and presumably active genomes than is observed in non-recovered symptomatic tissues (Fig. 3.3C and 3.3D, Appendix B, Fig. S2).

3.4.4 DRB3 interacts with DCL3 and AGO4 in *Nicotiana benthamiana* cells.

To our knowledge, the ability of DRB3 to physically associate with other dsRBM proteins has not been tested. We used bimolecular fluorescence complementation (BiFC) to examine *in vivo* interactions of DRB3 with DCL3, AGO4, DCL4, and DRB4 proteins (Hu et al., 2002). BiFC involves fusion of potentially interacting proteins with the N- or C-terminal portions of yellow fluorescent protein (YFP), and subsequent co-expression of the fusion proteins. Association of the fusion proteins reconstitutes YFP, producing a fluorescent signal that at once reveals the interaction and where it occurs in the cell. We used a transient system that involves co-infiltration of *N. benthamiana* leaves with mixtures of *Agrobacterium tumefaciens* cells harboring constructs that express the fusion proteins (Yang et al., 2007). All proteins were expressed as both N- and C-terminal fusions with both the N- and C-terminal portions of YFP. When interactions were detected, they were observed in all
combinations tested. Proteins were judged not to interact when signal was absent in all pairwise combinations. RFP-histone 2B was used to locate nuclei, and RFP-fibrillarin was used as a marker for the nucleolus in higher magnification images. In no case was a signal observed when any of the fusion proteins were individually expressed, or when co-expressed with empty vector or various negative control YFP fusion proteins.

In agreement with evidence linking DRB3 to the methylation pathway, we found that DRB3 and DCL3 interact exclusively in the nucleus (Fig. 3.4A; additional images for all tested proteins are presented in Appendix B: Fig. S3.) The DRB3:DCL3 signal was observed throughout the nucleus and was especially intense in punctate subnuclear spots, some but not all of which were associated with nucleoli (Fig. 3.4B). This is consistent with previous work that localized DCL3, AGO4 and siRNA to ribonucleoprotein processing centers (Cajal bodies) that can move into and out of the nucleolus (Pontes et al., 2006). By contrast, no evidence of interaction was observed when any combination of DRB3 and DCL4 fusion proteins was expressed. As a positive control, we confirmed that DRB4 interacts with DCL4, both of which have been shown to localize to the nucleus (Hiraguri et al., 2005; Nakazawa et al., 2007). DRB4:DCL4 complexes were similarly observed in nucleolus-associated bodies (Fig. 3.4B). Unlike DRB3:DCL3, however, DRB4:DCL4 complexes were also present in the cytoplasm, although the signal was relatively weak (Fig. 3.4A). Interaction between DRB4 and DCL3 was also detected, although in this case the signal was observed throughout the nucleus, including the nucleolus (Fig. 3.4). In addition, DRB3 and DRB4 were observed to interact throughout the cytoplasm and the nucleus, with the exception of nucleoli. In
addition, complexes were concentrated in small spots that resemble the sites of DRB3:AGO4 interaction (see below; Fig. 3.4B).

AGO4 has been reported to localize in two distinct subnuclear bodies, namely nucleolus-associated Cajal bodies and smaller AB-bodies, which also contain RNA POL V and DRM2 (Pontes et al., 2006; Li et al., 2006; Li et al., 2008). We observed that DRB3 interacts with AGO4 throughout the nucleus, and in most nuclei complexes were concentrated in small nuclear bodies that were not associated with nucleoli. These bodies are distinct from the larger bodies containing DRB3:DCL3 (Fig. 3.4B). We speculate that DRB3:DCL3 complexes preferentially accumulate in Cajal bodies, while DRB3:AGO4 complexes are enriched in AB-bodies. Thus, DRB3 may choreograph nuclear events by interacting with upstream (DCL3) and downstream (AGO4) pathway components in different locations. DRB3:AGO4 signal was also apparent in the cytoplasm (Fig. 3.4A). Surprisingly, DCL3:AGO4 complexes were distributed throughout the nucleus but were excluded from the nucleolus, were not concentrated in punctate spots, and were absent from the cytoplasm. The DCL4 protein did not appear to interact with AGO4 and produced at best a weak background signal. However, DRB4 showed interaction with AGO4 in the cytoplasm and throughout the nucleus, except for the nucleolus (Fig. 3.4).

These studies demonstrate *in vivo* physical associations between DRB3:DCL3 and DRB3:AGO4 that are consistent with functional cooperation between these proteins in the methylation pathway. The absence of interaction between DRB3 and DCL4 is in accord with a previously noted one-for-one specificity in dsRBM protein-Dicer
interactions. However, the DRB4:DCL3 interaction appears to be an exception to this general rule. The significance of this, and the DRB4:AGO4 and DRB3:DRB4 interactions, is at present unclear, but might suggest cooperation or functional redundancy between the DRB3 and DRB4 proteins.

While it is possible that the interactions observed in these BiFC experiments are mediated by bridging factors, *in vitro* studies have demonstrated that dsRBMs can sponsor direct protein-protein interactions (Chang and Ramos, 2005; Hiraguri et al., 2005). However, the presence of dsRBMs is not sufficient, as we observed no association between DRB3 and DCL4. In addition, we found that DRB3:DCL3 and DRB3:AGO4 complexes are enriched in distinct compartments, even though these proteins were present throughout the nucleus. Thus, it would appear that other as yet unidentified cellular factors influence where these complexes preferentially accumulate.

3.4.5 DRB3 is not required for the biogenesis of 24 nt siRNAs by DCL3.

While some dsRBM proteins do not significantly impact small RNA levels, others facilitate the biogenesis of small RNA species by the Dicers with which they associate. We asked whether DRB3 is necessary for the accumulation of 24 nt siRNA generated by DCL3 in response to BCTV. While it was previously shown that levels of 24 nt siRNA derived from an RNA virus are not diminished in *drb3* plants (Curtin et al., 2008), virus-specific siRNAs of this size class are considerably more abundant in geminivirus infected plants (Akbergenov et al., 2006; Blevins et al., 2006). RNA was isolated from wild-type, *drb3*, and *dcl3* plants inoculated with BCTV. Virus-infected *drb4*, *dcl4*, and *ago4* plants
were included in this study for comparison. RNA samples were fractionated on polyacrylamide gels and probed with a mixture of $^{32}$P-labeled oligonucleotides specific for the BCTV IR and the coat protein coding region.

As expected, dcl3 mutants lack 24 nt siRNAs corresponding to the BCTV genome, and dcl4 mutants showed much reduced 21 nt siRNA levels (Fig. 3.5). However, little change was apparent in the type and abundance of small RNA species present in the drb3, drb4, and ago4 mutant plants compared to wild-type plants (Col-0 or Ler-0). Thus, neither DRB3, DRB4, nor AGO4 are required for the synthesis and accumulation of geminivirus-derived siRNAs.

3.5 Conclusion

By taking advantage of the extreme sensitivity of methylation-deficient plants to geminiviruses, we have generated several lines of evidence that strongly link DRB3 with the nuclear methylation pathway that involves DCL3 and AGO4. 1) Like ago4 and other methylation-deficient mutants, drb3 plants show enhanced susceptibility to geminivirus infection. 2) Neither drb3, dcl3, or ago4 plants can recover from infection with BCTV L2 virus. By comparison, drb2, drb4, and drb5 plants are not hypersusceptible to geminiviruses and recover from infection. 3) The small amount of viral DNA obtained from asymptomatic, recovered shoots consists largely of hypermethylated viral genomes. By contrast, a larger proportion of hypomethylated viral genomes is evident in comparable, non-recovered tissue from drb3, dcl3, and ago4 plants. 4) DRB3 interacts with DCL3 and AGO4 in distinct subnuclear bodies. Finally, DRB3 does not
significantly affect the biosynthesis or accumulation of geminivirus-derived siRNAs. Thus DRB3 acts downstream of DCL3-mediated dsRNA processing and, by analogy with other dsRBM proteins, we speculate that it may be involved in AGO4-RISC loading.

The studies presented here underscore the utility of geminiviruses as models for the analysis of mechanisms controlling the methylation and epigenetic regulation of cellular chromatin, and lay the groundwork for further analysis of DRB3 function.
Figure 3.1 *Arabidopsis* *drb3* mutants show enhanced susceptibility to geminivirus infection. (A) Photographs illustrate CaLCuV disease symptoms in wild-type (Col-0 ecotype) and *drb3* mutant plants, shown 14 days post-inoculation. (B) BCTV symptoms shown 21 days post-inoculation. BCTV has an inherently longer latent period than CaLCuV. (C) Histograms show the percentage of cytosines methylated in CaLCuV IR DNA isolated from wild-type and *drb3* plants, as determined by bisulfite sequencing.
Figure 3.2 *Arabidopsis* *drb*3 and *dcl*3 mutants do not recover from infection with BCTV L2’ mutant virus. (A) The photographs show secondary tissue of wild-type (Col-0), *drb*3, *drb*4, *dcl*3, and *dcl*4 plants infected with BCTV L2’ virus. Note recovery (absence of symptoms) in wild-type, *drb*4, and *dcl*4 mutants, and severe disease symptoms in the *drb*3 and *dcl*3 mutants. Although deformation of floral tissue was not observed in secondary tissue of infected *drb*4 mutants, plants were occasionally stunted. An example of a stunted *drb*4 plant is shown in this figure. (B) Close-up views of the same infected plants.
Figure 3.3 Arabidopsis *drb3* and *dcl3* mutants fail to hypermethylate the viral genome. Methylation of BCTV *L2*^−^ IR DNA was assessed by bisulfite sequencing. (A) Histograms indicate the percentages of cytosine residues methylated in different sequence contexts in viral DNA obtained from secondary tissues of wild-type (Col-0 or Ler-0) or mutant plants. The *drb* and *dcl* mutants are in Col-0, whereas *ago4* is in the Ler-0 ecotype background. Cytosine methylation profiles of BCTV *L2*^−^ DNA obtained from (B) wild-type Col-0, (C) *drb3*, and (D) *dcl3* plants are shown. The dots represent all cytosines in the IR and are color coded according to sequence context (red: CG, blue: CNG, green: CHH). Filled circles indicate methylation, and each line represents the sequence of an individual clone. Most viral genomes are densely methylated in recovered tissue, and presumably these are repressed. A larger proportion of hypomethylated (or unmethylated) genomes are apparent in non-recovered tissue from *drb3* and *dcl3* mutants.
Figure 3.3 Arabidopsis drb3 and dcl3 mutants fail to hypermethylate the viral genome.
Figure 3.4 DRB3 independently interacts with DCL3 and AGO4 in distinct subnuclear bodies. BiFC analysis of the indicated DCL, DRB and AGO proteins in *N. benthamiana* epidermal cells was performed. Constructs expressing DCL3, DCL4, DRB3, DRB4 and AGO4 fused to the N- or C-terminal portion of YFP were delivered by agroinfiltration to *N. benthamiana* leaves. Cells were photographed 36 h postinfiltration using a confocal laser scanning microscope. RFP-histone 2B (RFP-H2B) and RFP-fibrillarin were used as markers for the nucleus and nucleolus, respectively. Protein combinations are indicated above each photograph. The photographs in (A) show lower-magnification views (20X) while (B) shows high magnification views (100X). Note that DRB3 interacts with DCL3 and AGO4 in distinct subnuclear locations.
**Figure 3.5** DRB3 is not required for biogenesis of geminivirus-derived siRNAs. RNA blot hybridization showing the accumulation of virus-specific siRNAs in BCTV-infected wild-type (Col-0 and Ler-0) and mutant plants. Ethidium bromide stained rRNA served as a loading control.
CHAPTER 4

IDENTIFICATION OF COMPONENTS REQUIRED FOR RECOVERY FROM GEMINIVIRUS DISEASE

The following chapter will be submitted for publication in its entirety, or contribute data to independent publications. As a chapter, this will be submitted, Raja, P.*, Wolf, J.*, Heard, I., Sanville B. and Bisaro, D.M. (*joint first authors). Genetic analysis is complete, and DNA methylation analysis has been carried out for a majority of the mutants. J. Wolf will carry out small RNA blot analysis and investigate histone methylation by ChIP in appropriate mutants.

4.1 Abstract

Wild-type *N. benthamiana* and *Arabidopsis* can recover from *BCTV L2* virus infection. Recovery occurs when asymptomatic axillary shoots arise following an initial infection. The secondary shoots contain low levels of viral DNA and are hypermethylated in a DCL3-DRB3-AGO4 dependent manner. Thus recovery has been previously shown
to be a sensitive and definitive method to identify components in the methylation pathway. In this study, we use the host recovery assay to determine whether non-CG methylation is crucial to recovery. Secondly, we exploit the geminivirus model system to examine mutants in plant-specific polymerases such as *nrpd* (*nrpd1a*; Pol IV) and *nrpe* (*nrpd1b*; Pol V) to evaluate which of these polymerases might be involved in siRNA directed methylation of geminivirus genomes. Thirdly, by testing *dcl* mutants that distinguish transcriptional gene silencing (TGS) and post-transcriptional gene silencing (PTGS), we ask whether TGS or PTGS is more important to recovery. Since previous studies indicated that a majority of methylation mutants that were hypersusceptible to geminiviruses showed reduced methylation at non-CG loci, we hypothesized that non-CG methylation was important to geminivirus defense and used the recovery assay to confirm this. Indeed, we found that mutants deficient in all non-CG methyltransferases (such as *ddc*) lost their ability to recover from infection by geminiviruses lacking the L2 protein.

We found that *pol V* mutants fail to recover from geminivirus disease, clearly implicating Pol V in the methylation pathway. The *pol IV* mutants, on the other hand, showed an interesting delayed recovery phenotype, suggesting that Pol IV is also involved in the methylation pathway but can be replaced by a redundant activity, likely Pol II. Finally, we found that while *dcl3* mutants do not recover, *dcl2/4* mutants deficient in PTGS show a mixture of recovered and symptomatic shoots, that we describe as mixed recovery. Further, the *dcl2/3/4* triple mutant showed an inability to recover that is unaccompanied by stunting, so lateral shoot extensions in this mutant appear to be longer than in the *dcl3* mutant which also shows a lack of recovery. The *dcl2/3/4* triple mutant surprisingly
remains methylation-competent, suggesting a previously undescribed role for DCL1 in methylation. Thus we have used the host recovery assay to identify the importance of non-CG methylation to host defense, identify important pathway components, and determine the role of TGS and PTGS in the process of recovery.

4.2 Introduction

Geminiviruses package small, circular single-stranded DNA genomes in a twin capsid. The ssDNA genome is amplified by rolling circle replication via a double-stranded DNA (dsDNA) intermediate that forms minichromosomes, which are targets of DNA and histone methylation leading to transcriptional gene silencing (TGS). Replication and transcription from viral templates is carried out by host polymerases, and geminivirus transcripts are subject to post-transcriptional gene silencing (PTGS) (Bisaro, 2006; Raja et al., 2010).

Plants have been reported to spontaneously recover from geminivirus disease under field conditions. Recovery is characterized by symptom remission correlated with low viral titers. Such natural recovery has previously been reported for members of the genus *Begomovirus* including *Pepper golden mosaic virus* (PepGMV), *Cucurbit leaf crumple virus* (CuLCrV) and *African cassava mosaic virus* (ACMV). Studies report different recovery phenotypes depending on the specific virus-host combination (Hagen et al., 2008; Rodriguez-Negrete et al., 2009; Chellappan et al., 2004). For instance, the begomovirus CuLCrV exhibits recovery in watermelon and cantaloupe. Recovered tissues were resistant to reinfection with CuLCrV, and contained hypermethylated viral
genomes relative to zucchini, a host where recovery does not occur (Hagen et al., 2008). Similar results were obtained in studies with PepGMV in pepper plants, where total virus-derived small RNA abundance was decreased in recovered tissue and recovery was correlated with hypermethylation of the viral genome and low viral titers (Rodríguez-Negrete et al., 2009).

Thus several studies have recently correlated the methylation pathway with recovery (Hagen et al., 2008, Rodríguez-Negrete et al., 2009). In previous work, we conclusively showed that methylation and transcriptional gene silencing (TGS) is an important host defense against geminiviruses, and that it is countered by geminivirus proteins AL2 and L2 that act as silencing suppressors via a mechanism that involves ADK inhibition and global reduction in methylation (Wang et al., 2005, Bisaro, 2006, Raja et al., 2008, Buchmann et al., 2009).

Begomovirus AL2 mutants are not systemically infectious because the AL2 protein, besides being a silencing suppressor, is also a transcriptional activator protein (TrAP) and is required for transactivation of the BR1 gene necessary for movement (Sunter and Bisaro, 1992). However, in the genus Curtovirus, L2 mutants are systemically infectious and inoculation of N. benthamiana or Arabidopsis with BCTV L2− null mutants (but not wild-type BCTV virus) results in recovery which is characterized by the absence of symptoms, reduced viral titers, and hypermethylation of the intergenic region of the viral genome. Thus, using a virus lacking its suppressor protein L2 enables us to recapitulate recovery in laboratory hosts such as N. benthamiana and Arabidopsis. We also previously established that DCL3, DRB3 and AGO4, involved
in the 24 nt siRNA pathway, are required for recovery (Raja et al., 2008, Chapter 3). Although the enhanced susceptibility of ago4 mutants to CaLCuV and BCTV was accompanied by relatively small reductions (15% to 25%) in viral genome methylation, ago4 mutants were unable to recover from a BCTV L2- virus infection. These experiments showed that AGO4 was required for recovery from geminivirus disease (Raja et al., 2008). Similarly, although functional redundancy resulted in only mild to moderate symptom enhancement in dcl3 mutants following geminivirus infection, dcl3 plants are unable to recover. Likewise, drb3 was identified as the DRB protein associated with the methylation pathway using the host recovery assay (Chapter 3). It has been established that AGO4, DCL3 and DRB3 are key components of the methylation pathway. Thus recovery is a definitive and sensitive tool to identify methylation pathway components and effectors.

In this study, our goal was to determine the relative roles of CG and non-CG methylation in the process of recovery, to evaluate the contribution of Pol IV and Pol V to RNA directed DNA methylation, and to reevaluate the roles of the Dicer-like enzymes in methylation to determine whether PTGS was also important to recovery. In order to address these questions, we initiated a genetic study of mutants in the methylation pathway. The study included plants lacking methylation pathway components in the following categories: methyltransferase mutants such as drm1/drm2 and cmt3, individually, and in combination in the ddc triple mutant; plant-specific RNA polymerases Pol IV (nrpd1a; nrpd) and Pol V (nrpd1b; nrpe), individually and in combination in the nrpd2a mutant; and Dicer-like enzymes (dcl2, dcl3, dcl4, dcl24,
The dcl1 mutants were excluded from the study owing to severe mutant phenotypes, which precluded analysis of virus symptoms.

Our findings confirmed that methylation plays an important role in recovery from geminivirus disease, and that this is countered by the BCTV L2 protein. We show that methylation-deficient mutant plants such as nrpe (Pol V), ago4, dcl3, and ddc lose the ability to recover. First, this demonstrates that non-CG methylation is required for recovery. Secondly, the absence of Pol V abolishes recovery, indicating that this polymerase plays an essential role in recovery. Thirdly, we confirm that the DCL3-mediated TGS pathway is required for mediating recovery, and that DCL2, DCL4 and the PTGS pathway is not required but plays a role in recovery. In addition, we propose that DCL1 may also play a role in the methylation pathway. Finally, we also describe unique and interesting phenotypes such as partial, mixed or delayed recovery, which require further investigation. In conclusion, we show that host recovery from geminivirus infection requires the Pol V-DCL3-AGO4 pathway.

4.3 Materials and Methods

4.3.1 Arabidopsis mutants

Mutants were obtained from the Arabidopsis Biological Research Center at The Ohio State University or from individuals. The following previously characterized seed stocks were used: kyp2 gl1-1 (CS6367/At5g13960) kryptonite/SuVH4 (Jackson et al., 2002); cmt3-7 (CS6365/At1g69770) chromomethylase (Lindroth et al., 2001); drm1
drm2 (CS6366/At1g28330 and At5g14620), domains rearranged methyltransferase (Cao and Jacobsen, 2002); ddc,drm1-2 drm2-2 cmt3-7 triple mutant (gift from Steve Jacobsen, nrd1a-3 (SALK_128428/At1g63020), nuclear RNA polymerase D 1A, Pol IV (Kanno et al., 2005); nrd1b-11 (SALK_029919/At2g40030), nuclear RNA polymerase D 1B, Pol V (Kanno et al., 2005); nrd2a-1 (SALK_095689/At3g23780), nuclear RNA polymerase D 2A, Pol IV and V (Herr et al., 2005); ddm1 (SALK_000590/At5g66750), decreased DNA methylation 1 (Jeddeloh et al., 1998); met1-7 (SALK_076522/At5g49160), decreased DNA methylation 2/methyltransferase-1, hemizygous due to seed abortion phenotype (Jeddeloh et al., 1998); adk1-1 (SALK_040957/At3g09820.1), adenosine kinase 1 (Xie et al., 2004); adk 2-1 (SALK_000565/At5g03300), adenosine kinase 2 (Young et al., 2006); dcl2-1 (SALK_064627/At3g03300), Dicer-like 2 (Xie et al., 2004); dcl3-1 (SALK_005512/At3g43920), Dicer-like 3 (Xie et al., 2004); dcl4-2 (GABI_160G05/At5g20320), Dicer-like 4 (Xie et al., 2005); dcl2-1 dcl3-1 dcl4-2, generated from stocks above by Dr. J.C. Carrington; dcl2-1 4-2 generated from stocks above by Dr. Keith Slotkin; rdr2-2 (SALK_059661/At4g11130), RDR polymerase 2 (Herr et al., 2005); and rdr6-11 (CS24285/At3g49500), RDR polymerase 6, silencing defective 1, suppressor of gene silencing 2 (Peragine et al., 2004).
4.3.2 Virus inoculation

Infection of Arabidopsis plants with CaLCuV and BCTV was carried out by agroinoculation using an overnight culture containing tandem copies of the virus as described previously (Raja et al., 2008), based on an earlier method used for N. benthamiana and sugar beet (Hormuzdi and Bisaro, 1993). Arabidopsis plants were inoculated within 5 days of bolting. Bolts were cut at the rosette, and inoculum applied to the freshly cut stem, which was then punctured with an insect pin multiple times. CaLCuV symptoms were observed, and symptomatic tissue was harvested 14 to 21 days postinoculation, while BCTV and BCTV L2' mutant virus symptoms were observed, and symptomatic inflorescence tissue was harvested 21 to 30 days postinoculation. This harvest was labeled ‘primary tissue’. Each sample comprised tissue pooled from four infected plants. After the primary harvest, plants were allowed to continue growing under the same conditions, and symptom development was recorded in secondary inflorescence tissue, which was then harvested as secondary tissue 14 to 21 days after the primary harvest.

4.3.3 5-azacytidine treatment

10 uM of 5-azacytidine along with 0.05% Silwet L-77 in an aqueous solution was sprayed onto N. benthamiana plants after primary harvest following BCTV L2- infection with a Paasche VLSTPRO double-action airbrush attached to a Husky (Home Depot, Inc.)
4-gallon air compressor set to 40 lb/in². The concentration of 5-azacytidine used was determined by carrying out initial experiments with 10 uM, 25 uM and 60 uM of 5-azacytidine and the lowest effective concentration was used. Control plants were sprayed with 0.05% Silwet in water. After being sprayed, plants were laid on damp paper towels under a dome to maintain high humidity for 24 h and then returned to the growth chamber. Plants were resprayed every 3-5 days for about 2 weeks, then observed for symptoms and harvested as secondary tissue.

4.3.4 DNA isolation and Southern Blot analysis

DNA was isolated from homogenized primary and secondary infected tissue using DNeasy columns (Qiagen), and 1 ug of genomic DNA from infected plants was restricted overnight. After digestion, 1% to 1.2% agarose gel electrophoresis was carried out, followed by overnight transfer onto Nytran Super Charge membrane. The blot was UV cross-linked (1,700 µJ) and hybridized overnight using a 32P-labeled full-length virus probe at 48°C in Ultrahyb hybridization buffer (Ambion). Random primer-based 32P labeling was carried out using Strip-EZ DNA (Ambion). Signal intensity was quantitated using a Phosphorimager (Bio-Rad Molecular Imager FX).

4.3.5 Bisulfite sequencing

The bisulfite sequencing method is based on Frommer et al. (Frommer, M. et al, 1992). The treatment was carried out as described previously (Raja et al., 2008). Briefly,
DNA isolated from primary and secondary plant tissue was restricted overnight using Scal enzyme, followed by overnight digestion with Proteinase K. Bisulfite treatment was then carried out using CT conversion reagent (EZ-DNA Methylation Gold; Zymo Research). The bisulfite-treated DNA was amplified using primers designed to the intergenic region of the BCTV genome. The PCR product was purified using Promega Wizard columns and TA cloned, and individual clones were sequenced at The Ohio State University Plant Microbe Genomics Facility. For conversion control, plasmids containing BCTV DNA were added to a vast excess of healthy plant DNA extract and treated with bisulfite reagent. Previously designed forward and reverse primers were used to amplify BCTV intergenic regions (IR) following bisulfite conversion: BCTV2640CF, GGGATATGTAAGAAATATG (forward), and BCTV147CR, TCTCCCTTCTTTATAACCAATCAAC (reverse).

4.4 Results

4.4.1 Methylation is required for recovery.

There is considerable genetic evidence that recovery utilizes the methylation pathway (Raja et al., 2008). We used two approaches to further confirm that methylation is required for recovery. First, we evaluated the consequence of treatment with 5-azacytidine on recovery from BCTV L2- infection. 5-azacytidine is a non-methylatable cytidine analog that incorporates into DNA and prevents methylation. Wild-type N. benthamiana plants were infected with BCTV L2- virus, which resulted in a
symptomatic primary infection as expected. After primary harvest, a test set of 16 plants was sprayed with 10 μM 5-azacytidine (in Silwet) every 3-5 days for about 2 weeks, during the period when new axillary shoots were developed. We found that 5-azacytidine treated plants did not recover, while control plants (16 plants) sprayed with water and Silwet alone recovered from disease (Isaac Heard, unpublished) (Figure 4.1A). Next, we tested mutants in the enzyme, adenosine kinase for their ability to recover from geminivirus disease. Adenosine kinase is required to sustain transmethylation reactions and serves to drive the methyl cycle by preventing accumulation of products of the S-adenosyl homocysteine hydrolase-catalyzed reaction that are inhibitory to methyltransferase reactions (Figure 1.5). We have previously shown that AL2 and L2 proteins suppress adenosine kinase resulting in an inhibition of viral genome methylation (Wang et al., 2003, Wang et al., 2005, Raja et al., 2008, Buchmann et al., 2009). Arabidopsis has two genes for adenosine kinase, adk1 and adk2 that show 92 per cent amino acid sequence identity. Both genes are expressed in Arabidopsis, but an adk1 adk2 double mutant is embryonic lethal (Moffat et al, 2000, Wang et al., 2003). We tested 32 plants each of the individual adk1 and adk2 mutants and found that Arabidopsis adk1 and adk2 mutants infected with BCTV L2-virus were unable to recover (Figure 4.1B). Thus, a pharmacological agent that prevents methylation, and adk mutants that inhibit the methyl cycle both prevent recovery in two different plant hosts. This indicates that methylation is required for recovery, and we proposed to dissect these requirements further to determine the context of cytosine methylation most important for recovery, the contribution of transcriptional and post-transcriptional gene silencing to recovery, and the
role of plant polymerases Pol IV and Pol V in initiating the methylation-based recovery process.

4.4.2 Non-CG methylation is necessary for recovery.

Methylation in plants can occur at CG, CNG, and CHH contexts (Chan et al., 2005). Methylation in these different contexts is catalyzed by specialized methyltransferases. DRM1 and DRM2 are the major de novo methyltransferases that act primarily in CNG and CHH contexts. In *Arabidopsis*, DRM2 is the only expressed gene. CMT3 carries out de novo and maintenance methylation at CNG loci, while MET1 carries out CG methylation. KYP2 (SuVH4) is the *Arabidopsis* H3K9 methyltransferase. We found previously that methyltransferase mutants show enhanced susceptibility to geminiviruses, and that most methylation reductions on the geminivirus genome in these mutants occurred at non-CG sites (readily apparent on sparsely methylated CaLCuV genomes), suggesting that methylation at non-CG sites is important for defense (Raja et al., 2008). In this study, methyltransferase mutants, such as *drm1*/*drm2*, *cmt3*, *ddc* (*drm1*/*drm2*/*cmt3* triple mutant) and *met1* were examined for their ability to recover from geminivirus disease. These mutants and their corresponding wild-type ecotypes (Ws-2, Ler-0 and Col-0) were infected with *BCTV L2* and tested for their ability to recover from disease. Disease symptoms were observed after primary infection, and as noted previously, the methyltransferase mutants showed more severe symptoms in response to infection. The primary shoots were then trimmed off and the secondary shoots were observed for symptoms. In wild-type plants, the secondary shoots showed an absence of
symptoms and recovered from disease, as previously observed. The plants were inoculated with wild-type BCTV and two \textit{BCTV L2}− null mutant viruses (\textit{L2-1} and \textit{L2-2}) (Hormuzdi and Bisaro, 1995). Primary infected tissue of wild-type and mutant plants was observed to show symptoms, with increased symptom severity noted in the \textit{drm2}, \textit{cmt3} and \textit{ddc} methyltransferase mutants as reported previously (Raja et al., 2008). After removal of primary infected tissue, secondary shoots of wild-type and mutant plants infected with wild-type BCTV showed severe symptoms, but wild-type plants infected with \textit{BCTV L2-1} or \textit{BCTV L2-2} mutant virus showed recovery. The mutants chosen in our study were assayed for their ability to recover when infected with \textit{BCTV L2}− virus. The BCTV IR in viral DNA isolated from secondary infected tissues was examined for cytosine methylation by bisulfite sequencing.

In 32 plants examined for each mutant type, we found that \textit{drm1/drm2} and \textit{cmt3} mutant plants showed mixed recovery, where about half the shoots recovered from disease, while half did not. By contrast, the \textit{ddc} mutant debilitated in both DRM2 and CMT3 activities, showed a total loss of recovery in all 32 plants analyzed. Recovered tissue isolated from wild-type plants infected with \textit{BCTV L2}− showed reduced viral titers, in contrast to the \textit{ddc} mutant where there was considerable virus accumulation, similar to what was earlier observed in the \textit{ago4} mutant (Figure 4.2). An analysis of cytosine methylation at the viral intergenic region revealed that symptomatic, non-recovered tissue from \textit{drm2} and \textit{cmt3} mutants showed a defect in hypermethylation of the intergenic region, a defect likely accentuated in the \textit{ddc} mutant, which awaits methylation analysis (Figure 4.3). We were unable to test the \textit{met1} mutant since homozygous \textit{met1-3} plants are
embryonic lethal. We tested hemizygous *met1-7* plants (32 plants each) and found that these plants harbored mild symptoms on secondary shoots, if at all, and appeared largely to have recovered. However, we cannot determine if this is because of a reduced effect of CG methylation on recovery, or whether the hemizygous condition is responsible for an attenuation of the phenotype.

Cytosine methylation in the intergenic region of the BCTV virus in recovered tissue occurs extensively in all sequence contexts – CG, CNG and CHH. Since the *ddc* mutant fails to recover, we conclude that that both CNG methylation due to CMT3 and CHH methylation mediated by DRM2 are necessary for recovery, but we cannot rule out a role for CG methylation that is maintained by MET1 (Figure 4.3).

4.4.3 Pol V is necessary for recovery; Pol IV plays an important role.

In previous studies, we found that *pol IV/V (nrpd2a)* double mutants were hypersusceptible to geminivirus infection (Raja et al., 2008). In this study, *pol IV (nrpd, nrpd1a)* and *pol V (nrpe, nrpd1b)* mutants were separately evaluated for susceptibility to geminiviruses. The mutations were in the largest unique subunits of both these polymerases. We found that *pol IV* and *pol V* mutants, individually and in combination, showed enhanced susceptibility to CaLCuV and BCTV (32 plants each were analyzed). When evaluated for the ability to recover from an infection with BCTV lacking L2 protein, we found that *pol V* mutants (*nrpe*) lost the ability to recover (all of 32 plants analyzed), while *pol IV* mutants (*nrpd*) showed what we describe as ‘delayed recovery’ (all of 32 plants analyzed). The unusual ‘delayed recovery’ observed in *pol IV* mutants
appeared to follow a gradient along each shoot, where flowers towards the rosette of the plant showed symptoms, and others progressively showed reduced symptom intensity along the shoot until the most apical flowers had no symptoms (Figure 4.4). This is an interesting and unique phenotype that requires further investigation and this study is being extended by Jamie Wolf. From what is currently known, we conclude that Pol V is essential for recovery, while Pol IV plays an important role, that is likely redundant with Pol II. The pol IV/V (nrpd2a) mutants tested showed mixed recovery (Table 4.1). This mutant was null for the second-largest subunit shared by Pol IV and Pol V, and that could be a potential reason for why the phenotype was less severe that pol V mutants mutated in the unique largest subunit. This awaits further testing with an nrpd1a/nrpd1b double mutant deficient in the largest catalytic subunit of both enzymes.

4.4.4 DCL3 is the most important Dicer-like enzyme for recovery, but DCL1, DCL2 and DCL4 can contribute

Arabidopsis Dicer-like enzymes: DCL1, DCL2, DCL3 and DCL4 act in partially overlapping, but essentially delineated pathways of silencing. DCL1 is involved in the miRNA pathway, DCL2 and DCL4 in the PTGS pathway, and DCL3 in the methylation pathway. Previously, in primary infections with CaLCuV and BCTV, we observed mild enhancement of floral symptoms in dcl2 and dcl3, but not dcl4 mutants (Raja et al., 2008). We also previously observed that among the dcl mutants, only dcl3 plants did not recover from BCTV L2− infection (Chapter 3).
Here, we further examined the role of DCL proteins in geminivirus defense by examining multiple mutants. We found that the \textit{dcl2/4} mutant showed only moderately enhanced susceptibility while the \textit{dcl2/3/4} mutant showed hypersusceptibility. Bisulfite sequencing was carried out on the BCTV genomes from \textit{dcl2/3/4} mutants (Figure 4.5). Surprisingly, the viral genome isolated from \textit{dcl2/3/4} mutants continued to harbor methylation, which is suggestive of a novel role for DCL1 in the methylation pathway, or a DCL independent mechanism for methylation.

We previously showed that the \textit{dcl3} mutant is the only one of the individual \textit{dcl} mutants that lacks the ability to recover, implicating the DCL3-mediated heterochromatin pathway in recovery. The \textit{dcl2} mutant showed severe primary symptoms, but partial recovery, i.e. the appearance of mild to moderate symptoms on secondary shoots. The \textit{dcl4} mutant, which is debilitating in DCL4, the main antiviral dicer involved in defense against RNA viruses and PTGS, can recover from geminivirus infection. Interestingly, the \textit{dcl2/4} double mutant, deficient in PTGS, shows mixed recovery, where some shoots lack symptoms and others do not, suggesting a role for PTGS in recovery. Interestingly, the \textit{dcl2/3/4} triple mutant does not recover, but differs from the non-recovered \textit{dcl3} mutants as it shows increased lateral shoot elongation with severe symptoms apparent in the flowers (Figure 4.6).

4.5 Discussion

We present further evidence that RNA-directed methylation is required for recovery from geminivirus disease. Mutants in \textit{Arabidopsis} that inactivate non-CG
cytosine methyltransferases (drml drm2 cmt3) and methylation pathway components (pol V, dcl3, drb3, ago4) lose the ability to recover from BCTV L2' infection (Table 4.1). We show that cytosine residues in the intergenic region of the BCTV L2' viral genomes are hypermethylated in recovered tissue from wild-type plants, but not in viral genomes isolated from methylation-deficient mutant plants that lose the ability to recover.

In this study, we found that ddc mutants that lack CHH and CNG methylation cannot recover from geminivirus disease, while drm2 mutants that lack CHH methylation or cmt3 mutants that lack CNG methylation only show mixed recovery. This suggests that either methyltransferase alone can independently condition recovery in an individual shoot, but not as efficiently as when both methyltransferases are present. Hence, non-CG methylation – both, CHH methylation mediated by DRM1/DRM2 as well as CNG methylation maintained by CMT3 – is required for total recovery. Further, hypermethylation in recovered tissue occurs at all sequence contexts in the intergenic region. It is likely that de novo methylation that targets the invading virus is mainly non-CG, which is then maintained through recovery in all sequence contexts.

This study showed that individual pol IV and pol V mutants are hypersusceptible to geminivirus infection, and that pol V mutants lose the ability to recover. This suggests that Pol V plays an important and unique role in generating a geminivirus-specific transcripts to either trigger methylation or direct it to the viral genome. In the pol IV mutant, interestingly, delayed recovery was observed. This suggests that Pol II, and perhaps Pol V, can act redundantly with Pol IV to generate a transcript that serves as a substrate for RDR activity (Figure 1.8). In support of this, Pol II, Pol IV and
Pol V have been shown to be present at the geminivirus intergenic region in ChIP studies (J. Wolf and J. Ostler, unpublished). However, it appears that Pol II and Pol IV cannot compensate for Pol V activity. The geminivirus model affords a good opportunity to test methylation-based selection of substrates for Pol II, Pol IV and Pol V in vivo, and study how these polymerases are recruited to the geminivirus genome. It has been shown that Pol V has three functions: First, it produces a non-coding transcript that can be converted by RDR2 to double-stranded RNA, which is then processed into siRNAs by DCL3, and secondly, Pol V also generates a scaffolding transcript upon which siRNAs dock, and thirdly, Pol V interacts with AGO4 through GW/WG residues and guides methylation (Wierzbicki et al., 2008). It is possible that this latter function, important for recovery, cannot be performed by Pol II or Pol IV.

The delayed recovery observed in pol IV mutants shows a gradient of symptoms, with severe symptoms on lateral shoots that arise first towards the bottom of the plant, and gradual recovery towards the top of the plant, with later shoots near the apex showing an absence of symptoms. This suggests that Pol IV participates in initiating recovery and that Pol II can partially compensate for Pol IV function in synthesizing transcripts that can serve as substrates for the silencing pathway. The appearance of a gradient could also suggest that silencing spread might not be efficient in these mutants, which is consistent with a role for Pol IV in this process (Dunoyer et al., 2007, Smith et al., 2007).

Natural recovery observed in PepGMV also interestingly shows a gradient even in wild-type plants (Rodriguez-Negrete et al., 2009). In this study, the largest difference between symptomatic and asymptomatic secondary shoots appeared to be in the amount
of virus-derived siRNAs that originated in the intergenic region, among all regions of the viral genome. In wild-type plants infected with BCTV L2-, however, there is uniform recovery throughout the plant. A study of virus-derived siRNAs in symptomatic and asymptomatic secondary tissue in pol IV mutants and symptomatic secondary tissue in pol V mutants has been initiated by J. Wolf. Further, it is proposed to test the role of Pol IV in spread of a silencing signal targeting geminiviruses through grafting experiments of symptomatic and asymptomatic secondary scions into naïve rootstocks of a virus promoter driven GFP reporter line. If Pol IV were involved in spread of the silencing signal, it is predicted that recovered asymptomatic apical shoots should be able to silence the GFP trangene in the rootstock, accompanied by an increase in viral siRNAs. However, symptomatic shoots derived from basal regions of the Pol IV plants showing delayed recovery may be compromised in their ability to do so. In summary, Pol V is required for recovery, while Pol IV is not essential, but plays an important role.

In Arabidopsis, primary infection data revealed that dcl2 and dcl3 showed moderately enhanced symptoms, but not dcl4 (Raja et al., 2008). Given the redundancy of the Dicer-like enzymes in producing virus-specific siRNAs, it was difficult to directly assign geminivirus defense to the DCL3 pathway. In recovery studies, however, we found that only dcl3 mutants cannot recover from geminivirus disease, definitively implicating methylation and the TGS pathway in geminivirus defense (Chapter 3).

We contrasted the dcl3 mutant deficient in TGS with the dcl2/4 double mutant that is defective in PTGS. The dcl2/4 mutant showed mixed recovery, suggesting that PTGS also plays a role in recovery (atleast in 50% of the shoots). This agrees with the
finding that both 21 and 24 nt virus-derived siRNAs are present in geminivirus recovered tissue (Rodriguez-Negrete et al., 2009). In this study, it was reported that most of the 21 nt siRNAs arose from the coding region, while most of the 24 nt siRNAs originated in the intergenic region. In the coding regions, the genes most targeted on DNA A were Rep, TrAP and REn, which are early and delayed early genes and hence would have abundant transcripts early in an infection that could have aberrant features. If PTGS were established in shoot primordia against these early viral mRNAs, then the virus would be efficiently silenced by PTGS, and could result in early host recovery. Understanding this will require measurement of transcripts, and siRNA profiles targeting each transcript, in symptomatic and asymptomatic secondary tissue.

Another possibility is that PTGS and TGS mutants differ in their ability to enable access to meristematic tissue and therefore, set up silencing in the lateral meristems so that secondary shoots that arise from meristems of plants mutated in the TGS pathway harbor severe symptoms and fail to recover. According to this model, in wild-type plants, there is little or no access of virus to the meristem, and PTGS is set up against the virus. PTGS and TGS then combine to generate the recovery phenotype. In TGS mutants, the functional PTGS pathway, can establish silencing in meristems, but this is inefficiently maintained in the absence of TGS, so the plants fail to recover. Since some shoots in drm1/drm2 and cmt3 mutants do recover, this suggests that each of these methyltransferases likely contributes independently towards recovery, and both are necessary for full recovery.
Another interesting finding from this study was that the \textit{dcl2/3/4} triple mutants were capable of methylating the viral genome. These mutants also showed a novel recovery phenotype where although there was lack of recovery, the axillary shoots could grow out and there was no stunting as was observed in the \textit{dcl3}, \textit{drb3} and \textit{ago4} mutants. We speculate that DCL1 generated miRNAs may play a redundant role in methylation in the \textit{dcl2/3/4} triple mutant, an observation obscured in the \textit{dcl3} mutant due to an effect of DCL2 or DCL4 in negatively regulating DCL1 (See Section 5.1 in Chapter 5).

This is the first systematic genetic analysis of components required for recovery. We found that Pol V is required for recovery and that Pol V function cannot be compensated by Pol II or Pol IV, so Pol V appears to play a unique role. Pol IV plants showed delayed recovery, suggesting that Pol II and/or Pol IV may act redundantly in generating substrates for RDR2, and that Pol IV may additionally be involved in spread of the silencing signal. We further found that DCL3 and the TGS pathway is essential for recovery. In addition, DCL2 and DCL4 mediated-PTGS also plays a role in recovery. In the absence of all other DCLs, DCL1 may also be able to direct methylation. From these studies, we conclude that the POL V-DCL3-AGO4-DRM2/CMT3 methylation pathway is important in establishing recovery. Thus the same pathway that is involved in silencing cellular transposons, also acts against exogenous DNA viruses, and the recovery assay is a powerful tool to help delineate the functions of different components in this pathway.
Figure 4.1 Methylation is required for recovery. (A) Secondary shoots from plants following treatment with the demethylating agent 5-azacytidine (10 μM) show a lack of recovery. (B) *adk1* and *adk2* mutants do not recover from geminivirus disease. This mimics the results seen in methylation mutants in *Arabidopsis*, and once again indicates that methylation is needed for recovery.
Figure 4.2 Non-CG methylation is required for recovery. The \textit{cmt3} (A) and \textit{drm1/drm2} (B) mutants show mixed recovery, suggesting that CHH and CNG methylation are both required for recovery. (C) The \textit{ddc} mutant failed to recover from \textit{BCTV L2} infection, indicating that non-CG methylation mediated by DRM1/2 and CMT3 are both independently required for recovery. (D) Contrast lack of symptoms in WT (recovered) plants, with mixed recovery (a mixture of shoots showing mild to no symptoms) in \textit{cmt3}, and \textit{drm1/drm2} mutants, and no recovery (severe symptoms) in \textit{ddc} mutants. The Southern Blot shown reveals low viral titers in secondary tissue from recovered Col-0 plants, but high viral titers in the non-recovered \textit{ddc} mutant.
Table 4.1. Methylation-deficient mutants are unable to recover from geminivirus disease

<table>
<thead>
<tr>
<th>Gene function and mutation(s)</th>
<th>Relative severity of disease$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Methylation pathway components</strong></td>
<td></td>
</tr>
<tr>
<td>nrpd2a (pol IV/V)</td>
<td>++/+ (Mixed recovery)</td>
</tr>
<tr>
<td>nrpd (pol IV)</td>
<td>+++/++/+ (Delayed recovery)</td>
</tr>
<tr>
<td>nrpe (pol V)</td>
<td>++++ (No recovery)</td>
</tr>
<tr>
<td>ago4</td>
<td>++++ (No recovery)</td>
</tr>
<tr>
<td>ddm1</td>
<td>+ (Recovery)</td>
</tr>
<tr>
<td>drd1</td>
<td>+/- (Recovery)</td>
</tr>
<tr>
<td><strong>Cytosine methyltransferases</strong></td>
<td></td>
</tr>
<tr>
<td>drm1, drm2</td>
<td>+++/+ (Mixed recovery)</td>
</tr>
<tr>
<td>cmt3</td>
<td>+++/+ (Mixed recovery)</td>
</tr>
<tr>
<td>ddc</td>
<td>++++ (No recovery)</td>
</tr>
<tr>
<td>met1</td>
<td>+ (Recovery)</td>
</tr>
<tr>
<td><strong>H3K9 methyltransferase, kyp2</strong></td>
<td></td>
</tr>
<tr>
<td>adk1</td>
<td>+++ (No recovery)</td>
</tr>
<tr>
<td>adk2</td>
<td>+++ (No recovery)</td>
</tr>
<tr>
<td><strong>Methyl cycle enzymes</strong></td>
<td></td>
</tr>
<tr>
<td>dcl2</td>
<td>++ (Partial recovery)</td>
</tr>
<tr>
<td>dcl3</td>
<td>++++ (No recovery)</td>
</tr>
<tr>
<td>dcl4</td>
<td>+ (Recovery)</td>
</tr>
<tr>
<td>dcl2 dcl4</td>
<td>++/+ (Mixed recovery)</td>
</tr>
<tr>
<td>dcl2 dcl3 dcl4</td>
<td>+++ (No recovery, no stunting)</td>
</tr>
<tr>
<td>rdr6</td>
<td>+ (Recovery)</td>
</tr>
<tr>
<td>rdr2</td>
<td>++/+ (Mixed recovery)</td>
</tr>
</tbody>
</table>

$^a$ Severity of disease symptoms in mutant Arabidopsis plants was based on the degree of floral deformation relative to wild-type plants of the appropriate ecotype (Col-0, Ler-0, or Ws-2) according to the following rating scale: +, typical, wild-type symptoms; ++, moderately enhanced; ++++, severe; ++++, very severe; ++/++, indicates a heterogeneous or ‘mixed’ phenotype where some flowers show symptoms, while others do not. The number of + in mixed recovery is indicative of the severity of symptoms on symptomatic shoots as noted previously.
Figure 4.3 Recovery is correlated with hypermethylation of the viral genome. Bisulfite sequencing of symptomatic, non-recovered tissue from BCTV L2 infected cmt3 mutants revealed loss of methylation associated with symptoms in mixed recovery phenotypes. Methylation was decreased in all sequence contexts, namely, CG, CNG and CHH.
Figure 4.4 (A) The pol IV mutant (nrpd1a/nrpd) shows delayed recovery. (B) The pol V mutant (nrpd1b/nrpe) does not recover. (C) Close-up photographs show delayed recovery along the same shoot of a pol IV mutant and lack of recovery in a pol V mutant.
Figure 4.5 The *dcl2/3/4* triple mutant is methylation-competent. The *dcl2/3/4* mutant shows moderately enhanced symptoms of BCTV infection, and bisulfite analysis on the BCTV IR from primary symptomatic tissue obtained from *dcl2/3/4* mutants reveals methylation comparable to wild-type levels.
**Figure 4.6** The *dcl3* mutant does not recover, but *dcl2/4* plants show mixed recovery, suggesting that both TGS and PTGS can condition recovery. The *dcl2/3/4* shows shoot extensions and a lack of stunting, unlike *dcl3*, but the symptoms on the flower continue to be severe, and reflect non-recovery.
CHAPTER 5

DISCUSSION

5.1 Methylation as a defense against geminiviruses.

Chapter 2 of this thesis establishes methylation as a defense against geminiviruses. Through genetic studies and subsequent methylation analysis of the viral genome isolated from an array of methylation mutants, we found that methylation-deficient plants are hypersusceptible to geminivirus infection. An evaluation of the viral genome showed that cytosines in the intergenic region are methylated in viral genomes isolated from infected plants and that methylation mutants show reductions in cytosine methylation. The greatest reductions were observed at non-CG sites. Further, these studies provided evidence for populations of viral minichromosomes carrying histone marks associated with active or repressed chromatin in infected plants. Lastly, we found that while wild-type plants can recover from infection with BCTV L2− virus, ago4 mutant plants cannot recover. Recovery is associated with symptom remission, low viral titers,
and hypermethylation of the geminivirus genome. In addition, recovery requires AGO4, a methylation pathway component. Early work showed that replication of in vitro methylated geminivirus in protoplasts is reduced compared to unmethylated templates (Brough et al., 1993). Reduced replication is mostly due to reduced transcription of the Rep protein, although there are also direct effects of methylation on replication (Ermack et al., 1993). Several studies have shown that geminiviruses, including ToLCV, TGMV, BCTV, CaLCuV and *Mung bean yellow virus*, are methylated during an infection (Raja et al., 2008; Bian et al., 2006; Seemanpillai et al., 2003). It has also been observed that various species of siRNAs are produced in a geminivirus infection, the most abundant of which is the 24 nt DCL3 generated class that guides methylation and silencing (Akbergenov R. et al., 2006; Blevins T. et al., 2006). Finally, it has been demonstrated that geminivirus AL2 and L2 proteins inhibit methylation and reverse TGS (Buchmann et al, 2008). These lines of evidence indicate that viral chromatin methylation is a host defense against geminiviruses (Raja et al., 2008).

In methylation-deficient plants showing enhanced susceptibility, methylation reductions were observed at certain sites of the geminivirus intergenic region corresponding to important regulatory regions. Most notably, methylation was reduced at the AL1 binding sites and other sites within the origin of replication (Figure 2.3). This suggests that there are specific sites that are important for pathogenicity. However, further work is needed to determine the consequences of methylation at these sites and to evaluate their effects on replication and/or transcription. Methylated and unmethylated
virus constructs at these key sites could be assayed for replication in protoplasts (Elmer et al., 1988; Sunter et al., 1990).

Further, this study was restricted to analyzing intergenic region methylation, where methylation of promoters is most likely to affect transcription. But it would also be interesting to study methylation across the geminivirus genome. Bisulfite sequencing may be used for such a comparative study provided primer bias is resolved and comparable for different primer sets (Warnecke P.M., 2002; Shen L., 2007) (see Appendix C). A parallel study would involve mapping different siRNA size classes to the viral genome. We predict that methylation and associated 24 nt siRNAs would be enriched at the intergenic region, and possibly the downstream promoter regions of DNA A (as discussed in 1.1.5).

Since methylation is required to condition recovery from disease, an analysis of methylation components required for recovery was initiated and is ongoing. This work is summarized in Chapter 4 of this thesis. In this study, we found that drm1/2 and cmt3 mutants showed mixed recovery when infected with BCTV L2⁻ virus, while ddc triple mutants that lack both drm2 and cmt3 failed to recover, indicating that non-CG methylation (CHH and CNG) is required for recovery.

The pol IV (nrpd1a, nrpd) and pol V (nrpd1b, nrpe) mutants both showed enhanced susceptibility to CaLCuV and BCTV infection. When tested for recovery upon BCTV L2⁻ infection, pol V mutants showed lack of recovery, while pol IV mutants showed delayed recovery, possibly owing to the presence of siRNAs from Pol II generated transcripts. Preliminary ChIP studies place Pol II, Pol IV and Pol V in the
intergenic region of the geminivirus genome (Wolf, J., Ostler, J., unpublished). If Pol IV/V is involved in directing geminivirus chromatin methylation, it would be interesting to study how these polymerases are recruited to the geminivirus genome, investigate their preference for methylated templates, and study virus-derived siRNAs, and work is currently underway in the lab to address these questions.

Viruses may have evolved to reduce the number of CGs, partly to avoid methylation. However, using in silico analysis, we found that the geminivirus intergenic region has the hallmarks of a CpG island that is prone to methylation (See Appendix D). It may be useful to map siRNAs from these regions, in particular.

The dcl3 mutant debilitated in TGS lost the ability to recover from a BCTV L2 infection, while a dcl2/4 double mutant deficient in PTGS showed partial recovery. This suggests a more important role for TGS in establishing and maintaining recovery. The dcl4 mutants can recover from geminivirus infection, so although DCL4 has been identified as the main antiviral dicer for RNA viruses, it is likely not the primary dicer involved in defense against geminiviruses. Another interesting observation from this study was that dcl2/3/4 triple mutants showed prolonged secondary shoot extensions from lateral meristems that culminated in flowers that failed to recover. This suggests the possibility that the dicers may regulate each other. DCL2 and/or 4 in the dcl3 mutant may negatively regulate DCL1, which results in lack of recovery, while in the dcl2/3/4 mutant, DCL1 can participate in some compensatory mechanism for DCL3 function. It has been reported that DCL1 negatively regulates DCL4 and DCL3 expression (Qu et al., 2008). It is also interesting to note that DCL1 and DCL3 are found in the same subnuclear
compartments in plant cell nucleoli (Pontes and Pikaard, 2008). Expression of DCLs in infected tissue may be examined by RT-PCR. Interestingly, viral genomes retrieved from the dcl2/3/4 triple mutant are methylated, suggesting that DCL1 may play a role, or that DCL generated siRNA-independent methylation mechanisms may exist within plants. Excitingly, there is new evidence that in the moss, *Physcomitrella patens*, miRNAs can guide DNA methylation leading to transcriptional gene silencing and that an abundance of miRNAs relative to its target can result in miRNAs guiding methylation rather than mRNA cleavage (Khraiwesh et al., 2010).

Recovery is not an all-or-none phenomenon, as we observed ‘partial recovery’ where symptoms persisted on new shoots, but were attenuated, or ‘mixed recovery’ where severe symptoms were present on some shoots and absent on others, or ‘delayed recovery’ where symptom severity varied along the same shoot. To sort through such epigenetic mosaics requires consideration of siRNA signaling and methylation maintenance within different tissue types of whole plants. In the short-term, it is proposed to test methylation of the viral intergenic region and coding regions, viral transcript levels and virus-specific siRNAs in these different types of tissue isolated from the same plant.

In summary, we conclude that geminiviruses are targeted by methylation as a host defense, leading to even more interesting questions regarding pathogenicity as well as the mechanisms of small RNA directed methylation in plants.
5.2 Using geminiviruses as a facile and efficient screen for new components of the methylation pathway.

Chapter 3 of this thesis describes work which demonstrates that geminiviruses may be used to screen for novel components or assign novel functions to known components of the methylation pathway. In this investigation, specifically, we sought to determine which members of the Arabidopsis dsRNA binding protein family (DRB2, 3, or 5) are associated with DCL3, the Dicer-like protein responsible for methylation-mediated silencing. We have previously shown that plants methylate geminivirus minichromosomes to effectively silence the virus. Methylation mutants are hypersusceptible to geminiviruses, and are unable to recover from geminivirus disease. In this study, we found that drb3 mutants are uniquely hypersusceptible to geminivirus infection with BCTV and CaLCuV. Methylation analysis of the viral intergenic region revealed that methylation is greatly reduced in the drb3 mutant. In addition, the drb3 mutant, like ago4 and dcl3, fails to recover from geminivirus disease. Lack of recovery in these mutants was associated with a lack of hypermethylation of viral DNA, contrary to wild-type plants which recover. Using bimolecular fluorescence complementation analysis, we also found that DRB3 interacts with DCL3 in plant nuclei, in punctate subnuclear bodies, that are frequently associated with nucleoli. This is consistent with the localization of these proteins to Cajal bodies. Further, DRB3 interacts with the RISC component, AGO4, in speckles that are not associated with nucleoli. Again, this is consistent with the localization of AGO4 to AB-bodies in plant cell nuclei. Since the production of 24 nt siRNAs was not affected in the drb3 mutant, DRB3 likely acts
downstream of siRNA generation, possibly by facilitating RISC loading of the 24 nt siRNA species into AGO4-containing RISC complexes.

The dsRNA binding domains have been previously reported to mediate protein-protein interactions. In our attempts to verify the DRB3:DCL3 interaction in a yeast two-hybrid system, unfortunately, yeast transformants containing the full-length DCL3 protein could not be obtained in our hands. A truncated DCL3 protein containing the RNase III domain and dsRNA binding domains did not interact with DRB3 in yeast. Since the dsRNA binding domain of other DRBs has been implicated in protein-protein interactions, this was surprising. This may imply a requirement for full-length protein for proper folding or interaction, or suggest a post-translational modification as a prerequisite for interaction, or imply that the interaction is bridged by RNA or another protein. Further work is needed to address these different scenarios. It has also been difficult to express full-length GST or His-tagged DCL3 protein in *E.coli* (Sizhun Li, personal communication). Efforts are currently underway to express HA-His tagged DCL3 protein in plants to carry out Far-western and co-immunoprecipitation experiments using antibodies to tagged versions of the proteins (Sizhun Li).

In this study, we found that the *drb3* mutant is not impaired in the production of virus-derived 24 nt siRNAs, suggesting that DRB3 acts downstream, likely in RISC loading of 24 nt siRNAs into AGO4-containing complexes. Our work and that of others also shows that DRB4 similarly appears to assist DCL4 but is not required for generation of 21 nt siRNAs (Curtin et al., 2008; Qu et al., 2008). Several DRBs known in other systems act downstream of siRNA generation, such as R2D2 in *Drosophila* or RDE4 in
C. elegans. In the case of R2D2, it has been determined that R2D2-Dicer 2 heterodimer serves to orient the siRNA duplex and therefore determine that only the antisense strand is incorporated into the RISC complex. The Dicer 2 is exchanged for AGO2 in the heterodimer which allows unwinding of the siRNA duplex (Tomari et al., 2004). RDE4 interacts with RDE1, a homolog of AGO2 in C. elegans. It has also been recently shown that DRB1 in plants acts similarly by orienting the miRNA duplex and enabling strand selection for RISC loading (Eamens et al., 2009). Further work is needed to identify how DRB3 may assist in RISC loading. This could be directly addressed using RNA-IP to determine if the AGO4/6 associated siRNAs are depleted in the drb3 mutant relative to wild-type plants. Since the AGO4 mutant showed no depletion of 24 nt siRNAs, this suggests a redundant role for AGO6, or the data may imply that double-stranded 24 nt siRNAs that are not loaded into RISC are somehow stabilized and may accumulate.

Using BiFC analysis, we found that DRB3 and DCL3 accumulate in subnuclear bodies that sometimes colocalize with the nucleolar marker, fibrillarin. DRB3 also interacts with AGO4 in subnuclear speckles that do not associate with nucleoli, but are consistent with AB-bodies, where AGO4 has been localized along with Pol V and DRM2. This suggests that DRB3 may shuttle siRNAs from DCL3 to AGO4. In addition, AGO4 is also found in nucleoli-associated Cajal bodies, but surprisingly does not appear to interact with DCL3 that has also been localized to Cajal bodies at this location, so we speculate that certain other factors must regulate their interaction. Such regulatory factors could be based on the dimerization state of the DCL or DRB proteins, or modifications to the siRNA such as methylation by HEN1, or perhaps the unwinding of siRNAs, or it may
be necessary that the AGO4-Pol V interaction must precede RISC loading. We could also speculate that DRB3 may bind the non-coding RNA generated by Pol V, and help recruit DRM2. In order to test these mechanisms, we must first determine if DRB3 binds 24 nt siRNAs in plant cells, and if so, ask whether bound siRNAs are methylated at the 3’ end in which case they would be resistant to β-elimination. We also propose to test if DRB3 interacts with DRM2 or Pol V, using BiFC analysis.

Using geminiviruses as sensitive probes, this study resolves the question of which Arabidopsis dsRNA binding protein binds to DCL3 to assist chromatin methylation via 24 nt siRNAs, and identifies for the first time, a function for the plant protein DRB3. It validates the use of geminiviruses as a tool to identify novel components of the chromatin methylation pathway in plants, and sets the stage for further investigation of DRB3 functions and its role in RNA silencing.

5.3 Exploring chromatin biology using the geminivirus minichromosome.

Various methylation mutants show a lack of recovery from geminivirus disease, and hypermethylation of the intergenic region of the viral genome. Viral genomes isolated from recovered tissues may be used to identify novel proteins that may associate with densely methylated DNA. Susceptibility to geminiviruses may also be used to screen methyl binding proteins involved in deciphering methyl marks.

Are H3K9 methylation marks associated with the geminivirus intergenic region exclusive to repressed genomes or do they occur in active genomes as well? What histone modifications are associated with repressed genomes in recovered tissue? Such questions
may be asked by performing serial ChIP to geminivirus genomes by sequentially using antibodies to different histone modifications (Xie and Grotewold, 2008).

The only known analysis of geminivirus minichromosome structure comes from ABMV. The geminivirus minichromosome consists of 11-13 nucleosomes with 11 representing open chromatin and 13 representing packed chromatin (Pilartz and Jeske, 2003). The composition of geminivirus minichromosomes can be manipulated experimentally using phenomenon such as recovery which should yield more packed minichromosomes, or using suppressors that facilitate and enable an open configuration of the geminivirus minichromosome. The small size and easily distinguished active and repressed genomes afford an excellent opportunity to use geminivirus minichromosomes as models for further analysis of chromatin structure and nucleosome modifications in infected cells.

In conclusion, this thesis presents work that establishes that methylation is an important host defense against geminiviruses and introduces geminiviruses as a tool to study chromatin methylation pathways.


APPENDICES
Appendix A: Viral genome methylation as an epigenetic defense against geminiviruses - Supplemental Data
Bisulfite sequencing of the BCTV intergenic region (IR)

The BCTV IR is shown below, where methylation was assayed using bisulfite sequencing. The Rep (L1) binding sites, the conserved hairpin loop sequences, and the L1 start codon (complementary sense) are indicated in bold and underlined. In total, 44 cytosine residues were tested in CG (red), CNG (pink) and CHH (grey) contexts.

Methylated cytosines are indicated in yellow. Unmethylated cytosines are converted to thymine (green). Numbers indicate the sum of methylated cytosines at each site.

Forward and reverse primers used to amplify the viral strand following bisulfite treatment:

BCTV2640CF:  GGGATATGTAAGAAATATG
BCTV147CR:  TCTCCCCTTCTATTAACCAATCAAC.
BCTV from wild-type *N. benthamiana.*
BCTV from wild-type Arabidopsis (Ler-0).

LerBCTV1  GGATATGAAGAATATGTTTTTCTTGTGAATAAGAATTTTAGAGATGAGCATTTGGA
LerBCTV2  GGATATGAAGAATATGTTTTTCTTGTGAATAAGAATTTTAGAGATGAGCATTTGGA
LerBCTV3  GGATATGAAGAATATGTTTTTCTTGTGAATAAGAATTTTAGAGATGAGCATTTGGA
LerBCTV4  GGATATGAAGAATATGTTTTTCTTGTGAATAAGAATTTTAGAGATGAGCATTTGGA
LerBCTV5  GGATATGAAGAATATGTTTTTCTTGTGAATAAGAATTTTAGAGATGAGCATTTGGA
LerBCTV6  GGATATGAAGAATATGTTTTTCTTGTGAATAAGAATTTTAGAGATGAGCATTTGGA
LerBCTV7  GGATATGAAGAATATGTTTTTCTTGTGAATAAGAATTTTAGAGATGAGCATTTGGA

LerBCTV1  GGATATGTAAGAAATATGTTTTTTGTTTGAATATGAAATTTTTTAGTAGGAGGCATTGGA
LerBCTV2  GGATATGTAAGAAATATGTTTTTTGTTTGAATATGAAATTTTTTAGTAGGAGGCATTGGA
LerBCTV3  GGATATGTAAGAAATATGTTTTTTGTTTGAATATGAAATTTTTTAGTAGGAGGCATTGGA
LerBCTV4  GGATATGTAAGAAATATGTTTTTTGTTTGAATATGAAATTTTTTAGTAGGAGGCATTGGA
LerBCTV5  GGATATGTAAGAAATATGTTTTTTGTTTGAATATGAAATTTTTTAGTAGGAGGCATTGGA
LerBCTV6  GGATATGTAAGAAATATGTTTTTTGTTTGAATATGAAATTTTTTAGTAGGAGGCATTGGA
LerBCTV7  GGATATGTAAGAAATATGTTTTTTGTTTGAATATGAAATTTTTTAGTAGGAGGCATTGGA

LerBCTV1  GTGATGAGGTGGTTAGATTAAAGGTTATATTAAGGTGGAAGGGTAATGATAAGATTTGTTGA
LerBCTV2  GTGATGAGGTGGTTAGATTAAAGGTTATATTAAGGTGGAAGGGTAATGATAAGATTTGTTGA
LerBCTV3  GTGATGAGGTGGTTAGATTAAAGGTTATATTAAGGTGGAAGGGTAATGATAAGATTTGTTGA
LerBCTV4  GTGATGAGGTGGTTAGATTAAAGGTTATATTAAGGTGGAAGGGTAATGATAAGATTTGTTGA
LerBCTV5  GTGATGAGGTGGTTAGATTAAAGGTTATATTAAGGTGGAAGGGTAATGATAAGATTTGTTGA
LerBCTV6  GTGATGAGGTGGTTAGATTAAAGGTTATATTAAGGTGGAAGGGTAATGATAAGATTTGTTGA
LerBCTV7  GTGATGAGGTGGTTAGATTAAAGGTTATATTAAGGTGGAAGGGTAATGATAAGATTTGTTGA

LerBCTV1  GTAATTATAAGGAGTTAGATGTGAAGGTTATTGAATGGAATTTTTTAGTAGGAGGCATTGGA
LerBCTV2  GTAATTATAAGGAGTTAGATGTGAAGGTTATTGAATGGAATTTTTTAGTAGGAGGCATTGGA
LerBCTV3  GTAATTATAAGGAGTTAGATGTGAAGGTTATTGAATGGAATTTTTTAGTAGGAGGCATTGGA
LerBCTV4  GTAATTATAAGGAGTTAGATGTGAAGGTTATTGAATGGAATTTTTTAGTAGGAGGCATTGGA
LerBCTV5  GTAATTATAAGGAGTTAGATGTGAAGGTTATTGAATGGAATTTTTTAGTAGGAGGCATTGGA
LerBCTV6  GTAATTATAAGGAGTTAGATGTGAAGGTTATTGAATGGAATTTTTTAGTAGGAGGCATTGGA
LerBCTV7  GTAATTATAAGGAGTTAGATGTGAAGGTTATTGAATGGAATTTTTTAGTAGGAGGCATTGGA

LerBCTV1  TATGGAATTAAATTTAAAAAATGGAATTATAAAAAGTGGAGAGTTAGAATGGAATTTTTTAGTAGGAGGCATTGGA
LerBCTV2  TATGGAATTAAATTTAAAAAATGGAATTATAAAAAGTGGAGAGTTAGAATGGAATTTTTTAGTAGGAGGCATTGGA
LerBCTV3  TATGGAATTAAATTTAAAAAATGGAATTATAAAAAGTGGAGAGTTAGAATGGAATTTTTTAGTAGGAGGCATTGGA
LerBCTV4  TATGGAATTAAATTTAAAAAATGGAATTATAAAAAGTGGAGAGTTAGAATGGAATTTTTTAGTAGGAGGCATTGGA
LerBCTV5  TATGGAATTAAATTTAAAAAATGGAATTATAAAAAGTGGAGAGTTAGAATGGAATTTTTTAGTAGGAGGCATTGGA
LerBCTV6  TATGGAATTAAATTTAAAAAATGGAATTATAAAAAGTGGAGAGTTAGAATGGAATTTTTTAGTAGGAGGCATTGGA
LerBCTV7  TATGGAATTAAATTTAAAAAATGGAATTATAAAAAGTGGAGAGTTAGAATGGAATTTTTTAGTAGGAGGCATTGGA
BCTV from *Arabidopsis kyp2* mutant plants.

| KypBCTV1 | GGATATGTAAGAATAATGTGTTTTTTGATTGATATAAGATAATTTGAGAGGACATATTGGAATTTTTGAA |
| KypBCTV2 | GGATATGTAAGAATAATGTGTTTTTTGATTGATATAAGATAATTTGAGAGGACATATTGGAATTTTTGAA |
| KypBCTV3 | GGATATGTAAGAATAATGTGTTTTTTGATTGATATAAGATAATTTGAGAGGACATATTGGAATTTTTGAA |
| KypBCTV4 | GGATATGTAAGAATAATGTGTTTTTTGATTGATATAAGATAATTTGAGAGGACATATTGGAATTTTTGAA |
| KypBCTV5 | GGATATGTAAGAATAATGTGTTTTTTGATTGATATAAGATAATTTGAGAGGACATATTGGAATTTTTGAA |
| KypBCTV6 | GGATATGTAAGAATAATGTGTTTTTTGATTGATATAAGATAATTTGAGAGGACATATTGGAATTTTTGAA |
| KypBCTV7 | GGATATGTAAGAATAATGTGTTTTTTGATTGATATAAGATAATTTGAGAGGACATATTGGAATTTTTGAA |
| KypBCTV8 | GGATATGTAAGAATAATGTGTTTTTTGATTGATATAAGATAATTTGAGAGGACATATTGGAATTTTTGAA |
| KypBCTV9 | GGATATGTAAGAATAATGTGTTTTTTGATTGATATAAGATAATTTGAGAGGACATATTGGAATTTTTGAA |
| KypBCTV10| GGATATGTAAGAATAATGTGTTTTTTGATTGATATAAGATAATTTGAGAGGACATATTGGAATTTTTGAA |

KypBCTV from Arabidopsis *kyp2* mutant plants.
BCTV from *Arabidopsis cmt3* mutant plants.

<table>
<thead>
<tr>
<th>CmtBCTV1</th>
<th>GGATATTGAAAGATATGTTTCTCTTGAATAAGAAATTGTTTAGAGGATTGAGGTTGGA</th>
</tr>
</thead>
<tbody>
<tr>
<td>CmtBCTV2</td>
<td>GGATATTGAAAGATATGTTTCTCTTGAATAAGAAATTGTTTAGAGGATTGAGGTTGGA</td>
</tr>
<tr>
<td>CmtBCTV3</td>
<td>GGATATTGAAAGATATGTTTCTCTTGAATAAGAAATTGTTTAGAGGATTGAGGTTGGA</td>
</tr>
<tr>
<td>CmtBCTV4</td>
<td>GGATATTGAAAGATATGTTTCTCTTGAATAAGAAATTGTTTAGAGGATTGAGGTTGGA</td>
</tr>
<tr>
<td>CmtBCTV5</td>
<td>GGATATTGAAAGATATGTTTCTCTTGAATAAGAAATTGTTTAGAGGATTGAGGTTGGA</td>
</tr>
<tr>
<td>CmtBCTV6</td>
<td>GGATATTGAAAGATATGTTTCTCTTGAATAAGAAATTGTTTAGAGGATTGAGGTTGGA</td>
</tr>
<tr>
<td>CmtBCTV7</td>
<td>GGATATTGAAAGATATGTTTCTCTTGAATAAGAAATTGTTTAGAGGATTGAGGTTGGA</td>
</tr>
</tbody>
</table>

CmtBCTV from *Arabidopsis* cmt3 mutant plants.
BCTV from *Arabidopsis ago4* mutant plants.
BCTV from secondary (2⁰) tissue of wild-type *Arabidopsis* (Ler-0).

<table>
<thead>
<tr>
<th>LernerBCTV</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>GGAATGTTAAGAATAATGTTTTTTGTTGATAAGAATTGTTAAGGAGGTATTGGA</td>
</tr>
<tr>
<td>2.2</td>
<td>GGAATGTTAAGAATAATGTTTTTTGTTGATAAGAATTGTTAAGGAGGTATTGGA</td>
</tr>
<tr>
<td>2.3</td>
<td>GGAATGTTAAGAATAATGTTTTTTGTTGATAAGAATTGTTAAGGAGGTATTGGA</td>
</tr>
<tr>
<td>2.4</td>
<td>GGAATGTTAAGAATAATGTTTTTTGTTGATAAGAATTGTTAAGGAGGTATTGGA</td>
</tr>
<tr>
<td>2.5</td>
<td>GGAATGTTAAGAATAATGTTTTTTGTTGATAAGAATTGTTAAGGAGGTATTGGA</td>
</tr>
<tr>
<td>2.6</td>
<td>GGAATGTTAAGAATAATGTTTTTTGTTGATAAGAATTGTTAAGGAGGTATTGGA</td>
</tr>
<tr>
<td>2.7</td>
<td>GGAATGTTAAGAATAATGTTTTTTGTTGATAAGAATTGTTAAGGAGGTATTGGA</td>
</tr>
<tr>
<td>2.8</td>
<td>GGAATGTTAAGAATAATGTTTTTTGTTGATAAGAATTGTTAAGGAGGTATTGGA</td>
</tr>
<tr>
<td>2.9</td>
<td>GGAATGTTAAGAATAATGTTTTTTGTTGATAAGAATTGTTAAGGAGGTATTGGA</td>
</tr>
<tr>
<td>2.10</td>
<td>GGAATGTTAAGAATAATGTTTTTTGTTGATAAGAATTGTTAAGGAGGTATTGGA</td>
</tr>
</tbody>
</table>

1 23 1 1 1 2 0 0
BCTV from secondary (2°) tissue of Arabidopsis ago4 mutant plants.
BCTV L2-1 from secondary (2nd) tissue of wild-type Arabidopsis (Ler-0).
(Asymptomatic tissue showing recovery from infection.)
BCTV L2-1 from secondary (2ⁿ) tissue of Arabidopsis ago4 mutant plants. (Symptomatic tissue; no recovery.)
BCTV L2-2 from secondary (2\textsuperscript{nd}) tissue of wild-type Arabidopsis (Ler-0).
(Asymptomatic tissue showing recovery from infection.)
BCTV L2-2 from secondary (2\textsuperscript{nd}) tissue of Arabidopsis ago4 mutant plants. (Symptomatic tissue; no recovery.)

\begin{verbatim}
ago4L2-22.1  GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.2  GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.3  GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.4  GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.5  GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.6  GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.7  GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.8  GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.9  GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.10 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.11 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.12 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.13 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.14 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.15 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.16 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.17 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.18 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.19 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.20 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.21 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.22 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.23 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.24 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.25 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.26 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.27 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.28 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.29 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.30 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.31 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.32 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.33 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.34 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.35 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.36 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.37 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.38 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.39 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
ago4L2-22.40 GATGTAAGAATATGTTTTTGTGTAATAAGAATTGAGCCATGGA
\end{verbatim}
BCTV conversion control. Plasmid DNA containing a BCTV IR insert was added to excess cellular DNA from mock inoculated *Arabidopsis*.

BCTV1: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV2: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV3: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV4: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV5: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV6: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV7: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV8: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV9: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV10: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA

**BCTV conversion control.** Plasmid DNA containing a BCTV IR insert was added to excess cellular DNA from mock inoculated *Arabidopsis*.

BCTV1: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV2: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV3: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV4: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV5: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV6: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV7: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV8: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV9: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV10: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA

**BCTV conversion control.** Plasmid DNA containing a BCTV IR insert was added to excess cellular DNA from mock inoculated *Arabidopsis*.

BCTV1: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV2: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV3: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV4: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV5: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV6: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV7: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV8: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV9: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV10: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA

**BCTV conversion control.** Plasmid DNA containing a BCTV IR insert was added to excess cellular DNA from mock inoculated *Arabidopsis*.

BCTV1: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV2: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV3: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV4: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV5: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV6: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV7: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV8: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV9: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA
BCTV10: GGATATGTAAGAATAATATTTTTGTTTAAGTATGGAATGAATTTGATGAGATTGGA

**BCTV conversion control.** Plasmid DNA containing a BCTV IR insert was added to excess cellular DNA from mock inoculated *Arabidopsis*.
Bisulfite sequencing of the CaLCuV intergenic region (IR).

The CaLCuV IR is shown below, where methylation was assayed using bisulfite sequencing. The Rep (AL1) binding sites, the conserved hairpin, and the AL1 (complementary sense) and coat protein start codons, are indicated in bold and underlined. In total, 79 cytosine residues were tested in CG (red), CNG (pink) and CHH (grey) contexts. Methylated cytosines are indicated in yellow. Unmethylated cytosines are converted to thymine (green).

Forward and reverse primers used to amplify the viral strand following bisulfite treatment:

CaLCuV2556CF: GGGGATATGTTAAGAATATATTTTG
CaLCuV359CR: TCCCCACCATAAAAACAC
CaLCuV from *N. benthamiana.*

<table>
<thead>
<tr>
<th>N</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>2</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>3</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>4</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>5</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>6</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>7</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>8</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>9</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>10</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>11</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>12</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
</tbody>
</table>

N=12

<table>
<thead>
<tr>
<th>N</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>2</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>3</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>4</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>5</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>6</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>7</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>8</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>9</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>10</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>11</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>12</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
</tbody>
</table>

N=12

<table>
<thead>
<tr>
<th>N</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>2</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>3</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>4</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>5</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>6</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>7</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>8</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>9</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>10</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>11</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>12</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
</tbody>
</table>

N=12

<table>
<thead>
<tr>
<th>N</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>2</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>3</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>4</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>5</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>6</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>7</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>8</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>9</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>10</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>11</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>12</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
</tbody>
</table>

N=12

<table>
<thead>
<tr>
<th>N</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>2</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>3</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>4</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>5</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>6</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>7</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>8</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>9</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>10</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>11</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
<tr>
<td>12</td>
<td>TAAAGAAAGATTTAGGGTTTTGTGGTA</td>
</tr>
</tbody>
</table>
**CaLCuV from wild-type Arabidopsis (Ler-0).**

<table>
<thead>
<tr>
<th>Ler01</th>
<th>TAAAGGAAGATTTAGGTTTTTGGCACA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ler02</td>
<td>TAAAGGAAGATTTAGGTTTTTGGCACA</td>
</tr>
<tr>
<td>Ler03</td>
<td>TAAAGGAAGATTTAGGTTTTTGGCACA</td>
</tr>
<tr>
<td>Ler04</td>
<td>TAAAGGAAGATTTAGGTTTTTGGCACA</td>
</tr>
<tr>
<td>Ler05</td>
<td>TAAAGGAAGATTTAGGTTTTTGGCACA</td>
</tr>
<tr>
<td>Ler06</td>
<td>TAAAGGAAGATTTAGGTTTTTGGCACA</td>
</tr>
<tr>
<td>Ler07</td>
<td>TAAAGGAAGATTTAGGTTTTTGGCACA</td>
</tr>
<tr>
<td>Ler08</td>
<td>TAAAGGAAGATTTAGGTTTTTGGCACA</td>
</tr>
<tr>
<td>Ler09</td>
<td>TAAAGGAAGATTTAGGTTTTTGGCACA</td>
</tr>
<tr>
<td>Ler10</td>
<td>TAAAGGAAGATTTAGGTTTTTGGCACA</td>
</tr>
<tr>
<td>Ler11</td>
<td>TAAAGGAAGATTTAGGTTTTTGGCACA</td>
</tr>
<tr>
<td>Ler12</td>
<td>TAAAGGAAGATTTAGGTTTTTGGCACA</td>
</tr>
</tbody>
</table>

N=12

<table>
<thead>
<tr>
<th>Ler01</th>
<th>TTTTGTAAATATTTAGGATAAGAAGGAGGAGCTTATATTTAGGA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ler02</td>
<td>TTTTGTAAATATTTAGGATAAGAAGGAGGAGCTTATATTTAGGA</td>
</tr>
<tr>
<td>Ler03</td>
<td>TTTTGTAAATATTTAGGATAAGAAGGAGGAGCTTATATTTAGGA</td>
</tr>
<tr>
<td>Ler04</td>
<td>TTTTGTAAATATTTAGGATAAGAAGGAGGAGCTTATATTTAGGA</td>
</tr>
<tr>
<td>Ler05</td>
<td>TTTTGTAAATATTTAGGATAAGAAGGAGGAGCTTATATTTAGGA</td>
</tr>
<tr>
<td>Ler06</td>
<td>TTTTGTAAATATTTAGGATAAGAAGGAGGAGCTTATATTTAGGA</td>
</tr>
<tr>
<td>Ler07</td>
<td>TTTTGTAAATATTTAGGATAAGAAGGAGGAGCTTATATTTAGGA</td>
</tr>
<tr>
<td>Ler08</td>
<td>TTTTGTAAATATTTAGGATAAGAAGGAGGAGCTTATATTTAGGA</td>
</tr>
<tr>
<td>Ler09</td>
<td>TTTTGTAAATATTTAGGATAAGAAGGAGGAGCTTATATTTAGGA</td>
</tr>
<tr>
<td>Ler10</td>
<td>TTTTGTAAATATTTAGGATAAGAAGGAGGAGCTTATATTTAGGA</td>
</tr>
<tr>
<td>Ler11</td>
<td>TTTTGTAAATATTTAGGATAAGAAGGAGGAGCTTATATTTAGGA</td>
</tr>
<tr>
<td>Ler12</td>
<td>TTTTGTAAATATTTAGGATAAGAAGGAGGAGCTTATATTTAGGA</td>
</tr>
</tbody>
</table>

N=12

<table>
<thead>
<tr>
<th>Ler01</th>
<th>TATTGTGTTTTGTTGGTTGTATGATATATAAAGGGGAAAGAGGGG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ler02</td>
<td>TATTGTGTTTTGTTGGTTGTATGATATATAAAGGGGAAAGAGGGG</td>
</tr>
<tr>
<td>Ler03</td>
<td>TATTGTGTTTTGTTGGTTGTATGATATATAAAGGGGAAAGAGGGG</td>
</tr>
<tr>
<td>Ler04</td>
<td>TATTGTGTTTTGTTGGTTGTATGATATATAAAGGGGAAAGAGGGG</td>
</tr>
<tr>
<td>Ler05</td>
<td>TATTGTGTTTTGTTGGTTGTATGATATATAAAGGGGAAAGAGGGG</td>
</tr>
<tr>
<td>Ler06</td>
<td>TATTGTGTTTTGTTGGTTGTATGATATATAAAGGGGAAAGAGGGG</td>
</tr>
<tr>
<td>Ler07</td>
<td>TATTGTGTTTTGTTGGTTGTATGATATATAAAGGGGAAAGAGGGG</td>
</tr>
<tr>
<td>Ler08</td>
<td>TATTGTGTTTTGTTGGTTGTATGATATATAAAGGGGAAAGAGGGG</td>
</tr>
<tr>
<td>Ler09</td>
<td>TATTGTGTTTTGTTGGTTGTATGATATATAAAGGGGAAAGAGGGG</td>
</tr>
<tr>
<td>Ler10</td>
<td>TATTGTGTTTTGTTGGTTGTATGATATATAAAGGGGAAAGAGGGG</td>
</tr>
<tr>
<td>Ler11</td>
<td>TATTGTGTTTTGTTGGTTGTATGATATATAAAGGGGAAAGAGGGG</td>
</tr>
<tr>
<td>Ler12</td>
<td>TATTGTGTTTTGTTGGTTGTATGATATATAAAGGGGAAAGAGGGG</td>
</tr>
</tbody>
</table>

N=12

<table>
<thead>
<tr>
<th>Ler01</th>
<th>AAAAGGATGATGAAATTAGTATAAGATTTTTTGGGTTTGGTGT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ler02</td>
<td>AAAAGGATGATGAAATTAGTATAAGATTTTTTGGGTTTGGTGT</td>
</tr>
<tr>
<td>Ler03</td>
<td>AAAAGGATGATGAAATTAGTATAAGATTTTTTGGGTTTGGTGT</td>
</tr>
<tr>
<td>Ler04</td>
<td>AAAAGGATGATGAAATTAGTATAAGATTTTTTGGGTTTGGTGT</td>
</tr>
<tr>
<td>Ler05</td>
<td>AAAAGGATGATGAAATTAGTATAAGATTTTTTGGGTTTGGTGT</td>
</tr>
<tr>
<td>Ler06</td>
<td>AAAAGGATGATGAAATTAGTATAAGATTTTTTGGGTTTGGTGT</td>
</tr>
<tr>
<td>Ler07</td>
<td>AAAAGGATGATGAAATTAGTATAAGATTTTTTGGGTTTGGTGT</td>
</tr>
<tr>
<td>Ler08</td>
<td>AAAAGGATGATGAAATTAGTATAAGATTTTTTGGGTTTGGTGT</td>
</tr>
<tr>
<td>Ler09</td>
<td>AAAAGGATGATGAAATTAGTATAAGATTTTTTGGGTTTGGTGT</td>
</tr>
<tr>
<td>Ler10</td>
<td>AAAAGGATGATGAAATTAGTATAAGATTTTTTGGGTTTGGTGT</td>
</tr>
<tr>
<td>Ler11</td>
<td>AAAAGGATGATGAAATTAGTATAAGATTTTTTGGGTTTGGTGT</td>
</tr>
<tr>
<td>Ler12</td>
<td>AAAAGGATGATGAAATTAGTATAAGATTTTTTGGGTTTGGTGT</td>
</tr>
</tbody>
</table>

N=12
CaLCuV from *Arabidopsis ago4* mutant plants.

Ago41  
Ago42  
Ago43  
Ago44  
Ago45  
Ago46  
Ago47  
Ago48  
Ago49  
Ago50  
Ago51  
Ago52  

N=12  

Ago41  
Ago42  
Ago43  
Ago44  
Ago45  
Ago46  
Ago47  
Ago48  
Ago49  
Ago50  
Ago51  
Ago52  

N=12  

Ago41  
Ago42  
Ago43  
Ago44  
Ago45  
Ago46  
Ago47  
Ago48  
Ago49  
Ago50  
Ago51  
Ago52  

N=12  

Ago41  
Ago42  
Ago43  
Ago44  
Ago45  
Ago46  
Ago47  
Ago48  
Ago49  
Ago50  
Ago51  
Ago52  

N=12  

Ago41  
Ago42  
Ago43  
Ago44  
Ago45  
Ago46  
Ago47  
Ago48  
Ago49  
Ago50  
Ago51  
Ago52  

N=12
Ago41       GGGATTAAATGGAATTTAATTAATATTAAATATGTTGCTGAAAGGTTAGTTAATTA
Ago42       GGGATTAAATGGAATTTAATTAATATTAAATATGTTGCTGAAAGGTTAGTTAATTA
Ago43       GGGATTAAATGGAATTTAATTAATATTAAATATGTTGCTGAAAGGTTAGTTAATTA
Ago44       GGGATTAAATGGAATTTAATTAATATTAAATATGTTGCTGAAAGGTTAGTTAATTA
Ago45       GGGATTAAATGGAATTTAATTAATATTAAATATGTTGCTGAAAGGTTAGTTAATTA
Ago46       GGGATTAAATGGAATTTAATTAATATTAAATATGTTGCTGAAAGGTTAGTTAATTA
Ago47       GGGATTAAATGGAATTTAATTAATATTAAATATGTTGCTGAAAGGTTAGTTAATTA
Ago48       GGGATTAAATGGAATTTAATTAATATTAAATATGTTGCTGAAAGGTTAGTTAATTA
Ago49       GGGATTAAATGGAATTTAATTAATATTAAATATGTTGCTGAAAGGTTAGTTAATTA
Ago50       GGGATTAAATGGAATTTAATTAATATTAAATATGTTGCTGAAAGGTTAGTTAATTA
Ago51       GGGATTAAATGGAATTTAATTAATATTAAATATGTTGCTGAAAGGTTAGTTAATTA
Ago52       GGGATTAAATGGAATTTAATTAATATTAAATATGTTGCTGAAAGGTTAGTTAATTA
N=12          00 1 1 1 01 10 00 02 0 1

Ago41       AGGTAAATGAGTGCTGTGGGTTTTATATAAGAATGTGATTTTAATTTATGTTTA
Ago42       AGGTAAATGAGTGCTGTGGGTTTTATATAAGAATGTGATTTTAATTTATGTTTA
Ago43       AGGTAAATGAGTGCTGTGGGTTTTATATAAGAATGTGATTTTAATTTATGTTTA
Ago44       AGGTAAATGAGTGCTGTGGGTTTTATATAAGAATGTGATTTTAATTTATGTTTA
Ago45       AGGTAAATGAGTGCTGTGGGTTTTATATAAGAATGTGATTTTAATTTATGTTTA
Ago46       AGGTAAATGAGTGCTGTGGGTTTTATATAAGAATGTGATTTTAATTTATGTTTA
Ago47       AGGTAAATGAGTGCTGTGGGTTTTATATAAGAATGTGATTTTAATTTATGTTTA
Ago48       AGGTAAATGAGTGCTGTGGGTTTTATATAAGAATGTGATTTTAATTTATGTTTA
Ago49       AGGTAAATGAGTGCTGTGGGTTTTATATAAGAATGTGATTTTAATTTATGTTTA
Ago50       AGGTAAATGAGTGCTGTGGGTTTTATATAAGAATGTGATTTTAATTTATGTTTA
Ago51       AGGTAAATGAGTGCTGTGGGTTTTATATAAGAATGTGATTTTAATTTATGTTTA
Ago52       AGGTAAATGAGTGCTGTGGGTTTTATATAAGAATGTGATTTTAATTTATGTTTA
N=12          00 0 0 0 0 0 0 0 0 0

Ago41       TTTGAAATGCTTAAGZGGGATGTZGTTGCTGTATGTTGATTTGATGZGGGGA
Ago42       TTTGAAATGCTTAAGZGGGATGTZGTTGCTGTATGTTGATTTGATGZGGGGA
Ago43       TTTGAAATGCTTAAGZGGGATGTZGTTGCTGTATGTTGATTTGATGZGGGGA
Ago44       TTTGAAATGCTTAAGZGGGATGTZGTTGCTGTATGTTGATTTGATGZGGGGA
Ago45       TTTGAAATGCTTAAGZGGGATGTZGTTGCTGTATGTTGATTTGATGZGGGGA
Ago46       TTTGAAATGCTTAAGZGGGATGTZGTTGCTGTATGTTGATTTGATGZGGGGA
Ago47       TTTGAAATGCTTAAGZGGGATGTZGTTGCTGTATGTTGATTTGATGZGGGGA
Ago48       TTTGAAATGCTTAAGZGGGATGTZGTTGCTGTATGTTGATTTGATGZGGGGA
Ago49       TTTGAAATGCTTAAGZGGGATGTZGTTGCTGTATGTTGATTTGATGZGGGGA
Ago50       TTTGAAATGCTTAAGZGGGATGTZGTTGCTGTATGTTGATTTGATGZGGGGA
Ago51       TTTGAAATGCTTAAGZGGGATGTZGTTGCTGTATGTTGATTTGATGZGGGGA
Ago52       TTTGAAATGCTTAAGZGGGATGTZGTTGCTGTATGTTGATTTGATGZGGGGA
N=12          2 01 1011 0 0 0

AMP1-2
<table>
<thead>
<tr>
<th>Kyp21</th>
<th>TAAAAGAAAAGATTAGGGNTTTGTGGGA</th>
<th>0 01 3 00 0 0 0 0 21</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kyp22</td>
<td>TATTGGTTTTTGGTGTAAATAATTTAGGA</td>
<td>0 01 3 00 0 0 0 0 21</td>
</tr>
<tr>
<td>Kyp23</td>
<td>TATTGGTTTTTGGTGTAAATAATTTAGGA</td>
<td>0 01 3 00 0 0 0 0 21</td>
</tr>
<tr>
<td>Kyp24</td>
<td>TATTGGTTTTTGGTGTAAATAATTTAGGA</td>
<td>0 01 3 00 0 0 0 0 21</td>
</tr>
<tr>
<td>Kyp25</td>
<td>TATTGGTTTTTGGTGTAAATAATTTAGGA</td>
<td>0 01 3 00 0 0 0 0 21</td>
</tr>
<tr>
<td>Kyp26</td>
<td>TATTGGTTTTTGGTGTAAATAATTTAGGA</td>
<td>0 01 3 00 0 0 0 0 21</td>
</tr>
<tr>
<td>Kyp27</td>
<td>TATTGGTTTTTGGTGTAAATAATTTAGGA</td>
<td>0 01 3 00 0 0 0 0 21</td>
</tr>
<tr>
<td>Kyp28</td>
<td>TATTGGTTTTTGGTGTAAATAATTTAGGA</td>
<td>0 01 3 00 0 0 0 0 21</td>
</tr>
<tr>
<td>Kyp29</td>
<td>TATTGGTTTTTGGTGTAAATAATTTAGGA</td>
<td>0 01 3 00 0 0 0 0 21</td>
</tr>
<tr>
<td>Kyp30</td>
<td>TATTGGTTTTTGGTGTAAATAATTTAGGA</td>
<td>0 01 3 00 0 0 0 0 21</td>
</tr>
<tr>
<td>Kyp31</td>
<td>TATTGGTTTTTGGTGTAAATAATTTAGGA</td>
<td>0 01 3 00 0 0 0 0 21</td>
</tr>
<tr>
<td>CaLCuV from Arabidopsis kyp2 mutant plants.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

N=12
CaLCuV from *Arabidopsis cmt3* mutant plants.

| Cmt31 | TAAAGAAAGTTTAAATGGTTTGAATAAGAGAGTCTCTCTAAACG |
| Cmt32 | TAAAGAAAGTTTAAATGGTTTGAATAAGAGAGTCTCTCTAAACG |
| Cmt33 | TAAAGAAAGTTTAAATGGTTTGAATAAGAGAGTCTCTCTAAACG |
| Cmt34 | TAAAGAAAGTTTAAATGGTTTGAATAAGAGAGTCTCTCTAAACG |
| Cmt35 | TAAAGAAAGTTTAAATGGTTTGAATAAGAGAGTCTCTCTAAACG |
| Cmt36 | TAAAGAAAGTTTAAATGGTTTGAATAAGAGAGTCTCTCTAAACG |
| Cmt37 | TAAAGAAAGTTTAAATGGTTTGAATAAGAGAGTCTCTCTAAACG |
| Cmt38 | TAAAGAAAGTTTAAATGGTTTGAATAAGAGAGTCTCTCTAAACG |
| Cmt39 | TAAAGAAAGTTTAAATGGTTTGAATAAGAGAGTCTCTCTAAACG |
| Cmt40 | TAAAGAAAGTTTAAATGGTTTGAATAAGAGAGTCTCTCTAAACG |
| Cmt41 | TAAAGAAAGTTTAAATGGTTTGAATAAGAGAGTCTCTCTAAACG |

**N=12**

| Cmt31 | TTTTTGTAATAATATTTAGGAATAGGGGAGAGGAGGTTTGTGGT |
| Cmt32 | TTTTTGTAATAATATTTAGGAATAGGGGAGAGGAGGTTTGTGGT |
| Cmt33 | TTTTTGTAATAATATTTAGGAATAGGGGAGAGGAGGTTTGTGGT |
| Cmt34 | TTTTTGTAATAATATTTAGGAATAGGGGAGAGGAGGTTTGTGGT |
| Cmt35 | TTTTTGTAATAATATTTAGGAATAGGGGAGAGGAGGTTTGTGGT |
| Cmt36 | TTTTTGTAATAATATTTAGGAATAGGGGAGAGGAGGTTTGTGGT |
| Cmt37 | TTTTTGTAATAATATTTAGGAATAGGGGAGAGGAGGTTTGTGGT |
| Cmt38 | TTTTTGTAATAATATTTAGGAATAGGGGAGAGGAGGTTTGTGGT |
| Cmt39 | TTTTTGTAATAATATTTAGGAATAGGGGAGAGGAGGTTTGTGGT |
| Cmt40 | TTTTTGTAATAATATTTAGGAATAGGGGAGAGGAGGTTTGTGGT |
| Cmt41 | TTTTTGTAATAATATTTAGGAATAGGGGAGAGGAGGTTTGTGGT |

**N=12**

| Cmt31 | TATTGTTTTGTTTGTGGTAATATATAGGTAAAAGTTATGGATAAGAGAGTCTCTCTAAACG |
| Cmt32 | TATTGTTTTGTTTGTGGTAATATATAGGTAAAAGTTATGGATAAGAGAGTCTCTCTAAACG |
| Cmt33 | TATTGTTTTGTTTGTGGTAATATATAGGTAAAAGTTATGGATAAGAGAGTCTCTCTAAACG |
| Cmt34 | TATTGTTTTGTTTGTGGTAATATATAGGTAAAAGTTATGGATAAGAGAGTCTCTCTAAACG |
| Cmt35 | TATTGTTTTGTTTGTGGTAATATATAGGTAAAAGTTATGGATAAGAGAGTCTCTCTAAACG |
| Cmt36 | TATTGTTTTGTTTGTGGTAATATATAGGTAAAAGTTATGGATAAGAGAGTCTCTCTAAACG |
| Cmt37 | TATTGTTTTGTTTGTGGTAATATATAGGTAAAAGTTATGGATAAGAGAGTCTCTCTAAACG |
| Cmt38 | TATTGTTTTGTTTGTGGTAATATATAGGTAAAAGTTATGGATAAGAGAGTCTCTCTAAACG |
| Cmt39 | TATTGTTTTGTTTGTGGTAATATATAGGTAAAAGTTATGGATAAGAGAGTCTCTCTAAACG |
| Cmt40 | TATTGTTTTGTTTGTGGTAATATATAGGTAAAAGTTATGGATAAGAGAGTCTCTCTAAACG |
| Cmt41 | TATTGTTTTGTTTGTGGTAATATATAGGTAAAAGTTATGGATAAGAGAGTCTCTCTAAACG |

**N=12**

| Cmt31 | TAAAGAAAGTTTAAATGGTTTGAATAAGAGAGTCTCTCTAAACG |
| Cmt32 | TAAAGAAAGTTTAAATGGTTTGAATAAGAGAGTCTCTCTAAACG |
| Cmt33 | TAAAGAAAGTTTAAATGGTTTGAATAAGAGAGTCTCTCTAAACG |
| Cmt34 | TAAAGAAAGTTTAAATGGTTTGAATAAGAGAGTCTCTCTAAACG |
| Cmt35 | TAAAGAAAGTTTAAATGGTTTGAATAAGAGAGTCTCTCTAAACG |
| Cmt36 | TAAAGAAAGTTTAAATGGTTTGAATAAGAGAGTCTCTCTAAACG |
| Cmt37 | TAAAGAAAGTTTAAATGGTTTGAATAAGAGAGTCTCTCTAAACG |
| Cmt38 | TAAAGAAAGTTTAAATGGTTTGAATAAGAGAGTCTCTCTAAACG |
| Cmt39 | TAAAGAAAGTTTAAATGGTTTGAATAAGAGAGTCTCTCTAAACG |
| Cmt40 | TAAAGAAAGTTTAAATGGTTTGAATAAGAGAGTCTCTCTAAACG |
| Cmt41 | TAAAGAAAGTTTAAATGGTTTGAATAAGAGAGTCTCTCTAAACG |

**N=12**
### CaLCuV from wild-type *Arabidopsis* (Ws-2).

<table>
<thead>
<tr>
<th>Reference</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ws21</td>
<td>TAAAGAAAAGATTAGGGTTT1GTGGGA</td>
</tr>
<tr>
<td>Ws22</td>
<td>TAAAGAAAAGATTAGGGTTT1GTGGGA</td>
</tr>
<tr>
<td>Ws23</td>
<td>TAAAGAAAAGATTAGGGTTT1GTGGGA</td>
</tr>
<tr>
<td>Ws24</td>
<td>TAAAGAAAAGATTAGGGTTT1GTGGGA</td>
</tr>
<tr>
<td>Ws25</td>
<td>TAAAGAAAAGATTAGGGTTT1GTGGGA</td>
</tr>
<tr>
<td>Ws26</td>
<td>TAAAGAAAAGATTAGGGTTT1GTGGGA</td>
</tr>
<tr>
<td>Ws27</td>
<td>TAAAGAAAAGATTAGGGTTT1GTGGGA</td>
</tr>
<tr>
<td>Ws28</td>
<td>TAAAGAAAAGATTAGGGTTT1GTGGGA</td>
</tr>
<tr>
<td>Ws29</td>
<td>TAAAGAAAAGATTAGGGTTT1GTGGGA</td>
</tr>
<tr>
<td>Ws30</td>
<td>TAAAGAAAAGATTAGGGTTT1GTGGGA</td>
</tr>
<tr>
<td>Ws31</td>
<td>TAAAGAAAAGATTAGGGTTT1GTGGGA</td>
</tr>
<tr>
<td>Ws32</td>
<td>TAAAGAAAAGATTAGGGTTT1GTGGGA</td>
</tr>
</tbody>
</table>

N=12
CaLCuV from *Arabidopsis drm1/2* mutant plants.

<table>
<thead>
<tr>
<th>Drm121</th>
<th>TAAAGGAAAAGATTTAGGGTTTTTGGGAA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drm122</td>
<td>TAAAGGAAAAGATTTAGGGTTTTTGGGAA</td>
</tr>
<tr>
<td>Drm123</td>
<td>TAAAGGAAAAGATTTAGGGTTTTTGGGAA</td>
</tr>
<tr>
<td>Drm124</td>
<td>TAAAGGAAAAGATTTAGGGTTTTTGGGAA</td>
</tr>
<tr>
<td>Drm125</td>
<td>TAAAGGAAAAGATTTAGGGTTTTTGGGAA</td>
</tr>
<tr>
<td>Drm126</td>
<td>TAAAGGAAAAGATTTAGGGTTTTTGGGAA</td>
</tr>
<tr>
<td>Drm127</td>
<td>TAAAGGAAAAGATTTAGGGTTTTTGGGAA</td>
</tr>
<tr>
<td>Drm128</td>
<td>TAAAGGAAAAGATTTAGGGTTTTTGGGAA</td>
</tr>
<tr>
<td>Drm129</td>
<td>TAAAGGAAAAGATTTAGGGTTTTTGGGAA</td>
</tr>
<tr>
<td>Drm130</td>
<td>TAAAGGAAAAGATTTAGGGTTTTTGGGAA</td>
</tr>
<tr>
<td>Drm131</td>
<td>TAAAGGAAAAGATTTAGGGTTTTTGGGAA</td>
</tr>
<tr>
<td>Drm132</td>
<td>TAAAGGAAAAGATTTAGGGTTTTTGGGAA</td>
</tr>
</tbody>
</table>

N=12

<table>
<thead>
<tr>
<th>Drm121</th>
<th>TTTTTGTAAATAATATTTAGGA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drm122</td>
<td>TTTTTGTAAATAATATTTAGGA</td>
</tr>
<tr>
<td>Drm123</td>
<td>TTTTTGTAAATAATATTTAGGA</td>
</tr>
<tr>
<td>Drm124</td>
<td>TTTTTGTAAATAATATTTAGGA</td>
</tr>
<tr>
<td>Drm125</td>
<td>TTTTTGTAAATAATATTTAGGA</td>
</tr>
<tr>
<td>Drm126</td>
<td>TTTTTGTAAATAATATTTAGGA</td>
</tr>
<tr>
<td>Drm127</td>
<td>TTTTTGTAAATAATATTTAGGA</td>
</tr>
<tr>
<td>Drm128</td>
<td>TTTTTGTAAATAATATTTAGGA</td>
</tr>
<tr>
<td>Drm129</td>
<td>TTTTTGTAAATAATATTTAGGA</td>
</tr>
<tr>
<td>Drm130</td>
<td>TTTTTGTAAATAATATTTAGGA</td>
</tr>
<tr>
<td>Drm131</td>
<td>TTTTTGTAAATAATATTTAGGA</td>
</tr>
<tr>
<td>Drm132</td>
<td>TTTTTGTAAATAATATTTAGGA</td>
</tr>
</tbody>
</table>

N=12

<table>
<thead>
<tr>
<th>Drm121</th>
<th>TATTTTTTTGTGTGTTTGTGGTA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drm122</td>
<td>TATTTTTTTGTGTGTTTGTGGTA</td>
</tr>
<tr>
<td>Drm123</td>
<td>TATTTTTTTGTGTGTTTGTGGTA</td>
</tr>
<tr>
<td>Drm124</td>
<td>TATTTTTTTGTGTGTTTGTGGTA</td>
</tr>
<tr>
<td>Drm125</td>
<td>TATTTTTTTGTGTGTTTGTGGTA</td>
</tr>
<tr>
<td>Drm126</td>
<td>TATTTTTTTGTGTGTTTGTGGTA</td>
</tr>
<tr>
<td>Drm127</td>
<td>TATTTTTTTGTGTGTTTGTGGTA</td>
</tr>
<tr>
<td>Drm128</td>
<td>TATTTTTTTGTGTGTTTGTGGTA</td>
</tr>
<tr>
<td>Drm129</td>
<td>TATTTTTTTGTGTGTTTGTGGTA</td>
</tr>
<tr>
<td>Drm130</td>
<td>TATTTTTTTGTGTGTTTGTGGTA</td>
</tr>
<tr>
<td>Drm131</td>
<td>TATTTTTTTGTGTGTTTGTGGTA</td>
</tr>
<tr>
<td>Drm132</td>
<td>TATTTTTTTGTGTGTTTGTGGTA</td>
</tr>
</tbody>
</table>

N=12

<table>
<thead>
<tr>
<th>Drm121</th>
<th>AAAAGGATTTAGGGTTTTTGGGAA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drm122</td>
<td>AAAAGGATTTAGGGTTTTTGGGAA</td>
</tr>
<tr>
<td>Drm123</td>
<td>AAAAGGATTTAGGGTTTTTGGGAA</td>
</tr>
<tr>
<td>Drm124</td>
<td>AAAAGGATTTAGGGTTTTTGGGAA</td>
</tr>
<tr>
<td>Drm125</td>
<td>AAAAGGATTTAGGGTTTTTGGGAA</td>
</tr>
<tr>
<td>Drm126</td>
<td>AAAAGGATTTAGGGTTTTTGGGAA</td>
</tr>
<tr>
<td>Drm127</td>
<td>AAAAGGATTTAGGGTTTTTGGGAA</td>
</tr>
<tr>
<td>Drm128</td>
<td>AAAAGGATTTAGGGTTTTTGGGAA</td>
</tr>
<tr>
<td>Drm129</td>
<td>AAAAGGATTTAGGGTTTTTGGGAA</td>
</tr>
<tr>
<td>Drm130</td>
<td>AAAAGGATTTAGGGTTTTTGGGAA</td>
</tr>
<tr>
<td>Drm131</td>
<td>AAAAGGATTTAGGGTTTTTGGGAA</td>
</tr>
<tr>
<td>Drm132</td>
<td>AAAAGGATTTAGGGTTTTTGGGAA</td>
</tr>
</tbody>
</table>

N=12
CaLCuV from wild-type Arabidopsis (Col-0).

Co101 TAAATGAAATGATTTAGGGTTTTGTGGTA
Co102 TAAATGAAATGATTTAGGGTTTTGTGGTA
Co103 TAAATGAAATGATTTAGGGTTTTGTGGTA
Co104 TAAATGAAATGATTTAGGGTTTTGTGGTA
Co105 TAAATGAAATGATTTAGGGTTTTGTGGTA
Co106 TAAATGAAATGATTTAGGGTTTTGTGGTA
Co107 TAAATGAAATGATTTAGGGTTTTGTGGTA
Co108 TAAATGAAATGATTTAGGGTTTTGTGGTA
Co109 TAAATGAAATGATTTAGGGTTTTGTGGTA
Co110 TAAATGAAATGATTTAGGGTTTTGTGGTA
Co111 TAAATGAAATGATTTAGGGTTTTGTGGTA
Co112 TAAATGAAATGATTTAGGGTTTTGTGGTA

N=12

Co101 TTTTTGAAAAATAATATTAGGATTAAGGGGAGGAAGAGAGAGTTTTTAAAT
Co102 TTTTTGAAAAATAATATTAGGATTAAGGGGAGGAAGAGAGAGTTTTTAAAT
Co103 TTTTTGAAAAATAATATTAGGATTAAGGGGAGGAAGAGAGAGTTTTTAAAT
Co104 TTTTTGAAAAATAATATTAGGATTAAGGGGAGGAAGAGAGAGTTTTTAAAT
Co105 TTTTTGAAAAATAATATTAGGATTAAGGGGAGGAAGAGAGAGTTTTTAAAT
Co106 TTTTTGAAAAATAATATTAGGATTAAGGGGAGGAAGAGAGAGTTTTTAAAT
Co107 TTTTTGAAAAATAATATTAGGATTAAGGGGAGGAAGAGAGAGTTTTTAAAT
Co108 TTTTTGAAAAATAATATTAGGATTAAGGGGAGGAAGAGAGAGTTTTTAAAT
Co109 TTTTTGAAAAATAATATTAGGATTAAGGGGAGGAAGAGAGAGTTTTTAAAT
Co110 TTTTTGAAAAATAATATTAGGATTAAGGGGAGGAAGAGAGAGTTTTTAAAT
Co111 TTTTTGAAAAATAATATTAGGATTAAGGGGAGGAAGAGAGAGTTTTTAAAT
Co112 TTTTTGAAAAATAATATTAGGATTAAGGGGAGGAAGAGAGAGTTTTTAAAT

N=12

Co101 TATTGTNTTGGGNTTTGGTGTATATATATAAAGAGCTTTGAGGAAATAGGGG
Co102 TATTGTNTTGGGNTTTGGTGTATATATATAAAGAGCTTTGAGGAAATAGGGG
Co103 TATTGTNTTGGGNTTTGGTGTATATATATAAAGAGCTTTGAGGAAATAGGGG
Co104 TATTGTNTTGGGNTTTGGTGTATATATATAAAGAGCTTTGAGGAAATAGGGG
Co105 TATTGTNTTGGGNTTTGGTGTATATATATAAAGAGCTTTGAGGAAATAGGGG
Co106 TATTGTNTTGGGNTTTGGTGTATATATATAAAGAGCTTTGAGGAAATAGGGG
Co107 TATTGTNTTGGGNTTTGGTGTATATATATAAAGAGCTTTGAGGAAATAGGGG
Co108 TATTGTNTTGGGNTTTGGTGTATATATATAAAGAGCTTTGAGGAAATAGGGG
Co109 TATTGTNTTGGGNTTTGGTGTATATATATAAAGAGCTTTGAGGAAATAGGGG
Co110 TATTGTNTTGGGNTTTGGTGTATATATATAAAGAGCTTTGAGGAAATAGGGG
Co111 TATTGTNTTGGGNTTTGGTGTATATATATAAAGAGCTTTGAGGAAATAGGGG
Co112 TATTGTNTTGGGNTTTGGTGTATATATATAAAGAGCTTTGAGGAAATAGGGG

N=12

CaLCuV from wild-type Arabidopsis (Col-0).
CaLCuV from *Arabidopsis adk2* mutant plants.

Adk21

```
TAAAGAAAAGATTTAGGGTTTTGTGGTTAATATATAATTAATATATAATTTTTTTG
```

Adk22

```
TAAAGAAAAGATTTAGGGTTTTGTGGTTAATATATAATTAATATATAATTTTTTTG
```

Adk23

```
TAAAGAAAAGATTTAGGGTTTTGTGGTTAATATATAATTAATATATAATTTTTTTG
```

Adk24

```
TAAAGAAAAGATTTAGGGTTTTGTGGTTAATATATAATTAATATATAATTTTTTTG
```

Adk25

```
TAAAGAAAAGATTTAGGGTTTTGTGGTTAATATATAATTAATATATAATTTTTTTG
```

Adk26

```
TAAAGAAAAGATTTAGGGTTTTGTGGTTAATATATAATTAATATATAATTTTTTTG
```

Adk27

```
TAAAGAAAAGATTTAGGGTTTTGTGGTTAATATATAATTAATATATAATTTTTTTG
```

Adk28

```
TAAAGAAAAGATTTAGGGTTTTGTGGTTAATATATAATTAATATATAATTTTTTTG
```

Adk29

```
TAAAGAAAAGATTTAGGGTTTTGTGGTTAATATATAATTAATATATAATTTTTTTG
```

Adk30

```
TAAAGAAAAGATTTAGGGTTTTGTGGTTAATATATAATTAATATATAATTTTTTTG
```

Adk31

```
TAAAGAAAAGATTTAGGGTTTTGTGGTTAATATATAATTAATATATAATTTTTTTG
```

Adk32

```
TAAAGAAAAGATTTAGGGTTTTGTGGTTAATATATAATTAATATATAATTTTTTTG
```

CaLCuV from *Arabidopsis* adk2 mutant plants.
CaLCuV conversion control. Plasmid DNA containing a CaLCuV genome insert was added to excess total cellular DNA extract from mock inoculated Arabidopsis.
CaLCuV from *N. benthamiana*.

**Primer set #2**

CaLCuV135CF: TTTGGGGAATATTAGGGGTAAAA
CaLCuV314CR: TCACATATCTTTATATATAACCCAC

```
CaLCuV from N. benthamiana.  
Primer set #2  
CaLCuV135CF:  TTTGGGGAATATTAGGGGTAAAA  
CaLCuV314CR:  TCACATATCTTTATATATAACCCAC  
Meth(N/12)  87  3        5        3   1      3 0      00          0

<table>
<thead>
<tr>
<th></th>
<th>NBC1</th>
<th>NBC2</th>
<th>NBC3</th>
<th>NBC4</th>
<th>NBC5</th>
<th>NBC6</th>
<th>NBC7</th>
<th>NBC8</th>
<th>NBC9</th>
<th>NBC10</th>
<th>NBC11</th>
<th>NBC12</th>
<th>NBC13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meth(N/12)</td>
<td>00 0 0 0 65 73 73 73 32 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>NBC1</th>
<th>NBC2</th>
<th>NBC3</th>
<th>NBC4</th>
<th>NBC5</th>
<th>NBC6</th>
<th>NBC7</th>
<th>NBC8</th>
<th>NBC9</th>
<th>NBC10</th>
<th>NBC11</th>
<th>NBC12</th>
<th>NBC13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meth(N/12)</td>
<td>6 1 49 1 5 6 7 6 72</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>NBC1</th>
<th>NBC2</th>
<th>NBC3</th>
<th>NBC4</th>
<th>NBC5</th>
<th>NBC6</th>
<th>NBC7</th>
<th>NBC8</th>
<th>NBC9</th>
<th>NBC10</th>
<th>NBC11</th>
<th>NBC12</th>
<th>NBC13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meth(N/12)</td>
<td>87 3 5 3 5 3 3 0 00 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

---

210
**CaLCuV conversion control #2.** Plasmid DNA containing a CaLCuV genome insert was methylated in vitro with CpG methylase, and added to excess total cellular DNA extract from mock inoculated *Arabidopsis*. In the top reference sequence, CG sites are indicated in red. Those sites modified by CpG methylase treatment were unconverted and were read as C (yellow). CNG sites, highlighted in blue on the reference sequence, were converted and read as T (green). CHH sites, in pink in the reference sequence, were nearly all converted as well.

**Primer set #2**

**CaLCuV135CF:** TTTGGGGAATATTAGGGGTAAAAA  
**CaLCuV314CR:** TCACATATCCTTATATAAAAAACCAC

<table>
<thead>
<tr>
<th>Primer Set</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>MpCaLCuV</td>
<td>CTTCTGGGAAACAGAAGGGGAAGACCGCATATGACAATATTAGGGATGTTGTGGGTTTATATAAAGATATGTG</td>
</tr>
<tr>
<td>MpCa11</td>
<td>xxxxtGCGGGAAATAGGSGAAAAGGGGAAGATCCCAAATATATTACGAATGGGCGG</td>
</tr>
<tr>
<td>MpCa12</td>
<td>xxxxtGCGGGAAATAGGSGAAAAGGGGAAGATCCCAAATATATTACGAATGGGCGG</td>
</tr>
<tr>
<td>MpCa13</td>
<td>xxxxtGCGGGAAATAGGSGAAAAGGGGAAGATCCCAAATATATTACGAATGGGCGG</td>
</tr>
<tr>
<td>MpCa14</td>
<td>xxxxtGCGGGAAATAGGSGAAAAGGGGAAGATCCCAAATATATTACGAATGGGCGG</td>
</tr>
<tr>
<td>MpCa15</td>
<td>xxxxtGCGGGAAATAGGSGAAAAGGGGAAGATCCCAAATATATTACGAATGGGCGG</td>
</tr>
<tr>
<td>MpCa16</td>
<td>xxxxtGCGGGAAATAGGSGAAAAGGGGAAGATCCCAAATATATTACGAATGGGCGG</td>
</tr>
<tr>
<td>MpCa17</td>
<td>xxxxtGCGGGAAATAGGSGAAAAGGGGAAGATCCCAAATATATTACGAATGGGCGG</td>
</tr>
<tr>
<td>MpCaLCuV</td>
<td>AnnnnnnGTCGTTCTGAGGGACACCTTAAATGAAATCTAAATGATCAGG</td>
</tr>
<tr>
<td>MpCa11</td>
<td>ATTTTTTTCTGCTGTTGGGTGAAAAGATTAATGAAATCTAAATGATCAGG</td>
</tr>
<tr>
<td>Mpca12</td>
<td>ATTTTTTTCTGCTGTTGGGTGAAAAGATTAATGAAATCTAAATGATCAGG</td>
</tr>
<tr>
<td>Mpca13</td>
<td>ATTTTTTTCTGCTGTTGGGTGAAAAGATTAATGAAATCTAAATGATCAGG</td>
</tr>
<tr>
<td>Mpca14</td>
<td>ATTTTTTTCTGCTGTTGGGTGAAAAGATTAATGAAATCTAAATGATCAGG</td>
</tr>
<tr>
<td>Mpca15</td>
<td>ATTTTTTTCTGCTGTTGGGTGAAAAGATTAATGAAATCTAAATGATCAGG</td>
</tr>
<tr>
<td>Mpca16</td>
<td>ATTTTTTTCTGCTGTTGGGTGAAAAGATTAATGAAATCTAAATGATCAGG</td>
</tr>
<tr>
<td>Mpca17</td>
<td>ATTTTTTTCTGCTGTTGGGTGAAAAGATTAATGAAATCTAAATGATCAGG</td>
</tr>
<tr>
<td>MpCaLCuV</td>
<td>CTTGACAAGGTAGCTATTAAGGGTAAGTGGCCCTATTAAAAAGACATG</td>
</tr>
<tr>
<td>mpc11</td>
<td>TTTGAAAGTTAGCTATTAAGGGTAAGTGGCCCTATTAAAAAGACATG</td>
</tr>
<tr>
<td>mpc12</td>
<td>TTTGAAAGTTAGCTATTAAGGGTAAGTGGCCCTATTAAAAAGACATG</td>
</tr>
<tr>
<td>mpc13</td>
<td>TTTGAAAGTTAGCTATTAAGGGTAAGTGGCCCTATTAAAAAGACATG</td>
</tr>
<tr>
<td>mpc14</td>
<td>TTTGAAAGTTAGCTATTAAGGGTAAGTGGCCCTATTAAAAAGACATG</td>
</tr>
<tr>
<td>mpc15</td>
<td>TTTGAAAGTTAGCTATTAAGGGTAAGTGGCCCTATTAAAAAGACATG</td>
</tr>
<tr>
<td>mpc16</td>
<td>TTTGAAAGTTAGCTATTAAGGGTAAGTGGCCCTATTAAAAAGACATG</td>
</tr>
<tr>
<td>mpc17</td>
<td>TTTGAAAGTTAGCTATTAAGGGTAAGTGGCCCTATTAAAAAGACATG</td>
</tr>
</tbody>
</table>
Appendix B: Arabidopsis double-stranded RNA binding protein

DRB3 partners with DICER-LIKE 3 and ARGONAUTE 4 in methylation-mediated defense against geminiviruses
Fig. B.1. Arabidopsis *drb3* and *dcl3* mutants do not recover from infection with *BCTV L2* virus. (A) Secondary shoots of wild-type (Col-0), *drb2, drb3, drb4*, and *drb5* plants infected with *BCTV L2* are shown. (B) Secondary shoots of wild-type (Col-0), *dcl2, dcl3*, and *dcl4* plants infected with *BCTV L2* are shown. The wild-type (Col-0) plants recover from infection, and present asymptomatic secondary shoots. The *drb3* and *dcl3* mutants fail to recover and secondary shoots continue to display symptoms of virus infection. The *drb2, drb4, drb5, dcl2*, and *dcl4* plants recover and secondary shoots show little evidence of disease symptoms. However, although secondary shoots lacked floral deformation, secondary shoots of infected *drb4* and *drb5* plants were sometimes stunted. Examples of stunted *drb4* and *drb5* plants are shown in this figure.
Fig. B.2. Cytosine methylation analysis of the BCTV L2′ genome in wild-type, ago4, drb3, drb4, dcl3 and dcl4 mutant plants. Dot plots depict cytosine methylation profiles of the viral intergenic region (IR). The 44 cytosines in the IR were analyzed by bisulfite sequencing and 12 different clones for each treatment are shown. Each filled circle represents a methylated cytosine, while open circles represent unmethylated cytosines. The dots are color-coded based on sequence context: CG (red), CNG (blue), CHH (green). (A) Wild-type Ler-0 plants recover from infection and the IR is hypermethylated (77% of total cytosines methylated). (B) Methylation is substantially reduced (to ~38%) in non-recovered ago4 plants. The ago4 mutation is in the Ler-0 ecotype background. (C) Wild-type Col-0 plants recover from infection and the IR is hypermethylated (63% of total cytosines methylated). The dot plots reveal that a majority of individual clones representing different genomes are densely methylated. (D) The drb3 plants fail to recover and are unable to accomplish hypermethylation of the viral genome. Total methylation in the IR is reduced to ~34%, and a larger proportion of viral genomes are unmethylated or hypomethylated. Methylation is reduced in all sequence contexts. (E) Viral genomes isolated from recovered drb4 plants show some reduction in total methylation relative to Col-0 (54% compared to 63%), but most genomes are densely methylated. (F) Methylation is substantially reduced (to 48%) in non-recovered dcl3 plants. (G) The dcl4 plants recover and methylation is similar to wild-type Col-0 (~63%). A majority of the viral genomes are hypermethylated.
Fig. B.2 Continued

C. Cytosine Methylation Profiles

BCTV L2- from wild-type (Col-0) plants (Recovered)

D. BCTV L2- from dbr3 mutant plants (Non-recovered)

E. BCTV L2- from dbr4 mutant plants (Recovered)

F. BCTV L2- from dcl3 mutant plants (Non-recovered)

G. BCTV L2- from dcl4 mutant plants (Recovered)
**Fig. B.3.** Interactions between selected DRB, DCL3 and AGO proteins using BiFC analysis in *N. benthamiana* epidermal cells. Constructs expressing the indicated DRB, DCL, and AGO proteins were fused to the N- or C-terminal portion of YFP and fusion proteins were co-expressed in *N. benthamiana* leaves. Constructs were delivered by agroinfiltration, and cells were photographed 36 h post-infiltration using a confocal laser scanning microscope. RFP-histone 2B and RFP-fibrillarin were used as markers for the nucleus and nucleolus, respectively. Protein combinations in each infiltration are indicated above each photograph, and 3 panels are shown for each combination at lower (20X) magnification and higher (100X) magnification. No interaction was observed between DRB3 and DCL4, or between DCL4 and AGO4, thus these are not presented. 

(A) DRB3 interacts with DCL3 exclusively in the nuclei of plant cells. The YFP signal representing a positive BiFC interaction colocalizes with the nuclear marker RFP-H2B. 

(B) Nuclei seen at 100X magnification reveals that the complexes are concentrated in subnuclear bodies that sometimes colocalize with the nucleolar marker, RFP-fibrillarin. We speculate that DRB3:DCL3 complexes are enriched in Cajal bodies, which have been reported to contain siRNA, DCL3 and AGO4 (13, 14, 15). 

(C) DRB4 also interacts with DCL3 exclusively in plant cell nuclei. 

(D) DRB4:DCL3 complexes are distributed throughout the nucleus, including nucleoli. 

(E) DRB4:DCL4 interactions occur in the cytoplasm and nucleus. 

(F) Within nuclei, DRB4 and DCL4 interact in punctate subnuclear bodies, some of which colocalize with the nucleolar marker RFP-fibrillarin. 

(G) DRB3 interacts with DRB4 in the nucleus and cytoplasm. 

(H) Upon closer inspection, DRB3:DRB4 complexes accumulate throughout the nucleus with the
exception of nucleoli, and appear to be concentrated in small spots that resemble the sites
of DRB3:AGO4 interaction (see L). (I) DCL3 interacts with AGO4 in nuclei. (J) DCL3:AGO4 interaction occurs throughout the nucleus, with the exception of nucleoli. (K) DRB3:AGO4 interactions occur in the nucleus and the cytoplasm. (L) DRB3:AGO4 complexes are enriched in small bodies that do not associate with nucleoli. This pattern is consistent with the observation that AGO4 localizes to subnuclear AB-bodies, along with Pol V and DRM2 (15). In addition, AGO4 has also been reported to exist in Cajal bodies, but we note that the DRB3:AGO4 complexes are concentrated at sites dissimilar from those observed with DRB3:DCL3 and DRB4:DCL4 (compare with B and F). (M) DRB4 interacts with AGO4 in nuclei. (N) DRB4:AGO4 complexes are found throughout the nucleus, except nucleoli.
Figure B.3

A  
(20x)  
YFP BiFC interaction  

RFP-H2B  

Merge  

B  
(100x)  
YFP BiFC interaction  

RFP-Fibrillarin  

Merge
Table B.1 Interactions of DRB and DCL proteins in the Yeast Two-hybrid System.

<table>
<thead>
<tr>
<th>Bait (pAS)</th>
<th>Prey (pACT)</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRB3</td>
<td>DCL4</td>
<td>++</td>
</tr>
<tr>
<td>DCL4</td>
<td>DRB3</td>
<td>+</td>
</tr>
<tr>
<td>DRB3</td>
<td>DRB3</td>
<td>-</td>
</tr>
<tr>
<td>DRB4</td>
<td>DRB4</td>
<td>++</td>
</tr>
<tr>
<td>DRB3</td>
<td>DRB4</td>
<td>+++</td>
</tr>
<tr>
<td>DRB4</td>
<td>DRB3</td>
<td>++</td>
</tr>
<tr>
<td>DCL4</td>
<td>DCL4</td>
<td>++</td>
</tr>
</tbody>
</table>

Continued

Table B.1 The bait proteins were expressed as GAL4-DNA binding domain fusions, and prey proteins as GAL4 activation domain fusions in yeast pJ694A cells. Growth on –Leu-Trp media indicated maintenance of bait and prey plasmids. Strong interaction was indicated by the ability of cells to grow on media lacking adenine, and weaker interactions by the ability of cells to grow on media lacking histidine supplemented with 45 mM 3-aminotriazole. Interaction symbols are as follows: - no interaction, + weak interaction, suboptimal growth on -His plates and no growth on -Ade plates, ++ good interaction, growth on -His plates and no growth on -Ade plates, +++ strong interaction, growth on -His-Ade plates.

Yeast two-hybrid analysis was carried out to verify interactions involving DRB3, DRB4 and DCL4. The results verified the following interactions: DRB3:DRB4, DRB4:DCL4,
DRB4:DRB4, and detected DRB3:DCL4 and DCL4:DCL4. DRB3:DCL4 and DCL4:DCL4 interactions were not detected *in planta* using BiFC.

Interactions involving DCL3 could not be detected using yeast two-hybrid since this protein proved recalcitrant to expression in yeast, and no transformants could be obtained. A shorter version of DCL3 containing RNaseIII and dsRBM domains only was constructed and tested in yeast. This truncated version of DCL3 was not found to interact in yeast with any DRB or DCL, suggesting that the dsRBM alone is not sufficient for interaction.

**Table B.1 Continued**

<table>
<thead>
<tr>
<th>Bait (pAS)</th>
<th>Prey (pACT)</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positive controls:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DRB4</td>
<td>DCL4</td>
<td>++</td>
</tr>
<tr>
<td>DCL4</td>
<td>DRB4</td>
<td>++</td>
</tr>
<tr>
<td>DRB4</td>
<td>DRB5</td>
<td>++</td>
</tr>
<tr>
<td>DRB5</td>
<td>DRB4</td>
<td>++</td>
</tr>
<tr>
<td><strong>Negative controls:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DCL4</td>
<td>DRB5</td>
<td>-</td>
</tr>
<tr>
<td>DRB5</td>
<td>DCL4</td>
<td>-</td>
</tr>
<tr>
<td>DRB3</td>
<td>DRB5</td>
<td>-</td>
</tr>
<tr>
<td>DRB5</td>
<td>DRB3</td>
<td>-</td>
</tr>
</tbody>
</table>
References for Appendix B:


Appendix C: Designing Primers for Bisulfite Analysis
In mammalian systems, since most methylation is confined to CG sites, primers for bisulfite analysis are typically situated in regions that lack CG sites, flanking the CG sites to be assayed for methylation. Further, primers are selected assuming conversion to T at all non-symmetric Cs, which also serves as an internal control for complete bisulfite conversion.

In plants, methylation can occur at any cytosine, in any context. Since we cannot eliminate all cytosines in primer design, we may minimize but cannot eliminate bias. Results have to be read within the context of each specific primer pair selected.

In our experiments, we used the following rules: Cs are minimized in the forward primer and Gs in the reverse primer, and where used, Cs (in the forward primer) and Gs (in the reverse primer) are excluded from the 3’ end (within 5 bases). For primer length of 18-21 nt, we allowed a maximum of 3 Cs or Gs. Product length of up to 500 bp was chosen (300 bp optimal). Primers were selected assuming conversion at Cs in template sequence. Degenerate bases were not used in the primers to maintain strand specificity of the methylation profile.

Since the publication of *Virus Genome Methylation as an Epigenetic Defense* (Chapter 2), a tool called Kismeth has been developed, which uses the same rules. This program further recommends use of degenerate bases in the final primer, Y (C/T) for forward primer, and R (A/G) for reverse primer.

References for Appendix D:

Appendix D: CpG islands in geminivirus genomes
**Tomato golden mosaic virus (TGMV) DNA A**


**Cabbage leaf curl virus (CaLCuV) DNA A**

Four CpG islands predicted – Island 1 (109 bp): (start-end position) 314-422 (AL1, CP), Island 2 (102 bp): 834-935 (AL1, CP), Island 3 (301 bp): 1270-1570 (AL1, AL2), Island 4: 1925-2089 (AL3).

**African cassava mosaic virus (ACMV) DNA A**


**Beet curly top virus (BCTV) Logan**

One CpG island predicted (150 bp): (start-end position) 2536-2685 (L1, L4).