

**Diversification of the SUVH6 Clade of Histone 3 Lysine 9 Methyltransferases
in Flowering Plants**

By

Jenna Elizabeth Kotak

B.S., Muhlenberg College, 2013

M.A., Brown University, 2017

Thesis

Submitted in partial fulfillment of the requirements for the degree of Doctor of
Philosophy in the
Division of Biology of Medicine at Brown University

PROVIDENCE, RHODE ISLAND
October 2020

© Copyright 2020 by Jenna E. Kotak

This dissertation by Jenna E. Kotak is accepted in its present form by the Department of Molecular Biology, Cell Biology and Biochemistry as satisfying the dissertation requirement for the degree of Doctor of Philosophy.

Date: _____

Dr. Judith Bender, Advisor

Recommended to the Graduate Council

Date: _____

Dr. Richard Freiman, Reader

Date: _____

Dr. Mark Johnson, Reader

Date: _____

Dr. Erica Larschan, Reader

Date: _____

Dr. Mary Gehring, Reader

Approved by the Graduate Council

Date: _____

Dr. Andrew Campbell,
Dean of the Graduate School

Curriculum Vitae
Jenna Kotak
jenna_kotak@brown.edu

EDUCATION

Doctor of Philosophy	Expected 10/2020
Molecular Biology, Cell Biology and Biochemistry	
Brown University, Providence, RI	
Thesis: “Diversification of the SUVH6 Clade of Histone 3 Lysine 9 Methyltransferases in Flowering Plants”	
Master of Arts	
Molecular Biology, Cell Biology and Biochemistry	05/2017
Brown University, Providence, RI	
Bachelor of Science	
Biochemistry, Mathematics with Honors	05/2013
Muhlenberg College, Allentown, Pennsylvania	

RESEARCH EXPERIENCE

Graduate Thesis	2014 – Present
Brown University, Providence, RI	
Advisor: Dr. Judith Bender	
Undergraduate Research	2011- 2013
Muhlenberg College, Allentown, PA	
Advisors: Dr. Amy Hark, Dr. Elizabeth McCain	

PUBLICATIONS

Kotak, J., Saisana, M., Gegas, V., Pechlivani, N., Kaldis, A., Papoutsoglou, P., Makris, A., Burns, J., Kendig, A. L., Sheikh, M., Kuschner, C. E., Whitney, G., Caiola, H., Doonan, J. H., Vlachonasios, K. E., McCain, E. R., & Hark, A. T. (2018). The histone acetyltransferase GCN5 and the transcriptional coactivator ADA2b affect leaf development and trichome morphogenesis in *Arabidopsis*. *Planta*, 248(3), 613–628.
<https://doi.org/10.1007/s00425-018-2923-9>

MICRO-PUBLICATIONS

Kotak, J., Kendig, A., Cann, K., Shaffer, J., Hark, A., & McCain, E. (2019). Disruption of the histone acetyltransferase GCN5 and the transcriptional coactivator ADA2b affect trichome density in *Arabidopsis thaliana*. *microPublication biology*, 2019,

PRESENTATIONS

Jenna Kotak, Judith Bender. Identifying the role of a conserved motif in *Arabidopsis* histone methyltransferase SUVH6 activity and specificity. Brown University Department Retreat Poster. 2018.

Jenna Kotak, Judith Bender. An amino-terminal motif necessary for H3K9 methyltransferase function. Brown University Department Data Seminar Series. 2018.

Jenna Kotak, Judith Bender. Investigating the role of a conserved motif in *Arabidopsis thaliana* SUVH5 and SUVH6 histone 3 lysine 9 methyltransferases. Cold Spring Harbor Conference on Chromatin and Epigenetics Poster. 2016.

Jenna Kotak, Judith Bender. Investigating the role of a novel regulator of an *Arabidopsis* histone methyltransferase in transcriptional silencing. Brown University Department Retreat Poster. 2016.

Jenna Kotak, Judith Bender. Pointing the way: a zinc-finger regulator of transcriptional silencing. Brown University Department Data Seminar Series. 2016.

Jenna Kotak, Judith Bender. Target acquired: Discovering how chromatin silencer proteins achieve locus specificity. Brown University Department Retreat Poster. 2015.

Jenna Kotak, Judith Bender. Identifying chromatin silencer interactions through a yeast two-hybrid screen. Brown University Department Retreat Poster. 2014.

Jenna Kotak, Elizabeth McCain, Amy Hark. The Effects of *gcn5* Mutations on Trichome Branching. Poster at the 15th Annual Undergraduate Research Symposium in the Chemical and Biological Sciences at the University of Maryland, Baltimore County. 2012.

AWARDS AND HONORS

MCB Department Retreat 1st Place Graduate Student Poster Award, Brown University (2018)

The Muhlenberg College Dean's Grant (2012)

Robert S. Ochner Award Winner, Muhlenberg College (2012)

Pi Mu Epsilon Math Honor Society Member (Spring 2012- Spring 2013)

James Vaughan Research Award Winner, Muhlenberg College (2011)

Level III Certified Tutor, College Reading and Learning Association (2011)

Dana Scholar Honors Program, Muhlenberg College (2009-2013)

COMMITTEES AND SERVICE

Student Member of MCBGP Admissions Committee, Brown University
(2015–2016)

Mentor, MCBGP First Year Mentoring Program, Brown University (2016-2018)

Organizer, MCBGP First Year mentoring program, Brown University
(2016-2017)

Logistics Coordinator, Graduate Women in Science and Education, Brown
University (2015-2017)

Student Liaison, Search Committee for Organic Chemistry Professor, Muhlenberg
College (2012)

TEACHING EXPERIENCE

Teaching Certificates:

Certificate IV Course Consultant, Sheridan Teaching Center, Brown
University, 2015-2016

Certificate I Discussion Leader, Sheridan Teaching Center, Brown University,
Fall 2016

Completed Certificate IV Course, Sheridan Teaching Center, Brown
University, Fall 2015 – Spring 2016

Completed Certificate I Course, Sheridan Teaching Center, Brown University,
Fall 2014 – Spring 2015

Course Instruction:

Teaching Assistant, Genetics, Brown University, Fall 2014

Teaching Assistant, HHMI Summer Research Course “Genome Explorers”,
Brown University, Summer 2014

Learning Assistant, Biology III, Muhlenberg College, Fall 2011, Fall 2012

Peer Tutor, Muhlenberg College, 2010-2013

Acknowledgements:

I give many thanks to all of the people within Brown MCB who helped me push my science forward. First, I thank Judith for scientific guidance and the many drafts she helped me with in not only crafting my thesis, but also in posters and presentations throughout my time in graduate school. I thank all of the members of my committee for their input and support throughout graduate school. The Plant Group at Brown has not only shared protocols and equipment but has been a source of knowledge about *Arabidopsis* techniques that were instrumental. I also extend many thanks to Elaine Butler, Ashley Walker, and Ray Windsor who all kept MCB running and took care of so many program activities behind the scenes. Additional thanks to Geoff Williams for teaching me to use the microscopes and answering my questions.

Many friends and family members have given me such support throughout this process. Friends I made at graduate school helped commiserate on bad days and I've loved sharing our milestones together. My crazy extended family always keeps me laughing, and my grandma especially has constantly told me how proud she is. My brothers Tom and James have been such a source of joy for me. They never fail to make me laugh with the littlest things, and a phone call with either of them can make my day.

I can't say enough about how wonderful my partner Bobby has been throughout this process. Not only did he spend several afternoons at the microscope giving his expert help, but he has always helped me put things into

perspective and has been such a patient and giving person. He shares in my joys and goes out of his way to make me smile when I need it.

My deepest appreciation goes out to my parents. They were my first science teachers in teaching us little kids about diabetes. My mom is such an endless source of love and support for me. Like she always has, she took care of me after my injury and I know her help allowed me to finish my degree. My dad instilled such a joy in education even when he was taking me to the library as a toddler. He's encouraged me to be ambitious and feel like I could do whatever I put my mind to. I wouldn't be here if it wasn't for the loving support and drive that both of my parents provided me.

Table of Contents

Title Page	i
Copyright Page	ii
Signature Page	iii
Curriculum Vitae	iv
Acknowledgements	vii
Table of Contents	ix
List of Tables and Figures	xii
 Thesis Abstract	 1
 Introduction	 2
Transposons threaten eukaryotic genome stability	3
H3K9 Methyltransferase Enzymes	5
5meC and H3K9me	8
Plant H3K9 methyltransferases	12
SUVH Evolution	14
SUVH4 function	19
SUVH4 and CMT2/CMT3 act in an amplifying loop	19
Interacting factors with SUVH4	22
H3K9me-binding proteins	23
ADCP1	23
IBM1 and Jumonji-C Proteins	25
SUVH6 clade proteins	27
Locus-specific targeting of SUVH5 and SUVH6	27
Biochemical Analysis of the SRA and SET domains	31
Summary	34
Figures	35
References	47

Chapter 2: An amino-terminal conserved motif controls <i>Arabidopsis</i> SUVH6 H3 Lysine 9 methyltransferase function through two distinct mechanisms	58
Abstract	59
Introduction	59
Results	63
Discussion	73
Methods	80
Figures	86
References	113
Chapter 3: The <i>Arabidopsis</i> SUVH5 and SUVH6 histone 3 lysine 9 methyltransferase amino-terminal extensions contribute to locus-specificity	117
Abstract	118
Introduction	119
Results	123
Discussion	132
Methods	138
Figures	143
References	158
Chapter 4: Conclusions and Perspectives	160
Evolution of the SUVH6 clade	161
The ACM encodes a nuclear localization signal	163
The ACM has a second role as a positive regulator of SUVH6 activity at a preferred target	166
Amino-terminal extensions are a conserved features of the SUVH6 clade	167
The SUVH6 clade amino-terminal extension contributes to locus-specific activity	170
Future characterization of preferred loci	173
SUVH6 clade proteins in other plants	175

Broader Perspectives	175
Figure	177
References	178

List of Tables and Figures

Chapter 1: Introduction

Figure 1: H3K9 methyltransferases have diversified interaction domains in multicellular organisms	35
Figure 2: Moss SUVH proteins likely result from an early retrotransposition of <i>SUVH4</i>	37
Figure 3: SUVH clades diversified after the evolution of lycophytic mosses	39
Figure 4: <i>Arabidopsis thaliana</i> expression analysis of <i>SUVH4</i> , <i>SUVH5</i> , and <i>SUVH6</i> in different tissues shows SUVH specificities in reproductive tissues	41
Figure 5: Tissue-specific expression analysis of SUVH4 clade versus SUVH6 clade proteins in other flowering plants	43

Chapter 2: An amino-terminal conserved motif controls *Arabidopsis* SUVH6 H3 Lysine 9 methyltransferase function through two distinct mechanisms

Figure 1: The ACM is specific to the SUVH6 clade and is positioned within amino-terminal extensions	86
Figure 2: The ACM is required for SUVH6 activity at the <i>PAI</i> loci	88
Figure 3: SUVH6 transgenes only weakly remethylate 25S rDNA repeats	90
Figure 4: SUVH6 nuclear localization is mediated through the ACM <i>K51 R52</i> residues	92
Figure 5: SUVH6 amino-terminal mutations do not affect sub-nuclear localization to chromocenters	94
Supplemental Figure 1: Angiosperm SUVH6 clade proteins have diverged sequences in the amino-terminal extensions	96
Supplemental Figure 2: The ACM sequence occurs in some SUVH proteins encoded by non-angiosperm plants	99
Supplemental Figure 3: SUVH4 and SUVH6 clade protein amino-terminal extensions are predicted to be disordered	101
Supplemental Figure 4: Epitope tags do not impair SUVH6 remethylation of <i>PAII-PAI4</i>	102
Supplemental Figure 5: Transgenic lines used in this analysis have similar or higher <i>SUVH6</i> transcript levels than endogenous <i>SUVH6</i> in a wild type background	104

Supplemental Figure 6: The <i>SUVH6</i> LDR deletion removes sequences predicted to be unstructured	106
Supplemental Figure 7: The SUVH6 ACM is sufficient for nuclear localization of GFP	108
Supplemental Figure 8: GFP-SUVH6 transgenes localize to mitotic chromosomes	109
Supplemental Figure 9: Summary of SUVH6 transgene add-back to 456 background results at loci tested in this chapter	110
Supplemental Table 1: Database IDs for named SUVH proteins in this study	111

Chapter 3: The *Arabidopsis* SUVH5 and SUVH6 histone 3 lysine 9 methyltransferase amino-terminal extensions contribute to locus-specificity

Figure 1: <i>SUVH</i> transgene constructs used to test SUVH6 clade amino-terminal extensions	143
Figure 2: SUVH5 transgenes are expressed at higher levels from the <i>SUVH6</i> promoter than the <i>SUVH5</i> promoter	144
Figure 3: <i>SUVH5</i> transgenes remethylate both <i>PAI</i> and 25S rDNA targets	145
Figure 4: <i>SUVH4</i> transgenes remethylate both <i>PAI</i> and 25S rDNA targets	147
Figure 5: The SUVH6 amino terminus restricts SUVH5 carboxy-terminal domain activity at the 25S rDNA repeats	149
Figure 6: The SUVH6 amino terminus does not restrict SUVH4 carboxy-terminal domain activity at <i>PAI2</i> or the 25S rDNA repeats	151
Figure 7: The SUVH5 amino terminus enhances SUVH6 carboxy-terminal domain activity at the 25S rDNA repeats	153
Figure 8: GFP-tagged SUVH5 is nuclear localized with enrichment in puncta	155
Supplemental Figure 1: Summary of transgene add-back phenotypes at the loci assayed in this chapter	157

Chapter 4: Conclusions and Perspectives

Figure 1: ACM consensus sequences resemble those of the SUVH4 consensus pre-SRA second helix	170
--	-----

Abstract of “Diversification of the SUVH6 Clade of Histone 3 Lysine 9 Methyltransferases in Flowering Plants”

by Jenna Kotak, Ph.D., Brown University, October 2010

Transposons are parasitic DNA sequences that pose threats to genome stability across a range of eukaryotes. To defend against transposition into essential genes and destabilizing transposon proliferation, eukaryotes regulate transposon transcription. Gene regulation is often achieved via chromatin modifications of histone proteins and DNA. Through methylation of histone 3 lysine 9 (H3K9), the eukaryotic cell can silence the transcription of transposons as a defense system. This fundamental H3K9 methylation is mediated by H3K9 methyltransferase enzymes that are responsible for efficient and precise modification of silenced regions. While various kingdoms use the same catalytic H3K9 methyltransferase domain, plants have evolved a novel H3K9 methyltransferase class of SUVH proteins. However, why the single SUVH protein present in green algae has expanded and diversified into a large family of proteins in the flowering plant *Arabidopsis thaliana* is unclear. In this work, I identify unique features of the SUVH6 clade of H3K9 methyltransferases and test the functionalities of these features through a genetic approach. I demonstrate that SUVH6 clade proteins contain longer, likely disordered amino-terminal extensions with a novel motif that function in nuclear localization and activity at preferred loci. My results provide insights into the diversification of chromatin-modifying enzymes and how these enzymes can gain new functions over evolution.

Chapter 1: Introduction

Written by JK with edits from Dr. Judith Bender

Transposons threaten eukaryotic genome stability

Transposons threaten eukaryotic cells by using cellular resources to replicate transposon sequences and create foreign transcripts and proteins. Left unchecked, transposable elements will propagate throughout the eukaryotic genome. RNA and DNA transposons encode transposase enzymes that allow the integration of a given sequence, or copied sequences, into the host genome via “copy and paste” or “cut and paste” mechanisms. McClintock’s discovery of mobile DNA described the deleterious effects of transposable elements on maize: she observed chromosome breaks and loss of gene functions upon transposition into new genomic positions (McClintock, 1950). Since this landmark discovery, the effects of transposition have been described across eukaryotic species.

Plant genomes in particular can have large proportions of transposons. In maize, transposons and transposon derived repeats occupy nearly 85% of the genome. There is wide variability in transposon composition across crop species, but estimates have determined the average crop genome is 50% transposable element derived (Vitte et al., 2014). Additionally, several environmental stressors can result in increased transposon expression, including heat (Tittel-Elmer et al., 2010, Sun et al., 2020). Understanding the biology of plant defenses to transposable elements is increasingly important as environmental stressors change and biologists engineer genetically modified organisms in new crops. In the model plant *Arabidopsis thaliana*, transposons only comprise about 20% of the genome (Vitte et al., 2014). However, by studying the eukaryotic defenses to

transposon expression in *Arabidopsis*, we can gain insights into plant genome defenses that can be applied to crops and other eukaryotes.

To defend the genome from transposon expression and subsequent propagation, eukaryotes have evolved mechanisms to silence the deleterious expression of transposable elements. Chromatin modifications are an essential mechanism by which eukaryotes regulate not only their own genomes, but also transposable elements. Covalent modifications to histone proteins and DNA act as a pattern of chemical modifications to promote or reduce expression of a sequence, often through binding of proteins to modified histones or DNA. By targeting heterochromatic histone modifications and DNA methylation to transposable element sequences, eukaryotes can diminish transposable element transcription (Miura et al., 2001, Zilberman et al., 2007). While repressive heterochromatin reduces transcription of genic sequences that can be activated in certain conditions, heterochromatin in transcriptional silencing of transposons and repeats operates through separate mechanisms (Zhang et al., 2007). A variety of histone modifications contribute to repressive heterochromatin, but at the core of silenced heterochromatin is the conserved methylation of histone 3 lysine 9 (H3K9).

Histone proteins form an octamer that DNA is wrapped around twice and allow for compaction and organization of the genome within the nucleus (Luger et al., 1997). Histone three has several lysine residues that can be post-translationally modified. In particular, H3K9 can be acetylated as a euchromatic modification or can be heterochromatically modified if methylated. Lysine

accepts three methyl groups, allowing for H3K9me1, H3K9me2, and H3K9me3 modifications, which can have individual chromatin states in different species (Krauss, 2008). Often interacting proteins can distinguish between binding H3K9me1 versus H3K9me3, demonstrating the diversity of regulatory functions allowed by even one lysine on one histone protein.

H3K9 Methyltransferase Enzymes

Eukaryotes have evolved H3K9 methyltransferase enzymes to provide a fundamental mechanism of transcriptional silencing. H3K9 methyltransferases were first identified in a *Drosophila melanogaster* screen for reactivation of silenced eye color genes (Tschierch et al., 1994). The Suppressor of Variation 3-9 (Su(var)3-9) protein was subsequently biochemically characterized as a methyltransferase of H3K9 substrates that functioned in transcriptional silencing (Schotta et al., 2002, Eskeland et al., 2004). Su(var)3-9 homologs were later identified and characterized in fungi, mammals, and plants. Although H3K9 methylation is present across an array of eukaryotes, the function and presence of H3K9me1, H3K9me2, and H3K9me3 varies across organisms (Krauss, 2008). In mammals, H3K9me3 is the predominant transcriptional silencing modification, while plants utilize H3K9me1 and H3K9me2 for transcriptional silencing.

At the core of all H3K9 methyltransferase proteins is the SET domain: a histone methyltransferase catalytic domain that can methylate lysines in various positions on different histone substrates. The SET domain binds to S-adenosyl-

methionine, which serves as the methyl-donor for the histone substrate (Min et al., 2002). While SET domains are found throughout methyltransferases for a variety of histone substrates, proteins can specifically methylate H3K9 substrates if the SET domain is paired with flanking pre-SET and post-SET domains. Crystal structures of H3K9 methyltransferases of *Schizosaccharomyces pombe* Clr4 and *Neurospora crassa* Dim-5 showed that the pre-SET domain cysteine residues coordinate three cofactor zinc ions (Min et al., 2002, Zhang et al., 2002). The post-SET domains is a 16-amino acid domain carboxy-terminal to the SET domain and contains three invariant cysteines that also bind a zinc ion (Zhang et al., 2003). Without these post-SET cysteines, catalytic activity is abolished (Zhang et al., 2002).

Organisms such as Cnidarians, the protist *Dictyostelium*, and fungus *Neurospora crassa*, contain H3K9 methyltransferases with just the catalytic pre-SET, SET, and post-SET domains (Krauss et al., 2006) (Figure 1). However, most H3K9 methyltransferases have evolved to pair an additional domain with the catalytic domains (Figure 1). Typically, the additional domains serve as interaction domains to target H3K9 methyltransferases to the appropriate regions of the genome through chromatin interactions. The most widespread domain paired with the catalytic domains is the chromodomain which binds H3K9me (Krauss et al., 2006). Based on the *Drosophila melanogaster* Su(var)3-9 H3K9 methyltransferase that contains a chromodomain, this family of proteins are termed the Su(var)3-9 proteins. Su(var) 3-9 H3K9 methyltransferases with a chromodomain are present in animals such as vertebrates, insects, platyhelminths,

nematodes and in the fungus *Schizosaccharomyces pombe* (Krauss et al., 2006) (Figure 1). Su(var)3-9 methyltransferases are proposed to processively propagate H3K9me by binding to the previously catalyzed H3K9me modification via the chromodomain (Bannister et al., 2001). Other H3K9 methyltransferases use different domains to bind to H3K9me. For example, mammalian G9a binds to H3K9me through the ankyrin repeats and mammalian SETDB1 binds H3K9me through the Tudor domain (Collins et al., 2008, Jurkowska et al., 2017).

H3K9 methyltransferases have diversified in different ways across eukaryotic kingdoms. Single-celled fungi such as *Schizosaccharomyces pombe* and *Neurospora crassa* each only contain one type of H3K9 methyltransferase protein (Figure 1), but *Saccharomyces cerevisiae* does not encode any H3K9 methyltransferases (Krauss et al., 2006). In contrast, multicellular animals and plants have larger numbers of H3K9 methyltransferases with diversified structures and functions. In animals, diversification of H3K9 methyltransferases has led to specialization in monomethylation, dimethylation, or trimethylation in different chromatin regions. The G9a and GLP class of H3K9 methyltransferases catalyzes H3K9me1 and H3K9me2 in transcriptional repression in euchromatin (Tachibana et al., 2002). In contrast, heterochromatic H3K9 methylation is catalyzed by enzymes in the SETDB1 class and the SUV39H1 and SUV39H2 class. SETDB1 catalyzes H3K9me1, while SUV39H1 and SUV39H2 use H3K9me1 substrates to catalyze H3K9me2 and H3K9me3 (Rea et al., 2000, Lehnertz et al., 2003, Loyola et al., 2009).

H3K9 methyltransferases can also be guided to targets by binding to methylated DNA. In organisms such as mammals and plants, 5-methyl cytosine (5meC) also functions in heterochromatin and can be paired to H3K9me through various mechanisms. For example, the animal SETDB1 proteins contains a Methyl CpG-binding domain (Figure 1). In the plant SUVH proteins there is a 5meC-binding domain amino-terminal to the Pre-SET, SET, and Post-SET domains as well: the SET and Ring Associated (SRA) domain.

5meC and H3K9me

In mammals, 5meC and H3K9me function through interconnected pathways. CpG methylation catalyzed by DNA methyltransferase DNMT3 is promoted by the presence of SUV39H1 and SUV39H2-mediated H3K9me3 at pericentromeres (Lehnertz et al., 2003). Additionally, the mammalian SRA domain protein UHRF1 has enhanced, but not dependent 5meC-binding in the presence of H3K9me (Zhao et al., 2016). Mammalian 5meC largely occurs in the CpG context, although non-CpG methylation is present at very low levels in mammalian neuronal tissues and pluripotent cells (He and Ecker, 2015).

Plants use non-CpG methylation much more extensively than mammals with different pathways controlling CpG and non-CpG methylation. In *Arabidopsis*, CpG methylation is the predominant DNA methylation mark and can be found in both euchromatic and heterochromatic regions (Cokus et al., 2008). Conversely, non-CpG methylation is found mainly in heterochromatic

regions. Both types of 5meC are highest at pericentromeric regions, but CpG methylation is also found within gene body regions of some protein-coding genes (Cokus et al., 2008). Non-CpG methylation is further classified into CHG and CHH methylation, where H is A, C, or T. CHG methylation exhibits tight correlation to H3K9me2, with only 10% of sites showing CHG methylation without H3K9me2 (Bernatavichute et al., 2008). CHH methylation is also well correlated with H3K9me2, but a third of CHH methylation occurs without H3K9me2. In contrast, CpG methylation is not correlated with H3K9me2 (Bernatavichute et al., 2008).

To initiate both CpG and non-CpG methylation on targeted sequences, small interfering RNAs (siRNAs) from those sequences are transcribed and used to target proteins in a process called RNA-directed DNA Methylation (RdDM). At the core of this process are two plant-specific RNA Polymerases, RNA Polymerase IV and V (Herr et al., 2005, Wierzbicki et al., 2009). RNA Pol IV transcribes regions to be transcriptionally silenced and transcripts are then cleaved by a Dicer protein into 24-nucleotide siRNAs and loaded onto Argonaute proteins (Wierzbicki et al., 2009). RNA Polymerase V transcribes long RNAs that the siRNA-Argonaute complexes bind to in a sequence-specific manner (Law et al., 2010). The key enzyme, DNA methyltransferase DRM2, then arrives via binding to Argonaute proteins and is able to methylate cytosines in CpG, CHG, or CHH contexts (Zhong et al., 2014, Cao and Jacobsen, 2002).

Though initiation is required for newly inserted transposons and repeats, maintenance is required perform 5meC over cellular divisions. *Arabidopsis*

maintenance is performed through two genetically distinct pathways that breakdown along CpG and non-CpG methylation. CpG methylation is maintained by the MET1 DNA methyltransferase. MET1 is composed of a DNMT replication foci domain, two bromo and adjacent homology (BAH) domains, and a cytosine methyltransferase domain. BAH domains have been well documented in binding to H3K9 methylation, but this function has not been experimentally determined for the MET1. *met1* mutants display a drastic reduction in CpG methylation with little effect on CHH and CHG methylation (Kankel et al., 2003, Aufsatz et al., 2004). MET1 requires the SRA 5meC-binding domain proteins VIM1, VIM2, or VIM3 for activity. *met1* and *vim1/2/3* mutants target the same set of targets and show similar levels of gene expression changes, indicating that they likely function through the same pathway (Shook and Richards, 2014). VIM1, VIM2, and VIM3 are functionally redundant and each contain a PHD domain, two RING domains, and an SRA domain, with functional E3 ubiquitin ligase activity through the RING and PHD domains (Woo et al., 2007, Woo et al., 2008, Kraft et al., 2008). Through the binding of the VIM SRA domain to hemi-methylated CpG contexts, MET1 is recruited to establish 5meC on the unmethylated cytosines in the newly-synthesized strand of DNA. VIM proteins may indirectly or transiently interact with MET1, or VIM-mediated ubiquitination of histones may recruit MET1 (Kim et al., 2014).

Non-CpG methylation is mostly maintained via Chromomethylase (CMT) proteins. The *Arabidopsis* genome encodes three CMT proteins, with *CMT1* encoding a truncated protein and CMT2 and CMT3 encoding functional proteins.

CMT3 was identified first and catalyzes 5meC at CHG sites (Bartee et al., 2001, Cao & Jacobsen, 2002). Until recently it was thought that much of CHH methylation was due to constant *de novo* methylation via the RdDM pathway with some minor contribution from CMT3 (Cao et al., 2002). However, plants with mutations to CMT3 and RdDM cytosine methyltransferases, *cmt3 drm drm2*, maintain a significant amount of CHH methylation, indicating that another cytosine methyltransferase must contribute to CHH methylation (Zemach et al., 2013). CMT2 was identified as a catalytically active cytosine methyltransferase that specifically plays a role in CHH methylation maintenance, with a minor role in CHG methylation maintenance (Gouil and Baulcombe, 2016, Zemach et al., 2013). However, the RdDM pathway does maintain 5meC on a separate population of CHH targets than CMT2 does, indicating that CHH methylation is indeed catalyzed by both CMT2 and repeated RdDM (Stroud et al., 2014). Mutant analyses demonstrate that CMT2 is most active at longer transposons and heterochromatic regions, while DRM2 is most active at shorter transposons on chromosome arms (Zemach et al., 2013, Gouil and Baulcombe, 2016).

Like MET1, CMT2 and CMT3 contain a BAH domain and DNA methyltransferase domain, but also include a chromodomain (Du et al., 2012). Strikingly, mutation of either the BAH domain or the chromodomain is sufficient to abolish CMT3 function *in vivo*, but do not affect catalytic activity *in vitro*, indicating that CMT3 requires H3K9me binding for function (Du et al., 2012). Both CMT2 and CMT3 can bind to H3K9me1, but bind more strongly to H3K9me2 (Du et al., 2012, Stroud et al., 2014). *In vitro* cytosine methylation

studies show that CMT2 prefers unmethylated DNA substrates, while CMT3 prefers hemi-methylated DNA substrates (Stroud et al., 2014).

Plant H3K9 methyltransferases

Like animals, plants have diversified H3K9 methyltransferases into classes that differ in chromatin recognition domains and catalytic specificities. In plants, there are two broad classes of H3K9 methyltransferases that both contain the pre-SET, SET, and post-SET domains. These two classes are the Su(var)3-9 Homolog (SUVH) and SUVH Related (SUVR) proteins (Figure 1). In the model plant *Arabidopsis thaliana*, SUVH proteins have further evolved into four SUVH sub-classes. The expansion of H3K9 methyltransferase proteins in plants provides an opportunity to understand the diversification and functions of chromatin modifying enzymes.

The SUVR family is an ancient protein class that has existed since the evolution of green algae. Green algae *Chlamydomonas reinhardtii* contains a single SUVR protein, while flowering plant *Arabidopsis* encodes five SUVR proteins (Casas-Mollano et al., 2007). Though all five *Arabidopsis* SUVR proteins contain pre-SET and SET domains, only SUVR4 has demonstrated H3K9 methyltransferase abilities. SUVR proteins lack the 5meC-binding SRA domain that is present in SUVH proteins and diversified different domains to be paired to the pre-SET, SET, and post-SET domains. SUVR1, SUVR2, and SUVR4 have evolved a novel amino-terminal plant-specific WIYLD domain

(Figure 1, Thorstensen et al., 2006, Rahman et al., 2014). The WIYLD domain binds histone 2B mono-ubiquitin modifications (Veiseth et al., 2011), showing another example of alternative domain pairings with the pre-SET, SET, and post-SET domains.

SUVR1 and SUVR2 are related proteins that do not have methyltransferase activity, but do have transcriptional silencing functions in the initiation of 5meC (Thorstensen et al., 2006, Han et al., 2014). SUVR2 is capable of binding to H3 through the non-catalytic SET domain and binds to SNF2 chromatin remodelers, perhaps allowing for chromatin remodeler recruitment to promote transcriptional silencing via blocking transcriptional machinery or allowing binding of silencing factors (Han et al., 2014). SUVR5 also functions in transcriptional silencing, but likely not through catalyzing H3K9me. *suvr5* mutants demonstrate some H3K9me changes at transposons and pericentromeres, but do not demonstrate any catalysis *in vitro* (Caro et al., 2012).

SUVR4 is the only SUVR protein with demonstrated H3K9 methylation *in vitro*, but this methylation only occurs in the presence of free ubiquitin and on H3K9me1 and H3K9me2 substrates, not unmethylated H3K9 (Veiseth et al., 2011). The model of SUVR4 activity is that when ubiquitin is removed from histone 2, this enhances activity of SUVR4. While animal genomes show significant levels of H3K9me3, *Arabidopsis* H3K9me3 is debated, and if present then rare. Though SUVR4 has the ability to catalyze H3K9me2, it must contribute a very minor amount of H3K9me2 because mutation of other H3K9 methyltransferases in *suvh4 suvh5 suvh6* triple mutants show a loss of 98% of

H3K9me2 sites (Li et al., 2018). Additionally, because SUVH4 does not catalyze H3K9me1, its activity is dependent on the H3K9me1 of the SUVH clade proteins. SUVH proteins contribute to the overwhelming majority of H3K9me, so we have focused our study of *Arabidopsis* H3K9 methyltransferases on the SUVH clade.

SUVH Evolution

Like the SUVH proteins, SUVH proteins have diversified and expanded from a single gene in chlorophytic algae. Within *Arabidopsis*, there are 4 SUVH sub-clades: the SUVH4 clade, SUVH6 clade, SUVH2/9 clade, and SUVH1 clade. SUVH4 has been identified as the basal SUVH protein in several phylogenetic analyses and is the only SUVH clade present in mosses and green algae (Baumbusch et al., 2001, Zhang et al., 2012). Phylogenetic analysis based on the SET domain alone shows that the SUVH1 clade, SUVH2/9 clade, and SUVH6 clade are all present in monocots and dicot angiosperms, but not in the vascular moss *Selaginella moellendorffii* and other non-vascular plants (Zhang et al., 2012). The presence of introns has informed the understanding of SUVH evolution. *Arabidopsis* SUVH4 contains fourteen introns, while the other *Arabidopsis* SUVH proteins have fewer than two introns per sequence, often outside of the coding sequence (Baumbusch et al., 2001). This pattern of at least one SUVH4 gene containing more than ten coding-sequence introns is conserved throughout flowering plants, the lycophyte *Selaginella*, the bryophytic moss *Physcomitrella patens*, and the green algae *Chlamydomonas reinhardtii* SET3 gene (Figure 2).

This intron pattern has led to the hypothesis that the SUVH4 mRNA underwent one or more retrotransposition events followed by gene duplications to diversify the other SUVH clade members (Baumbusch et al., 2001).

Previous analyses have only identified diversified SUVH clades in flowering plants. However, I have used recent gymnosperm genome sequences to determine that numerous SUVH proteins exist in gymnosperms, including SUVH2/9 clade proteins and SUVH6 clade proteins (Figure 3). Additionally, gymnosperm SUVH6 clade proteins share unique SUVH6 clade features (See Chapter 2). Together this suggests that the expansion of the SUVH2/9 and SUVH6 clades prior to the evolution of seed plants, preceding evolution of the flowering plants.

Two recently sequenced fern genomes from *Azolla filiculoides* and *Salvinia cucullata* also encode diversified SUVH proteins. *Azolla* encodes six SUVH proteins, and *Salvinia* encodes seven SUVH proteins. In both ferns, two genes contain many introns, consistent with the basal SUVH4 gene, while the other proteins contain zero or one intron (Figure 2). Maximum likelihood-based phylogeny places one of the SUVH proteins from each fern species into the SUVH6 clade (Figure 3). Therefore, diversification of SUVH proteins into the SUVH6 clade likely occurred with ferns and prior to the evolution of seed plants.

Mosses *Selaginella* and *Physcomitrella* encode three and four SUVH proteins, respectively (Zhang et al., 2012). Previous phylogenetic studies based on SET domains place these additional SUVH proteins within the SUVH4 clade

(Zhang et al., 2012). Phylogenetic analysis based on the full protein sequences with gymnosperms and ferns concurs with this finding (Figure 3). In support of this result, intron counts show that each of these four genes contains fewer than two introns (Figure 2). This pattern indicates that at least one retrotransposition event occurred prior to the diversification of mosses.

In vitro studies have given insights into the diversification of SUVH functions; *Arabidopsis* SUVH catalytic domain truncations from each clade were tested for H3K9 methyltransferase activity and only SUVH4 and SUVH6 clade proteins SUVH5 and SUVH6 have demonstrated catalytic activity (Ebbs & Bender, 2006). However, the SUVH1 and SUVH2/9 clades use the SRA domain to identify 5meC and contribute to gene regulation through other mechanisms. The *Arabidopsis* SUVH2 and SUVH9 proteins lack post-SET domains, accounting for the lack of methyltransferase function *in vitro* (Johnson et al., 2008). However, SUVH2 and SUVH9 are critical proteins in the 5meC-initiating RdDM pathway. SUVH2 and SUVH9 bind 5meC and recruit downstream RdDM factors (Johnson et al., 2014, Jing et al., 2016). The SUVH1 clade also lacks H3K9 methyltransferase activity but has evolved a novel role in gene regulation. Recent studies demonstrate that SUVH1, SUVH3, and three DNA-J domain proteins form a complex and bind methylated cytosines through activity of the SRA domains (Zhao et al., 2019, Harris et al., 2018, Xiao et al., 2019). The SUVH1 complex is enriched at gene promoters that contain 5meC and increases transcription of those genes. Though this complex also binds to transposons, the SUVH1/3-DNAJ complex does not appear to activate silenced regions (Harris et

al., 2018). Other *Arabidopsis* SUVH1 clade proteins SUVH7 and SUVH8 currently have unknown roles.

Several models can be put forth to hypothesize why SUVH proteins expanded and diversified after the evolution of mosses. One idea is that as genome size increased, more SUVH proteins were required to transcriptionally silence the expanded genome. The plant average 2C size shows a large expansion in ferns: algae 2.6 Gb, bryophytes 1.8 Gb, lycophytes 2.3 Gb, ferns 23.0 Gb, gymnosperms 36.0 Gb, and angiosperms 10.1 Gb (Kew.org). Whole genome duplications are a major mechanism of plant genome size increases, which would increase both SUVH proteins and transposons in a genome. Transposon composition often increases with genome size (Kim, 2017), underscoring this relationship and the potential need for plants to evolve more SUVH proteins with genome size increases. Because my phylogenetic tree places a fern SUVH protein within the SUVH6 clade (Figure 3), but not within lycophytes, this evidence supports the hypothesis that the SUVH6 clade emerged to address evolved large genome sizes.

SUVH6 clades could have evolved with changing environments for land plants. However, expression of *Arabidopsis* SUVH6 clade genes does not appear to change with cold, drought, ultraviolet light, wounds, oxidative damage, or heat (AtGenExpress). There are minor expression increases observed for SUVH6 in response to salt stress in roots but not shoot tissues (AtGenExpress). Due to the poor stress-induced responses of SUVH6 clade proteins, diversification of SUVH

proteins in response to environmental change is not the most supported hypothesis.

Another idea for why plants diversified SUVH clades after the evolution of mosses is for evolving tissue diversity. For the *Arabidopsis* H3K9 methyltransferase SUVH proteins, we observe fairly low expression for SUVH4, SUVH5, and SUVH6 in most somatic tissues. SUVH4 shows particular enrichment in egg cells, embryos, and meristems (Figure 4). SUVH5 and SUVH6 show a spike in expression within the endosperm. SUVH5 shows fairly low expression in most tissues but shows enrichment in the sperm and pollen tissues (Figure 4). Angiosperm chromatin undergoes decondensation in the gametogenesis of both pollen and ovaries. Due to reactivation of transcription of transposons and repeats in the pollen vegetative cell or ovule central cell, it has been hypothesized that siRNAs are generated and transported to the gametophyte (Ibarra et al., 2012). Immunostaining of pollen shows that sperm nuclei contain strong H3K9me2 modifications, but the vegetative nucleus lacks nearly all H3K9me2 (Schoft et al., 2009). Upon fertilization, the embryo undergoes chromatin reprogramming where H3 from the sperm and egg are actively removed and replaced with newly synthesized H3 (Ingouff et al., 2010)

To identify reproductive tissue expression patterns of the SUVH6 clade in other species, I used the CoNekT database of RNA-sequencing data. Dicot *Solanum lycopersicum* and monocots *Oryza sativa* and *Zea mays* all had cell-type specific data available (Figure 5). Notably, sperm expression was high for a SUVH6 clade protein in all three species. However, a previous characterization of

strawberry expression demonstrates that the only strawberry SUVH6 clade protein is largely absent from pollen in RNA-sequencing data, indicating that this pattern of SUVH6 clade expression enrichment in sperm may not be universal (Gu et al., 2016). Like *Arabidopsis*, both *Oryza sativa* and *Zea mays* have enriched SUVH4 expression within the egg cell, although egg cell data was unavailable for *Solanum lycopersicum* (Figure 5). Additionally, while meristems show high SUVH4 expression for *Arabidopsis*, *Oryza*, and *Solanum*, in *Zea mays* a SUVH6 clade protein had highest expression within the meristem (Figure 5). Endosperm data was not consistently available in the species investigated, but *Zea mays* does not show enrichment of any SUVH protein in the endosperm, indicating that this may not be a universal pattern. All of the plants assessed appear to have SUVH4 clade and SUVH6 clade proteins enriched in the same tissues of the egg, sperm, and meristem, but the individual SUVH clade patterns may be species-specific. Together, this data highlights the importance of SUVH H3K9 methyltransferases in gametophytes and meristems but does not reveal a consistent pattern to understand SUVH6 clade diversification.

SUVH4 function

SUVH4 and CMT2/CMT3 act in an amplifying loop

Both CMT3 and SUVH4 were identified through the same genetic screens for 5meC defects at a reporter gene (Bartee et al., 2001, Malagnac et al., 2002).

The Bender laboratory identified SUVH4 and CMT3 with an EMS-mutagenesis

screen for suppressors of DNA methylation of tryptophan biosynthetic gene *PAI2* (Malagnac et al., 2002, Bartee et al., 2001). The Jacobsen laboratory identified SUVH4 (also termed KRYPTONITE) with an EMS-mutagenesis screen for suppressors of hypermethylated cytosines on the *SUPERMAN* gene that result in a floral organ number phenotype (Jackson et al., 2002). Both *suvh4* and *cmt3* mutations resulted in a drastic decrease in CHG methylation and a moderate decrease in CHH methylation at the *PAI2* locus, with little effect on CpG methylation, demonstrating that SUVH4 activity is critical for proper CHG 5meC (Bartee et al., 2001). However, *cmt3* mutants have a stronger CHG defect than *suvh4* mutants because SUVH5 and SUVH6 function as H3K9 methyltransferases with overlapping target loci. Further work has revealed that differentially methylated regions in *cmt3* mutants overlap almost entirely with *suvh456* targets (Du et al., 2012). CMT2-mediated CHH methylation also accounts for most *suvh456* differentially methylated CHH sites (Stroud et al., 2014).

To determine if the SUVH4 SRA domain acted as a 5meC reader domain, *in vitro* gel-shift studies showed that the SUVH4 SRA domain is capable of binding to 5meC (Johnson et al., 2007). Crystal structures of SRA domains bound to substrate DNA revealed a base-flipping mechanism where the SRA stabilizes bound DNA and forms a pocket for a single 5meC nucleotide (Rajakumara et al., 2011). Additionally, SUVH4 crystal structures indicate that two alpha helices form amino-terminal to the SRA domain and fold between the SRA and catalytic domains (Du et al., 2014). These pre-SRA domain helices make critical contacts with both the bound DNA and SET domain and are required for SRA mediated

5meC-binding (Johnson et al., 2007). Though the pre-SRA helices show poorer amino acid conservation than the other SUVH domains, the structure of two helices is a conserved feature (Li et al., 2018, Chapter 2).

The SUVH4 domain biochemical data and *cmt2/ cmt3* mutant H3K9me defects have demonstrated the interconnectedness of 5meC and H3K9me. This has led to the model where CMT2 and CMT3 chromodomains bind methylated H3K9 residues and methylate nearby cytosines, while SUVH4, SUVH5, and SUVH6 bind to methylated cytosines and methylate H3K9 on adjacent nucleosomes. These paired activities create a positive feedback loop to maintain both types of transcriptional silencing. Furthermore, the tight connection between H3K9me and non-CpG 5meC has also informed ideas about how replicating DNA reestablishes both modifications in *Arabidopsis*. One model proposes that hemi-methylated 5meC sites on each strand of DNA after semi-conservative DNA replication allows for SUVH recruitment through SRA domain 5meC-binding. Then, H3K9me is reestablished and further recruits CMT2/CMT3 (Saze et al., 2008). Alternative models consider the deposition of histone transmission to daughter strands: parental histones can stochastically segregate as intact octamers to daughter strands or can split octamers with newly synthesized histone proteins (Saze et al., 2008). Evidence for either model of histone segregation has been found in other systems and is currently unclear in this context (Ramachandran & Henikoff, 2015). In either case, H3K9me will be present on some octamers, allowing for CMT2/3 binding, 5meC modification, and subsequent H3K9me by SUVH proteins.

Interacting factors with SUVH4

Immunoprecipitation and mass spectrometry analysis of SUVH4 has not yielded any interacting proteins, even histones (Du et al., 2014). This finding has led to the idea that SUVH4 binds to histones transiently and is unable to be captured with this method. Similarly, immunoprecipitation and mass spectrometry analysis of CMT3 shows that it does stably associate with nucleosomes or SUVH proteins (Du et al., 2012). In addition, several groups have attempted to use chromatin immunoprecipitation sequencing (ChIP-seq) with SUVH proteins to detect SUVH-bound target loci without success, suggesting that the interaction of SUVH4 with chromatin is transient. However, immunoprecipitation and mass spectrometry of methylated DNA fragments has pulled down both SUVH4 and SUVH6 (Harris et al., 2018). This result suggests that the association of SUVHs with methylated DNA is more stable than the association of SUVH proteins with H3.

To date, the only experimentally confirmed *Arabidopsis* protein that interacts with SUVH4, SUVH5, and SUVH6 is Histone Deacetylase 6 (HDA6) (Yu et al., 2017). SUVH5 was identified as an interacting protein in an HDA6 immunoprecipitation and mass spectrometry screen (Yu et al., 2017). Though SUVH5 was the only SUVH identified in the pull-down experiment, SUVH4 and SUVH6 were also able to interact with HDA6 in bimolecular complementation assays. HDA6 interacts with a carboxy-terminal construct of SUVH5 containing the Pre-SET, SET, and Post-SET domains. HDA6 also interacts with MET1, but *hda6* mutants show 5meC hypomethylation at a large number of CHG and CHH

sites, with minor hypomethylation at CpG sites (Yang et al., 2020). *hda6* mutants also have decreased levels of H3K9me2 at several transposons (Yang et al., 2020). HDA6 mostly functions at SUVH4/ SUVH5/ SUVH6-targeted transposons, but only a subset of SUVH4/ SUVH5/ SUVH6-targeted sites are affected in *hda6* mutants, indicating HDA6 interaction is not a predominant mechanism of SUVH recruitment (Yang et al., 2020). HDA6 could remove acetyl groups from H3K9ac at a subset of SUVH targets and recruit SUVH4, SUVH5, and SUVH6 to replace that euchromatic modification with H3K9me2. Alternatively, SUVH4/ SUVH5/ SUVH6 binding to 5meC could recruit HDA6 to remove H3K9ac in order to catalyze H3K9me.

H3K9me-binding proteins

ADCP1

Until recently, CMT2 and CMT3 were the only known H3K9me2-binding proteins in *Arabidopsis*. However, biochemical screens have led to the recent identification of H3K9me2-reader Agenet Domain Containing Protein 1, ADCP1 (also termed AGDP1) (Zhao et al., 2018, Zhang et al., 2018). ADCP1 is composed of three pairs of Agenet domains and *in vitro* experiments show that each pair of Agenet domains is able to bind to H3K9me1 and H3K9me2 with little binding to unmethylated H3K9 (Zhang et al., 2018). *adcp1* hypomorph mutants show a marked reduction in H3K9me2, particularly at pericentromeres and transposons longer than 4 kilobases, indicating that ADCP1 binding also

promotes H3K9me (Zhang et al., 2018, Zhao et al., 2018). Additionally, *adcp1* plants display 5meC hypomethylation at CHG and CHH sites, but have little effect on CpG methylation, indicating a role with SUVH4/ SUVH5/ SUVH6 and CMT2/ CMT3 (Zhang et al., 2018).

How loss of ADCP1 function contributes to loss of H3K9me remains to be understood. Immunoprecipitation-mass spectrometry experiments show that ADCP1 does not stably interact with SUVH4, SUVH5, SUVH6, CMT2, CMT3, or chromosome remodeler DDM1 (Zhang et al., 2018), disputing ADCP1 function as a SUVH co-factor. However, there could be transient interactions not captured in this experiment, or ADCP1 could regulate SUVH function through an indirect mechanism. ADCP1 exhibits phase separation *in vitro*, which is negatively regulated by the first pair of Agenet domains (Zhao et al., 2018). Therefore, ADCP1 could create a phase separated chromatin boundary that can then recruit either SUVH or CMT proteins.

An initial experiment to screen for proteins with the ability to bind to H3K9me₂ on an array found that related protein ADCP2 was able to bind to H3K9me₂ to a weaker degree than ADCP1. Therefore, ADCP2 might also play a role in H3K9me and 5meC regulation *in vivo* (Zhao et al., 2018). Bioinformatics approaches reveal that there is a family of *Arabidopsis* Agenet domain proteins, most with unstudied functions (TAIR). Forthcoming experiments should identify if other ADCP proteins have novel or redundant functions with ADCP1. Though nearly all of methylation changes in *adcp1* mutant occurred at SUVH4/ SUVH5/ SUVH6 targets, those changes only comprised about 10% of SUVH4/ SUVH5/

SUVH6 sites. Therefore, ADCP family redundancies are possible and could contribute to H3K9me at a greater portion of H3K9me sites.

IBM1 and Jumonji-C Proteins

Many heterochromatin pathways contain positive feedback loops to propagate transcriptional silencing. Of equal importance is the ability of the cell to restrict silencing modifications from active genes, especially on gene-rich chromosome arms. 5meC demethylases ROS1, DEMETER, and DEMETER-Like proteins have been identified and characterized as DNA glycosylases, with DEMETER having endosperm-specific activities in chromatin reprogramming (Gehring et al., 2009, Penterman et al., 2007). In 2008, IBM1 (or JMJ25) was identified as a counterpart to the SUVH4/ SUVH5/ SUVH6 pathway. Using a forward genetics approach, *ibm1* was identified as a mutation that rescued encroaching methylation from an adjacent transposon into the cell cycle gene *BONSAI* (Saze et al., 2008). *ibm1* mutants displayed increased CHG and CHH methylation on the border of the transposon and *BONSAI* locus. The hypermethylation and associated phenotypes were suppressed by *suvh4* and *cmt3* mutations (Saze et al., 2008). Genome-wide studies show that the predominant hypermethylation in *ibm1* is at CHG sites with very minor effects on CHH and no significant changes to CpG methylation (Miura et al., 2009). *ibm1* hypermethylation is also independent of mutations to RdDM components, further

strengthening the idea that IBM1 functions in the CMT2/ CMT3 and SUVH4/ SUVH5/ SUVH6 pathway (Miura et al., 2009).

IBM1 has a single domain at the carboxy-terminus: a Jumonji C domain (jmc). Jumonji-C domains have been found to be histone demethylases in other species, including rice (Sun et al., 2008). To test if IBM1 directly demethylates H3K9 rather than acting at a different point in the CMT/ SUVH pathway, *in vitro* biochemical evidence that IBM1 has demethylase activity is needed. However, transfected IBM1 in tobacco cells results in a reduction in H3K9me1 and H3K9me2 (Inagaki et al., 2010). Additionally, mutation of residues that coordinate an iron ion for demethylase activity results in no H3K9me2 reduction.

There are five genes most related to *IBM1/ JMJ25*: *JMJ24*, *JMJ26*, *JMJ27*, *JMJ28*, and *JMJ29* (Sun and Zhou, 2008). Except for *JMJ26*, all of these proteins have evolved to have one or two RING fingers amino terminal to the jmc domain (TAIR). Although *JMJ26*, *JMJ28*, and *JMJ29* remain to be tested, *JMJ27* has *in vitro* demethylase activity specific to H3K9me1 and H3K9me2 (Dutta et al., 2017). The role of the RING fingers in *JMJ27* activity is unclear. RING fingers can be E3 ubiquitin ligases, so it will be exciting to learn what the targets of *JMJ27* are and how that contributes to its H3K9 demethylase activity. Additionally, *JMJ27* is the closest relative to IBM1, so future work will elucidate whether *JMJ27* and IBM1 have overlapping or distinct activities throughout the genome (Sun and Zhou, 2008).

Investigation into the remaining family member JMJ24 has shown a separate role for JMJ24 in anti-silencing. Because JMJ24 has lost iron-binding residues in the catalytic *jmjC* domain, JMJ24 does not have H3K9 demethylase activity, but can still bind H3, H3K9me1, and H3K9me2 (Kabelitz et al., 2016). However, JMJ24 has *in vitro* E3 ubiquitin ligase activity (Kabelitz et al., 2016). Intriguingly, *in vitro* and *in vivo* evidence show JMJ24, but not IBM1, binds, ubiquitinates, and destabilizes CMT3 (Deng et al., 2015). This ubiquitination of CMT3 likely leads to its proteasomal degradation and provides a potential mechanism for the observed increase in CHG methylation and H3K9me2 in gene bodies in *jmj24* plants (Deng et al., 2015). It is unclear how much of the H3K9me2 increase in *jmj24* plants is simply due to CMT3 accumulation or due to another role for JMJ24 in functioning as a possible reader of H3K9me (Audonnet et al., 2017).

SUVH6 clade proteins

Locus-specific targeting of SUVH5 and SUVH6

Initial *in vitro* assays demonstrated that SUVH4, SUVH5, and SUVH6 were able to methylate histone substrates, but mutant phenotypes demonstrate that the three H3K9 methyltransferases do not contribute equally to H3K9me (Ebbs & Bender, 2006). SUVH4 is the major contributor to H3K9me for targets of SUVH4, SUVH5, and SUVH6. Due to the likely transient association of SUVH4, SUVH5, and SUVH6 with H3, genome-wide studies of the SUVH proteins have

used ChIP-seq with H3K9me2 antibodies or bisulfite sequencing of 5meC in *suvh* mutant backgrounds to infer changes in the absence of specific SUVH proteins rather than directly performing SUVH ChIP-seq. H3K9me2 ChIP-seq experiments show about a 90% decrease in peaks between a *suvh4* mutant and wildtype; approximately a 10% decrease in peaks between *suvh5* and wild type, and negligible differences between *suvh6* and wildtype (Li et al., 2018).

Though SUVH5 and SUVH6 maintain minor levels of genome-wide H3K9me, they do maintain significant levels of H3K9me at specific targets. This targeted activity is not generally transposon type-specific; genome-wide assays of transposon families indicate only the LTR/ Gypsy family of transposons show significant levels of 5meC at CHH sites in *suvh4* mutants compared to *suvh4 suvh5 suvh6* triple mutants (Stroud et al., 2013). An additional target where SUVH5 and SUVH6 have a major role is at the 45S rDNA repeats (Chandrasekhara et al., 2016). *Arabidopsis* encodes two Nucleolar Organizing Regions (NORs) on the short arms of chromosomes two and four. The two NORs encode hundreds of ribosomal DNA repeats, but NOR2 is heavily silenced with H3K9me2 (Chandrasekhara et al., 2016). While CHG methylation in the 45S rDNA promoter follows the pattern of SUVH4 having the strongest role, *suvh5 suvh6* lines show stronger defects than *suvh4* mutants at CpG methylation sites, indicating that SUVH5 and SUVH6 contribute to the majority H3K9me2 at CpG sites at this target (Pontvianne et al., 2012).

Genome-wide bisulfite sequencing of *suvh4* mutants reveals drastic drops in pericentromeric CTG and CAG methylation, but have only moderate defects in

CCG methylation, where the outer cytosine is methylated (Gouil et al., 2016). This pattern suggests that SUVH5 and/ or SUVH6 may be contributing a significant amount to H3K9me at CCG sites. Additionally, *suvh5* and *suvh6* single mutants both show some hypomethylation at CHH sites genome-wide, though not nearly as pronounced as in *suvh4* single mutants (Li et al., 2018). One study found that *suvh5* mutants showed no genome-wide CHG methylation changes but did show a surprising slight genome-wide CpG methylation decrease (Li et al., 2018). While SUVH4 mediates most H3K9me at both CHG and CHH sites, SUVH5 and SUVH6 have greater contributions to H3K9me at CHH sites than CHG sites.

Because SUVH4, SUVH5, and SUVH6 have overlapping targets, genome-wide experiments on *suvh5* or *suvh6* mutants do not reveal the full scope of SUVH5 or SUVH6 H3K9me (Ebbs et al., 2005, Ebbs & Bender, 2006). In order to discern the H3K9me contribution of SUVH6 individually, experiments need to be performed on *suvh4 suvh5* double mutant lines where SUVH6 is the only SUVH H3K9 methyltransferase contributing to the phenotype. Similarly, *suvh4 suvh6* double mutants will reveal the H3K9me contributions of SUVH5. Unfortunately, H3K9me2 ChIP-seq experiments have not yet been performed on *suvh4 suvh5* and *suvh4 suvh6* lines to capture a genome-wide understanding of the individual contributions of SUVH5 and SUVH6. In spite of this limitation, assays of individual transposons and repeats have been performed on *suvh4 suvh5* and *suvh4 suvh6* lines and reveal that SUVH5 and SUVH6 have locus-specific preferences (Ebbs & Bender, 2006).

There is rich evidence about differential contributions from the SUVH4, SUVH5, and SUVH6 at the *PAI* loci. The *PAI* genetic reporter system used by the Bender laboratory to recover mutations in *SUVH4* and *CMT3* also revealed that SUVH5 and SUVH6 make locus-specific contributions to H3K9me (Ebbs & Bender, 2006). In the Wassilewskija (Ws) ecotype of *Arabidopsis*, the phosphoanthranilate synthase isomerase (*PAI*) tryptophan biosynthetic genes have undergone recent duplications resulting in an inverted repeat of the *PAI-PAI4* sequences. (Bender & Fink, 1995, Melquist et al., 2002). A distal promoter drives transcription of *PAI1* resulting in *PAI1* transcripts that occasionally form a *PAI1-PAI4* hairpin RNA that is processed by dicer proteins into small RNAs, triggering RdDM silencing initiation (Enke et al., 2011). However, RdDM silencing is initiated not only on the *PAI1-PAI4* inverted repeat, but also at the unlinked *PAI2* and *PAI3* duplications (Chapter 2 Supplemental Figure 4). The distal *PAI1* promoter is unmodified and allows continued transcription of *PAI1-PAI4*, subsequent PAI protein production, and tryptophan biosynthesis. In Ws plants, a *pai1* missense mutation results in a hypomorphic PAI1 protein, but plants survive due to tryptophan biosynthesis through residual PAI1 activity and/or a small amount of PAI2 protein produced from the silenced *PAI2* gene (Bartee & Bender, 2001). These *pai1* plants have reduced size and fertility, plus a blue fluorescent phenotype under ultraviolet (UV) light, providing a reporter for changes in *PAI2* silencing. Therefore, forward screens and transgene screens can use blue fluorescence as an output of cytosine methylation at the *PAI2* gene.

Strikingly, SUVH6 maintains high levels of H3K9me at *PAI1-PAI4* but has weak contributions to H3K9me at *PAI2* (Ebbs & Bender, 2006). *PAI1* and *PAI2* have nearly identical DNA sequences but SUVH6 strongly contributes to H3K9me only at the former. This provides a key finding that sequence specificity in SRA 5meC-binding is not the only contributor to SUVH5 and SUVH6 locus-specific activity. In contrast, SUVH5 has weak contributions to H3K9me at both *PAI1-PAI4* and *PAI2* loci (Ebbs et al., 2005). However, SUVH5 is not merely a weaker enzyme because SUVH5 makes significant contributions to H3K9me at the *Ta3* retrotransposon (At1g37110) and *Mu1* DNA transposon (At4G08680) (Ebbs et al., 2005, Ebbs & Bender, 2006). Conversely, SUVH6 has a very minor H3K9me contribution to *Ta3* and *Mu1*. Assays of 5meC at CCG sites in *Ta3* and the *PAI* repeats demonstrate that these locus-specific patterns are still reflected (Ebbs & Bender, 2006). Therefore, locus-specific differences hold even when only observing changes at CCG sites and are not merely due to differences in assessment of CHG or CHH sites. Together, these data indicate that while both SUVH5 and SUVH6 function at only a subset of SUVH4 targets, SUVH5 and SUVH6 have different locus-specific preferences.

Biochemical Analysis of the SRA and SET domains

Biochemical analyses of SRA and SET domain structure and function have revealed slight differences between SUVH4, SUVH5, and SUVH6, with SUVH5 having the most diverged characteristics. *In vitro* gel shift studies show

that the SUVH4 and SUVH6 pre-SRA and SRA domains are capable of binding to methylated cytosine in the CHG and CHH contexts (Johnson et al., 2007). SUVH5 shows the same ability to bind specifically to 5meC and not unmethylated cytosines but differs in binding methylated CpG as well as CHH and CHG (Rajakumara et al., 2011). This result is in agreement with the finding that *suvh5* mutants show a minor decrease in CpG methylation (Li et al., 2018).

SUVH binding to distinct trinucleotides has been investigated further with isothermal titration calorimetry with purified SUVH protein truncations beginning at the pre-SRA domain. SUVH6 can bind methylated CpG, CAT, CAG, and CCG substrates equally well and with approximately equivalent binding affinities as SUVH4 (Li et al., 2018). SUVH5 shows weaker binding affinities for nearly all cytosine contexts tested (Li et al., 2018). SUVH6 and SUVH4 SRA 5meC binding is similar, but SUVH6 clearly does not contribute as much H3K9 methylation. Meanwhile, *suvh5* has a slight hypomethylation defect at CpG sites, but that defect is not reflected in the binding affinity to methylated CpG *in vitro*. Because these binding affinities are not congruous with mutant phenotypes, this indicates that *in vitro* assays of SRA truncations are not fully reflective of the SUVH full enzyme and/ or SUVH activity in the context of chromatin. This discrepancy could be due to factors outside of SRA binding that influence SUVH5 substrate preferences.

One way that SUVH5 and SUVH6 do differ from SUVH4 in these *in vitro* assays is in the activity on hemi-methylated substrates (Li et al., 2018). SUVH4 shows much weaker binding to hemi-methylated substrates than fully methylated

DNA while SUVH5 SUVH6 bound hemi-methylated and fully methylated DNA with equal affinity. If these *in vitro* results reflect function *in vivo*, perhaps SUVH5 and SUVH6 play more significant roles in directing H3K9me to hemi-methylated regions at DNA replication forks.

In vitro assays have also tested differences in SUVH4, SUVH5, and SUVH6 H3K9me catalysis using truncations containing the SRA and catalytic domains and lacking amino-terminal extensions. *In vitro*, SUVH4 displays strong monomethylase activity on unmethylated H3 substrates, but only slight activity on H3K9me1 substrates to catalyze H3K9me2 (Du et al., 2014). SUVH6 displays a similar pattern of preferential monomethylase activity (Li et al., 2018). Thus far the activity of SUVH5 for monomethylation versus dimethylation has not been tested *in vitro*.

SUVH4 and SUVH6 do not demonstrate significant biochemical differences *in vitro* that can explain why SUVH4 has greater genome-wide activity than SUVH6. Additionally, these experiments used nearly all leaf tissues where *SUVH4* and *SUVH6* have similar expression levels (Figure 4), indicating that limited expression of *SUVH6* does not account for limited SUVH6 activity relative to SUVH4. SUVH5 may have weaker *in vitro* biochemical properties and has lower expression levels than SUVH6 (Figure 4) but maintains higher levels of H3K9me at *Ta3* and *Mu1* than SUVH6 (Ebbs & Bender, 2006). The biochemical data for SUVH5 and SUVH6 do not provide clear mechanisms for how 5meC-binding or H3K9me catalysis underlie locus-specificity.

Summary

SUVH4, SUVH5, and SUVH6 are the major H3K9 methyltransferases in *Arabidopsis*. SUVH4, the most ancestral SUVH protein, maintains a majority of genome-wide H3K9me, although all three proteins have overlapping targets.

Unlike SUVH4, SUVH5 and SUVH6 have locus-specific preferences for H3K9me activity. To how the SUVH6 clade diversified from the SUVH4 clade, in this thesis I have characterized SUVH6 clade-specific features and tested the functions of these features with a transgene complementation strategy. In chapter two, I present evidence of a conserved, novel motif that functions both as a nuclear localization signal and to promote SUVH6 function at preferred loci. In chapter three, I demonstrate that the amino-terminal extensions in SUVH5 and SUVH6 contain sequences critical for locus-specificity with amino-terminal swapped transgenes. My results define SUVH6 clade features and provide evidence that SUVH6 clade amino-terminal sequences function in regulating the SUVH6 clade proteins.

Figures

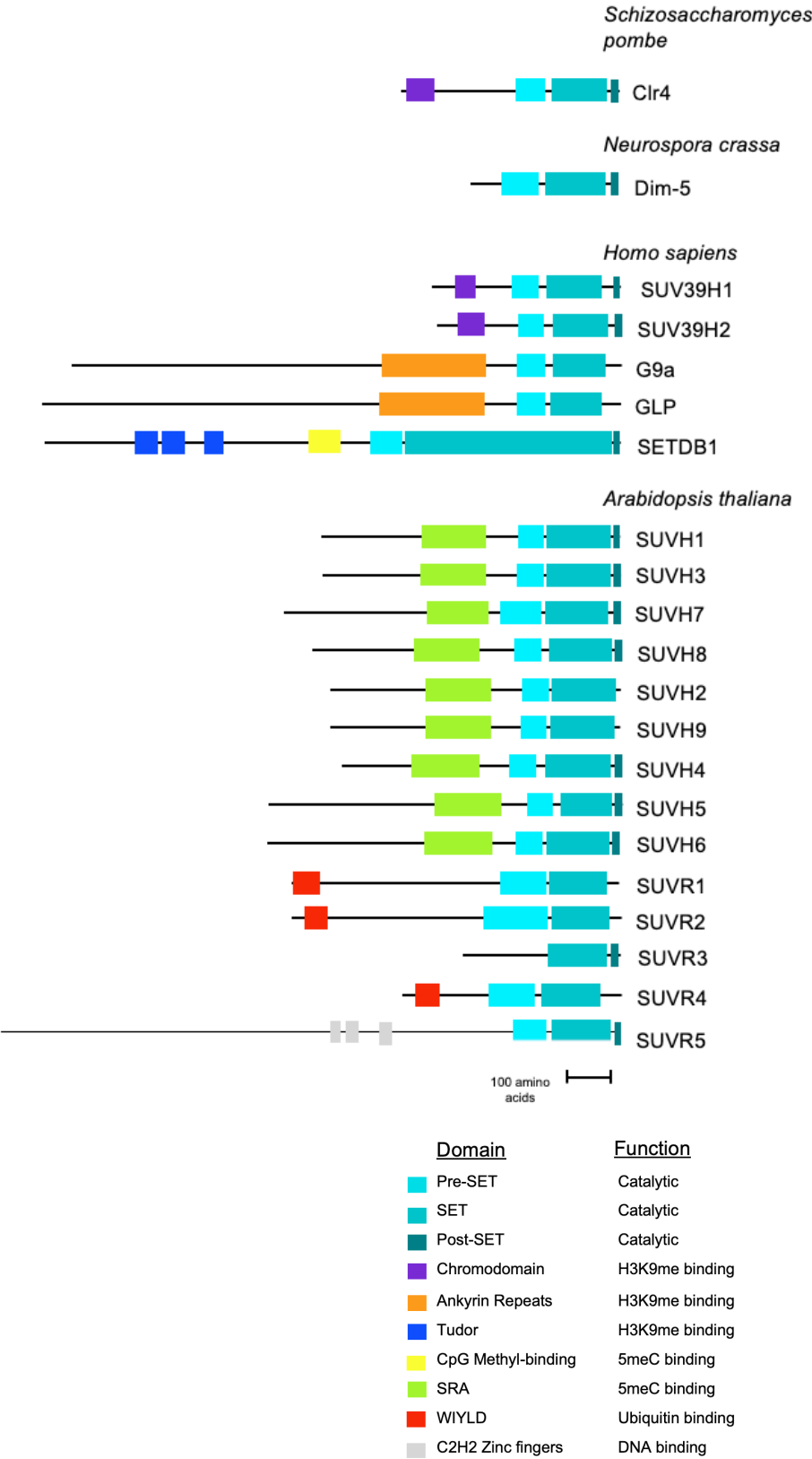


Figure 1: H3K9 methyltransferases have diversified interaction domains in multicellular organisms. Domains drawn to scale as categorized by Expasy Prosite. Sequences obtained from NCBI or TAIR. Domain colors listed and described in Introduction.

Species	SUVH	CDS Introns	UTR introns
<i>Chlamydomonas reinhardtii</i>	CrSET3	16	0
<i>Physcomitrella patens</i>	PpSUVH4a	13	0
	PpSUVH4b	13	0
	PpSUVH4d	0	1
	PpSUVH4c	0	3
<i>Selaginella moellendorffii</i>	SmSUVH4a	12	1
	SmSUVH4c	0	2
	SmSUVH4d	0	1
<i>Azolla filiculoides</i>	Af0303.g063783	13	0
	Af0001.g000963	14	0
	Af0215.g058438	0	0
	Af0188.g056787	0	0
	Af0046.g030118	0	0
	Af0116.g046348	0	0
	Af0007.g011217	1	0
<i>Oryza sativa</i>	OsSUVH4	13	1
	OsSUVH5a	0	0
	OsSUVH5b	1	1
	OsSUVH5c	1	2
<i>Zea mays</i>	ZmSUVH4a	15	0
	ZmSUVH4b	13	1
	ZmSUVH5a	0	1
	ZmSUVH5b	1	1
	ZmSUVH5c	1	0
<i>Vitis vinifera</i>	VvSUVH4a	14	1
	VvSUVH4b	13	1
	VvSUVH5a	0	2
	VvSUVH5b	0	3
	VvSUVH6	0	2
<i>Arabidopsis thaliana</i>	AtSUVH4	11	2
	AtSUVH5	0	2
	AtSUVH6	0	1

Figure 2: Moss SUVH proteins demonstrate likely originated from an early retrotransposition of *SUVH4*. Coding sequence (CDS) introns and untranslated region (UTR) introns separated. Intron numbers as calculated from NCBI, FernBase, or Phytozome. Intron counts unavailable for gymnosperms. At= *Arabidopsis thaliana*, Vv= *Vitis vinifera*, Os= *Oryza sativa*, Zm= *Zea mays*, Af= *Azolla filiculoides*, Sm= *Selaginella moellendorffi*, Pp= *Physcomitrella patens*, Cr= *Chlamydomonas reinhardtii*.

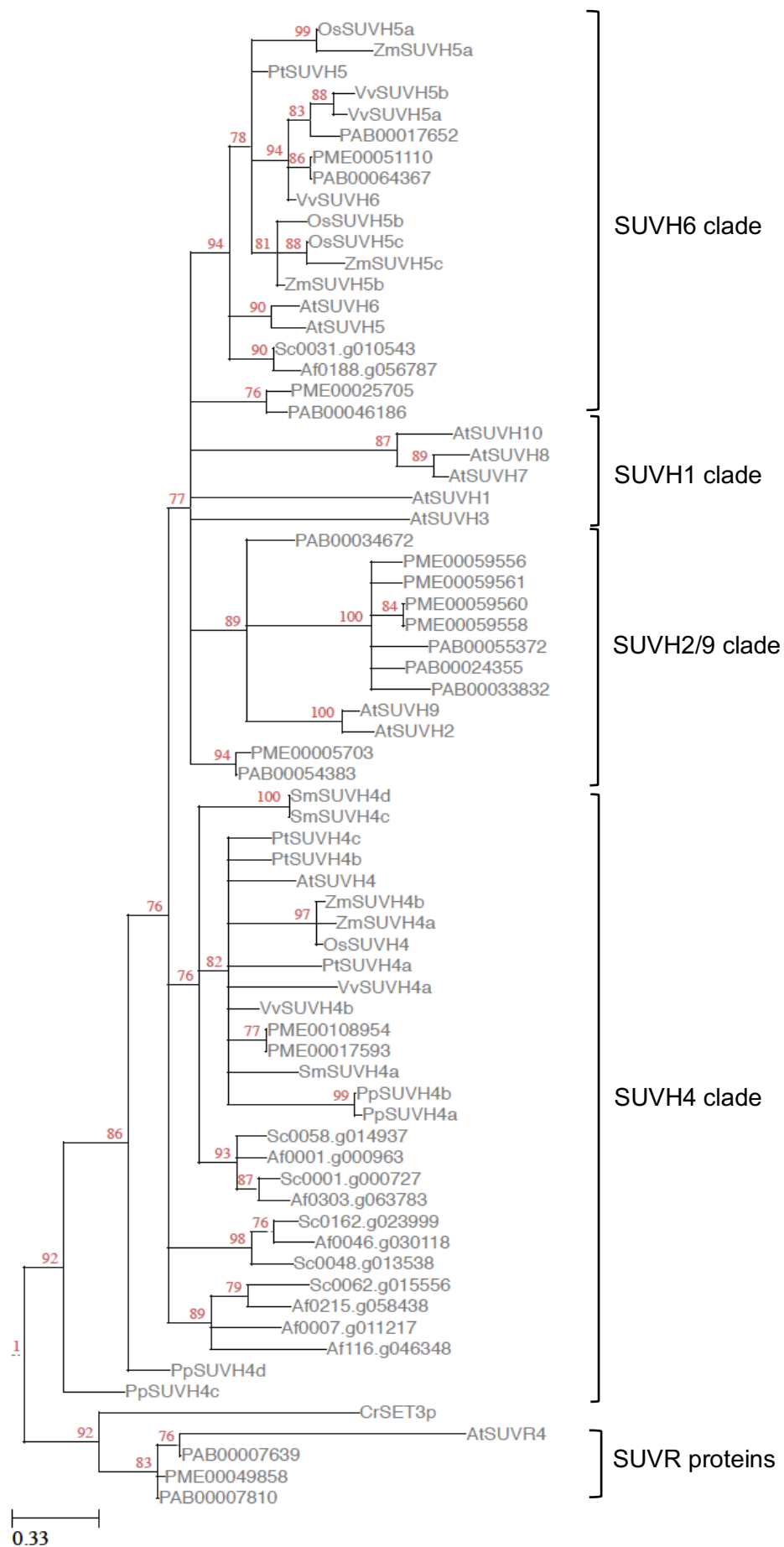


Figure 3: SUVH clades diversified after the evolution of lycophytic mosses.

Maximum likelihood tree generated with Phylogeny.fr with MUSCLE alignment followed by PhyML. Tree visualized with ETE Toolkit Tree Viewer. Bootstraps listed on branches with cutoffs at 70%. At= *Arabidopsis thaliana*, Vv= *Vitis vinifera*, Os= *Oryza sativa*, Zm= *Zea mays*, PAB= *Picea abies*, PME= *Pseudotsuga menziesii*, Af= *Azolla filiculoides*, Sc= *Salvinia cucullata*, Sm= *Selaginella moellendorffi*, Pp= *Physcomitrella patens*, Cr= *Chlamydomonas reinhardtii*.

	SUVH6	SUVH5	SUVH4
Flower (carpels)	10.27	4.93	19.95
Flower (stigmatic tissue)	7.72	3.15	6.22
Pedicel	6.75	2.41	8.12
Axis of the inflorescence	7.26	4.05	15.03
Siliques	6.57	3.03	6.42
Siliques (senescent)	3.27	0.13	0.21
Pods of siliques	7.62	2.1	4.21
Pods of siliques (senescent)	8.55	0.87	1.01
Embryo	10.63	16.72	23.85
Endosperm	18.43	17.9	8.28
Seedling	6.99	1.32	5.51
Seedling (etiolated)	6.32	1.56	4.51
Meristem	4.61	3.26	23.82
Stem (internode)	5.71	2.9	5.64
Stem (internode, senescent)	9.84	0.52	4.68
Stems	5.91	2.69	9.49
Leaf (rosette)	8.3	0.36	1.8
Epidermis cells	2.09	0.51	1.33
Mature guard cells	0.92	0.26	0.5
Root (differentiation zone)	8.17	1.2	3.62
Root (elongation zone)	8.65	1.77	5.43
Root (meristematic zone)	5.16	6.31	22.44
Root (apex)	6.91	2.81	8.26
Root (stele)	0.05	0.0	0.03
Root (QC cells)	10.07	0.0	21.78
Root (tip)	3.84	3.59	8.99

	SUVH6	SUVH5	SUVH4
Egg cell	4.29	1.27	30.95
Ovule	9.31	7.12	23.01
Uninucleate microspore (Col-0)	5.32	18.29	10.33
Uninucleate microspore (Ler-0)	8.08	12.66	5.61
Pollen (bicellular, Col-0)	5.96	13.99	6.6
Pollen (late bicellular, Col-0)	6.41	5.48	1.73
Pollen (bicellular, Ler-0)	7.39	13.77	5.09
Pollen (tricellular, Col-0)	9.22	2.02	0.6
Pollen (tricellular, Ler-0)	7.03	2.37	0.49
Pollen (mature, Col-0)	7.14	1.1	0.25
Pollen (mature, Ler-0)	4.81	1.67	0.39
Sperm	1.52	15.23	3.96
Pollen tube (Col-0)	2.25	1.9	8.56
Flower (receptacles)	9.16	3.67	9.39
Flower (floral buds)	14.43	2.96	11.83
Flower (sepals)	8.47	1.87	3.56
Flower (petals)	9.17	2.06	1.81
Flower (stamen filaments)	7.29	2.18	2.28
Flower (anthers)	11.29	3.35	3.13

Figure 4: *Arabidopsis thaliana* expression analysis of *SUVH4*, *SUVH5*, and *SUVH6* in different tissues shows SUVH specificities in reproductive tissues. Values expressed as read per million, with red color indicating high expression. RNA-sequencing data obtained and visualized from CoNekT at <https://evorepro.sbs.ntu.edu.sg>.

A

	SISUVH4	SISUVH6c	SISUVH6d	SISUVH6a	SISUVH6b
Ovule (0-1 DPA)	11.76	0.01	0.06	10.12	36.68
Ovule (2 & 5 DPA)	23.73	0.01	0.02	6.7	22.74
Ovary	8.74	0.01	0.27	5.96	8.11
Microspore	3.69	28.9	89.75	1.59	0.65
Generative cell	0.25	17.28	0.45	87.83	0.59
Pollen	0.34	0.01	0.01	2.08	0.0
Sperm cell	0.11	5.38	0.56	19.57	0.47
Pollen grain	0.25	0.1	0.0	3.75	0.01
Pollen tube	0.4	0.03	0.0	3.57	0.0
Pollen tube (1.5h)	0.98	0.11	0.0	3.84	0.0
Pollen tube (3h)	0.86	0.09	0.0	3.83	0.0
Pollen tube (9h)	0.86	0.04	0.0	3.8	0.0
Pistil	6.29	0.03	0.13	5.32	17.2
Style (before flowering)	2.86	0.0	0.06	4.69	0.65
Style (after flowering)	0.95	0.0	0.08	5.33	0.07
Style (pollinated, after flowering)	0.99	0.03	0.06	4.8	0.08
Flower (buds)	14.9	0.2	0.53	5.09	6.28
Flowers	2.23	0.04	0.08	3.82	0.95
Ovary wall/pericarp (0-1 DPA)	11.89	0.0	0.01	6.97	5.15
Ovary wall/pericarp (2 & 5 DPA)	16.13	0.0	0.01	4.41	2.54
Pericarp/Exocarp (4 DPA)	14.86	0.02	0.33	4.01	7.08
Pericarp	6.7	0.0	0.41	9.16	0.0
Pericarp (5 DPA)	19.7	0.0	0.1	3.79	4.16
Pericarp (7 DPA)	14.95	0.01	0.16	3.93	4.17
Pericarp (10 DPA)	10.89	0.01	0.17	3.72	3.69
Fruit (mature green)	4.31	0.0	0.01	11.8	0.13
Fruit (breaker stage)	1.95	0.01	0.04	7.27	0.0
Septum (7 DPA)	18.75	0.03	0.41	4.22	11.19

	SISUVH4	SISUVH6c	SISUVH6d	SISUVH6a	SISUVH6b
Septum (10 DPA)	10.34	0.0	0.32	4.73	4.57
Septum/Seed (4 DPA)	14.5	0.03	0.33	4.05	11.92
Seed (5 DPA)	19.79	0.05	0.21	4.91	20.93
Seed (7 DPA)	17.11	0.14	0.28	4.88	15.89
Seed (10 DPA)	15.13	0.14	0.24	5.23	10.71
Seeds (10 DPA)	2.4	0.03	0.03	1.04	8.04
Seeds (20 DPA)	18.26	0.06	0.04	8.74	10.79
Seeds (30 DPA)	1.71	0.0	0.08	7.17	1.53
Seeds (Pk-stem)	0.99	0.0	0.31	6.33	0.0
Seeds (Br-stem)	1.22	0.01	0.43	5.32	0.03
Seeds (Pk-equatorial)	1.4	0.0	0.52	5.89	0.08
Seeds (Pk-stylar)	1.38	0.01	0.74	7.01	0.03
Seeds (Br-stylar)	1.15	0.03	0.45	4.41	0.03
Seeds	1.04	0.0	0.33	4.66	0.12
Seeds (RR)	1.34	0.0	0.41	7.62	0.03
Seeds (LR)	0.97	0.0	0.53	7.06	0.0
Seeds (MG-equatorial)	1.26	0.02	0.23	5.89	0.44
Hypocotyl	8.52	0.0	0.05	7.27	0.21
Seedling (leaves)	0.64	0.0	0.01	1.19	0.02
Seedling (roots)	9.46	0.0	0.2	11.27	0.04
Callus	8.32	0.08	0.1	3.12	0.11
Meristem (vegetative)	82.76	0.0	0.92	4.26	17.97
Meristem	72.93	0.0	1.04	4.76	10.69
Apical meristem	9.75	0.0	0.1	3.29	0.81
Leaf	4.14	0.0	0.07	4.21	0.34
Root	6.31	0.0	0.08	2.98	0.06
Root (differentiation zone)	5.34	0.0	0.04	2.27	0.0
Root (elongation zone)	5.84	0.0	0.0	2.16	0.02
Root (meristematic zone)	20.99	0.0	0.03	2.41	0.04
Root hair cells	4.35	0.0	0.09	2.07	0.0

B

	OsSUVH4	OsSUVH5a	OsSUVH5c	OsSUVH5b
Ovule	42.06	30.36	12.64	15.51
Ovary	9.79	10.07	3.7	6.53
Egg cell	23.83	12.77	2.38	7.29
Pollen tricellular	0.03	3.17	0.11	11.03
Pollen	0.74	7.35	0.12	12.58
Sperm	0.18	21.7	0.19	46.89
Flower (buds)	6.01	13.47	1.75	10.33
Flower	4.25	13.28	2.68	11.66
Panicle	16.47	25.24	10.06	10.9
Milk grains	1.26	3.4	0.46	3.02
Seeds	4.76	5.09	1.14	4.94
Seedling	6.02	14.72	1.13	9.49
Seedling (12-day-old, DongJin)	3.87	12.99	0.51	10.9
Seedling (12-day-old, DongJin, mutant)	6.9	11.64	0.5	9.51
Seedling (22-days-old, 22C)	1.75	6.66	0.89	6.03
Seedling (22-days-old, 30C)	0.93	6.69	0.65	5.23
Meristem	36.49	25.04	9.0	13.67
Stems	2.42	8.42	0.51	5.84
Callus	7.06	20.78	2.22	10.4
Leaf	2.04	8.8	2.28	5.78

C

	ZmSUVH5c	ZmSUVH5a	ZmSUVH5b	ZmSUVH4b
Egg cell	3.31	5.57	16.01	32.16
Ovary	1.45	1.71	0.28	1.51
Ovule	1.84	7.0	4.37	6.03
Microspore	1.87	7.16	7.14	0.55
Bicellular microspore	0.2	1.3	1.84	0.0
Pollen (mature)	0.38	0.28	5.96	0.05
Pollen tube	0.98	0.63	6.08	0.09
Sperm cell	10.85	6.43	37.87	0.82
Pollen tube	0.98	0.63	6.08	0.09
Tassels (primordia)	5.78	20.74	10.78	15.22
Tassels (anther)	5.01	11.69	14.19	5.53
Tassels	2.15	9.68	13.76	4.78
Silk	0.42	4.76	5.02	3.93
Ear	0.97	10.6	5.37	8.79
Ear (immature)	2.42	9.11	4.1	5.86
Ear (inflorescence)	3.61	11.39	4.99	10.71
Ear (primordia)	1.77	22.57	15.25	11.26
Zygote (12 HAP)	0.82	5.52	17.43	19.23
Zygote (24 HAP)	1.05	3.97	17.45	25.93
Daughter cell (apical)	0.16	8.47	17.93	28.6
Daughter cell (basal)	0.83	3.89	12.39	15.29
Embryo	4.26	13.0	5.09	16.19
Nucellus	17.0	16.07	8.54	24.52
Endopserm	0.07	3.07	5.75	1.33
Kernel	0.58	5.76	7.13	4.32
Kernel (germinating)	0.49	6.72	4.54	6.03
Seeds	1.48	10.52	6.96	9.1
Embryo Sac	3.63	9.79	9.29	11.95
Pericarp and aleurone	1.28	4.55	2.64	6.24
Coleoptile (5 day)	0.76	16.08	10.88	13.88

	ZmSUVH5c	ZmSUVH5a	ZmSUVH5b	ZmSUVH4b
Seedling	0.15	9.82	5.51	6.05
Internode	0.06	9.21	5.73	10.93
Sheath	0.06	17.5	14.43	11.45
Shoot (apex)	0.74	10.99	4.45	8.08
Shoot	0.07	6.78	4.46	4.19
Whole plant	0.75	14.42	8.52	5.02
Vegetative Meristem & Surrounding Tissue	0.25	7.47	2.89	5.39
Meristem (SAM with Laser capture)	3.67	27.35	4.88	10.92
Apical meristem	3.43	21.76	12.94	12.81
Stem and meristem	0.47	15.0	6.25	7.28
Stems	0.22	14.95	8.49	8.73
Stem (Internode)	0.01	9.48	6.33	11.87
Leaf	0.23	7.15	5.69	6.85
Leaf (apex)	1.09	14.8	9.9	9.17
Leaf (bundle sheath)	0.0	4.99	2.2	3.98
Leaf (mesophyll)	0.01	6.55	2.61	3.11
Leaf (seedlings)	0.0	3.7	3.13	1.51
Leaf and kernel	0.0	4.19	2.53	2.67
Ear (leaf)	0.0	6.5	4.6	4.45
Root	0.13	8.83	5.12	7.6
Root (elongation zone)	0.0	6.4	3.71	0.79
Root (meristematic zone)	0.45	10.76	5.25	4.81
Root (stele)	0.02	11.38	7.96	6.98
whole crown root (20-30 mm)	0.14	9.71	5.42	8.2
Secondary root	0.11	4.46	3.59	2.27
Root, root hairs removed	0.18	10.13	6.24	6.92
Root (maturation zone)	0.38	4.59	2.17	2.37

Figure 5: Tissue-specific expression analysis of SUVH4 clade versus SUVH6 clade proteins in other flowering plants. RNA sequencing data from (a) dicot tomato (*Solanum lycopersicum*), (b) monocot rice (*Oryza sativa*), and (c) monocot corn (*Zea mays*). Values expressed as read per million, with red highlighting high expression tissues. ZmSUVH4a sequence not available on CoNekT. Tomato SUVH proteins obtained from NCBI annotations. RNA-sequencing data obtained and visualized from CoNekT at <https://evorepro.sbs.ntu.edu.sg>.

References

- Audonnet, L., Shen, Y., & Zhou, D. (2017). JMJ24 antagonizes histone H3K9 demethylase IBM1 / JMJ25 function and interacts with RNAi pathways for gene silencing. *Gene Expression Patterns*, 25–26, 1–7.
<https://doi.org/10.1016/j.gep.2017.04.001>
- Aufsatz, W., Mette, M. F., Matzke, A. J. M., & Matzke, M. (2004). The role of MET1 in RNA-directed de novo and maintenance methylation of CG dinucleotides. *Plant Molecular Biology*, 54, 793–804.
<https://doi.org/10.1007/s11103-004-0179-1>
- Bainard, J. D., Newmaster, S. G., Budke, J. M., Centre, D., Canada, A., Road, A., & Current, S. (2020). Genome size and endopolyploidy evolution across the moss phylogeny. *Annals of Botany*, 125, 543–555.
<https://doi.org/10.1093/aob/mcz194>
- Bannister, a J., Zegerman, P., Partridge, J. F., Miska, E. a, Thomas, J. O., Allshire, R. C., & Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature*, 410(MARCH), 120–124.
<https://doi.org/10.1038/35065138>
- Bartee, L., & Bender, J. (2001). Two Arabidopsis methylation-deficiency mutations confer only partial effects on a methylated endogenous gene family, 29(10), 2127–2134.
- Bartee, L., Malagnac, F., & Bender, J. (2001). Arabidopsis cmt3 chromomethylase mutations block non-CG methylation and silencing of an endogenous gene. *Genes and Development*, 15(14), 1753–1758.
<https://doi.org/10.1101/gad.905701>
- Baumbusch, L. O., Thorstensen, T., Krauss, V., Fischer, a, Naumann, K., Assalkhou, R., ... Aalen, R. B. (2001). The Arabidopsis thaliana genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. *Nucleic Acids Research*, 29(21), 4319–4333.
- Bender, J., & Fink, G. R. (1995). Epigenetic Control of an Endogenous Gene Family Is Revealed by a Novel Blue Fluorescent Mutant of Arabidopsis. *Cell*, 83, 725–734.
- Bernatavichute, Y. V., Zhang, X., Cokus, S., Pellegrini, M., & Jacobsen, S. E. (2008). Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. *PLoS ONE*, 3(9).
<https://doi.org/10.1371/journal.pone.0003156>

- Boavida, Leonor C et al. (2011). Whole Genome Analysis of Gene Expression Reveals Coordinated Activation of Signaling and Metabolic Pathways During Pollen-Pistil Interactions in Arabidopsis. *Plant physiology* 155(4): 2066–2080. Web.
- Borges, Filipe et al. (2008). Comparative Transcriptomics of Arabidopsis Sperm Cells. *Plant physiology* 148(2): 1168–1181. Web.
- Cao, X., & Jacobsen, S. E. (2002). Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 16491–16498. <https://doi.org/10.1073/pnas.162371599>
- Caro, E., Stroud, H., Greenberg, M. V. C., Bernatavichute, Y. V., Feng, S., Groth, M., ... Jacobsen, S. E. (2012). The SET-Domain Protein SUVR5 Mediates H3K9me2 Deposition and Silencing at Stimulus Response Genes in a DNA Methylation-Independent Manner. *PLoS Genetics*, 8(10), e1002995. <https://doi.org/10.1371/journal.pgen.1002995>
- Casas-mollano, J. A., Dijk, K. Van, Eisenhart, J., & Cerutti, H. (2007). SET3p monomethylates histone H3 on lysine 9 and is required for the silencing of tandemly repeated transgenes in Chlamydomonas. *Nucleic Acids Research*, 35(3), 939–950. <https://doi.org/10.1093/nar/gkl1149>
- Chandrasekhara, C., Mohannath, G., Blevins, T., Pontvianne, F., & Pikaard, C. S. (2016). Chromosome-specific NOR inactivation explains selective rRNA gene silencing and dosage control in Arabidopsis. *Genes and Development*, 30(2), 177–190. <https://doi.org/10.1101/gad.273755.115>
- Cokus, S. J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschield, C. D., ... Jacobsen, S. E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature*, 452(13), 215–219. <https://doi.org/10.1038/nature06745>
- Collins, R. E., Northrop, J. P., Horton, J. R., Lee, D. Y., Zhang, X., Stallcup, M. R., & Cheng, X. (2008). The ankyrin repeats of G9a and GLP histone methyltransferases are mono- and dimethyllysine binding modules, 15(3), 245–250. <https://doi.org/10.1038/nsmb.1384>
- Deng, S., Jang, I., Su, L., Xu, J., & Chua, N. (2016). JMJ24 targets for proteasomal degradation in Arabidopsis. *Genes and Development*, 30(1), 1–6. <https://doi.org/10.1101/gad.274647.115.2010>
- Dong, H., Liu, D., Han, T., Zhao, Y., Sun, J., & Lin, S. (2015). Diversification and evolution of the SDG gene family in Brassica rapa after the whole

- genome triplication. *Nature Publishing Group*, 5(16851), 1–14.
<https://doi.org/10.1038/srep16851>
- Du, J., Johnson, L. M., Groth, M., Feng, S., Hale, C. J., Li, S., ... Jacobsen, S. E. (2014). Mechanism of DNA methylation-directed histone methylation by KRYPTONITE. *Molecular Cell*, 55(3), 495–504.
<https://doi.org/10.1016/j.molcel.2014.06.009>
- Du, J., Zhong, X., Bernatavichute, Y. V., Stroud, H., Feng, S., Caro, E., ... Jacobsen, S. E. (2012). Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell*, 151(1), 167–180. <https://doi.org/10.1016/j.cell.2012.07.034>
- Dutta, A., Choudhary, P., Caruana, J., & Raina, R. (2017). JMJ27, an Arabidopsis H3K9 histone demethylase, modulates defense against *Pseudomonas syringae* and flowering time. *The Plant Journal*, 91(2), 1015–1028.
<https://doi.org/10.1111/tpj.13623>
- Ebbs, M. L., Bartee, L., & Bender, J. (2005). H3 lysine 9 methylation is maintained on a transcribed inverted repeat by combined action of SUVH6 and SUVH4 methyltransferases. *Molecular and Cellular Biology*, 25(23), 10507–10515. <https://doi.org/10.1128/MCB.25.23.10507>
- Ebbs, M. L., & Bender, J. (2006). Locus-specific control of DNA methylation by the Arabidopsis SUVH5 histone methyltransferase. *The Plant Cell*, 18(5), 1166–1176. <https://doi.org/10.1105/tpc.106.041400.1>
- Eskeland, R., Czermin, B., Bonaldi, T., Regula, T., & Imhof, A. (2004). The N-Terminus of Drosophila SU (VAR) 3 - 9 Mediates Dimerization and Regulates Its Methyltransferase Activity †, 3740–3749.
<https://doi.org/10.1021/bi035964s>
- Gehring, M., Bubbs, K. L., & Henikoff, S. (2009). Extensive Demethylation of Repetitive Elements During Seed Development Underlies Gene Imprinting. *Science*, 324(June), 1447–1451.
- Gouil, Q., & Baulcombe, D. C. (2016). DNA Methylation Signatures of the Plant Chromomethyltransferases. *PLoS Genetics*, 12(12), 1–17.
<https://doi.org/10.1371/journal.pgen.1006526>
- Gu, T., Han, Y., Huang, R., Mcavoy, R. J., & Li, Y. (2016). Identification and characterization of histone lysine methylation modifiers in *Fragaria vesca*. *Scientific Reports*, 6(23581), 1–13. <https://doi.org/10.1038/srep23581>
- Han, Y.-F., Dou, K., Ma, Z.-Y., Zhang, S.-W., Huang, H.-W., Li, L., ... He, X.-J. (2014). SUVH2 is involved in transcriptional gene silencing by associating

- with SNF2-related chromatin-remodeling proteins in Arabidopsis. *Cell Research*, 24(12), 1445–1465. <https://doi.org/10.1038/cr.2014.156>
- Harris, C. J., Scheibe, M., Wongpalee, S. P., Liu, W., Zhong, Z., Yen, L., ... Jacobsen, S. E. (2018). A DNA methylation reader complex that enhances gene transcription. *Science*, 2(6419), 1182–1186. <https://doi.org/10.1126/science.aar7854>
- He, Y., & Ecker, J. R. (2015). Non-CG Methylation in the Human Genome. *Annual Review of Genomic and Human Genetics*, 16, 15–77. <https://doi.org/10.1146/annurev-genom-090413-025437>
- Herr, A. J., Jensen, M. B., Dalmay, T., & Baulcombe, D. C. (2005). RNA polymerase IV directs silencing of endogenous DNA. *Science (New York, N.Y.)*, 308(2002), 118–120. <https://doi.org/10.1126/science.1106910>
- Honys, David, and David Twell. (2004). Transcriptome Analysis of Haploid Male Gametophyte Development in Arabidopsis. *Genome biology* 5(11): R85. Web.
- Huerta-Cepas, J., Serra, F., & Bork, P. (2016). ETE 3: Reconstruction, analysis and visualization of phylogenomic data. *Mol Biol Evol.* <https://doi.org/10.1093/molbev/msw046>
- Ibarra, C. A., Feng, X., Schoft, V., Hsieh, T.-F., Uzawa, R., Rodrigues, J. A., ... Zilberman, D. (2012). Active DNA Demethylation in Plant Companion Cells Reinforces Transposon Methylation in Gametes. *Science*, 337(September), 1360–1364.
- Inagaki, S., Miura-kamio, A., Nakamura, Y., Lu, F., Cui, X., Cao, X., ... Kakutani, T. (2010). Autocatalytic differentiation of epigenetic modifications within the Arabidopsis genome. *The EMBO Journal*, 29(20), 3496–3506. <https://doi.org/10.1038/emboj.2010.227>
- Ingouff, M., Rademacher, S., Holec, S., Xin, N., Readshaw, A., Foo, S. H., ... Berger, F. (2010). Zygotic Resetting of the HISTONE 3 Variant Repertoire Participates in Epigenetic Reprogramming in Arabidopsis. *Current Biology*, 20(23), 2137–2143. <https://doi.org/10.1016/j.cub.2010.11.012>
- Jackson, J. P., Lindroth, A. M., Cao, X., & Jacobsen, S. E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature*, 416(April), 556–560. <https://doi.org/10.1038/nature731>
- Johnson, L. M., Du, J., Hale, C. J., Bischof, S., Feng, S., Chodavarapu, R. K., ... Jacobsen, S. E. (2014). SRA- and SET-domain-containing proteins link RNA

- polymerase V occupancy to DNA methylation. *Nature*, 507(7490), 124–128.
<https://doi.org/10.1038/nature12931>
- Johnson, L. M., Bostick, M., Zhang, X., Kraft, E., Henderson, I., Callis, J., & Jacobsen, S. E. (2007). The SRA Methyl-Cytosine-Binding Domain Links DNA and Histone Methylation. *Current Biology*, 17, 379–384.
<https://doi.org/10.1016/j.cub.2007.01.009>
- Johnson, L. M., Law, J. a., Khattar, A., Henderson, I. R., & Jacobsen, S. E. (2008). SRA-domain proteins required for DRM2-mediated de novo DNA methylation. *PLoS Genetics*, 4(11).
<https://doi.org/10.1371/journal.pgen.1000280>
- Jurkowska, R. Z., Qin, S., Kungulovski, G., Tempel, W., Liu, Y., Bashtrykov, P., ... Jeltsch, A. (2017). H3K14ac is linked to methylation of H3K9 by the triple Tudor domain of SETDB1. *Nature Communications*, 1–13.
<https://doi.org/10.1038/s41467-017-02259-9>
- Kabelitz, T., Brzezinka, K., Friedrich, T., Górka, M., Graf, A., Kappel, C., & Bäurle, I. (2016). A JUMONJI Protein with E3 Ligase and Histone H3 Binding Activities Affects Transposon Silencing in Arabidopsis 1, 171(May), 344–358. <https://doi.org/10.1104/pp.15.01688>
- Kankel, M. W., Ramsey, D. E., Stokes, T. L., Flowers, S. K., Haag, J. R., Jeddeloh, J. A., ... Richards, E. J. (2003). Arabidopsis MET1 Cytosine Methyltransferase Mutants. *Genetics*, 163(3), 1109–1122.
- Kim, J., Hee, J., Richards, E. J., Min, K., & Woo, H. R. (2014). Arabidopsis VIM Proteins Regulate Epigenetic Silencing by Modulating DNA Methylation and Histone Modification in Cooperation with MET1. *Molecular Plant*, 7(9), 1470–1485. <https://doi.org/10.1093/mp/ssu079>
- Kim, N. S. (2017). The genomes and transposable elements in plants : are they friends or foes ? *Genes & Genomics*, 39(4), 359–370.
<https://doi.org/10.1007/s13258-017-0522-y>
- Kraft, E., Bostick, M., Jacobsen, S. E., Callis, J., Biology, C., Avenue, O. S., ... Angeles, L. (2008). ORTH / VIM proteins that regulate DNA methylation are functional ubiquitin E3 ligases. *The Plant Journal*, 56(5), 704–715.
<https://doi.org/10.1111/j.1365-313X.2008.03631.x>
- Krauss, V. (2008). Glimpses of evolution: Heterochromatic histone H3K9 methyltransferases left its marks behind. *Genetica*, 133(1), 93–106.
<https://doi.org/10.1007/s10709-007-9184-z>

- Krauss, V., Fassl, A., Fiebig, P., Patties, I., & Sass, H. (2006). The evolution of the histone methyltransferase gene Su(var)3-9 in metazoans includes a fusion with and a re-fission from a functionally unrelated gene. *BMC Evolutionary Biology*, 6(18), 1–15. <https://doi.org/10.1186/1471-2148-6-18>
- Law, J. a., Ausin, I., Johnson, L. M., Vashisht, A. a., Zhu, J. K., Wohlschlegel, J. a., & Jacobsen, S. E. (2010). A Protein Complex Required for Polymerase V Transcripts and RNA- Directed DNA Methylation in Arabidopsis. *Current Biology*, 20(10), 951–956. <https://doi.org/10.1016/j.cub.2010.03.062>
- Lehnertz, B., Ueda, Y., Derijck, A. A. H. A., Braunschweig, U., Perez-burgos, L., Kubicek, S., ... Peters, A. H. F. M. (2003). Suv39h -Mediated Histone H3 Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin. *Current Biology*, 13, 1192–1200. [https://doi.org/10.1016/S0960-9822\(03\)00432-9](https://doi.org/10.1016/S0960-9822(03)00432-9)
- Li, F., Brouwer, P., Carretero-paulet, L., Cheng, S., Vries, J. De, Delaux, P., ... Pryer, K. M. (2018). Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nature Plants*, 4(July), 460–472. <https://doi.org/10.1038/s41477-018-0188-8>
- Li, X., Harris, C. J., Zhong, Z., Chen, W., Liu, R., Jia, B., ... Du, J. (2018). Mechanistic insights into plant SUVH family H3K9 methyltransferases and their binding to context-biased non-CG DNA methylation. *Proceedings of the National Academy of Sciences*, 115(37), 201809841. <https://doi.org/10.1073/pnas.1809841115>
- Loyola, A., Tagami, H., Bonaldi, T., Roche, D., Quivy, J. P., Nakatani, Y., ... Almouzni, G. (2009). The HP1a–CAF1–SetDB1-containing complex provides H3K9me1 for Suv39-mediated K9me3 in pericentric heterochromatin. *EMBO Reports*, 10(7). <https://doi.org/10.1038/embo.2009.90>
- Luger, K., Mäder, a W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389, 251–260. <https://doi.org/10.1038/38444>
- Malagnac, F., Bartee, L., & Bender, J. (2002). An Arabidopsis SET domain protein required for maintenance but not establishment of DNA methylation. *EMBO J*, 21(24), 6842–6852.
- McClintock, B. (1950). The Origin and Behaviour of Mutable Loci in Maize. *Proc. Natl Acad. Sci. USA*, 36(6), 344–355. <https://doi.org/10.1073/pnas.36.6.344>

- Min, J., Zhang, X., Cheng, X., Grewal, S. I. S., & Xu, R. (2002). Structure of the SET domain histone lysine methyltransferase Clr4. *Nature Structural Biology*, 9(11). <https://doi.org/10.1038/nsb860>
- Miura, A., Nakamura, M., Inagaki, S., Kobayashi, A., Saze, H., & Kakutani, T. (2009). An Arabidopsis jmjC domain protein protects transcribed genes from DNA methylation at CHG sites. *The EMBO Journal*, 28(8), 1078–1086. <https://doi.org/10.1038/emboj.2009.59>
- Miura, A., Yonebayashi, S., Watanabe, K., Toyama, T., Shimada, H., & Kakutani, T. (2001). Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. *Nature*, 411(May), 212–214. <https://doi.org/10.1038/35075612>
- Penterman, J., Zilberman, D., Huh, J. H., Ballinger, T., Henikoff, S., & Fischer, R. L. (2007). DNA demethylation in the Arabidopsis genome. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 6752–6757. <https://doi.org/10.1073/pnas.0701861104>
- Pontvianne, F., Blevins, T., Chandrasekhara, C., Feng, W., Stroud, H., Jacobsen, S. E., ... Pikaard, C. S. (2012). Histone methyltransferases regulating rRNA gene dose and dosage control in Arabidopsis. *Genes and Development*, 26(9), 945–957. <https://doi.org/10.1101/aad.182865.111>
- Qian, Y., Xi, Y., & Cheng, B. (2014). Identification and characterization of the SET domain gene family in maize. *Mol Biol Rep*, 41(3), 1341–1354. <https://doi.org/10.1007/s11033-013-2980-x>
- Qin, F. J., Sun, Q. W., Huang, L. M., Chen, X. S., & Zhou, D. X. (2010). Rice SUVH histone methyltransferase genes display specific functions in chromatin modification and retrotransposon repression. *Molecular Plant*, 3(4), 773–782. <https://doi.org/10.1093/mp/ssq030>
- Rahman, M. A., Kristiansen, P. E., Veiseth, S. V., Andersen, J. T., Yap, K. L., Zhou, M., ... Aalen, R. B. (2014). The Arabidopsis Histone Methyltransferase SUVH4 Binds Ubiquitin via a Domain with a Four-Helix Bundle Structure. *Biochemistry*, 53, 2091–2100. <https://doi.org/10.1021/bi401436h>
- Rajakumara, E., Law, J. A., Simanshu, D. K., Voigt, P., Johnson, L. M., Reinberg, D., ... Jacobsen, S. E. (2011). A dual flip-out mechanism for 5mC recognition by the Arabidopsis SUVH5 SRA domain and its impact on DNA methylation and H3K9 dimethylation in vivo. *Genes and Development*, 25, 137–152. <https://doi.org/10.1101/gad.1980311>

- Ramachandran, S., & Henikoff, S. (2015). Replicating nucleosomes. *Science Advances*, 1(7), 1–10. <https://doi.org/10.1126/sciadv.1500587>
- Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z., Schmid, M., ... Jenuwein, T. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature*, 406(10), 593–599. <https://doi.org/10.1038/35020506>
- Saze, H. (2008). Epigenetic memory transmission through mitosis and meiosis in plants. *Seminars in Cell & Developmental Biology*, 19, 527–536. <https://doi.org/10.1016/j.semcdb.2008.07.017>
- Saze, H., Shiraishi, A., Miura, A., & Kakutani, T. (2008). Control of Genic DNA Methylation by a jmjC Domain-Containing Protein in *Arabidopsis thaliana*. *Science*, 319(January), 462–465.
- Schoft, V. K., Chumak, N., Mosiolek, M., Slusarz, L., Komnenovic, V., Brownfield, L., ... Tamaru, H. (2009). Induction of RNA-directed DNA methylation upon decondensation of constitutive heterochromatin. *EMBO Reports*, 10(9), 1015–1021. <https://doi.org/10.1038/embor.2009.152>
- Schotta, G., Ebert, A., Krauss, V., Fischer, A., Hoffmann, J., Rea, S., ... Reuter, G. (2002). Central role of *Drosophila* SU (VAR) 3 ± 9 in histone H3-K9 methylation and heterochromatic gene silencing. *The EMBO Journal*, 21(5). <https://doi.org/10.1093/emboj/21.5.1121>
- Schmid, M., Davison, T., Henz, S. *et al.* A gene expression map of *Arabidopsis thaliana* development. *Nat Genet* 37, 501–506 (2005). <https://doi.org/10.1038/ng1543>
- Shook, M. S., & Richards, E. J. (2014). VIM proteins regulate transcription exclusively through the MET1 cytosine methylation pathway. *Epigenetics*, 9(7), 980–986. <https://doi.org/10.4161/epi.28906>
- Sievers F., Wilm A., Dineen D., Gibson T.J., Karplus K., Li W., Lopez R., McWilliam H., Remmert M., Söding J., Thompson J.D. and Higgins D.G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539 <https://doi.org/10.1038/msb.2011.75>
- Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., ... Jacobsen, S. E. (2014). Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. *Nature Structural & Molecular Biology*, 21(1), 64–72. <https://doi.org/10.1038/nsmb.2735>

- Stroud, H., Greenberg, M. V. C., Feng, S., Bernatavichute, Y. V., & Jacobsen, S. E. (2013). Comprehensive Analysis of Silencing Mutants Reveals Complex Regulation of the Arabidopsis Methylome. *Cell*, 152(1–2), 352–364. <https://doi.org/10.1016/j.cell.2012.10.054>
- Sun, L., Jing, Y., Liu, X., Li, Q., Xue, Z., Cheng, Z., ... Qian, W. (2020). Heat stress-induced transposon activation correlates with 3D chromatin organization rearrangement in Arabidopsis. *Nature Communications*, 11(1886). <https://doi.org/10.1038/s41467-020-15809-5>
- Sun, Q., & Zhou, D. (2008). Rice jmjC domain-containing gene JMJ706 encodes H3K9 demethylase required for floral organ development. *Proc. Natl Acad. Sci. USA*, 105(36). <https://doi.org/10.1073/pnas.0805901105>
- Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., Ohki, M., ... Shinkai, Y. (2002). G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes and Development*, 16(3), 1779–1791. <https://doi.org/10.1101/gad.989402.also>
- Thorstensen, T., Fischer, A., Sandvik, S. V., Johnsen, S. S., Grini, P. E., Reuter, G., & Aalen, R. B. (2006). The Arabidopsis SUV4 protein is a nucleolar histone methyltransferase with preference for monomethylated H3K9. *Nucleic Acids Research*, 34(19), 5461–5470. <https://doi.org/10.1093/nar/gkl687>
- Tittel-elmer, M., Bucher, E., Broger, L., Mathieu, O., Paszkowski, J., & Vailant, I. (2010). Stress-Induced Activation of Heterochromatic Transcription. *PLoS Genetics*, 6(10), 1–11. <https://doi.org/10.1371/journal.pgen.1001175>
- Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korgel, G., & Reuter, G. (1994). The protein encoded by the Drosophila position- combines domains of antagonistic regulators of homeotic gene complexes. *The EMBO Journal*, 13(16), 3822–3831. <https://doi.org/10.1002/j.1460-2075.1994.tb06693.x>
- Qin, Yuan et al.(2009) Penetration of the Stigma and Style Elicits a Novel Transcriptome in Pollen Tubes, Pointing to Genes Critical for Growth in a Pistil. *PLoS genetics* 5(8): e1000621. Web.
- Veiseth, S. V., Rahman, M. a., Yap, K. L., Fischer, A., Egge-Jacobsen, W., Reuter, G., ... Thorstensen, T. (2011). The SUV4 histone lysine methyltransferase binds ubiquitin and converts H3K9me1 to H3K9me3 on transposon chromatin in Arabidopsis. *PLoS Genetics*, 7(3). <https://doi.org/10.1371/journal.pgen.1001325>

- Vitte, C., Fustier, M., Alix, K., & Tenaillon, M. I. (2014). The bright side of transposons in crop evolution. *Briefings in Functional Genomics*, 13(4). <https://doi.org/10.1093/bfpg/elu002>
- Wierzbicki, A. T., Ream, T. S., Haag, J. R., & Pikaard, C. S. (2009). RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nature Genetics*, 41(5), 630–634. <https://doi.org/10.1038/ng.365>
- Woo, H. R., Dittmer, T. a., & Richards, E. J. (2008). Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in Arabidopsis. *PLoS Genetics*, 4(8). <https://doi.org/10.1371/journal.pgen.1000156>
- Woo, H. R., Pontes, O., Pikaard, C. S., & Richards, E. J. (2007). VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. *Genes and Development*, 21(3), 267–277. <https://doi.org/10.1101/gad.1512007.Bestor>
- Xiao, X., Zhang, J., Li, T., Fu, X., Satheesh, V., Niu, Q., ... Lei, M. (2019). A group of SUVH methyl-DNA binding proteins regulate expression of the DNA demethylase ROS1 in Arabidopsis. *Journal of Integrative Plant Biology*, 61(2), 110–119. <https://doi.org/10.1111/jipb.12768>
- Yang, J., Yuan, L., Yen, M., Zheng, F., Ji, R., Peng, T., ... Liu, X. (2020). SWI3B and HDA6 interact and are required for transposon silencing in Arabidopsis. *The Plant Journal*, 102, 1–14. <https://doi.org/10.1111/tpj.14666>
- Yu, C., Tai, R., Wang, S., Yang, P., Luo, M., Yang, S., ... Wu, K. (2017). HISTONE DEACETYLASE6 Acts in Concert with Histone Methyltransferases SUVH4 , SUVH5 , and SUVH6 to Regulate Transposon Silencing. *The Plant Cell*, 29(8), 1970–1983. <https://doi.org/10.1105/tpc.16.00570>
- Zemach, A., Kim, M. Y., Hsieh, P. H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., ... Zilberman, D. (2013). The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell*, 153(1), 193–205. <https://doi.org/10.1016/j.cell.2013.02.033>
- Zhang, C., Du, X., Tang, K., Yang, Z., Pan, L., Zhu, P., ... Zhu, J. (n.d.). Arabidopsis AGDP1 links H3K9me2 to DNA methylation in heterochromatin. *Nature Communications*, (2018). <https://doi.org/10.1038/s41467-018-06965-w>
- Zhang, L., & Ma, H. (2012). Complex evolutionary history and diverse domain organization of SET proteins suggest divergent regulatory interactions. *New*

- Phytologist*, 195, 248–263. <https://doi.org/10.1111/j.1469-8137.2012.04143.x>
- Zhang, X., Germann, S., Blus, B. J., Khorasanizadeh, S., Gaudin, V., & Jacobsen, S. E. (2007). The Arabidopsis LHP1 protein colocalizes with histone H3 Lys27 trimethylation. *Nature Structural & Molecular Biology*, 14(9), 869–871. <https://doi.org/10.1038/nsmb1283>
- Zhang, X., Tamaru, H., Khan, S. I., Horton, J. R., Keefe, L. J., Selker, E. U., & Cheng, X. (2002). Structure of the Neurospora SET Domain Protein DIM-5, a Histone H3 Lysine Methyltransferase, *III*, 117–127.
- Zhang, X., Yang, Z., Khan, S. I., Horton, J. R., Tamaru, H., Selker, E. U., & Cheng, X. (2003). Structural Basis for the Product Specificity of Histone Lysine Methyltransferases. *Molecular Cell*, 12(1), 177–185. [https://doi.org/10.1016/s1097-2765\(03\)00224-7](https://doi.org/10.1016/s1097-2765(03)00224-7)
- Zhao, Q., Zhang, J., Chen, R., Wang, L., Li, B., Cheng, H., ... Wong, J. (2016). Dissecting the precise role of H3K9 methylation in crosstalk with DNA maintenance methylation in mammals. *Nature Communications*, (7), 12464. <https://doi.org/10.1038/ncomms12464>
- Zhao, Q., Lin, R., Li, L., Chen, S., & He, X. (2019). A methylated-DNA-binding complex required for plant development mediates transcriptional activation of promoter methylated genes. *Journal of Integrative Plant Biology*, 61(2), 120–139. <https://doi.org/10.1111/jipb.12767>
- Zhao, S., Cheng, L., Gao, Y., Zhang, B., Zheng, X., Wang, L., ... Sun, Q. (2019). Plant HP1 protein ADCP1 links multivalent H3K9 methylation readout to heterochromatin formation. *Cell Research*, 29, 54–66. <https://doi.org/10.1038/s41422-018-0104-9>
- Zhong, X., Du, J., Hale, C. J., Gallego-Bartolome, J., Feng, S., Vashisht, A. a., ... Jacobsen, S. E. (2014). Molecular mechanism of action of plant DRM de novo DNA methyltransferases. *Cell*, 157(5), 1050–1060. <https://doi.org/10.1016/j.cell.2014.03.056>
- Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T., & Henikoff, S. (2007). Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genetics*, 39(1), 61–69. <https://doi.org/10.1038/ng1929>

Chapter 2: An amino-terminal conserved motif controls *Arabidopsis* SUVH6

H3 Lysine 9 methyltransferase function through two distinct mechanisms

Jenna Kotak and Judith Bender

This is currently an unpublished chapter. Experiments were performed by JK.

Manuscript written by JK and JB.

Abstract

In plants, H3K9 methylation (H3K9me) and 5-methyl cytosine (5meC) are paired heterochromatin modifications that mediate transcriptional silencing of transposons and repeats. *Arabidopsis thaliana* H3K9 methyltransferases SUVH4, SUVH5, and SUVH6 bind 5meC and catalyze H3K9me, while CMT 5meC methyltransferases bind H3K9me and catalyze 5meC. SUVH4 is the major contributor to H3K9me. SUVH5 and SUVH6 are members of a separate SUVH6 clade that evolved from the SUVH4 clade and make more specialized contributions to H3K9me. We identified amino-terminal structural features unique to the SUVH6 clade versus the SUVH4 clade. We then used a SUVH6 transgene complementation strategy to test the functional consequences of mutations in these features. We found that a conserved 21-amino acid motif contributes to two aspects of SUVH6 function: basic residues function in nuclear localization, while a conserved aromatic residue is necessary for full activity at a SUVH6 preferred locus. This unique conserved functional motif provides insights into the diversification of SUVH functions in plants.

Introduction

In eukaryotic genomes, genome stability and gene function are threatened by non-genic sequences such as transposable elements and repeats. Left unchecked, transposable elements and repeats can be transcribed and subsequently expand throughout the genome, resulting in mutations and

rearrangements (McClintock, 1950). In addition, transcription and translation of non-genic sequences wastes cellular resources. To defend against these deleterious consequences, eukaryotes target transposons and repeats for heterochromatin modifications that silence transcription. Although heterochromatin-mediated transcriptional silencing is a fundamental genome defense mechanism described in animals, fungi, and plants, many aspects of how heterochromatin modifications are initiated and maintained on appropriate target regions remain to be elucidated.

Throughout eukaryotic kingdoms, methylation of histone 3 lysine 9 (H3K9me) is a fundamental mediator of transcriptional silencing and is catalyzed by enzymes with conserved H3K9 methyltransferase domains (Krauss, 2008, Schotta et al., 2003). To further enforce silencing, plants evolved a mechanism where H3K9me is maintained in an amplification loop with cytosine methylation (5meC) (Johnson et al., 2007). In this amplification loop, SUVH H3K9 methyltransferases are localized to regions of the genome carrying 5meC through a central SET and Ring finger Associated (SRA) domain that selectively flips out and binds to a 5meC nucleotide (Du et al., 2014). Correspondingly, chromomethyltransferase (CMT) clade cytosine methyltransferases are localized to regions of the genome carrying H3K9me through a central H3K9me-binding chromodomain (Du et al., 2012). CMT methyltransferases preferentially modify cytosines in the symmetric and asymmetric trinucleotide contexts CHG and CHH (where H = A, C, T), although they also contribute to dinucleotide CG methylation (Du et al., 2012, Stroud et al., 2014). Together, SUVH and CMT

enzymes efficiently maintain H3K9me and dense 5meC on heterochromatic target sequences, including highly repetitive arrays of pericentromeric sequences and ribosomal RNA-encoding genes, transposons, and other sequence duplications.

The SUVH4 H3K9 methyltransferase clade, named for the first functionally characterized member in *Arabidopsis thaliana*, initially arose in green algae (Zhang et al., 2012). However, over the evolution of angiosperms, the basal SUVH4 clade has diversified into three additional clades, also named for *Arabidopsis* proteins in each clade: SUVH6, SUVH2/9 and SUVH1 (Baumbusch et al., 2001, Zhang et al., 2012). The SUVH2/9 and SUVH1 clade proteins lack H3K9 methyltransferase activity, but can bind 5meC through the central SRA domain, thus serving as readers of 5meC (Johnson et al., 2008). The SUVH2/9 clade proteins act in the RNA-directed DNA methylation pathway, which initiates 5meC on new target sequences and contributes to the maintenance of 5meC in CHH contexts (Johnson et al., 2008). The SUVH1 clade proteins form a 5meC reader complex that binds 5meC of euchromatic transposable elements and promotes transcription of nearby genes (Harris et al., 2018). Like SUVH4 clade proteins, SUVH6 clade proteins have both H3K9 methyltransferase activity and functional 5meC-binding SRA domains (Ebbs & Bender, 2006, Johnson et al., 2007, Li et al., 2018). The SUVH6 clade is most closely related to the SUVH4 clade (Zhang et al., 2012).

Arabidopsis thaliana encodes a single *SUVH4* gene plus two SUVH6 clade genes *SUVH5* and *SUVH6*. Based on analysis of *suvh4*, *suvh5*, and *suvh6* mutants, SUVH4 contributes to most mono-methyl (H3K9me1) and dimethyl

(H3K9me2) modifications genome-wide (Li et al., 2018). SUVH5 and SUVH6 contribute lower levels of these modifications with locus-specific preferences (Ebbs & Bender 2006). The *suvh4* mutant is strongly depleted for H3K9me and associated 5meC at most target regions, whereas a *suvh5 suvh6* double mutant displays similar H3K9me and associated 5meC patterning to wild type at most target regions (Li et al., 2018, Pontvianne et al., 2012). Furthermore, analysis of *suvh4 suvh6* versus *suvh4 suvh5* double mutants shows that in each case, the remaining functional SUVH5 or SUVH6 H3K9 methyltransferase maintains different residual modification patterns (Ebbs & Bender, 2006). For example, the *suvh4 suvh5* mutant maintains similar levels of H3K9me1, H3K9me2, and 5meC to wild type at a transcribed inverted repeat gene duplication *PAII-PAI4*, whereas the *suvh4 suvh5* mutant is strongly depleted for these *PAII-PAI4* modifications, similarly to a *suvh4 suvh5 suvh6* mutant. These patterns indicate that *PAII-PAI4* is a preferred target sequence for SUVH6 over SUVH5. However, at an unlinked transcriptionally silenced *PAI* sequence duplication *PAI2*, there is strong depletion of H3K9me and associated 5meC in a *suvh4* mutant, indicating minimal additional contributions from either SUVH5 or SUVH6 at *PAI2*. Therefore, factors beyond the primary sequence of the target locus underlie the locus-specific preference of SUVH6 for the *PAII-PAI4* inverted repeat versus the *PAI2* duplication.

It remains unknown why *Arabidopsis* SUVH5 and SUVH6 have more restricted H3K9 methyltransferase activity than SUVH4, and more generally why angiosperms diversified the SUVH6 clade from the SUVH4 clade. To differences

that distinguish between the SUVH4 clade and SUVH6 clade, we used sequence comparisons to identify structural features unique to the SUVH6 clade. We then tested the function of these features by mutational analysis of a *SUVH6* transgene added back into a *suvh4 suvh5 suvh6* mutant background and assayed remodification of the SUVH6 preferred target locus *PA11-PA14*. This analysis identified a 21-amino acid conserved motif near the amino-terminus of SUVH6 that controls two different functions: basic residues function in nuclear localization, whereas a conserved aromatic residue is dispensable for nuclear localization but is necessary for remodification at the preferred *PA11-PA14* locus. These results support a model where the SUVH6 clade diversified from the SUVH4 clade by acquisition of this unique amino-terminal regulatory sequence.

Results

A short, conserved motif embedded in a long, disordered amino-terminal extension is characteristic of the SUVH6 clade of H3K9 methyltransferases

To identify amino acid sequence features that differentiate the SUVH6 clade from the SUVH4 clade of plant H3K9 methyltransferases, we used alignment and motif-finder programs to analyze sequences from angiosperms that had been previously assigned to the SUVH4 or SUVH6 clade based on phylogenetic analysis of the histone methyltransferase catalytic SET domains (Zhang et al., 2012). In particular, we focused on SUVH4 and SUVH6 clade

proteins from monocot and eudicot species with well-annotated genomes to reduce any partial sequences fragments.

In addition to the known functional 5meC-binding SRA and catalytic domains, we found a conserved 21-amino acid motif that occurs one to three times in the amino-terminal extensions of SUVH6 clade members, but not SUVH4 clade members (Figure 1A, Supplemental Figure 1). We searched for this amino-terminal conserved motif (ACM) in other angiosperm proteins using motif-finder programs FIMO and Genome Net. We detected significant matches in SUVH6 clade members and SUVH6-related truncations but did not detect significant matches consistently in any other functional class of proteins, including other SUVH clades (data not shown). The consensus ACM contains a cluster of four basic amino acids at the amino-terminal end, similar to a monopartite nuclear localization sequence (NLS) (Figure 1B). However, many SUVH6 clade proteins have fewer than four basic amino acids in this cluster (Figure 1C). The retention of the full 21-amino acid ACM across members of the SUVH6 clade suggests that this motif is important for SUVH6 function.

To better understand when the ACM emerged in the evolution of the SUVH6 clade, we analyzed SUVH proteins encoded by the bryophyte *Physcomitrella patens* and lycophyte *Selaginella moellendorffii*. We found that one of four SUVH proteins in *Physcomitrella* contains a diverged ACM, but none of the three SUVH proteins in *Selaginella* contains an ACM (Supplemental Figure 2). Therefore, the ACM may have emerged early in the evolution of plant SUVH proteins, but it did not become fixed as a necessary functional component

until the emergence of the SUVH6 clade. We also identified SUVH6-like proteins that contain ACMs in gymnosperms *Picea abies* and *Pseudotsuga menziesii* and in ferns *Azolla filiculoides* and *Salvinia culculata* (Supplemental Figure 2), indicating that the SUVH6 clade emerged during the evolution of monilophytes.

We also found that the majority of SUVH6 clade members have longer amino-terminal extensions than the SUVH4 clade members (Figure 1A). Upstream of the pre-SRA conserved alpha helices, SUVH6 clade members have amino-terminal extensions ranging from 269- 801 amino acids long. Similarly, the ACM-containing SUVH proteins have amino-terminal extensions ranging from 255-674 amino acids long in gymnosperms, and 164- 942 amino acids long in ferns (Supplemental Figure 2). In contrast, SUVH4 clade members have amino-terminal extensions ranging from 84- 224 amino acids long. The exception to this pattern is *Prunus persica* PrSUVH5, which has only 23 amino acids in the amino-terminal extension and lacks an ACM (Figure 1A). However, PrSUVH5 also lacks a carboxy-terminal post-SET domain necessary for H3K9 methyltransferase activity, suggesting that PrSUVH5 is a truncated non-functional gene duplication. Aside from the ACM, SUVH6 clade member amino-terminal sequences are not conserved with each other (Supplemental Figure 1). In addition, the divergent amino-terminal sequences are predicted to be highly disordered in *Arabidopsis* and other angiosperms (Figure 1D, Supplemental Figure 3). Overall, our analysis indicates that ACM sequences embedded in long disordered amino-terminal extensions are specific to the SUVH6 clade.

The ACM is necessary for full activity of *Arabidopsis* SUVH6

To test the functions of amino-terminal sequence features characteristic of the SUVH6 clade, we developed a transgene complementation strategy to assay *Arabidopsis thaliana* SUVH6 activity. In this assay we introduced *SUVH6* transgene variants into a *suvh4 suvh5 suvh6* (*456*) triple mutant background that is strongly depleted for H3K9me and associated 5meC genome-wide, including at the *PAI* gene duplications (Ebbs & Bender, 2006). As previously reported, *456* mutants maintain partial transcriptional silencing through other 5meC pathways and are viable and morphologically normal (Ebbs et al., 2005). We used a DNA gel blot assay with the 5meC-sensitive restriction enzyme *HincII* to assess 5meC patterns on the Wassilewskija (Ws) *PAI* gene duplications as a means to determine whether the *SUVH6* transgene was able to complement the *suvh6* deficiency (Supplemental Figure 4A). We previously determined that *PAI* 5meC detected by the *HincII* *PAI* DNA gel blot assay accurately reflects dense 5meC patterning as measured by bisulfite sequencing, and H3K9me1 and H3K9me2 as measured by chromatin immunoprecipitation with modification-specific antibodies (Ebbs & Bender, 2006). The *PAI* DNA gel blot simultaneously assays the SUVH6 preferred target *PAI1-PAI4* and the SUVH6 non-preferred target *PAI2*. Therefore, the *HincII* *PAI* DNA gel blot assay provides a sensitive and facile means to detect 5meC associated with H3K9me alterations at a preferred and non-preferred site.

To first determine whether a wild type *SUVH6* transgene could restore similar *PAI* 5meC patterns in a *456* background to the patterns observed for

endogenous SUVH6 in a *suvh4 suvh5* background, we generated transgenic lines with either native or epitope-tagged *SUVH6* driven by its native promoter. For the epitope-tagged constructs, we introduced either a 22-amino acid 3XFLAG epitope tag or a 238-amino acid green fluorescent protein (GFP) epitope tag just after codon 11 at the amino-terminus of SUVH6, 38 amino acids upstream of the ACM. We then determined *PAI* 5meC patterns with the *HincII* DNA gel blot assay (Supplemental Figure 4B). For each transgene construct, we identified several independent transgenic lines with genetic segregation of the transgene diagnostic of a single insertion site to avoid transgene silencing due to high copy numbers (Methods). We then used *SUVH6* transcript analysis to identify lines with similar or higher expression levels compared to endogenous *SUVH6* for further analysis (Supplemental Figure 5).

In each *PAI* sequence, *HincII* cleaves at a single site near the translational start codon at 5' ATGTCAACAG 3' sites, monitoring 5meC in the CAG and CTG contexts. Thus, at the *PAI1-PAI4* inverted repeat, *HincII* digestion can yield four sizes of cleavage products depending on whether the internal sites in both, one, or neither of the inverted repeat *PAI* sequences are modified by 5meC (Supplemental Figure 4A). At *PAI2*, *HincII* can yield two sizes of cleavage products depending on whether the single internal site is modified by 5meC. Another more diverged sequence duplication *PAI3* lacks an internal *HincII* site, and thus yields only a single cleavage product from flanking unmodified sites. *HincII* *PAI* DNA gel blot analysis of 456 transgenic lines transformed with *SUVH6*, *3XFLAG-SUVH6*, and *GFP-SUVH6* transgenes all displayed cleavage

patterns diagnostic of dense 5meC at *PAII-PAI4* but not at *PAI2*, similar to the patterns observed in a *suvh4 suvh5* mutant (Supplemental Figure 4B). Therefore, wild type *SUVH6* transgenes can restore modifications at the preferred target *PAII-PAI4* and this activity is not impaired by the insertion of *3XFLAG* or *GFP* amino-terminal epitope tags. Additionally, both native and epitope-tagged *SUVH6* transgenes do not promote over-methylation of non-preferred target *PAI2*. These results indicate that transgenic SUVH6 protein can restore H3K9me-associated 5meC on the *PAI* duplications with similar locus-specific patterns to the patterns maintained by endogenous SUVH6.

We next tested two amino-terminal sequence deletion mutant variants of *3XFLAG-SUVH6* in the *456* transgene complementation assay: deletion of the central most conserved 16-amino acids in the ACM (codons 51-66, ΔACM), and deletion of the long disordered region (LDR) between the ACM and the conserved pre-SRA alpha helices (codons 104-264 ΔLDR) (Figure 2A, Supplemental Figure 6). Using the *PAI HincII* DNA gel blot assay, we found that lines expressing the ΔACM construct displayed only slightly inhibited *PAII-PAI4 HincII* digestion patterns relative to the *456* parental strain, diagnostic of little remethylation of *PAII-PAI4*. (Figure 2). Therefore, the ΔACM mutation causes loss of SUVH6 function at this locus. In contrast, lines expressing the ΔLDR construct displayed strongly inhibited *PAII-PAI4 HincII* digestion patterns similar to wild type *SUVH6* control lines and a *suvh4 suvh5* control strain, diagnostic of full or nearly full remethylation of *PAII-PAI4*. Therefore, the ΔLDR mutation causes at most a weak impairment of SUVH6 function. None of the lines

expressing either the ΔACM or ΔLDR mutant constructs displayed alterations in the *PAI2 HincII* digestion patterns relative to the parental 456 strain, indicating that these mutations do not promote hyperactivity at a non-preferred SUVH6 target.

To further dissect ACM function, we tested two additional *SUVH6* transgene constructs carrying ACM missense mutations: mutation of the basic amino acids lysine and arginine at the third and fourth positions of the ACM to alanines (*K51A R52A*), and mutation of the aromatic amino acid phenylalanine at the thirteenth position of the ACM to alanine (*F61A*). The KR to AA missense mutations were designed to disrupt a cluster of basic residues that could constitute a nuclear import signal. The F to A missense mutation was designed to test function of the conserved sequences downstream of the basic cluster. Using the *HincII PAI* DNA gel blot assay, we found that lines expressing the *K51A R52A* construct displayed only slightly inhibited *PAI1-PAI4 HincII* digestion patterns relative to the 456 parental strain, diagnostic of little remethylation of *PAI1-PAI4*. (Figure 2B). Lines expressing the *F61A* construct consistently displayed *PAI1-PAI4 HincII* digestion patterns intermediate between the 456 parental strain and the wild type *SUVH6* control lines, diagnostic of partial remethylation of *PAI1-PAI4*. Therefore, both the basic cluster *K51A R52A* missense mutations and the *F61A* missense mutation impair SUVH6 function, with the basic cluster mutations likely causing a stronger defect. Transgene remodification results are summarized in Supplemental Figure 9.

To determine the effect of SUVH6 amino-terminal mutations on activity at a larger silenced region, we developed a new DNA gel blot assay to assess restoration of 5meC in non-CG contexts at 25S ribosomal RNA-encoding (rDNA) repeats in the *456* background, diagnostic of chromatin modifications generated by the CMT/SUVH pathway. In *Arabidopsis*, 18S, 5.8S, and 25S ribosomal RNA genes (rDNAs) lie in arrays of hundreds of copies on both chromosome 2 and chromosome 4, but the chromosome 2 copies are silenced by modification with H3K9me and 5meC whereas the chromosome 4 copies are unmodified and expressed (Chandrasekhara et al., 2015). In our DNA blot assay, we cleaved genomic DNA with the restriction enzyme *AvaII*, which is sensitive to 5meC at either the inner or outer positions of the recognition sequence 5' GGWCC 3', where W indicates A or T. *AvaII* cleaves six times in the 25S rDNA gene covered by our probe, generating a series of fragments ranging from 24 bp to 2.9 kb plus higher molecular weight products depending on the density of 5meC and monitoring 5meC in various CHH and CHG contexts. We used a qualitative assessment of the proportion of signal from the highest molecular weight species to the lower molecular weight bands within each sample to score 5meC patterning in different *suvh* mutants.

Relative to wild type Ws, we found that the proportion of the highest molecular weight 25S rDNA *AvaII* digestion species was partially reduced in a *suvh4 suvh6* mutant, more strongly reduced in a *suvh4 suvh5* mutant, and most strongly reduced in a *456* mutant (Figure 3B). These patterns indicate that across the 25S rDNA repeats, SUVH4 and SUVH5 maintain the majority of H3K9me-

associated 5meC, with only a low contribution from SUVH6. However, due to the highly repetitive nature of the 25S rDNA sequences, even the low contribution from SUVH6 can be detected above residual 5meC patterning in a *456* background. We observed similar *Ava*II digestion pattern trends in *suvh* mutants using an 18S rDNA probe (data not shown).

We next tested the ability of wild type and mutant *SUVH6* transgene constructs to restore 5meC at the SUVH6 non-preferred target 25S rDNA using the *Ava*II DNA gel blot assay. We found that lines expressing either wild type or mutant *SUVH6* constructs displayed slightly increased levels of the highest molecular weight species diagnostic of restored 5meC relative to the *456* parental strain, with similar cleavage patterns to the *suvh4 suvh5* mutant control (Figure 3). Therefore, all of the amino-terminal mutations including the Δ ACM mutation are likely to only partially impair SUVH6 function. However, none of the transgenic lines displayed inhibited *Ava*II cleavage patterns to the higher levels observed in the wild type or *suvh4 suvh6* controls. Therefore, none of the transgene constructs promotes over-methylation of the 25S rDNA repeats.

The ACM controls SUVH6 function through two distinct mechanisms

In order to determine if the amino-terminal deletion mutant constructs affect SUVH6 localization, we used fluorescence confocal microscopy analysis of live root tips in *456* lines expressing wild type versus mutant GFP epitope-tagged *SUVH6* transgene constructs (Figure 4). Lines expressing wild type *GFP-SUVH6*

displayed nuclear-localized GFP signals consistent with SUVH6 function as a chromatin-modifying enzyme, and consistent with our observation that the GFP epitope tag does not interfere with full remethylation of *PAI1-PAI4* (Supplemental Figure 4B). Similarly, lines expressing the *GFP-ΔLDR* mutant construct displayed nuclear-localized GFP signals (Figure 4). In contrast, lines expressing the *GFP-ΔACM* mutant construct displayed predominantly cytoplasmic GFP signals. Additionally, transgenic 456 lines expressing a truncation transgene that encodes a GFP fusion to only the first 67 amino acids of SUVH6, ending at consensus ACM residue 19, display a nuclear GFP signal (Supplemental Figure 7). Therefore, the ACM is sufficient to promote nuclear localization. However, the GFP signal for the truncation construct was consistently significantly weaker than the signal from full-length *GFP-SUVH6*, perhaps because the truncation protein has reduced stability.

We further analyzed localization patterns with lines expressing *GFP-SUVH6* constructs carrying the *K51A R52A* or *F61A* ACM missense mutations (Figure 4). Lines expressing the *GFP-K51A R52* mutant construct displayed predominantly cytoplasmic GFP signals similarly to the *GFP-ΔACM* lines. Therefore, the *K51A R52A* missense mutations are sufficient to impair nuclear localization. In contrast, lines expressing the *GFP-F61A* mutant construct displayed predominantly nuclear GFP signals similarly to the wild type control lines (Figure 4).

To determine whether the *F61A* mutation impairs sub-nuclear localization of SUVH6, we stained root tips with the DNA-binding fluorescent dye DAPI and

determined both DAPI and GFP signals using fluorescence confocal microscopy for the GFP constructs with nuclear localized patterns (Figure 5). In lines expressing wild type *GFP-SUVH6* constructs, DAPI staining revealed densely stained puncta corresponding to pericentromeric repetitive sequences that are enriched for H3K9me and associated 5meC (chromocenters; Fransz, et al., 2002). The GFP signals colocalized with the DAPI-stained chromocenters, consistent with SUVH6 localization to chromatin. Lines expressing either the *GFP-ALDR* or the *GFP-F61A* mutant constructs displayed similar patterns to wild type *GFP-SUVH6*, with colocalization of DAPI staining and GFP signals, including enrichment at chromocenters (Figure 5, Supplemental Figure 8). In rare cells undergoing mitosis, both *GFP-K51A R52A* and *GFP-F61A* lines displayed GFP signals corresponding to condensed chromosomes, similar to that observed for the wild type construct (Supplemental Figure 8). Therefore, the *GFP-K51A R52A* and *GFP-F61A* mutant proteins are able to associate with chromatin after nuclear envelope breakdown. Overall, imaging analysis reveals that the F61A mutation does not generally impair chromatin localization. Instead, the *F61A* defect in *PAII-PAI4* remethylation (Figure 2) could be due to impaired chromatin sub-localization and/ or impaired activity at this preferred target.

Discussion

Plant SUVH H3K9 methyltransferases have diversified over evolution from chlorophytic algae through angiosperms to mediate transcriptional silencing

of transposable elements, repeats, and other non-genic sequences (Zhang et al., 2012). SUVH proteins contain carboxy-terminal catalytic domains conserved with H3K9 methyltransferases from animal and fungal species, plus a central 5meC-binding SRA domain that enables enriched localization at heterochromatic target sequences. Genetic and biochemical analysis of *Arabidopsis* SUVH proteins indicates that members of the basal SUVH4 clade and the diversified SUVH6 clade retain both catalytic and 5meC binding functions, but SUVH4 maintains the majority of H3K9me genome-wide, whereas SUVH5 and SUVH6 make more specialized contributions (Ebbs & Bender, 2006). Nonetheless, the broad conservation of SUVH6 clade members in flowering plants indicates that they provide important components in the network of proteins that control transcriptional silencing. How SUVH6 clade methyltransferases integrate with SUVH4 methyltransferases remains an open question.

To gain insights into diversification of the SUVH6 clade from the SUVH4 clade, we conducted mutational analysis of amino-terminal sequence features unique to SUVH6 methyltransferases, using *Arabidopsis* SUVH6 as a model clade member. Our analysis was enabled by our previous identification of a unique SUVH6 preferred target region, the *PAI1-PAI4* inverted repeat, which is transcribed across from a distal promoter to produce small RNA signals for H3K9me and associated 5meC across the inverted repeat *PAI1-PAI4* sequences plus matching sequences in singlet gene *PAI2* and *PAI3* duplications elsewhere in the genome. We used a transgene complementation strategy to test the ability of different SUVH6 amino-terminal mutant variants to restore H3K9me-associated

5meC to *PAII-PAI4* in a *456* mutant background. This analysis revealed that the ACM, which is unique to SUVH6 clade proteins, contributes to SUVH6 function through two distinct mechanisms.

Deletion of the ACM or the *K51A R52A* double missense mutations in the ACM basic residue cluster impair SUVH6 nuclear localization (Figure 4) and remethylation of *PAII-PAI4* (Figure 2). Furthermore, a *GFP-SUVH6* construct truncated just downstream of the ACM is sufficient for nuclear localization of GFP (Supplemental Figure 7). Therefore, the cluster of four basic residues in the SUVH6 ACM could constitute a classic monopartite NLS (Kosugi et al., 2009). However, our analysis of ACM sequences in different plant species shows that not every ACM contains a four basic-residue cluster (Figure 1C, Supplemental Figure 2C). In addition, the consensus ACM (Figure 1B) does not match other previously identified plant NLS motifs (Chang et al., 2013). Moreover, the *K51A R52A* displays function in a 25S rDNA repeat remethylation assay (Figure 3). Therefore, SUVH6 clade members could contain more than one NLS and some clade members could have evolved away from dependence on the NLS in the ACM in favor of other NLS motifs.

Our analysis of the conserved aromatic residue in the ACM through the SUVH6 *F61A* mutation reveals a second function of the ACM. The *F61A* mutation does not affect nuclear localization or localization to chromocenters and mitotic chromosomes as detected with a GFP epitope-tagged fusion construct (Figure 4, Figure 5, Supplemental Figure 8), but the mutation impairs remethylation of *PAII-PAI4* (Figure 2). In addition, the *F61A* mutation does not

promote over-methylation of SUVH6 non-preferred targets *PAI2* or the 25S rDNA repeats (Figure 2, Figure 3). These observations argue against the *F61A* mutation affecting either nuclear localization or disrupting general chromatin localization as the mechanism underlying reduced *PAI1-PAI4* remethylation. The *F61A* mutation could instead disrupt ACM function in stabilizing SUVH6 protein accumulation. However, because we did not observe drastic changes in GFP intensities of epitope-tagged fusion protein signals and similar transgene expression ranges for wild type versus *F61A* mutant transgenic lines (Figure 4, Figure 5, Supplemental Figure 5), the *F61A* mutation would at most have a partial effect on steady-state levels of SUVH6. In addition, we consistently observed full *PAI1-PAI4* remethylation for control wild type *SUVH6* transgenic lines over a three-fold range of transgene expression levels, indicating that the *PAI1-PAI4* remethylation by SUVH6 is relatively insensitive to variations in expression levels (data not shown). Another possibility is that the *F61A* mutation impairs SUVH6 chromatin sub-localization, below the level of detection with our GFP-SUVH6 imaging, such that SUVH6 accumulates less efficiently at *PAI1-PAI4*. A third possibility is that the *F61A* mutation impairs a mechanism that promotes optimal catalytic activity. For example, F61 could function as a critical residue for an interaction, either in *trans* with another protein or small molecule, or in *cis* with another domain of SUVH6, to promote SUVH6 activity, at least at the preferred target *PAI1-PAI4*. Truncated recombinant SUVH6 protein with only the carboxy-terminal catalytic sequences is sufficient for H3K9 methyltransferase activity on histones *in vitro* (Ebbs et al., 2005, Li et al., 2018), but the ACM could

enhance this activity at preferred targets *in vivo* by improving reaction kinetics or processivity on a chromatin substrate. Although the mechanism underlying the *F61A* mutant phenotype remains to be elucidated, our observation that the ACM sequences beyond the basic cluster do not affect nuclear localization indicates that the ACM acts as a compact module that optimizes SUVH6 activity at both the level of nuclear localization and at the level of promoting chromatin sub-localization and/ or methyltransferase function within the nucleus.

Although *Arabidopsis* SUVH6 contains only one ACM, other SUVH6 clade members contain up to three ACMs (Figure 1, Supplemental Figure 2). Multiple ACMs could function additively to mediate localization and activity. However, because the ACM occurs in the context of unconserved and likely disordered sequences (Figure 1D, Supplemental Figure 3), duplicated ACMs could reflect a relatively high rate of mutations and rearrangements in this part of the protein relative to the SRA and catalytic domains, with only one of the ACMs maintaining function.

In contrast to deletion and missense mutations in the ACM, the Δ *LDR* mutation removing the majority of amino-terminal disordered sequences conferred at most a subtle impairment of SUVH6-mediated remethylation in transgene complementation assays (Figure 2, Figure 3). Nonetheless, our observation that long expanses of unconserved disordered sequences are a signature of SUVH6 clade members (Figure 1D, Supplemental Figure 3) argues that these sequences contribute to optimal SUVH6 function. One possibility is that these sequences provide a flexible linker for the ACM. However, the shorter

amino-terminal sequences of SUVH4 clade members are also predicted to be disordered despite the absence of an ACM in this clade (Supplemental Figure 3). An alternate possibility is that amino-terminal disordered sequences allow dynamic phase separation changes that concentrate SUVH6 at heterochromatin targets. For example, the *Arabidopsis* H3K9me2 binding protein ADCP1 undergoes phase separation *in vitro* and has been proposed to use phase separation compartments to recruit SUVH proteins (Zhao et al., 2019). Additionally, mammalian SUV39H1 undergoes phase separation *in vitro* and thus may use phase separated compartments to isolate heterochromatic target regions from the rest of the genome (Wang et al., 2019). However, the lack of substantial remethylation defects for the *ALDR* mutant construct (Figure 2, Figure 3) indicates that SUVH6 can still access the preferred target *PAII-PAI4* and the non-preferred target 25S rDNA repeats, even with removal of most of the disordered sequences potentially capable of phase separation (Supplemental Figure 6).

Our identification of an ACM in one of the SUVH proteins encoded by the bryophytic moss *Physcomitrella patens* but not in SUVH proteins encoded by the lycophytic spikemoss *Selaginella moellendorffii* suggests an evolutionary path for the emergence of the SUVH6 clade (Supplemental Figure 2). In this path, the initial ACM acquired in some bryophyte SUVH sequences could have provided a redundant NLS and/or activity-enhancing function. From here, in some lineages the ACM would be lost while redundant functional sequences were retained (as in *Selaginella*), while in other lineages the ACM would be retained while redundant functional sequences were lost as in ferns, gymnosperms, and angiosperms. Our

analysis of fern genomes indicates that by this stage in plant evolution and on through the emergence of gymnosperms and angiosperms, the sequence features of one or more ACMs in the context of a long, unconserved amino-terminal extension were fixed as important for SUVH6 function (Figure 1A, Supplemental Figure 2). Future elucidation of how the ACM regulates SUVH6 function beyond nuclear localization could provide the key to understanding differences between the SUVH6 clade and the SUVH4 clade proteins that define their non-redundant roles in maintaining plant heterochromatin.

Methods

Bioinformatic Analysis

Angiosperm SUVH sequences and annotations were described in Zhang et al., 2012. *Physcomitrella patens* and *Selaginella moellendorffii* SUVH sequences were retrieved using NCBI BLAST analysis with *Arabidopsis* SUVH4 (AtSUVH4) protein sequence as the query. Gymnosperm SUVH sequences were identified using PLAZA Gymnosperm BLAST with AtSUVH4 protein sequence as the query. Fern SUVH sequences were identified through BLAST of AtSUVH4 protein sequence to the fern genomes at FernBase. Previously-identified angiosperm sequences were obtained from Phytozome and NCBI (Zhang et al., 2012). Supplemental Table 1 contains sequence information and sources.

Angiosperm SUVH4 and SUVH6 clade protein sequences were used for ACM identification. The consensus ACM sequence was constructed with MEME motif finder. ACM sequences indicated by dark pink boxes in Figure 1A have significance scores $p < e-10$, whereas divergent ACM sequences indicated by light pink boxes in Figure 1A have significance scores of $e-5 > p > 5e-10$. Searches for other proteins containing the consensus ACM were performed with FIMO on the available species used in this study (Angiosperms, *Selaginella*, *Physcomitrella*, and *Chlamydomonas*) and against all reference genomes in Genome Net. Angiosperm SUVH6 clade proteins were aligned with Clustal Ω and visualized with Jalview. Disorder predictions were performed with Phyre2

and DISOPRED3. Domain positions were identified as characterized by Expasy Prosite. The two pre-SRA alpha helices were manually added based on Phyre2 alignments to AtSUVH4 and AtSUVH6 crystal structure pre-SRA helices.

Transgene constructs and plant strains

Transgenes were cloned into the pCAMBIA1300 plant transformation vector, which encodes a hygromycin resistance gene as a transgene selectable marker. The *SUVH6* promoter was amplified by PCR on genomic DNA as a 1.7 kb fragment extending from an endogenous *SalI* site at the distal end to an introduced *Bam*HI site at the proximal end immediately upstream of the SUVH6 translational start codon. The *SUVH6* coding region was amplified by PCR on genomic DNA as a 2.4 kb fragment extending from an introduced *Bam*HI site immediately upstream of the translational start codon to an introduced *Eco*RI site downstream of the translational stop codon. To allow insertion of epitope tags, a *Sna*BI site was introduced into the SUVH6 coding region with silent mutations at codon 11 using oligonucleotide-directed mutagenesis. A synthetic DNA cassette encoding the 3XFLAG 22 amino acid epitope tag (Sigma-Aldrich) was ligated into the *Sna*BI site to make the 3XFLAG-SUVH6 in-frame fusion. A PCR product encoding the 238 amino acid GFP epitope tag amplified from a pGGC014 GreenGate vector template (Lampropoulos et al., 2013) was inserted into the *Sna*BI site to make the GFP-SUVH6 in-frame fusion. PCR-mediated mutagenesis was used to engineer the ACM deletion from codons 51-66 and the LDR deletion

from 104-264 in the *SUVH6* coding sequence (Lee et al., 2004). *K51A R52A* and *F61A* mutations were introduced to the *SUVH6* coding sequence using oligonucleotide-directed mutagenesis. The *SUVH6* plant transformation constructs consist of the 1.7 kb *SalI*- *Bam*HI promoter fragment ligated to the *Bam*HI- *Eco*RI coding region fragment ligated to an *Eco*RI- *Sac*I 300 bp nopaline synthase transcriptional termination cassette inserted into the polylinker of pCAMBIA1300. The truncation construct to test whether the ACM is sufficient for nuclear localization (Supplemental Figure 7) was amplified by PCR from the full-length *GFP-SUV6* template to include just the GFP fusion in the context of *SUVH6* amino acids 1-67, driven by the *SUVH6* promoter in the pCAMBIA1300 vector.

The *pai1 suvh4 suvh5 suvh6* (456) strain was used as the parental background for all transgenic lines (Ebbs et al., 2006). Control double mutants *suvh4 suvh5* and *suvh4 suvh6* also contained the *pai1* mutation (Ebbs & Bender, 2006). Floral dip transformation via *Agrobacterium tumefaciens* was performed for each transgene into the 456 background as described (Clough & Bent, 1998). First generation transgenic T1 seedlings were identified by screening for hygromycin resistance. Second generation T2 self-fertilization progeny of each T1 plant were screened for 3:1 segregation of hygromycin resistance diagnostic of a single heterozygous transgene insertion site in the T1 parent. T3 progeny of individual T2 plants were tested for hygromycin resistance to identify true-breeding lines homozygous for the transgene. Homozygous transgenic lines were used for genomic DNA, RNA,

and microscopy analysis, except for primary T1 transgenic lines expressing a *GFP-SUVH6 1-67* truncation construct shown in Supplemental Figure 7.

RNA Analysis

RNA was prepared from leaf tissue from four-week-old homozygous transgenic lines, with two separate RNA preparations made for each line (Melquist et al., 1999). RNA gel blots were prepared in duplicate with one blot probed with β -tubulin as a loading control. The second blot was probed with a SUVH6 sequence. Most transgenic lines were probed against the SUVH6 amino-terminal region, but because the ΔLDR line lacks that sequence, a SUVH6 probe complementary to the SUVH6 SRA was used for this construct. The SUVH6 amino-terminal probe covers 560 bp from the translational start codon. The SUVH6 SRA 441 bp probe was amplified by PCR, using the forward primer 5' GTAGTCGGAACAAAGTAAAGGAAACTC 3', and the reverse primer 5' GGTCTTCAGGCTCTTTCAGTTCC 3'.

DNA Gel Blot Analyses

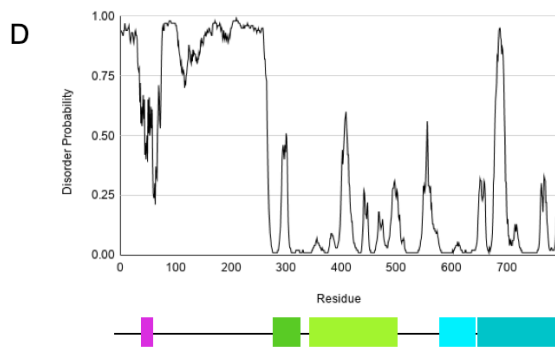
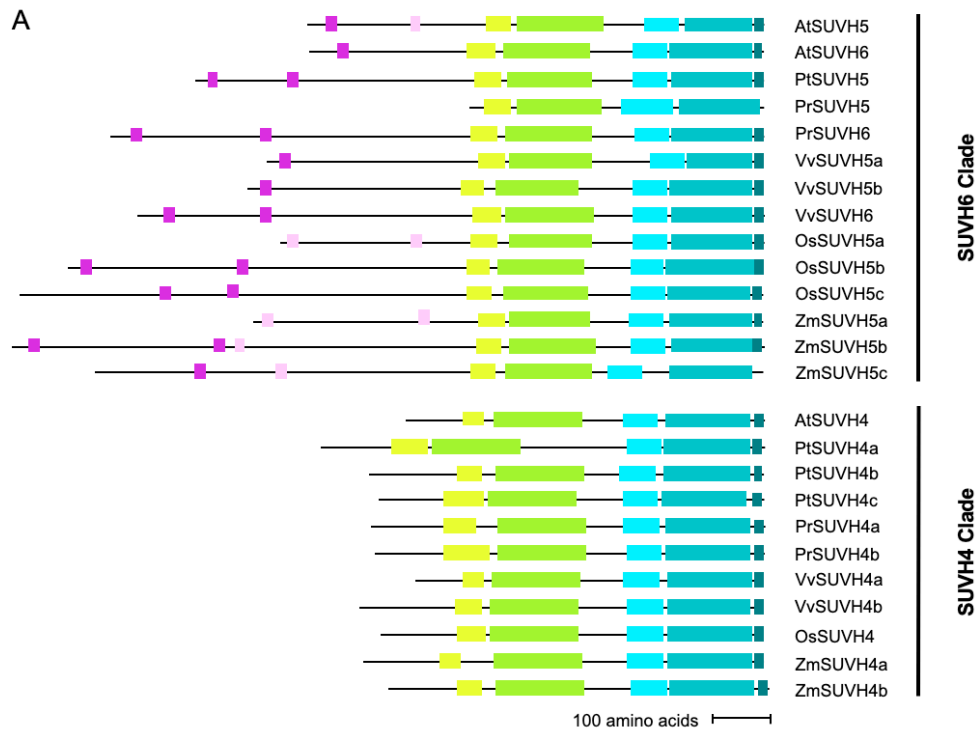
Genomic DNA was extracted from four-week-old plants as described (Melquist et al., 1999). 5meC remethylation at the *PAI* loci was assayed using DNA gel blots (Melquist et al., 1999). Briefly, genomic DNA from transgenic lines was cleaved with *HincII*, run on an agarose gel, depurinated in acid, and transferred to

GeneScreen Plus (Perkin Elmer) membranes overnight. The *PAI* probe is an 0.7 kb internal *PstI* fragment from a *PAII* cDNA (Melquist et al., 1999). 5meC remethylation at the 25S rDNA repeats was similarly assayed using genomic DNA cleaved with *AvaII*. Membranes were probed with a 998bp 25S probe amplified via PCR with forward primer 5' ACCCCAGGTCAGGCGGGATTACC 3' and reverse primer 5' CGATAGAACTCGCGTCCGAGCTCC 3'.

Microscopy

Seedlings were grown on plain Plant Nutrient Medium with 0.5% sucrose and 0.75% agar petri plates sealed with Parafilm for 8-9 days. Live seedlings were gently pulled from the agar plates, mounted in water, and imaged on the Zeiss LSM 800 Confocal Laser Scanning microscope. DAPI staining was performed on nuclear localized samples with 1 µg/ml DAPI in a 0.1% Triton X-100 solution. Roots were soaked in DAPI for 30 minutes on a slide, washed with water three times for 10 minutes, and mounted in water. Unstained whole root tips were imaged at 20X magnification and DAPI-stained whole root tips were imaged at 40X magnification. The DAPI signal was collected in the 410-470 nm range and the GFP signal was collected in the 495-599nm range with auto-exposure. For the *GFP-SUVH6 1-67* samples shown in Supplemental Figure 7, 15 first generation T1 transgenic seedlings were examined. For this construct the GFP signal was substantially weaker than for full-length *GFP-SUVH6* and required a laser intensity of 50% rather than the 10% used for all other constructs. Out of the 15

T1 seedlings, 11 had a GFP signal above background levels with a nuclear localization pattern (three representative seedlings shown in Supplemental Figure 7), whereas four seedlings had no GFP signal above background.



C

AtSUVH6	KFKRRKVFVRDFFPGCGSRA	*
AtSUVH5 1	KSKRFKVAASEFSPDFGSIT	*
AtSUVH5 2	ELGRTISVGRDLSNMGSKF	*
PtSUVH5 1	KFKRRRVSAVRDFFPGCGPLA	*
PtSUVH5 2	YPPRRRISAVRDFFPGCGPNA	*
PrSUVH6 1	KLKRRRISAVRDFFPGCGRFA	*
PrSUVH6 2	YPPRRRVSAVRDFFLLCGRNV	*
VvSUVH5a	NFNRPVDAVRDWPGRYGKFI	*
VvSUVH5b	KFRPKVDVRDWPRLCGKFT	*
VvSUVH6 1	RYKRRKVSARDFFPGCGPLA	*
VvSUVH6 2	YPPRRKISARDFFPGCGRNA	*
OsSUVH5a 1	GPRRYKGLVPWRFPQGFVRFP	*
OsSUVH5a 2	LPWRFPQVGFKRSFSKAFCSDS	*
OsSUVH5b 1	GVKRRAVSAKRSWPLCGRFP	*
OsSUVH5b 2	PPKRRRAISARKFPDRCGRAA	*
OsSUVH5c 1	PPKRRRVSAKRQFPDPFGRDS	*
OsSUVH5c 2	LPMRRTVSATDRFTAGCGRDA	*
ZmSUVH5a 1	GVRRCKILVPWRFPQGFVRQP	*
ZmSUVH5a 2	VPWRFPQARYKRFSDFAGSNN	*
ZmSUVH5b 1	RLTPSRVSAKRSWPPCGGRFP	*
ZmSUVH5b 2	LPKRRRAVSAVRKFPPGCGRTI	*
ZmSUVH5b 3	DGVLVSPISRFAPSFGMSA	*
ZmSUVH5c 1	PPKRRRAVSVTRQFPFASRWTS	*
ZmSUVH5c 2	RFFPKPEIISTRTLLPGHGK	*

Figure 1: The ACM is specific to the SUVH6 clade and is positioned within amino-terminal extensions. (A) Angiosperm SUVH4 and SUVH6 clade protein domain structures. Dark pink boxes indicate ACM sequences with $p < e-10$ and light pink boxes indicate diverged ACM sequences with $e-5 < p < e-10$ (Methods). Light green boxes indicate the pre-SRA domain, dark green boxes indicate the SRA domain, light teal boxes indicate the pre-SET domain, medium teal boxes indicate the SET domain, and dark teal boxes indicate the post-SET domain. At= *Arabidopsis thaliana*, Pt= *Populus trichocarpa*, Pr= *Prunus persica*, Vv= *Vitis vinifera*, Os= *Oryza sativa*, Zm= *Zea mays*. (B) Angiosperm ACM consensus sequence with residue weight indicative of frequency. (C) Angiosperm ACM sequences from proteins shown in A. Asterisks indicate ACMs in significant agreement with the consensus ACM, corresponding to dark pink boxes in A. (D) Disorder prediction plot generated with DISOPRED3 for AtSUVH6.

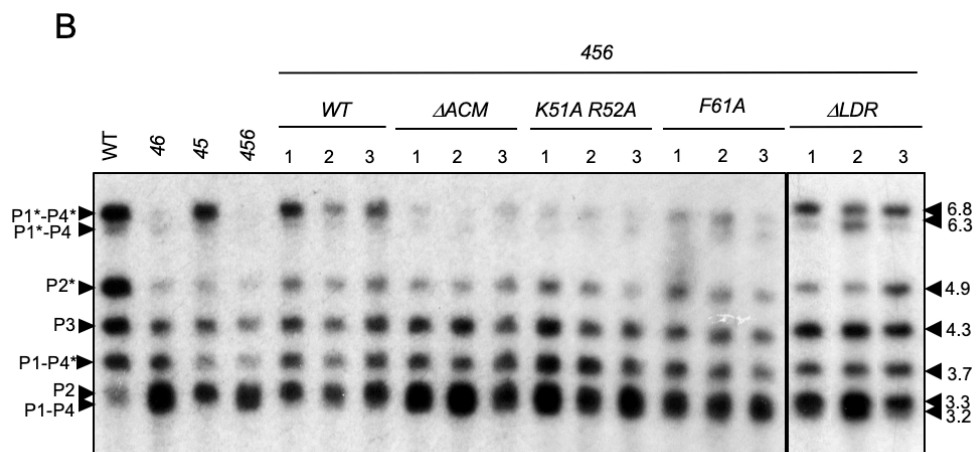
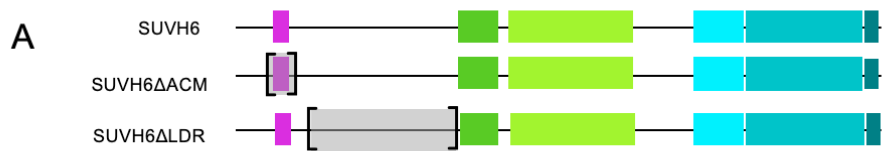
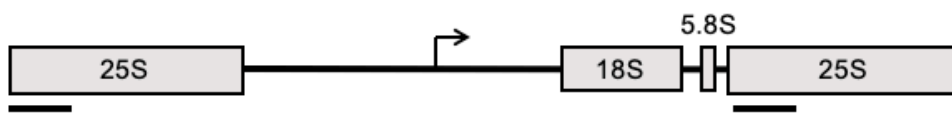


Figure 2: The ACM is required for SUVH6 activity at the *PAI1-PAI4*. (A) *SUVH6* constructs are shown with grey boxes indicating deleted regions. SUVH6 domains are color-coded as in Figure 1A. (B) *PAI HincII* DNA gel blot analysis of *SUVH6* transgene construct lines in the *456* background. Left margin labels indicate *PAI HincII* digestion products corresponding to *PAI1-PAI4* = P1-P4, *PAI2* = P2, and *PAI3* = P3, with asterisks indicating bands corresponding to 5meC-inhibited cleavage products. Right margin labels indicate sizes of digestion products in kb. All transgene constructs consist of the *SUVH6* promoter driving the *SUVH6* coding sequence (Methods), with WT indicating wild type *SUVH6* and the other constructs containing the indicated deletion or missense mutations in the row above the numbered lanes. WT label on the leftmost lane indicates wild type Ws, *45* indicates *pai1 suvh4 suvh5*, and *46* indicates *pai1 suvh4 suvh6* control strains. Numbered lanes indicate independent transgenic lines for each construct.

A



B

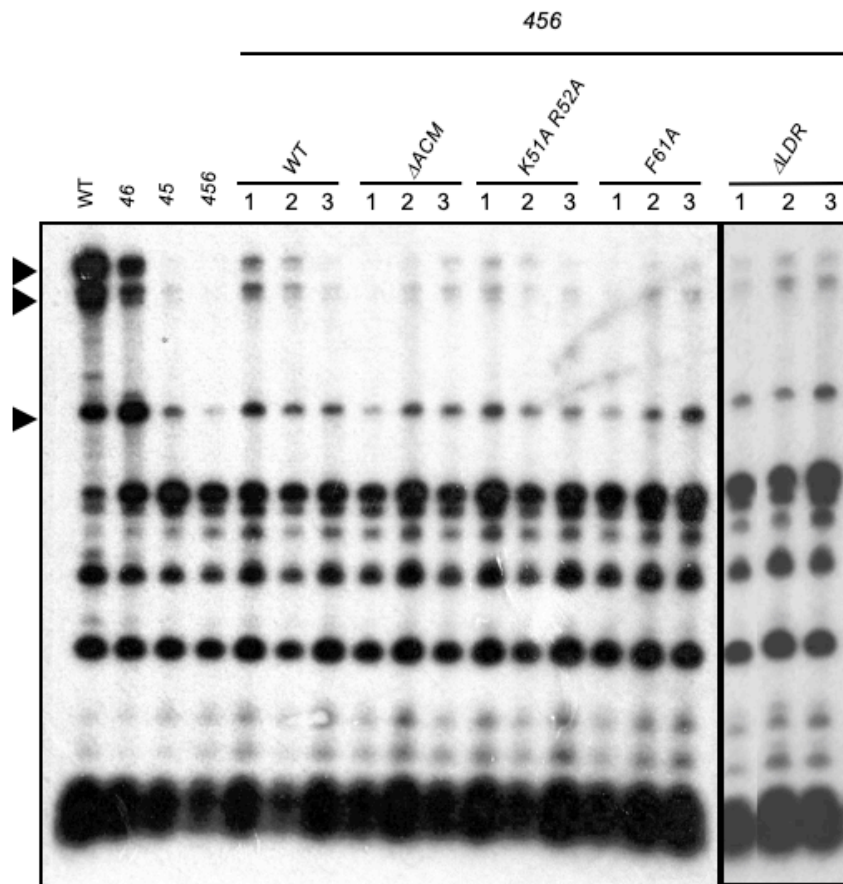


Figure 3: SUVH6 transgenes only weakly remethylate 25S rDNA repeats. (A) 45S rDNA repeat drawn with 18S, 5.8S, and 25S sequences labeled. The arrow indicates the transcriptional start site. The black bar indicates the position of the 25S rDNA probe. (B) 25S rDNA repeat *Ava*II DNA gel blot analysis of *SUVH6* transgene construct lines in the *456* background. Black arrowheads indicate bands corresponding to 5meC-inhibited cleavage products. Labels for transgene constructs and control strains are as in Figure 2. Sample numbers correspond to sample numbers in Figure 2.

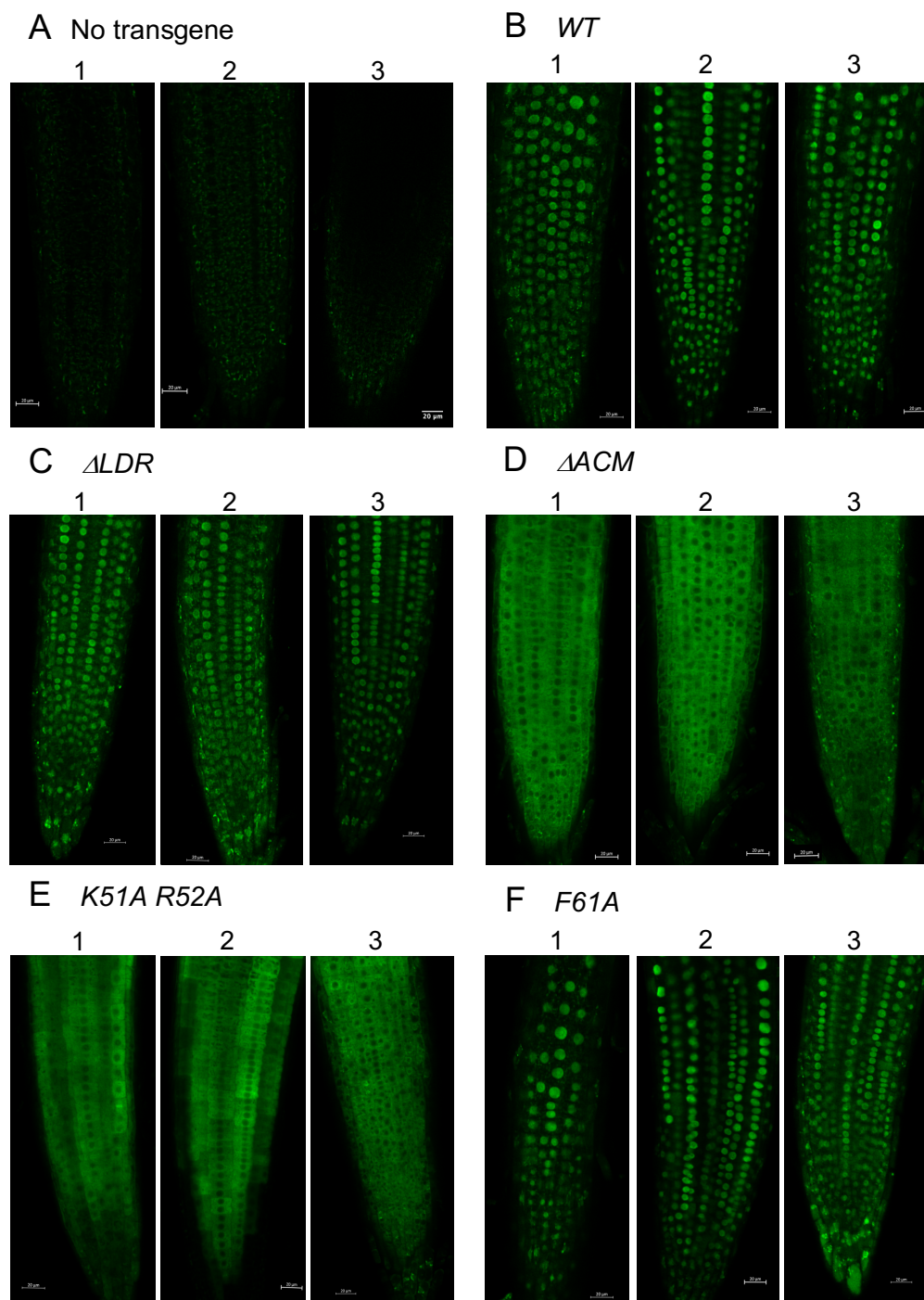


Figure 4: SUVH6 nuclear localization is mediated through the ACM *K51 R52* residues. Live root tissues from three independent transgenic lines per construct imaged at 20X magnification in (A) *456* parental background, or *456* with transgene constructs (B) *GFP-SUVH6*, (C) *GFP-SUVH6 Δ LDR*, (D) *GFP-SUVH6 Δ ACM*, (E) *GFP-SUVH6 K51A R52A*, (F) *GFP-SUVH6 F61A*.

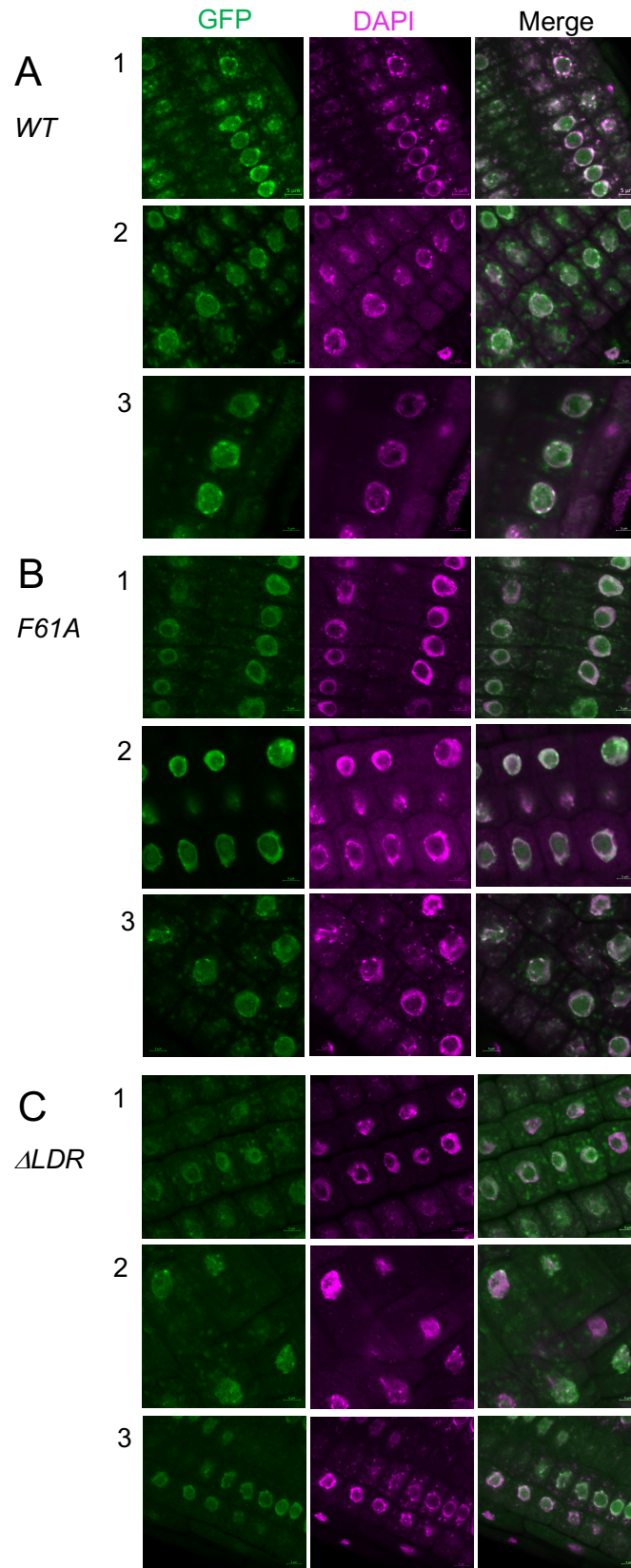
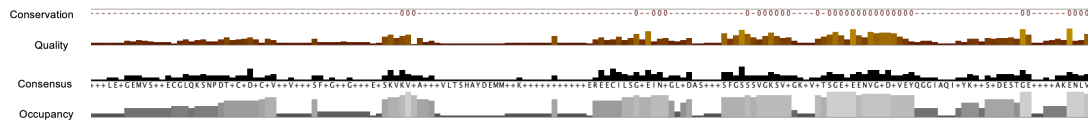


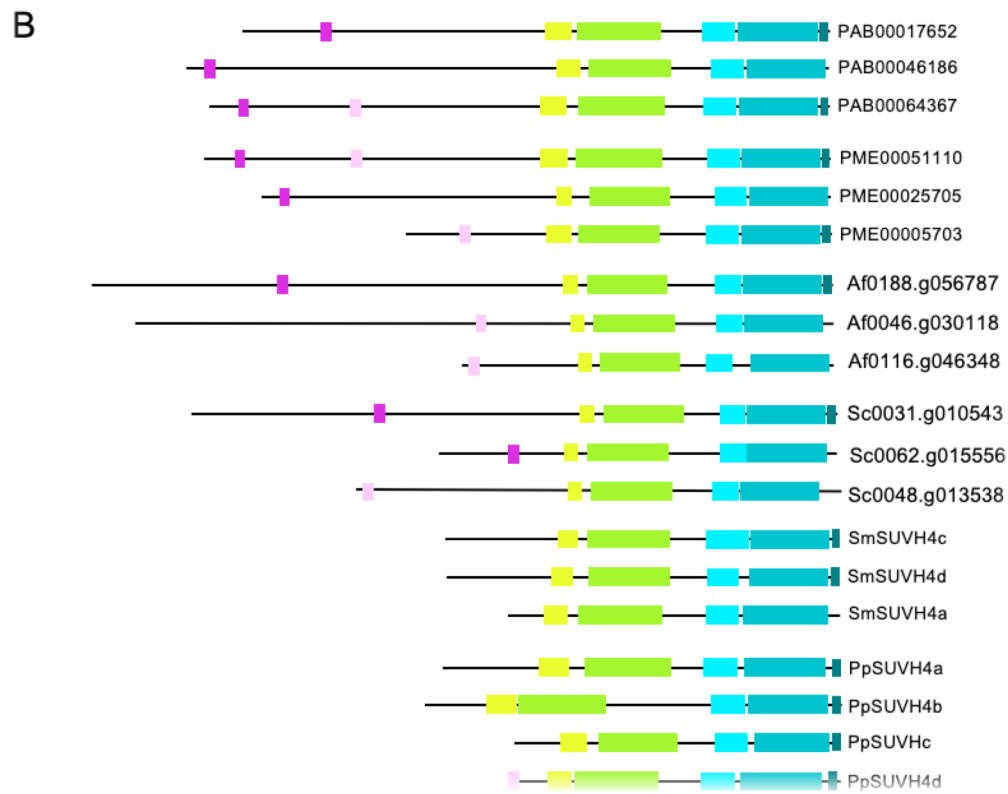
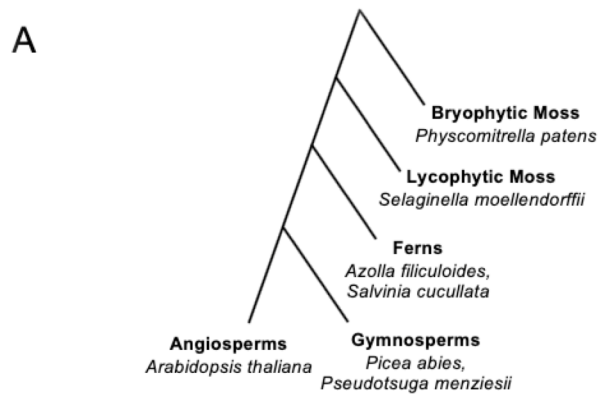
Figure 5: SUVH6 amino-terminal mutations do not impair chromatin localization.

DAPI-stained root tissues imaged at 40X magnification for nuclear localized constructs. DAPI signal is magenta and GFP signal is green. Three independent transgenic lines per construct are shown, with sample numbers corresponding to the sample numbers in Figure 4. (A) *GFP-SUVH6*, (B) *GFP-SUVH6 Δ LDR*, (C) *GFP-SUVH6 F61A*.





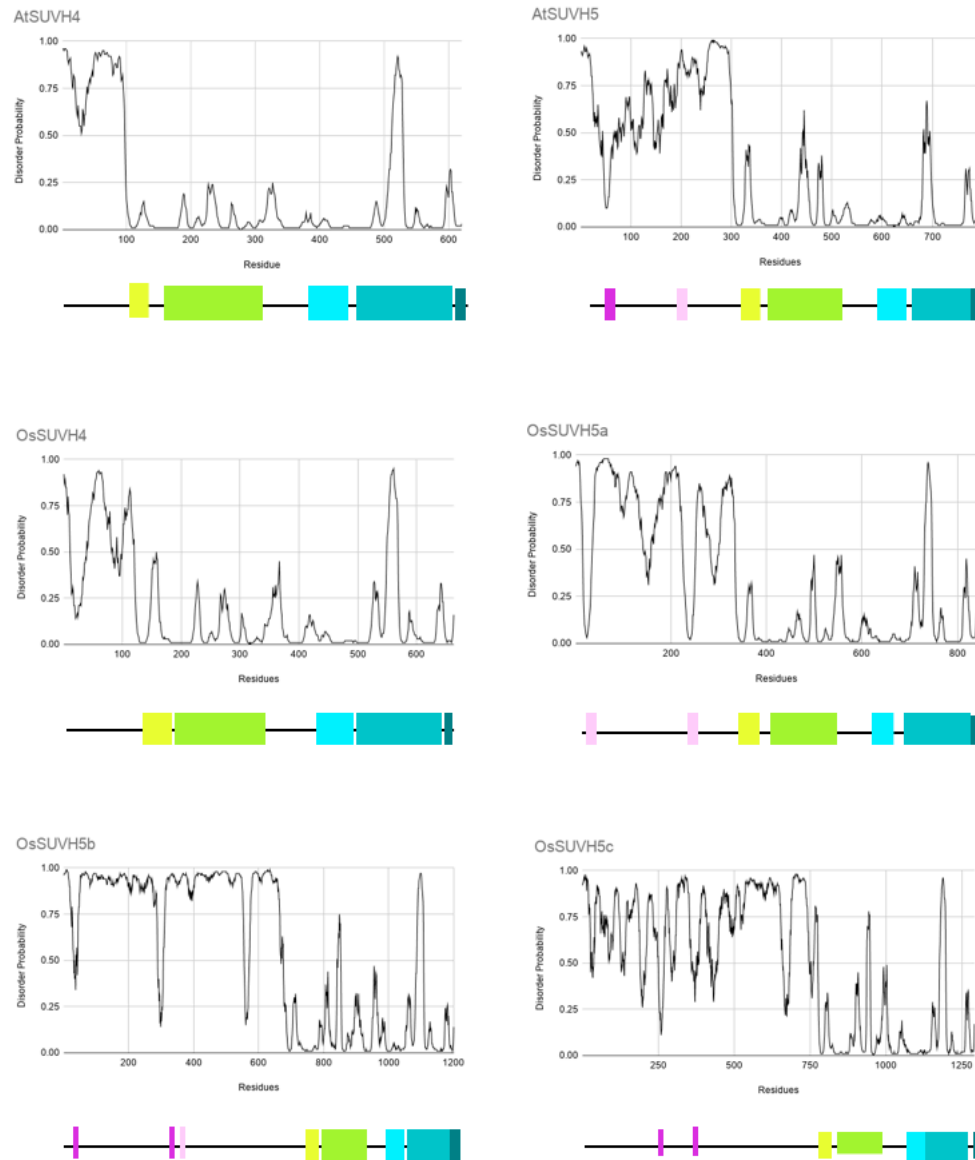
Supplemental Figure 1: Angiosperm SUVH6 clade proteins have diverged sequences in the amino-terminal extensions. SUVH6 proteins aligned with the degree of consensus indicated below the alignment. Domains are drawn below the alignment in the same color code as described in Figure 1, based on the AtSUVH6 domain boundaries and ACM site.



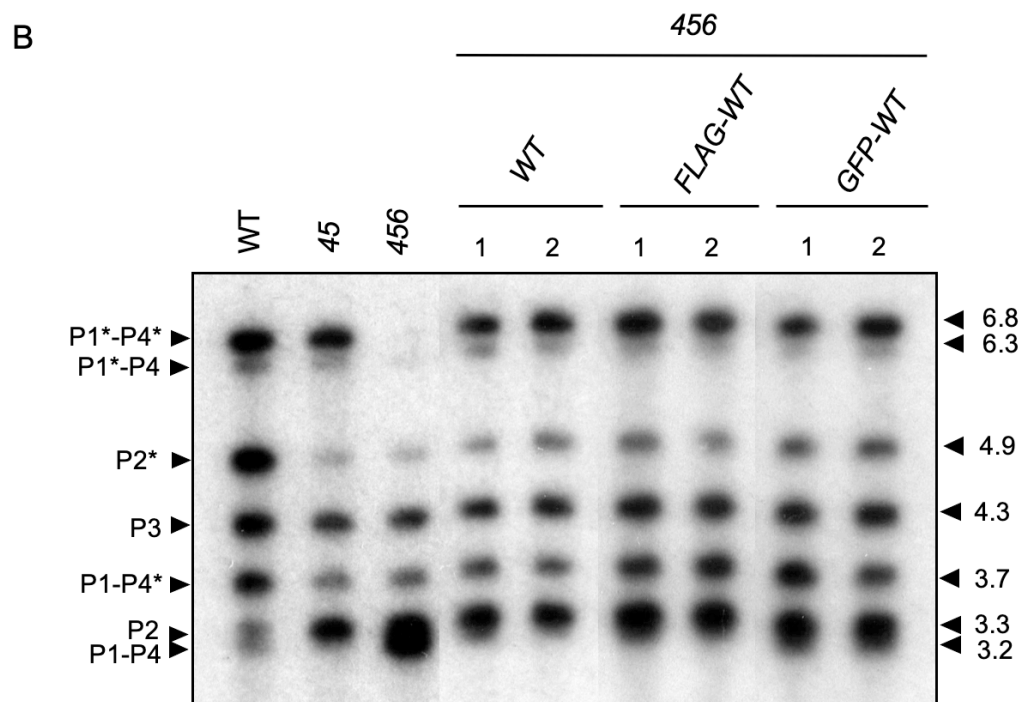
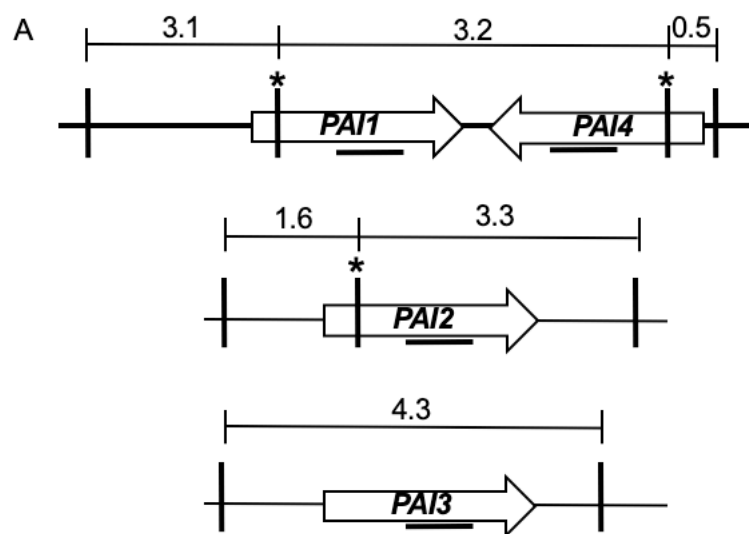
C

ID	Sequence	P-value	ID	Sequence	P-value
PAB00017652	KLQKPIRPSAVRDFPLGCGRYA	1.37e-19	Af0188.g056787	GKRKRLRSSAIRDYPEGCGRIS	8.16e-16
PAB00046186	AKSKKS RPSAFRDFPVGCGRNA	2.05e-23	Af0046.g030118	RVKKAQRKSAIRDLPDVARSV	7.51e-5
PAB00064367	PKPKKRKPSAIRDFPPGCGRRA	7.06e-27	Af0116.g046348	LPMKLQRSTAVRDYDPDGCVR	4.12e-10
PAB00064367	IKFSPRPVSACRDFPFGCGRNA	2.17 e-13	Sc0031.g010543	GNRKRQRSSAVRDYPEGCARIS	4.26e-15
PME00051110	PNPKKRKPSAIRDFPLGCGRLA	4.10e-25	Sc0062.g015556	RASKSRRRS AVRDFFAGSFKPG	8.73e-13
PME00051110	IKFPSRLVSACRDFPFGCGRNA	9.16 e-15	Sc0048.g013538	MRNGKPRPCAVRDFFHYLPV	4.37e-5
PME00025705	AKSKKS RPSAFRDFPLGCGRNA	1.17e-23	PpSUVH4d	MDRDKSTVAVRDYPPGCCRLP	3.71e-12
PME00005703	SKEKRAKRS AFRNSVKSFGKNN	1.81 e-5			

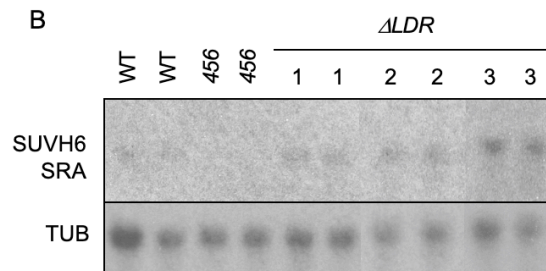
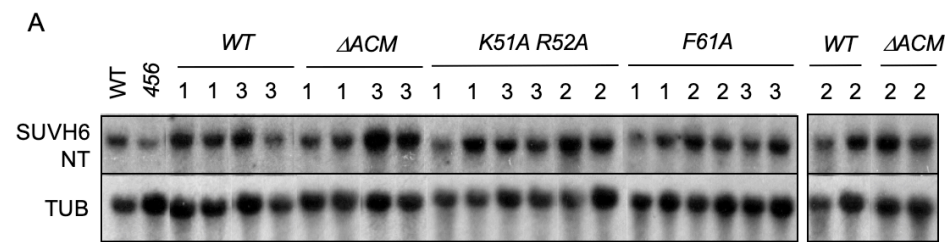
Supplemental Figure 2: The ACM sequence occurs in some SUVH proteins encoded by non-angiosperm plants. (A) Simplified evolutionary tree of bryophytes, lycophytes, gymnosperms, and angiosperms indicating relationships for species used in this analysis. (B) Pre-angiosperm SUVH proteins with ACMs. Domains are drawn to scale with the same color code as in Figure 1A. For *Physcomitrella patens* (Pp) and *Selaginella moellendorffii* (Sm), all the SUVH proteins encoded in the genome are shown. For the gymnosperms, *Picea abies* (PAB) and *Pseudotsuga menziesii* (PME), and ferns, *Azolla filiculoides* (Af) and *Salvinia cucullata* (Sc), only SUVH proteins including an ACM as identified by MEME are shown. PAB and PME proteins are listed by ID number in PLAZA. Af and Sc IDs as described in FernBase.



Supplemental Figure 3: SUVH4 and SUVH6 clade protein amino-terminal extensions are predicted to be disordered. Disorder probability as calculated by DISOPRED3 at each residue. Dicot *Arabidopsis thaliana* (At) and monocot *Oryza sativa* (Os) SUVH4 and SUVH6 clade proteins are shown. Domains are color-coded as in Figure 1A.



Supplemental Figure 4: Epitope tags do not impair SUVH6 remethylation of *PAI1-PAI4*. (A) *PAI* loci *HincII* restriction map with *HincII* sites drawn as vertical bars. Asterisks indicate sites that are demethylated in a *456* line versus the wild type. Fragment sizes are listed above each map in kb. The *PAI* probe is indicated by horizontal bar below each map. (B) *PAI HincII* DNA gel blot analysis of *SUVH6* transgene construct lines in the *456* background. Left and right margin labels are as in Figure 2. All transgene constructs consist of the *SUVH6* promoter driving the *SUVH6* coding sequence (Methods), with WT indicating wild type *SUVH6* and the other constructs containing the indicated epitope tag in the row above the numbered lanes. WT label on the leftmost lane indicates wild type Ws, and *45* indicates *pai1 suvh4 suvh5*.



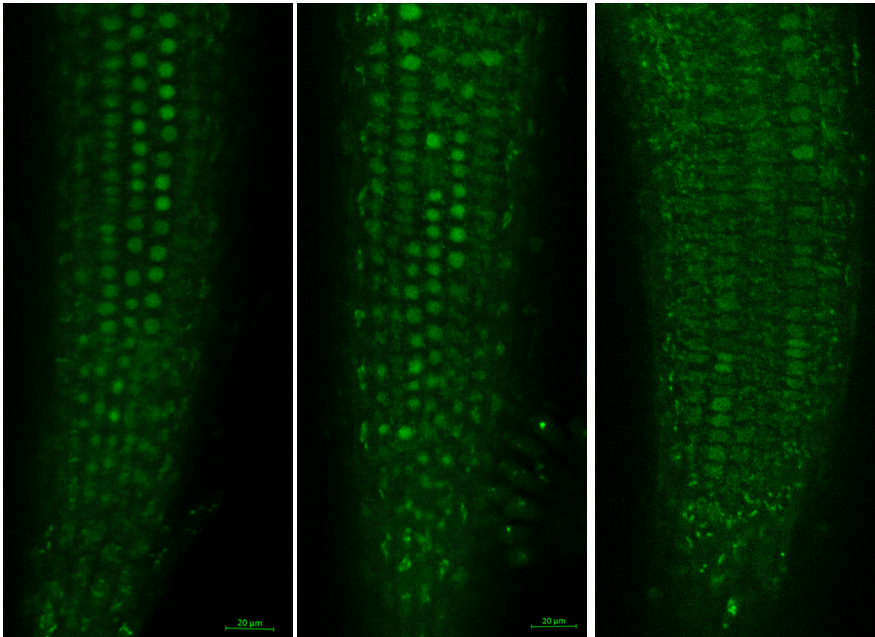
Supplemental Figure 5: Transgenic lines express similar or higher *SUVH6* transcript levels than endogenous *SUVH6* in a wild type background. RNA gel blot analysis with SUVH6 or Tubulin probes are shown. Transgene construct and control strain labels are as in Figure 2. Two independent RNA extractions were assayed for each strain. (A) Samples probed with a SUVH6 amino-terminal probe (SUVH6NT). Transgene-expressed SUVH6 transcript levels were estimated by calculating band intensities using ImageJ software, dividing SUVH6 band intensity by the Tubulin band intensity to determine a normalized value of SUVH6 transcript levels, and then subtracting the intensity calculated for the 456 parental strain from the total intensity calculated in the transgenic sample. All lines were within about a two-fold range of transcript levels. The *suvh6* allele in the 456 strain is an insertional disruption near the 3' end of the transcript in the region encoding the SET domain. This insertion mutation does not result in an obvious change in transcript length, but does decrease *suvh6* transcript accumulation relative to wild type. (B) Samples probed with a SUVH6 SRA probe (SUVH6SRA). Because the *suvh6* insertion mutant transcripts are not detected above background with this weaker probe, the *SUVH6SRA/TUB* transcript level ratios were directly compared between wild type and transgenic samples, with all three *ALDR* lines showing higher *SUVH6* transcript levels than wild type.

AtSUVH6

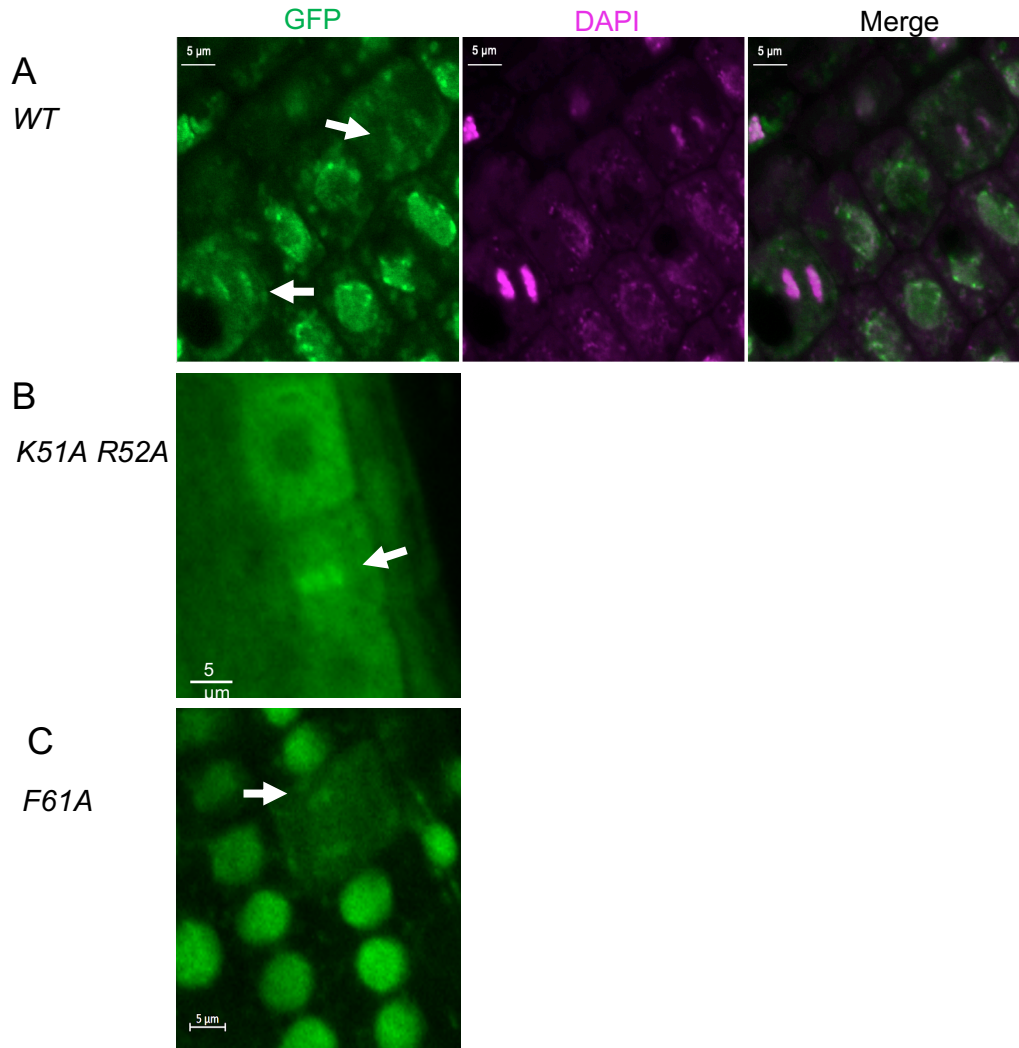


Supplemental Figure 6: The *SUVH6 LDR* deletion removes sequences predicted to be unstructured. Phyre2 structural predictions are shown under the SUVH6 amino acid sequence, with domains color-coded as in Figure 1A. The amino acids deleted in the ΔACM mutant are highlighted in pink and the amino acids deleted in the ΔLDR mutant are highlighted in grey. The key at the bottom of the figure indicates structural element graphic code and color code for degree of confidence from Phyre2 analysis.

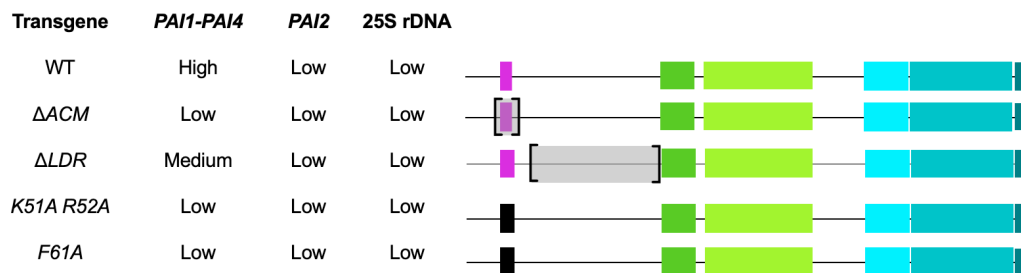
SUVH6 1-67



Supplemental Figure 7: The SUVH6 ACM is sufficient for nuclear localization of GFP. Live roots from three independent T1 seedlings expressing the *SUVH6-GFP 1-67* truncation transgene in the *456* background were imaged using confocal fluorescence microscopy, with green indicating the GFP signal.



Supplemental Figure 8: GFP-SUVH6 transgenes co-localize with mitotic chromosomes. (A) DAPI-stained root tissue from a wild type GFP-SUVH6 line in the *456* background imaged at 40X magnification. DAPI signal is magenta and GFP signal is green. (B) Live root tissue from a *GFP-SUVH6 K51A R52A* line imaged at 40X magnification for GFP. (C) Live root tissue from a *GFP-SUVH6 F61A* line imaged at 40X magnification for GFP. White arrows indicate mitotic cells.



Supplemental Figure 9: Summary of SUVH6 transgene add-back to 456

background results at loci tested in this chapter. Qualitative high, medium, or low remethylation phenotypes are indicated for each construct as illustrated in Figure 2 and Figure 3.

	SUVH (per Zhang)	Database name	NT length
DICOTS			
<i>Arabidopsis thaliana</i> (At)			
	4	TAIR	100
	5	TAIR	305
	6	TAIR	268
<i>Vitis vinifera</i> (Vv)			
	4a	GSVIVT01001083001	85
	4b	GSVIVT01032986001	171
	5a	GSVIVP00014173001	359
	5b	GSVIVP00014209001	367
	6	GSVIVP00013493001	567
<i>Populus trichocarpa</i> (Pt)			
	4a	POPTR_0003s16130	223
	4b	POPTR_0002s23900	151
	4c	Poptr1_1.409038_LG_XIV_7062 013_7067797_-1	120
	5	POPTR_0003s08130	476
<i>Prunus persica</i> (Pr)			
	4a	ppa005549m	112
	4b	ppa002574m	111
	5	ppa021683m	23
	6	ppa000541m	627
MONOCOTS			
<i>Zea mays</i> (Zm)			
	4a	GRMZM2G054380	130
	4b	GRMZM2G336909	110
	5a	GRMZM2G021044	381
	5b	GRMZM2G300955	800
	5c	GRMZM2G074094	647
<i>Oryza sativa</i> (Os)			
	4	LOC_OS01g70220	123
	5a	LOC_OS04g45990	336
	5b	LOC_OS09g19830	684
	5c	LOC_OS08g30910	778
OTHERS			
<i>Selaginella moellendorffi</i> (Sm)		Zhang's sequence IDs are out of date. Hits listed here are from a SUVH4 BLAST to <i>Selaginella</i>	
	4c	LOC9635524	198
	4d	LOC9648636	192
	4a	LOC9637374	65
<i>Physcomitrella patens</i> (Pp)			
	4a	Pp1s469_7V6	178
	4d	Pp1s9_150V6	74
	6	BLAST of phytozome	84
	4b	Pp1s190_36V6	211
<i>Chlamydomonas reinhardtii</i> (Cr)	SET3p	AAV84356.1	28

Supplemental Table 1: Database IDs for named SUVH proteins used in this study.

Where possible, SUVH names are as assigned by Zhang et al., 2012. Other SUVH proteins were identified through BLAST as described in Methods. ID numbers are retrieved from Phytozome or NCBI.

References

- 1001 Genomes Consortium. (2016). 1,135 Genomes Reveal the Global Pattern of Polymorphism in *Arabidopsis thaliana* Resource 1,135 Genomes Reveal the Global Pattern of Polymorphism in *Arabidopsis thaliana*. *Cell*, 166(2), 481–491. <https://doi.org/10.1016/j.cell.2016.05.063>
- Bartee, L., Malagnac, F., & Bender, J. (2001). *Arabidopsis cmt3* chromomethylase mutations block non-CG methylation and silencing of an endogenous gene. *Genes and Development*, 15(14), 1753–1758. <https://doi.org/10.1101/gad.905701>
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., ... Noble, W. S. (2009). MEME SUITE : tools for motif discovery and searching. *Nucleic Acids Research*, 37(May), 202–208. <https://doi.org/10.1093/nar/gkp335>
- Castro, E. De, Sigrist, C. J. A., Gattiker, A., Bulliard, V., Langendijk-genevaux, P. S., Gasteiger, E., ... Hulo, N. (2006). ScanProsite : detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins, 34, 362–365. <https://doi.org/10.1093/nar/gkl124>
- Chandrasekhara, C., Mohannath, G., Blevins, T., Pontvianne, F., & Pikaard, C. S. (2016). Chromosome-specific NOR inactivation explains selective rRNA gene silencing and dosage control in *Arabidopsis*. *Genes and Development*, 30(2), 177–190. <https://doi.org/10.1101/gad.273755.115>
- Chang, C., Couñago, R. M., Williams, S. J., & Boden, M. (2013). The distribution of different classes of nuclear localization signals, 2324. <https://doi.org/10.4161/psb.25976>
- Clough, S. J., & Bent, A. F. (1998). Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant Journal*, 16(June 1998), 735–743. <https://doi.org/10.1046/j.1365-313X.1998.00343.x>
- Du, J., Johnson, L. M., Groth, M., Feng, S., Hale, C. J., Li, S., ... Jacobsen, S. E. (2014). Mechanism of DNA methylation-directed histone methylation by KRYPTONITE. *Molecular Cell*, 55(3), 495–504. <https://doi.org/10.1016/j.molcel.2014.06.009>
- Du, J., Zhong, X., Bernatavichute, Y. V., Stroud, H., Feng, S., Caro, E., ... Jacobsen, S. E. (2012). Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell*, 151(1), 167–180. <https://doi.org/10.1016/j.cell.2012.07.034>

- Ebbs, M. L., Bartee, L., & Bender, J. (2005). H3 lysine 9 methylation is maintained on a transcribed inverted repeat by combined action of SUVH6 and SUVH4 methyltransferases. *Molecular and Cellular Biology*, 25(23), 10507–10515. <https://doi.org/10.1128/MCB.25.23.10507>
- Ebbs, M. L., & Bender, J. (2006). Locus-specific control of DNA methylation by the Arabidopsis SUVH5 histone methyltransferase. *The Plant Cell*, 18(5), 1166–1176. <https://doi.org/10.1105/tpc.106.041400.1>
- Fransz, P., Jong, J. H. De, Lysak, M., Castiglione, M. R., & Schubert, I. (2002). Interphase chromosomes in Arabidopsis are organized as well defined chromocenters from which euchromatin loops emanate.
- Harris, C. J., Scheibe, M., Wongpalee, S. P., Liu, W., Zhong, Z., Yen, L., ... Jacobsen, S. E. (2018). A DNA methylation reader complex that enhances gene transcription. *Science*, 2(6419), 1182–1186. <https://doi.org/10.1126/science.aar7854>
- Johnson, L. M., Bostick, M., Zhang, X., Kraft, E., Henderson, I., Callis, J., & Jacobsen, S. E. (2007). The SRA Methyl-Cytosine-Binding Domain Links DNA and Histone Methylation. *Current Biology*, 17, 379–384. <https://doi.org/10.1016/j.cub.2007.01.009>
- Johnson, L. M., Law, J. a., Khattar, A., Henderson, I. R., & Jacobsen, S. E. (2008). SRA-domain proteins required for DRM2-mediated de novo DNA methylation. *PLoS Genetics*, 4(11). <https://doi.org/10.1371/journal.pgen.1000280>
- Jones DT, Cozzetto D. (2015). DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics*. 2015 Mar 15;31(6):857-63. doi: 10.1093/bioinformatics/btu744. Epub 2014 Nov 12
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10(6), 845–858. <https://doi.org/10.1038/nprot.2015-053>
- Kosugi, S., Hasebe, M., Matsumura, N., Takashima, H., Miyamoto-Sato, E., Tomita, M., & Yanagawa, H. (2009). Six classes of nuclear localization signals specific to different binding grooves of importin α . *Journal of Biological Chemistry*, 284(1), 478–485. <https://doi.org/10.1074/jbc.M807017200>
- Krauss, V. (2008). Glimpses of evolution: Heterochromatic histone H3K9 methyltransferases left its marks behind. *Genetica*, 133(1), 93–106. <https://doi.org/10.1007/s10709-007-9184-z>

- Lampropoulos, A., Sutikovic, Z., Wenzl, C., Maegele, I., Lohmann, J. U., & Forner, J. (2013). GreenGate - A Novel, Versatile, and Efficient Cloning System for Plant Transgenesis. *PLoS ONE*, 8(12), e83043. <https://doi.org/10.1371/journal.pone.0083043>
- Lee, J., Lee, H. J., Shin, M., & Ryu, W. (2004). Versatile PCR-mediated insertion or deletion mutagenesis. *Benchmarks*, 36(3), 398–400.
- Li, X., Harris, C. J., Zhong, Z., Chen, W., Liu, R., Jia, B., ... Du, J. (2018). Mechanistic insights into plant SUVH family H3K9 methyltransferases and their binding to context-biased non-CG DNA methylation. *Proceedings of the National Academy of Sciences*, 115(37), 201809841. <https://doi.org/10.1073/pnas.1809841115>
- McClintock, B. (1950). The Origin and Behaviour of Mutable Loci in Maize. *Proc. Natl Acad. Sci. USA*, 36(6), 344–355. <https://doi.org/10.1073/pnas.36.6.344>
- Melquist, S., Luff, B., & Bender, J. (1999). Arabidopsis *PAI* Gene Arrangements , Cytosine Methylation and Expression, 4.
- Pontvianne, F., Blevins, T., Chandrasekhara, C., Feng, W., Stroud, H., Jacobsen, S. E., ... Pikaard, C. S. (2012). Histone methyltransferases regulating rRNA gene dose and dosage control in Arabidopsis. *Genes and Development*, 26(9), 945–957. <https://doi.org/10.1101/aad.182865.111>
- Proost, S., Bel, M. Van, Vaneechoutte, D., Peer, Y. Van De, Mueller-roeber, B., & Vandepoele, K. (2015). PLAZA 3.0 : an access point for plant comparative genomics Dirk Inz e. *Nucleic Acids Research*, 43(October 2014), 974–981. <https://doi.org/10.1093/nar/gku986>
- Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, <https://imagej.nih.gov/ij/>, 1997-2018.
- Schotta, G., Ebert, A., & Reuter, G. (2003). SU (VAR) 3-9 is a conserved key function in heterochromatic gene silencing, 149–158.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., ... So, J. (2011). Fast , scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7(539). <https://doi.org/10.1038/msb.2011.75>
- Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., ... Jacobsen, S. E. (2014). Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nature Structural & Molecular Biology*, 21(1), 64–72. <https://doi.org/10.1038/nsmb.2735>

- Wang, L., Gao, Y., Zheng, X., Liu, C., Dong, S., Li, R., & Zhang, G. (2019). Histone Modifications Regulate Chromatin Compartmentalization by Contributing to a Phase Article Histone Modifications Regulate Chromatin Compartmentalization by Contributing to a Phase Separation Mechanism, 646–659. <https://doi.org/10.1016/j.molcel.2019.08.019>
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2 — a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9), 1189–1191. <https://doi.org/10.1093/bioinformatics/btp033>
- Zhang, L., & Ma, H. (2012). Complex evolutionary history and diverse domain organization of SET proteins suggest divergent regulatory interactions. *New Phytologist*, 195, 248–263. <https://doi.org/10.1111/j.1469-8137.2012.04143.x>
- Zhao, S., Cheng, L., Gao, Y., Zhang, B., Zheng, X., Wang, L., ... Sun, Q. (2019). Plant HP1 protein ADCP1 links multivalent H3K9 methylation readout to heterochromatin formation. *Cell Research*, 29, 54–66. <https://doi.org/10.1038/s41422-018-0104-9>

**Chapter 3: The *Arabidopsis* SUVH5 and SUVH6 histone 3 lysine 9
methyltransferase amino-terminal extensions contribute to locus-specificity**

Jenna Kotak and Judith Bender

This is a currently unpublished chapter. Experiments were performed by JK. Brown undergraduate Agnes Cheng helped with genetic and DNA gel blot analysis of some transgenic lines in this chapter as an Undergraduate Research and Teaching Award-funded project in Summer 2019. Manuscript written by JK and JB.

Abstract

Histone 3 lysine 9 methylation (H3K9me) is a fundamental heterochromatic modification across eukaryotes, with functions in transcriptional silencing of non-gene sequences like transposons and repeats. In *Arabidopsis thaliana*, H3K9 methyltransferases SUVH4, SUVH5, and SUVH6 catalyze H3K9 methylation on transposons and repeats throughout the genome. SUVH4 is the basal H3K9 methyltransferase that contributes the majority of H3K9me, whereas the more recently evolved SUVH5 and SUVH6 methyltransferases each contribute lower levels of H3K9me with locus-specific preferences. We previously determined that a long amino-terminal extension with an embedded short conserved motif is a signature feature of the SUVH6 clade versus the SUVH4 clade. Here, we used a transgene add-back strategy with chimeric SUVH constructs to test whether the SUVH6 amino-terminal extension contributes to SUVH6 target preferences. We found that the SUVH6 amino-terminal extension can alter target preferences in the context of SUVH5 but not SUVH4 carboxy-terminal domains. Our results support a model where the amino-terminal extension and carboxy-terminal domains co-evolved within the SUVH6 clade to create a means of regulating H3K9 methyltransferase activity to be most efficient at a subset of targets.

Introduction

Chromatin modifications on histone proteins underlie critical regulatory mechanisms in eukaryotic genomes. Activating modifications enhance precise spatiotemporal transcription of genes, whereas heterochromatic modifications repress transcription of non-genic sequences including transposable elements and repeats, protecting the host cell from mutations and rearrangements caused by proliferation of non-genic sequences. Therefore, proper localization and activity of histone modifying enzymes along chromatin is critical for regulation of normal cellular processes and for genome stability.

Histone 3 lysine 9 methylation (H3K9me) is a fundamental heterochromatin modification that functions in transcriptional silencing in animals, fungi, and plants (Krauss, 2008). H3K9 methyltransferases contain a well conserved catalytic SET domain, flanked by conserved pre-SET and post-SET domains that provide selectivity for H3K9 (Min et al., 2002, Zhang et al., 2002). In plant SUVH H3K9 methyltransferases, these conserved catalytic domains are paired with a SET and RING finger-associated (SRA) domain that selectively binds to 5-methyl-cytosine (5meC), another heterochromatic modification. Reciprocally, plant chromomethyltransferase (CMT) enzymes selectively bind to H3K9me through a chromodomain (Du et al., 2012). Together, SUVH and CMT enzymes promote H3K9me and 5meC in an amplification loop, with CMT enzymes preferentially methylating cytosines in CHG and CHH contexts (where H is A, T, or C). Combined with other 5meC pathways, the SUVH/CMT pathway provides dense modifications genome-wide on

heterochromatin targets including pericentromeric repeats, silenced ribosomal RNA gene repeats (rDNA), and transposable elements (Stroud et al., 2013).

Two clades of SUVH H3K9 methyltransferases contribute to H3K9me: the SUVH4 clade and SUVH6 clade. The SUVH4 clade contains the progenitor SUVH protein present since the evolution of chlorophytic algae and evolved other SUVH clades throughout the diversification of plants (Zhang et al., 2012). The SUVH6 clade emerged more recently during the evolution of ferns (Chapter 2). In the model plant *Arabidopsis thaliana*, there is one SUVH4 clade protein, SUVH4, and two SUVH6 clade proteins, SUVH5 and SUVH6. Although *Arabidopsis* SUVH4, SUVH5, and SUVH6 share target regions genome-wide, SUVH4 maintains the majority of H3K9me with nearly 90% of H3K9me lost in *suvh4* mutants, whereas *suvh5* or *suvh6* mutants show negligible H3K9me deficiencies (Li et al., 2018). However, single locus analysis of *suvh4 suvh5* and *suvh4 suvh6* double mutants versus a *suvh4 suvh5 suvh6* triple mutant reveals that SUVH5 and SUVH6 contribute to H3K9me at lower levels than SUVH4 across a spectrum of genomic targets, and with different locus-specific preferences between SUVH5 and SUVH6 at a subset of targets (Ebbs & Bender, 2006). In particular, analysis of a *suvh4 suvh5* mutant over a range of transposable element and repeat targets showed that SUVH6 contributes only low levels of H3K9me-associated 5meC at all of the surveyed targets except at an inverted repeat gene duplication. This inverted repeat, *PAI1-PAI4*, is transcribed across from a distal upstream promoter to create small RNA signals that promote 5meC and H3K9me on duplicated *PAI* sequences. In a *suvh4 suvh5* mutant background, SUVH6 alone is capable of

maintaining high levels of H3K9me and associated 5meC at *PAI1-PAI4* similar to levels observed in the wild type where SUVH4, SUVH5, and SUVH6 are all present for *PAI1-PAI4* H3K9me. In contrast, at an unlinked transcriptionally silenced *PAI2* duplication, the majority of H3K9me and associated 5meC is maintained by SUVH4, with only minor contributions from SUVH6 or SUVH5.

Biochemical analysis of H3K9 methyltransferase activity *in vitro* for SUVH4 versus SUVH6 catalytic domains does not reveal substantial differences that would account for the restricted activity of SUVH6 *in vivo* (Du et al., 2014, Li et al., 2018). Additionally, SUVH4 and SUVH6 SRA domains bind to 5meC-modified DNA fragments with similar efficiency (Li et al., 2018). Moreover, the observation that SUVH6 acts preferentially at *PAI1-PAI4* versus *PAI2* despite nearly identical primary sequences with similar 5meC patterning in a wild type background argues that the *PAI1-PAI4* preference is not due to sequence context preferences for 5meC binding by the SUVH6 SRA domain (Ebbs & Bender, 2006). Finally, SUVH4 and SUVH6 transcripts accumulate to comparable levels in most tissues in *Arabidopsis* (Bar eFP Browser, Chapter 1 Figure 4). Therefore, neither isolated domain activities nor expression patterns account for the broad activity of SUVH4 compared to the locus-specific preferred activity of SUVH6.

SUVH5 and SUVH6 share a high degree of sequence identity through the 5meC-binding and catalytic domains, although SUVH5 demonstrates weaker SRA 5meC-binding in most 5meC contexts *in vitro* (Ebbs et al., 2005, Li et al., 2018). However, SUVH5 has less restricted locus-specific preferences than SUVH6 *in vivo*. Analysis of a *suvh4 suvh6* versus a *suvh4 suvh5* mutant showed

that SUVH5 has stronger activity than SUVH6 at transposable element targets *Ta3* and *Mu1* in leaf tissue (Ebbs & Bender, 2006), despite SUVH5 having approximately 15-fold lower transcript levels than SUVH6 in mature leaves (Bar eFP Browser, Chapter 1 Figure 4). Furthermore, we found that SUVH5 has higher activity than SUVH6 at silenced 25S rDNA repeats (Chapter 2). However, at other targets including *PAI2*, silenced 5S rDNA repeats, and pericentromeric repeats, both SUVH5 and SUVH6 have low activity (Ebbs & Bender, 2006). The only locus thus far identified where SUVH6 alone mediates higher H3K9me levels than SUVH5 alone is at the *PAI1-PAI4* locus. Therefore, although both SUVH5 and SUVH6 have overall lower activity than SUVH4, SUVH5 has a broader range of preferred targets compared to SUVH6.

We previously determined that the SUVH6 clade differs from the SUVH4 clade by a long amino-terminal extension upstream of the SRA domain, consisting of unconserved disordered sequences surrounding one to three short amino-terminal conserved motifs (ACMs) (Chapter 2). SUVH6 contains a single ACM that is highly related to the ACM consensus sequence whereas SUVH5 contains two ACMs with more diverged sequences. We therefore hypothesized that the SUVH6 amino-terminal extension could contribute to the locus-specific preference differences between SUVH6 and SUVH5 and/or the restricted activity of the SUVH6 clade methyltransferases versus SUVH4. To test this hypothesis, we used a transgene add-back strategy with swap constructs between the SUVH6 amino-terminal extension and SUVH5 or SUVH4 carboxy-terminal domains into a *suvh4 suvh5 suvh6 (456)* mutant background that is depleted for H3K9me and

associated 5meC genome-wide. We found that the SUVH6 amino-terminal extension contributes to locus-specific preferences in the context of SUVH5 but not SUVH4 carboxy-terminal domains. These results support a model where the SUVH6 clade evolved to include amino-terminal regulation of the carboxy-terminal 5meC binding and/or catalytic activities.

Results

A transgene add-back approach to assess SUVH6 versus SUVH5 or SUVH4 locus-specific activity

We previously determined that a *SUVH6* transgene with or without an amino-terminal epitope tag in the *456* background can restore H3K9me-associated 5meC with similar patterns to a *suvh4 suvh5* mutant at the *PAI* loci and at the 25S rDNA repeats (Chapter 2). To further identify the segments of the *SUVH6* transgene that contribute to these locus-specific remethylation patterns, we tested transgene constructs with swaps between *SUVH6* and *SUVH5* or *SUVH4* sequences (Figure 1). As the first step for this approach, we determined the *PAI* loci and 25S rDNA repeat remethylation patterns for control *SUVH5* and *SUVH4* transgenes in a *456* background.

We tested H3K9me-associated 5meC patterns at the *PAI* loci using a *PAI* *HincII* DNA gel blot assay as a facile and sensitive means to assess H3K9me-associated 5meC in non-CG contexts. We previously established that the *PAI* *HincII* assay accurately reflects 5meC patterns determined by bisulfite sequencing

and H3K9me1 and H3K9me2 patterns determined by chromatin immunoprecipitation (Ebbs et al., 2005). The 5meC-sensitive enzyme *HincII* cleaves once in each *PAI* duplication near the translational start codon, monitoring 5meC in the CAG and CTG contexts. For the *PAI1-PAI4* inverted repeat duplication, the two internal *HincII* sites have the potential to yield four different fragment sizes depending on whether neither, one, or both internal sites are modified by 5meC and inhibited from cleavage. For the *PAI2* duplication, the single internal *HincII* site has the potential to yield two different fragment sizes depending on whether the site is modified by 5meC and inhibited from cleavage. Another diverged *PAI* sequence duplication *PAI3* lacks an internal *HincII* site, yielding only a single fragment due to cleavage at external unmodified sites regardless of *PAI3* 5meC patterning. Thus, the *PAI HincII* DNA gel blot provides 5meC patterning information at *PAI1-PAI4* and *PAI2*, with *PAI3* serving as a DNA loading control. We previously determined that SUVH4 and SUVH6 make redundant contributions to *PAI1-PAI4* H3K9me and 5meC, with only a minor contribution from SUVH5, but SUVH4 alone makes the major contribution to *PAI2* H3K9me and 5meC with only minor contributions from SUVH5 and SUVH6 (Ebbs & Bender, 2006).

Similarly, we tested H3K9me-associated 5meC patterns at the 25S rDNA repeats using an *AvaII* DNA gel blot assay. The 25S rDNA repeats are present in arrays of hundreds of copies on chromosome 2 and on chromosome 4, with only the chromosome 2 arrays silenced by 5meC and H2K9me (Chandrasekhara et al., 2016). The 5meC-sensitive enzyme *AvaII* cleaves multiple times in each 25S

rDNA repeat to yield a ladder of band sizes with the highest molecular weight fragments diagnostic of inhibited cleavage due to 5meC. We previously determined that the majority of H3K9me-associated 5meC at the 25S rDNA repeats is controlled by SUVH4 and SUVH5, with a minimal contribution from SUVH6 (Chapter 2).

To assess the ability of *SUVH5* control transgenes to remethylate the *PAI* loci and 25S rDNA repeats, we generated two constructs: *SUVH5* driven by its native promoter (*5P:SUVH5*), and *SUVH5* driven by the *SUVH6* promoter (*6P:SUVH5*) as a control for swap constructs driven by the *SUVH6* promoter (Figure 1). We tested individual transgenic lines in the *456* background for transgene genetic segregation patterns diagnostic of a single-insertion locus to avoid transgene silencing due to high copy numbers (Methods), and then further analyzed at least two independent single-insertion locus lines for each construct. We found that the *5P:SUVH5* constructs had substantially lower transcript levels than the *6P:SUVH5* constructs (Figure 2), consistent with endogenous *SUVH5* having lower transcript levels than endogenous *SUVH6* in leaf tissue (BAR eFP Browser, Chapter 1 Figure 4).

In the *PAI HincII* DNA gel blot assay, the *5P:SUVH5* lines displayed a minor increase in the highest molecular weight bands diagnostic of dense *PAI1-PAI4* 5meC remethylation, intermediate between a *suvh4 suvh6* control strain and wild type (Figure 3A, 3B). The *5P:SUVH5* lines also displayed a minor increase in the higher molecular weight band diagnostic of *PAI2* remethylation relative to the *456* parental strain versus wild type. However, the *6P:SUVH5* lines displayed

similar levels of the highest molecular weight bands diagnostic of dense *PAII-PAI4* remethylation to wild type, along with band patterns diagnostic of weak *PAI2* remethylation. Together these results indicate that the degree of *PAII-PAI4* remethylation correlates with *SUVH5* transgene expression level. Furthermore, these results suggest that the low activity of endogenous *SUVH5* at the *PAII-PAI4* locus is due at least in part to low expression levels. Therefore, the *PAII-PAI4* locus does not serve as a reporter of a non-preferred *SUVH5* locus, especially in a transgene add-back assay with varied expression levels across different transgenic lines. In contrast, we previously observed consistent patterns of strong *PAII-PAI4* remethylation and weak *PAI2* remethylation across multiple independent *SUVH6* control transgenic lines in the *456* background, indicating that the *SUVH6* preference for *PAII-PAI4* is relatively insensitive to variation in transgene expression levels (Chapter 2). Transgenic line results are summarized in Supplemental Figure 1.

In the 25S rDNA *AvaII* DNA gel blot assay, we found that most of the *5P:SUVH5* and the *6P:SUVH5* lines displayed increased ratios of the high molecular weight bands relative to the fourth band (a product of an uninhibited *AvaII* cleavage site), diagnostic of 5meC remethylation of the *456* parental strain (Figure 3C, 3D). Both *5P:SUVH5* and *6P:SUVH5* lines displayed 5meC remethylation patterns similar to the patterns in the *suvh4 suvh6* control strain (Figure 3C, 3D). Therefore, the 25S rDNA repeats can be used as reporters of *SUVH5* preferred target region, with little variability across *SUVH5* transgene expression levels.

To assess the ability of *SUVH4* control transgenes to remethylate the *PAI* loci and 25S rDNA repeats, we generated two constructs: genomic *SUVH4* driven by its native promoter (*4P:SUVH4*) and a 3XFLAG epitope-tagged *SUVH4* driven by the *SUVH6* promoter (*6P:SUVH4*) as a control for swap constructs with *SUVH6* (Figure 1). In the *PAI* *HincII* assay, both the *4P:SUVH4* and the *6P:SUVH4* lines displayed similar levels of the high molecular weight bands diagnostic of dense *PAI1-PAI4* and *PAI2* 5meC remethylation, similar to the wild type control (Figure 4A, 4B). Therefore, the *SUVH4* transgenes can strongly remethylate H3K9me at the *PAI* loci, recapitulating the *PAI* 5meC patterns maintained by endogenous *SUVH4* in a *suvh5 suvh6* mutant strain (Ebbs & Bender, 2006). These results are consistent with our previous observation that a *4P:SUVH4* transgene in a *suvh4* mutant background can strongly remethylate the *PAI* loci (Malagnac et al., 2002). In addition, the similar remethylation patterns for *4P* versus *6P* constructs is consistent with the endogenous *SUVH4* and *SUVH6* genes expressing similar transcript levels in vegetative tissues (Bar eFP Browser, Chapter 1 Figure 4). Moreover, these results indicate that the 3XFLAG epitope tag in the *6P:SUVH4* construct does not interfere with *SUVH4* activity. We subsequently focused on the *6P:SUVH4* transgenic lines as controls for swaps between *SUVH6* and *SUVH4* sequences.

In the 25S rDNA *AvaII* DNA gel blot assay, we found that all of three *6P:SUVH4* lines displayed increased levels of the high molecular weight bands diagnostic of remethylation relative to the *456* parental strain, and similar to the levels in the *suvh5 suvh6* control strain (Figure 4C). Therefore, the *6P:SUVH4*

construct mediates strong remethylation of the 25S rDNA repeats, recapitulating the activity of endogenous SUVH4 at these repeats in a *suvh5 suvh6* mutant strain.

Overall, our analysis indicates that the 25S rDNA repeats can be used to discriminate between SUVH6-like weak remethylation (Chapter 2) versus SUVH5-like and SUVH4-like strong remethylation. In contrast, the *PAI* remethylation assay does not allow reliable discrimination between SUVH6-like and SUVH4-like strong remethylation at *PAI1-PAI4* and SUVH5-like weaker activity at *PAI1-PAI4* due to expression-dependent variation across *SUVH5* transgenic lines. However, *PAI2* can be used to discriminate SUVH4-like strong remethylation from SUVH6-like weak remethylation.

The SUVH6 amino-terminal extension restricts 25S rDNA repeat remethylation patterns mediated by SUVH5 but not SUVH4 carboxy-terminal domains

To test whether the amino-terminal extension of SUVH6 contributes to locus-specific preferences, we generated transgene constructs with the *SUVH6* promoter and the 3XFLAG epitope-tagged SUVH6 amino-terminus upstream of the pre-SRA domain (amino acids 1-264) fused to the carboxy-terminal sequences that mediate 5meC binding (pre-SRA and SRA domains) and catalysis (pre-SET, SET, and post-SET domains) from SUVH5 (*6P:SUVH6/5*) or SUVH4 (*6P:SUVH6/4*) (Figure 1). We then generated multiple independent single-locus

transgenic lines in the *456* background as described for the control lines (Methods) and assessed H3K9me-associated 5meC remethylation patterns using the *PAI* and 25S rDNA DNA gel blot assays.

For the *6P:SUVH6/5* lines in the *PAI HincII* assay, we found that the majority of lines displayed similar digestion patterns to *6P:SUVH6* control lines, with band patterns diagnostic of strong *PAI1-PAI4* remethylation but weak *PAI2* remethylation (Figure 5A). Therefore, the fusion junction in the SUVH6/5 swap constructs does not abrogate SUVH activity. However, as discussed above, the *PAI1-PAI4* locus does not allow reliable discrimination of SUVH6 versus SUVH5 locus-specific preferences, and the *PAI2* locus is a weak target for both SUVH5 and SUVH6. In the 25S rDNA *AvaII* DNA gel blot assay, we found that all of the *6P:SUVH6/5* lines displayed similar patterns to the *6P:SUVH6* control lines diagnostic of weak remethylation, and in contrast to the stronger remethylation patterns in the *6P:SUVH5* control lines (Figure 5B). Therefore, the 25S rDNA assay suggests that in the SUVH6/5 swap construct, the SUVH6 amino-terminal extension restricts the ability of the SUVH5 carboxy-terminal domains to remethylate the 25S rDNA repeats.

For the *6P:SUVH6/4* lines in the *PAI HincII* assay, we found that all of the lines displayed similar digestion patterns to *6P:SUVH4* control transgenic lines and a wild type control, with band patterns diagnostic of strong *PAI1-PAI4* and *PAI2* remethylation (Figure 6A). These patterns contrast with the weak *PAI2* remethylation patterns in the *6P:SUVH6* control lines (Figure 5A). Therefore, the *PAI* assay suggests that in the *SUVH6/4* construct, the SUVH6 amino-terminal

extension does not restrict the ability of the SUVH4 carboxy-terminal domains to remethylate *PAI2* or *PAII-PAI4*. In the 25S rDNA *AvaII* DNA gel blot assay, we found that all of the *6P:SUVH6/4* lines displayed similar patterns to the *6P:SUVH4* and *suvh5 suvh6* controls, with patterns diagnostic of strong 25S rDNA remethylation (Figure 6B). Therefore, as at the *PAI* loci, the SUVH6 amino-terminal extension does not restrict the SUVH4 carboxy-terminal sequences from acting at the 25S repeats.

The SUVH5 amino-terminal extension allows increased remethylation of the 25S rDNA repeats by the SUVH6 carboxy-terminal domains

To further investigate the role of the amino-terminal extension in modulating SUVH6 activity, we generated transgenic constructs with the *SUVH6* promoter driving expression of a swap construct with the *SUVH5* amino-terminal extension fused to the *SUVH6* carboxy-terminal domains beginning at the pre-SRA domain (*6P:SUVH5/6*), reciprocal to the *6P:SUVH6/5* construct (Figure 1). We then tested multiple independent lines expressing the *6P:SUVH5/6* construct in the *456* background. In the *PAI HincII* assay, we found that *6P:SUVH5/6* lines displayed similar digestion patterns to *6P:SUVH5* and *6P:SUVH6* control lines, with band patterns diagnostic of strong *PAII-PAI4* remethylation but weak *PAI2* remethylation (Figure 7A). Therefore, as with the *SUVH6/5* swap construct, the fusion junction in the *SUVH5/6* swap construct does not abrogate SUVH activity at *PAII-PAI4*. In the 25S rDNA *AvaII* DNA gel blot assay, we found that

6P:SUVH5/6 lines displayed similar patterns to the *6P:SUVH5* control lines diagnostic of strong remethylation (Figure 7B). The ability of the *6P:SUVH5/6* lines to promote strong 25S rDNA remethylation, unlike *6P:SUVH6* control lines, suggests that the substitution of the SUVH5 amino-terminal extension improves the ability of the SUVH6 carboxy-terminal domains to act at the 25S rDNA repeats.

SUVH5 has similar sub-nuclear localization patterns to SUVH6

We previously determined that functional amino-terminal GFP-tagged SUVH6 construct in the *456* background displays a pattern of nuclear localization with bright nuclear puncta plus a large region that excludes the GFP signal (Chapter 2). The nuclear puncta correspond to regions of dense DAPI staining, which have been defined as concentrated regions of pericentromeric and silenced rDNA repeat heterochromatin (chromocenters, Fransz et al., 2002). The large excluded region likely corresponds to the nucleolus, where the active copies of rDNA sequences are transcribed. To determine whether there are differences between the SUVH6 and SUVH5 localization patterns, we conducted a similar analysis of an amino-terminal GFP-tagged SUVH5 construct driven by the SUVH6 promoter in the *456* background (Figure 8). We found that this *6P::GFP-SUVH5* construct displays nuclear localization, with bright puncta plus an excluded region corresponding to the nucleolus (Figure 8). Therefore, the overall chromatin localization patterns are similar between SUVH5 and SUVH6 and are consistent with SUVH functions as chromatin-modifying enzymes.

Discussion

In eukaryotes, H3K9me is precisely maintained on non-genic sequences like transposable elements and repeats to prevent transcription and subsequent proliferation. Plants have evolved a system of dual H3K9me and 5meC heterochromatin modifications to reinforce transcriptional silencing. Members of the basal SUVH4 clade of plant H3K9 methyltransferases contain carboxy-terminal catalytic domains and a central 5meC-binding SRA domain. Proteins belonging to the more recently evolved SUVH6 clade contain similar catalytic and SRA domains, but also include long amino-terminal extensions consisting of unconserved sequences predicted to be structurally disordered and one to three ACMs (Chapter 2). In *Arabidopsis*, the sole SUVH4 clade member SUVH4 provides the majority of H3K9me deposition throughout the genome, whereas the SUVH6 clade members SUVH5 and SUVH6 contribute lower levels of H3K9me at the same target regions as SUVH4, with different preferences for a subset of target regions. It is not understood why SUVH5 and SUVH6 have more restricted activity than SUVH4, or why SUVH5 and SUVH6 have different preferred targets to each other.

Because the amino-terminal extensions structurally differentiate the SUVH6 clade from the SUVH4 clade and *Arabidopsis* SUVH5 and SUVH6 differ in ACM number and placement, we hypothesized that the amino-terminal extension could contribute to the unique properties of the SUVH6 clade. To test this hypothesis, we used a transgene add-back strategy with SUVH amino-terminal extension swaps, with focus on the SUVH6 amino-terminal extension.

We assessed whether chimeric SUVH proteins displayed preferential activity at the *PAI* loci, which provide both a SUVH6 preferred target *PAI1-PAI4* and a SUVH6 non-preferred target *PAI2* (Ebbs & Bender, 2006), and the 25S rDNA repeats, which provide a SUVH6 non-preferred target but a SUVH5 preferred target (Chapter 2).

In developing our transgene add-back assays, we surveyed a number of SUVH target regions in addition to the *PAI* loci and the 25S repeats, including the *Ta3* and *Mu1* transposable elements that were previously determined to be SUVH4 and SUVH5 preferred targets and SUVH6 non-preferred targets (Ebbs & Bender, 2006). We found that even in SUVH4 and SUVH5 control transgenic lines, *Ta3* was inefficiently remethylated, at least during the first few generations after the *SUVH* transgene was introduced into the *456* background (data not shown). Similarly, a previous analysis of a *SUVH4* transgene added back to a *suvh4* mutant background showed that overall the genome was only partially remethylated, possibly due to a loss of the amplification loop with 5meC that helps reset SUVH4 activity at some target regions (Du et al., 2014). We also found that although the *Mu1* sequences could be remethylated by *SUVH4*, *SUVH5*, and *SUVH6* transgenes, the density of remethylation varied with transgene expression level, making the *Mu1* sequences unsuitable reporters for locus-specific remethylation across transgenic lines with a range of expression levels (data not shown). Similarly, we found that although *PAI1-PAI4* is a non-preferred target for endogenous SUVH5, the locus becomes a better target in transgenic lines with increased SUVH5 expression (Figure 2, Figure 3,

Supplemental Figure 1), so that *PAII-PAI4* mainly serves as a control locus for *SUVH* transgene activity. However, the 25S rDNA repeats can be strongly remethylated during the first transgenic generation and do not display progressive increases in 5meC density in lines that begin with low levels over subsequent generations of inbreeding (data not shown). In addition, the 25S rDNA repeats are relatively insensitive to variations in *SUVH5* transgene expression levels (Figure 2, Figure 3). Therefore, the 25S rDNA repeats provide accurate reporters for the intrinsic activity of *SUVH* constructs in the 456 add-back assay.

Our analysis of SUVH5 and SUVH6 amino-termini swap construct activity at the 25S rDNA repeats suggests that the amino-termini contribute to preferred or non-preferred activity at this region. We found that wild type SUVH5 and a SUVH5/6 amino-terminal swap construct restore similar substantial levels of H3K9me-associated 5meC on the 25S repeats, whereas wild type SUVH6 and the reciprocal SUVH6/5 swap construct promote only low levels of 25S rDNA remethylation (Figure 3C and 3D, Figure 5B, Figure 7B). However, the wild type SUVH6 and SUVH6/5 constructs promote strong remethylation of the SUVH6 preferred target *PAII-PAI4* (Figure 3A, Figure 5A), indicating that the weak activity at the 25S repeats for these constructs is not due to an inactive transgene construct. In contrast, the SUVH6/4 amino-terminal swap construct promotes similar broad remethylation patterns to SUVH4 controls, including at the SUVH6 non-preferred targets *PAI2* and the 25S rDNA repeats (Figure 4, Figure 6). Overall, our results show that the SUVH6 amino-terminal extension alters the locus-specific preferences of the closely related SUVH5 carboxy-terminal

domains but possibly not the more diverged SUVH4 carboxy-terminal domains. The contribution of the amino-terminal extension could depend on having compatible SUVH5 or SUVH6 carboxy-terminal domains, suggesting an interaction between the domains such as the amino-terminus making direct contacts with the carboxy-terminus to mediate the locus-specificity mechanism. By this model, in the context of the incompatible SUVH4 carboxy-terminal domains, the SUVH6 amino-terminal extension would have no effect due to absence of the interaction interface. However, it is possible that the SUVH6/4 construct is processed at the mRNA or protein level to cleave off the SUVH6 amino-terminus. Future analysis of the 3XFLAG-epitope tagged SUVH6/4 protein product and transcript size will elucidate if transcript and protein processing is intact.

Using transgene add-back constructs into the *456* background, we previously determined that in the context of the SUVH6 carboxy-terminal domains, the SUVH6 amino-terminal extension promotes strong remethylation at *PAII-PAI4* (Chapter 2). A missense mutation (F61A) in the single SUVH6 ACM strongly impairs *PAII-PAI4* remethylation, and deletion of the long disordered region between the ACM and the carboxy-terminal domains slightly reduces *PAII-PAI4* remethylation in some transgenic lines. However, in the context of SUVH5 carboxy-terminal domains, the SUVH6 amino-terminal extension becomes restrictive for remethylation of the 25S repeats (Figure 5B). A possible model to reconcile these positive versus negative effects of the SUVH6 amino-terminal extension is that within the SUVH6 clade, the amino-terminal extension

has co-evolved with the carboxy-terminal domains to create an overall protein structure that enables preferential localization to and/or activity at certain targets. In this view, the SUVH6 amino-terminus would be compatible with either the SUVH5 or SUVH6 carboxy-terminal domains to create an overall protein with preference for *PAIL-PAI4* but not the 25S repeats. Conversely, the SUVH5 amino-terminus would be compatible with either the SUVH5 or SUVH6 carboxy-terminal domains to create an overall protein with preference for the 25S repeats.

Comparing the amino-termini of SUVH5 and SUVH6, the differences in ACM sequence and arrangements could account for the differences in locus-specific preferences. SUVH6 contains a single ACM closely related to the consensus sequences from flowering plants, whereas SUVH5 contains two ACMs that are more diverged from the consensus (Chapter 2). Alternatively, the long disordered regions between the ACMs and the carboxy-terminal domains could contribute to SUVH5 versus SUVH6 differences. In particular, a recent proteomic analysis found that SUVH6 has a single phosphorylation site at serine 248 in the amino-terminal extension, while SUVH5 has two phosphorylation sites at serine 211 and serine 224 in the amino-terminal extension (Mergner et al., 2020). Although the context of these serine residues is not conserved across the SUVH6 clade members, perhaps differences in overall phosphorylation or other post-translational modifications of amino-terminal extension residues could modulate the biophysical properties of the extension, affecting flexibility or phase transitions for the long disordered region.

A comparison of GFP-SUVH6 and GFP-SUVH5 sub-nuclear localization patterns in the 456 mutant background shows similar colocalization to chromocenters and exclusion from the nucleolus (Chapter 2, Figure 8). These patterns reveal that the more restricted activity of SUVH5 and SUVH6 than SUVH4 at the pericentromeric repeats (Ebbs & Bender, 2006) and 25S rDNA sequences that constitute the chromocenters are not due to SUVH5 and SUVH6 exclusion from the chromocenters. Previous *in vitro* analysis of isolated SRA and catalytic domain activities does not show major differences that would account for the restricted activity of SUVH5 and SUVH6 versus SUVH4 *in vivo* (Du et al., 2014, Li et al., 2018,). However, our results support the view that in the context of the full enzyme with amino-terminal extension, and in the context of chromatin substrates with 5meC-modified DNA wrapped around nucleosomes, the SUVH5 and SUVH6 proteins include a regulatory module that allows efficient H3K9 methyltransferase activity at only a subset of target regions. This regulated activity for the SUVH6 clade H3K9 methyltransferases versus the broad activity of the SUVH4 clade methyltransferases could give SUVH6 clade members unique roles in controlling heterochromatin in specific developmental or environmental conditions that became important during the evolution of fern, gymnosperm, and angiosperm plant lineages.

Methods

Plant material and transgenes

Homozygous *pai1 suvh4 suvh5 suvh6* mutants were used as a parental background for transgenic introduction (Ebbs & Bender, 2006). Control double mutants *suvh4 suvh5* and *suvh4 suvh6* also included the *pai1* mutation (Ebbs & Bender, 2006). Transgenes were cloned into the pCAMBIA1300 plant transformation vector, which contains a hygromycin resistance gene for transgenic plant selection. Transgene structures are shown in Figure 1. The *4P:SUVH4* genomic fragment was described in Malagnac et al., 2002, consisting of a 11.7 kb *SaII* genomic sequence containing 6.3 kb of promoter, 4.4 kb of exonic/intronic coding region, and 1.0 kb of downstream sequence. For this study, this *SUVH4* genomic fragment was cloned into the polylinker *SaII* site of pCAMBIA1300. *6P:SUVH6* constructs with or without a 3XFLAG epitope tag inserted at codon 11 were as described in Chapter 2. To clone the 3XFLAG epitope-tagged *6P:SUVH4* control construct, oligonucleotide-directed mutagenesis of the *SUVH4* cDNA was used to generate a *Bam*HI site just upstream of the translational start codon and a silent *Sna*BI site for insertion of a 3XFLAG epitope at codon 4. PCR-mediated insertion and subcloning were used to introduce the 3XFLAG epitope tag in the context of the genomic *SUVH4* 11.7 kb *SaII* fragment. The *SUVH6* promoter was swapped for the *SUVH4* promoter as a 1.7 kb *SaII*- *Bam*HI fragment. The resulting *SaII*- *Bam*HI *6P:Bam*HI-*SaII* *3XFLAG-SUVH4* genomic chimeric fragment was cloned into the polylinker *SaII* site of pCAMBIA1300 to create the final *6P:SUVH4* construct. To generate the

6P:SUVH6/4 swap construct, the 3XFLAG epitope-tagged *6P::SUVH6* control was oligo-mutagenized to create a silent *BspEI* site at codons 265 and 266, just upstream of the pre-SRA domain. The 3XFLAG epitope-tagged *4P:SUVH4* genomic construct was oligo-mutagenized to create a *BspEI* site at codons 96 and 97, just upstream of the pre-SRA domain, and to destroy an endogenous *BspEI* site at codons 337 and 338 with a silent mutation. A three-piece ligation was performed to build a chimeric construct with a *SalI*- *BspEI* *6P:3XFLAG-SUVH6* amino-terminal extension sequences plus *BspEI*- *SalI* *SUVH4* carboxy-terminal genomic sequences into the *SalI* polylinker site of pCAMBIA1300 to create the final *6P:SUVH6/4* swap construct. The *5P:SUVH5* transgene consists of a genomic region including 1.6 kb of promoter sequence, with a *BamHI* site introduced by oligo-mutagenesis just upstream of the translational start codon. This fragment was cloned into the pCAMBIA1300 polylinker with a 300 bp nopaline synthase transcriptional stop cassette inserted downstream to create the final *5P:SUVH5* transgene. For the *6P:SUVH5* construct, 1.7 kb of *SUVH6* promoter sequences on a *SalI*- *BamHI* fragment were swapped for the *SUVH5* promoter sequences in the *5P:SUVH5* construct. For the 3XFLAG epitope-tagged *6P:SUVH6/5* and *6P:SUVH5/6* constructs, the *SUVH5* coding sequence was oligo-mutagenized to create a silent *BspEI* site at codons 302 and 303, just upstream of the pre-SRA domain, and oligo-mutagenized to destroy an endogenous *BspEI* site at codons 575-577 with a silent mutation. *SUVH6* with the corresponding *BspEI* site just upstream of the pre-SRA domain is as described for the *6P:SUVH6/4* construct above. *SUVH5* and *SUVH6* amino-terminal extension

sequences were swapped as *Bam*HI- *Bsp*EI fragments. Epitope tags were introduced into *SUVH5* by creating a silent *Sna*BI site via oligo-mutagenesis for tag at codon 3. The *GFP-SUVH5* construct shown in Figure 8 was made by inserting the GFP cassette from vector pGGC014 into the *Sna*BI site via PCR-mediated insertion (Lampropoulos et al., 2013, Lee et al., 2004). We also made a 3XFLAG epitope-tagged version of *SUVH5* using the amino-terminal *Sna*BI site. However, analysis of *6P:3XFLAG-SUVH5* constructs versus *6P:SUVH5* untagged constructs in the *PAI Hinc*II assay showed impaired *PAI1-PAI4* remethylation for the 3XFLAG-tagged constructs. Therefore, we pursued only untagged *SUVH5* constructs for functional studies.

Transgenes were introduced into the *456* background using *Agrobacterium tumefaciens*-mediated floral dips (Clough & Bent, 1998). First generation T1 seedlings were identified by hygromycin resistance. T2 generation seedlings were screened for 3:1 hygromycin resistance segregation to identify transgenic lines with single-locus insertions. T3 generation progeny of individual T2 plants were tested for hygromycin resistance to identify lines homozygous for the transgene insertion.

RNA Gel Blot Analysis

RNA was extracted from four-week-old plants of homozygous transgenic lines and used for gel-blot analysis as previously described (Melquist et al., 1999). Duplicate blots were prepared, with one membrane probed with β -tubulin as a

loading control, and the other membrane probed with a *SUVH5* amino-terminal extension probe. The *SUVH5* amino-terminal extension probe corresponds to sequences coding for amino acids 1-193.

DNA Gel Blot Analysis

With the exception of constructs containing *SUVH4* sequences used in Figure 4 and Figure 6, genomic DNA was prepared from four-week-old leaves from homozygous, single insertion locus transgenic lines as previously described (Melquist et al., 1999). For *4P:SUVH4*, *6P:SUVH4*, and *6P:SUVH6/4* transgenes, individual leaves harvested from approximately 12 hygromycin-resistant T2 generation plants per line including both homozygous and heterozygous transgenic plants were pooled and used for DNA extraction with urea lysis/phenol extraction miniprep method.

DNA gel blots were performed as previously described (Melquist et al., 1999). The *PAI* probe corresponds to an internal 0.7 kb *PstI* fragment from a *PAII* cDNA. The 25S rDNA probe corresponds to 998bp of 25S rDNA sequence (Chapter 2).

Microscopy

8-day-old hygromycin resistant T1 generation seedlings of *6P:GFP-SUVH5* lines in the *456* background were imaged with the Zeiss LSM 800 Confocal Laser

Scanning microscope in the 495-599nm range with auto-exposure. Live seedlings were mounted in water and roots were immediately imaged. Of 15 T1 seedlings examined, 10 displayed nuclear GFP signals, with three representative lines shown in Figure 8. Of the remaining T1 seedlings, four had no GFP signal above background likely due to transgene silencing, and one had a weak cytoplasmic signal, perhaps due to a transgene rearrangement in this atypical line.

Figures

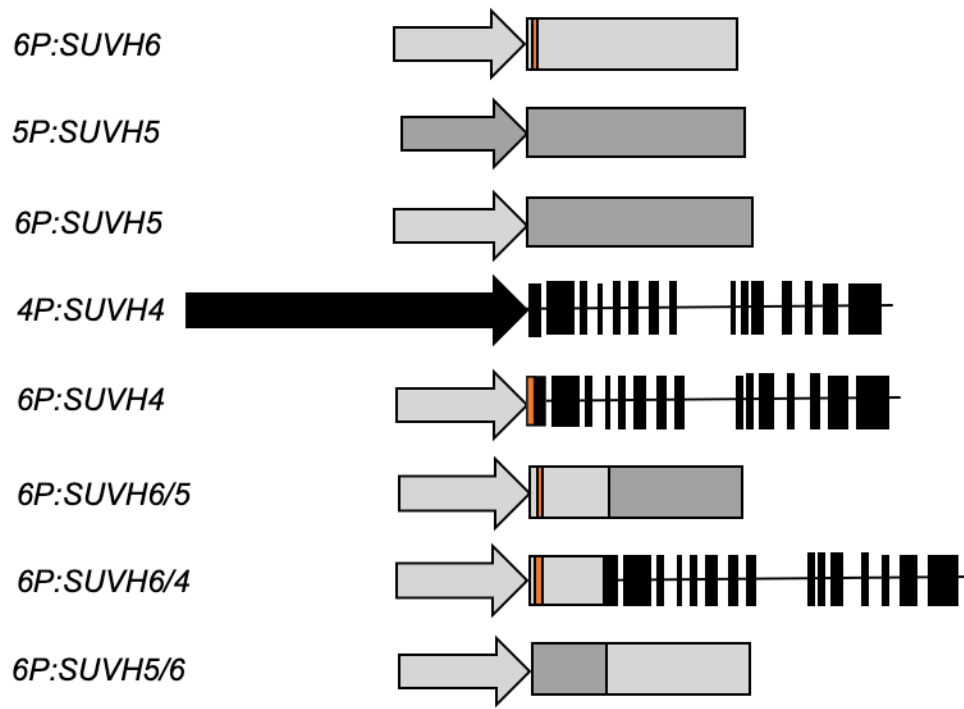


Figure 1: *SUVH* transgene constructs used to test SUVH6 clade amino-terminal extensions. Transgenes diagrams show SUVH6 sequences in light gray, SUVH5 sequences in medium gray, and SUVH4 sequences in black. 3XFLAG epitope tags are shown in orange. Arrows represent promoter sequences. Exonic coding regions are shown as rectangles, and introns are shown as lines. The endogenous *SUVH5* and *SUVH6* genes lack coding sequences introns, whereas the endogenous *SUVH4* gene contains 14 coding sequence introns.

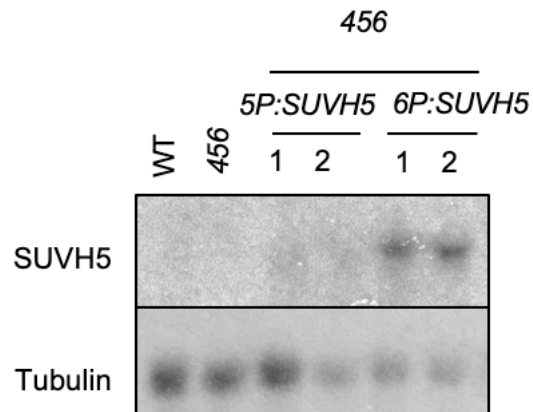


Figure 2: SUVH5 transgenes are expressed at higher levels from the *SUVH6* promoter than the *SUVH5* promoter. RNA gel blots were performed with either a *SUVH5* amino-terminal extension probe, or tubulin as a loading control. Two independent transgenic lines for each construct in the *456* background are shown, with the same line numbers as for the samples in Figure 3. Transgene structures are shown in Figure 1.

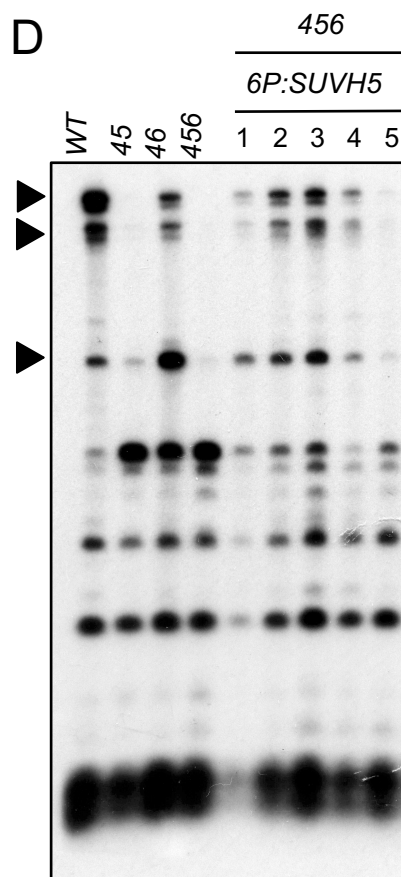
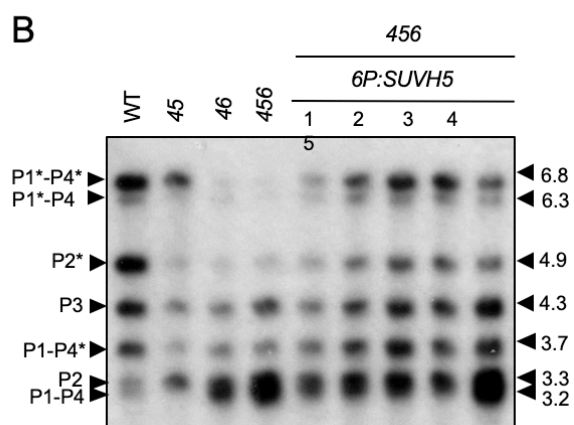


Figure 3: *SUVH5* transgenes remethylate both *PAI* and 25S rDNA targets. *HincII* DNA gel blots for (A) *5P:SUVH5* or (B) *6P:SUVH5* transgenic lines in the *456* background. Left margin labels indicate *PAI HincII* digestion products corresponding to *PAI1-PAI4* = P1-P4, *PAI2* = P2, and *PAI3* = P3, with asterisks indicating bands corresponding to 5meC-inhibited cleavage products. Right margin labels indicate sizes of digestion products in kb. 25S rDNA repeat *AvaII* DNA gel blots for (C) *5P:SUVH5* or (D) *6P:SUVH5* transgenic lines in the *456* background. Black arrowheads indicate bands corresponding to 5meC-inhibited cleavage products. Transgene structures are shown in Figure 1. WT indicates wild type Ws, *45* indicates *pai1 suvh4 suvh5*, and *46* indicates *pai1 suvh4 suvh6* control strains. Numbered lanes indicate independent transgenic lines for each construct.

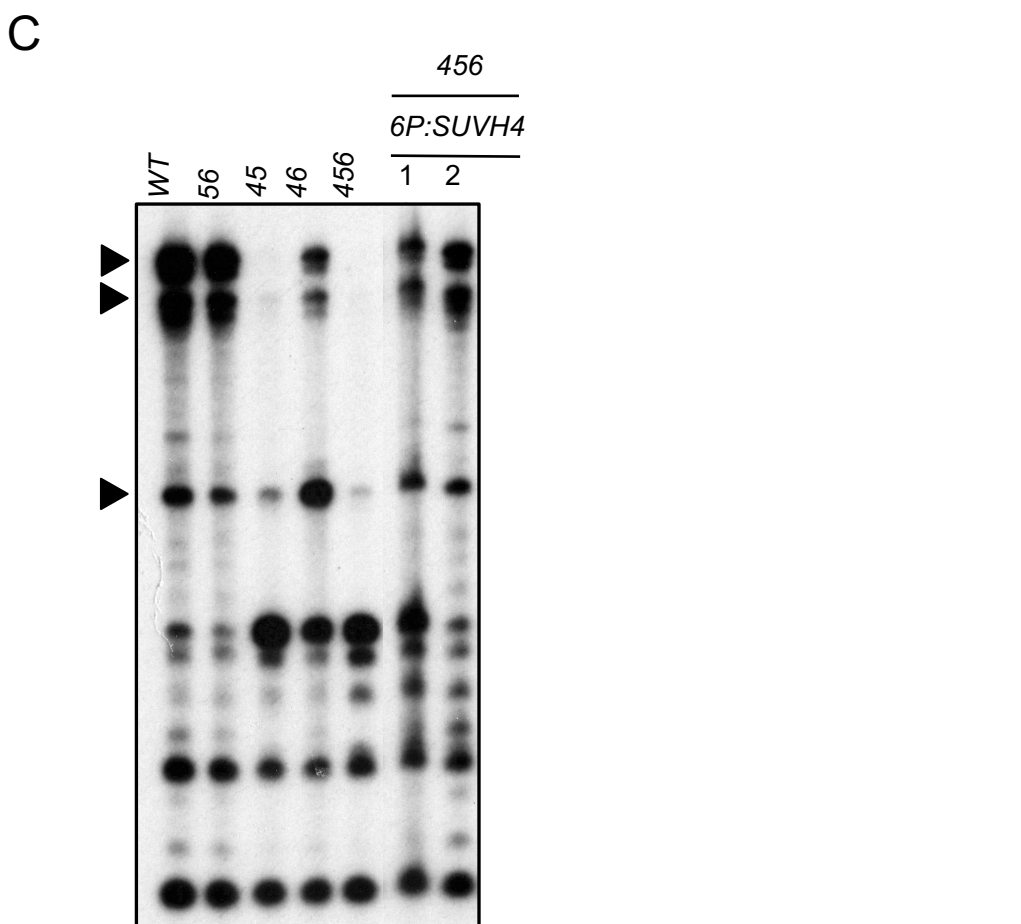
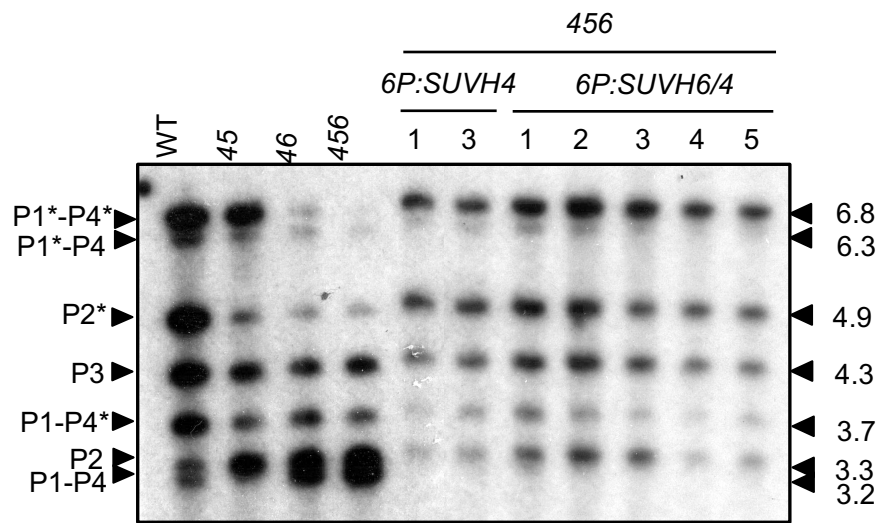


Figure 4: *SUVH4* transgenes remethylate both *PAI* and 25S rDNA targets. *PAI* *HincII* DNA gel blots for (A) *4P:SUVH4* or (B) 3XFLAG epitope-tagged *6P:SUVH4* lines in the *456* background. Left and right margin labels are as described in Figure 3. (C) 25S rDNA repeat *AvaII* DNA gel blots for 3XFLAG epitope-tagged *6P:SUVH4* lines in the *456* background. Black arrowheads indicate bands corresponding to 5meC-inhibited cleavage products. Transgene structures are shown in Figure 1. WT indicates wild type Ws, *45* indicates *pai1 suvh4 suvh5*, *46* indicates *pai1 suvh4 suvh6*, and *56* indicates *pai1 suvh5 suvh6* control strains. Numbered lanes indicate independent transgenic lines for each construct.

Figure 5: The SUVH6 amino terminus restricts SUVH5 carboxy-terminal domain activity at the 25S rDNA repeats. (A) *PAI HincII* DNA gel blot of 3XFLAG epitope-tagged *6P:SUVH6* control lines and 3XFLAG epitope-tagged *6P:SUVH6/5* swap construct lines in the *456* background. Left and right margin labels are as described in Figure 3. (B) 25S rDNA repeats *AvaII* DNA gel blot with untagged *6P:SUVH5* control lines, 3XFLAG epitope-tagged *6P:SUVH6* control lines, and 3XFLAG epitope-tagged *6P:SUVH6/5* swap construct lines in the *456* background. Black arrowheads indicate bands corresponding to 5meC-inhibited cleavage products. Transgene structures are shown in Figure 1. WT indicates wild type *Ws*, *45* indicates *pai1 suvh4 suvh5*, and *46* indicates *pai1 suvh4 suvh6* control strains. Numbered lanes indicate independent transgenic lines for each construct.

A



B

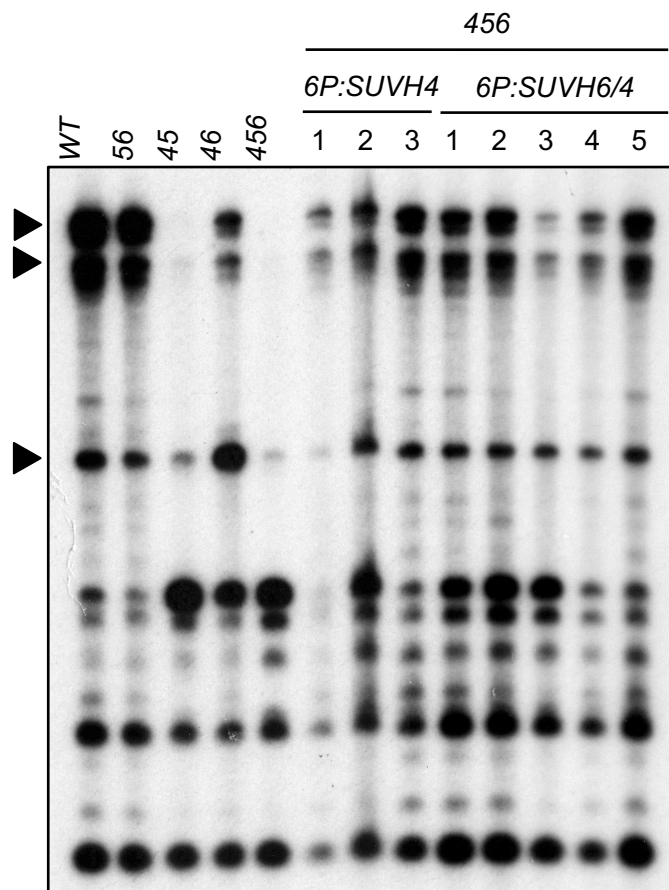


Figure 6: The SUVH6 amino terminus does not restrict SUVH4 carboxy-terminal domain activity at *PAI2* or the 25S rDNA repeats. (A) *PAI HincII* DNA gel blot of 3XFLAG epitope-tagged *6P:SUVH4* control lines and 3XFLAG epitope-tagged *6P:SUVH6/4* swap construct lines in the *456* background. Left and right margin labels are as described in Figure 3. (B) 25S rDNA repeat *AvaII* DNA gel blot with 3XFLAG epitope-tagged *6P:SUVH4* control lines and 3XFLAG epitope-tagged *6P:SUVH6/4* swapped lines. Black arrowheads indicate bands corresponding to 5meC-inhibited cleavage products. Transgene structures are shown in Figure 1. WT indicates wild type Ws, *45* indicates *pai1 suvh4 suvh5*, *46* indicates *pai1 suvh4 suvh6*, and *56* indicates *pai1 suvh5 suvh6* control strains. Numbered lanes indicate independent transgenic lines for each construct.

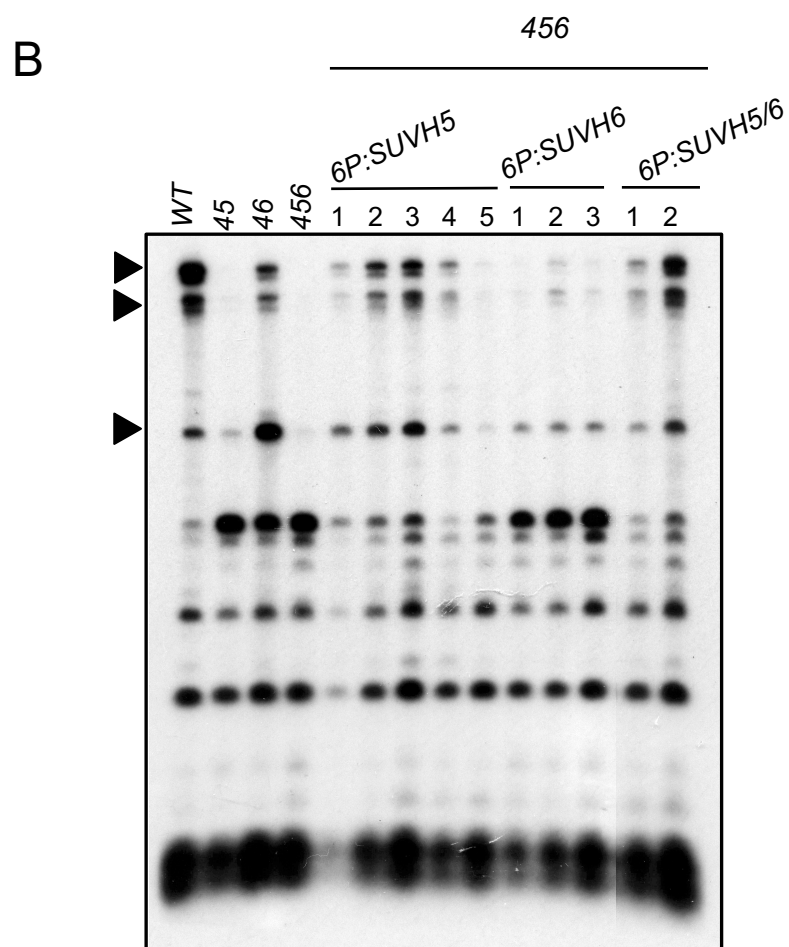
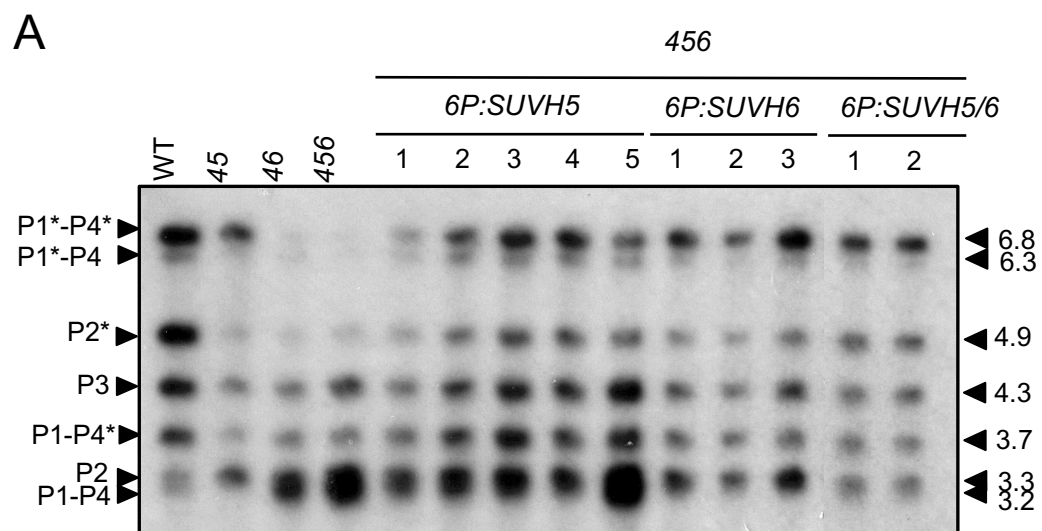


Figure 7: The SUVH5 amino terminus enhances SUVH6 carboxy-terminal domain activity at the 25S rDNA repeats. (A) *PAI HincII* DNA gel blot of untagged *6P:SUVH5*, untagged *6P:SUVH6* control lines, and untagged *6P::SUVH5/6* swap construct lines in the *456* background. Left and right margin labels are as described in Figure 3. (B) 25S rDNA repeat *AvaII* DNA gel blot with untagged *6P:SUVH5* and *6P:SUVH6* control lines, and untagged *6P:SUVH5/6* swap construct lines in the *456* background. Black arrowheads indicate bands corresponding to 5meC-inhibited cleavage products. Transgene structures are shown in Figure 1. WT indicates wild type *Ws*, *45* indicates *pail suvh4 suvh5*, and *46* indicates *pail suvh4 suvh6* control strains. Numbered lanes indicate independent transgenic lines for each construct.

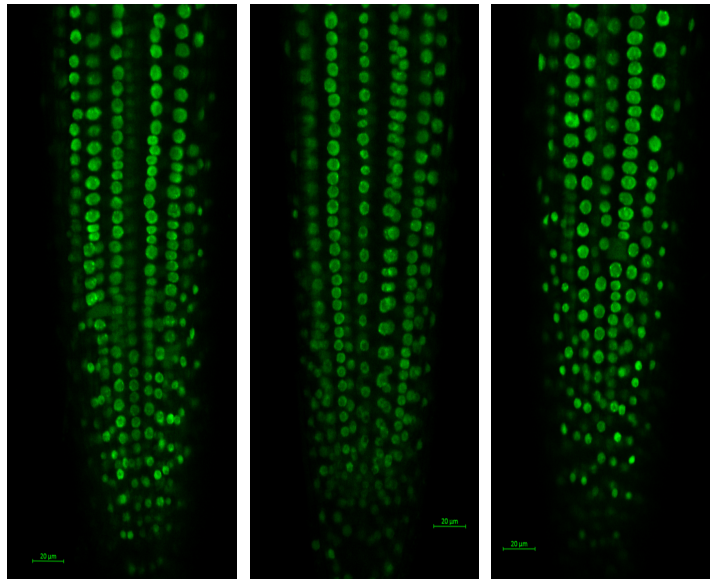
A

GFP-SUVH5

1

2

3



B GFP-SUVH5

1

2

3

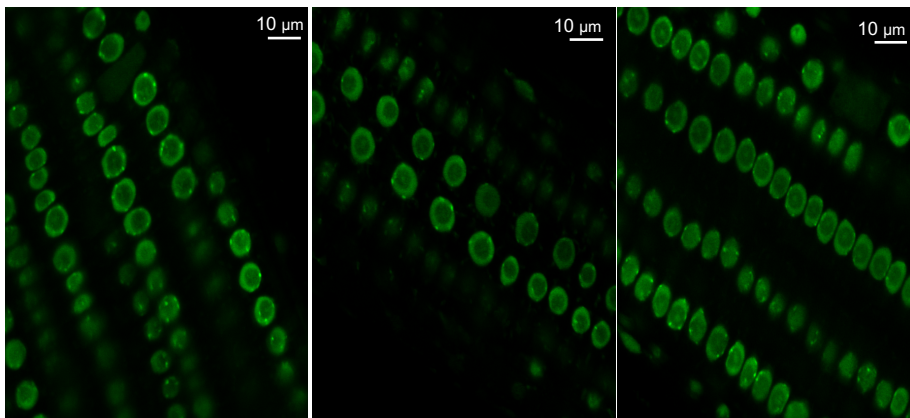
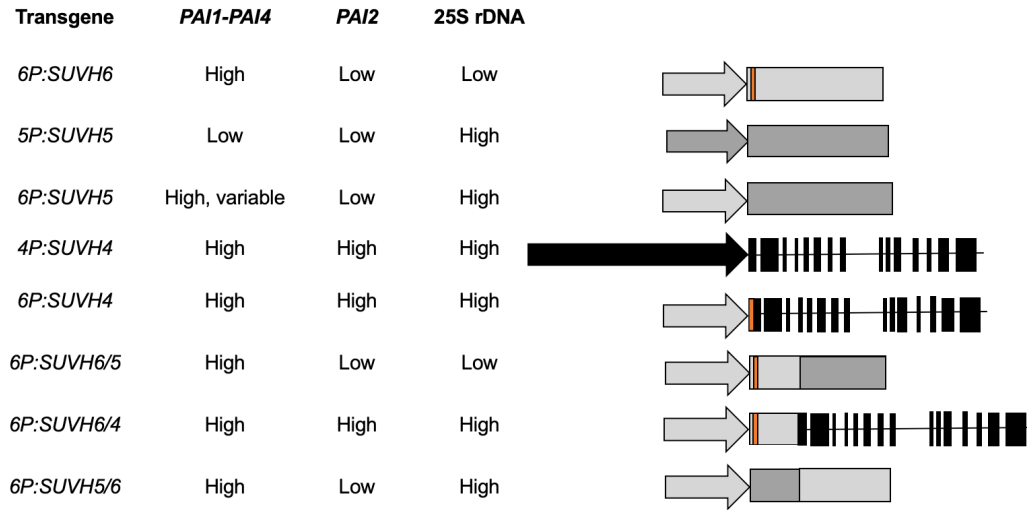


Figure 8: GFP-tagged SUVH5 is nuclear localized with enrichment in puncta. Live root tips from epitope-tagged *6P:GFP-SUVH5* constructs in the *456* background were imaged using confocal fluorescence microscopy at (A) 20X or (B) 40X magnifications. Numbers indicate independent transgenic lines.



Supplemental Figure 1: Summary of transgene add-back phenotypes at the loci assayed in this chapter. Orange bar indicates 3XFLAG epitope tag. Black indicates *SUVH4* sequences, Dark gray indicates *SUVH5* sequences, and light gray indicates *SUVH6* sequences.

References

- Chandrasekhara, C., Mohannath, G., Blevins, T., Pontvianne, F., & Pikaard, C. S. (2016). Chromosome-specific NOR inactivation explains selective rRNA gene silencing and dosage control in Arabidopsis. *Genes and Development*, 30(2), 177–190. <https://doi.org/10.1101/gad.273755.115>
- Clough, S. J., & Bent, A. F. (1998). Floral dip: A simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant Journal*, 16 (June 1998), 735–743. <https://doi.org/10.1046/j.1365-313X.1998.00343.x>
- Du, J., Johnson, L. M., Groth, M., Feng, S., Hale, C. J., Li, S., ... Jacobsen, S. E. (2014). Mechanism of DNA methylation-directed histone methylation by KRYPTONITE. *Molecular Cell*, 55(3), 495–504. <https://doi.org/10.1016/j.molcel.2014.06.009>
- Ebbs, M. L., Bartee, L., & Bender, J. (2005). H3 lysine 9 methylation is maintained on a transcribed inverted repeat by combined action of SUVH6 and SUVH4 methyltransferases. *Molecular and Cellular Biology*, 25(23), 10507–10515. <https://doi.org/10.1128/MCB.25.23.10507>
- Ebbs, M. L., & Bender, J. (2006). Locus-specific control of DNA methylation by the Arabidopsis SUVH5 histone methyltransferase. *The Plant Cell*, 18(5), 1166–1176. <https://doi.org/10.1105/tpc.106.041400.1>
- Fransz, P., Jong, J. H. De, Lysak, M., Castiglione, M. R., & Schubert, I. (2002). Interphase chromosomes in Arabidopsis are organized as well defined chromocenters from which euchromatin loops emanate.
- Johnson, L. M., Bostick, M., Zhang, X., Kraft, E., Henderson, I., Callis, J., & Jacobsen, S. E. (2007). The SRA Methyl-Cytosine-Binding Domain Links DNA and Histone Methylation. *Current Biology*, 17, 379–384. <https://doi.org/10.1016/j.cub.2007.01.009>
- Krauss, V. (2008). Glimpses of evolution: Heterochromatic histone H3K9 methyltransferases left its marks behind. *Genetica*, 133(1), 93–106. <https://doi.org/10.1007/s10709-007-9184-z>
- Lampropoulos, A., Sutikovic, Z., Wenzl, C., Maegerle, I., Lohmann, J. U., & Forner, J. (2013). GreenGate - A Novel, Versatile, and Efficient Cloning System for Plant Transgenesis. *PLoS ONE*, 8(12), e83043. <https://doi.org/10.1371/journal.pone.0083043>
- Lee, J., Lee, H. J., Shin, M., & Ryu, W. (2004). Versatile PCR-mediated insertion or deletion mutagenesis. *Benchmarks*, 36(3), 398–400.

- Li, X., Harris, C. J., Zhong, Z., Chen, W., Liu, R., Jia, B., ... Du, J. (2018). Mechanistic insights into plant SUVH family H3K9 methyltransferases and their binding to context-biased non-CG DNA methylation. *Proceedings of the National Academy of Sciences*, 115(37), 201809841. <https://doi.org/10.1073/pnas.1809841115>
- Malagnac, F., Bartee, L., & Bender, J. (2002). An Arabidopsis SET domain protein required for maintenance but not establishment of DNA methylation. *EMBO J*, 21(24), 6842–6852.
- Melquist, S., Luff, B., & Bender, J. (1999). Arabidopsis *PAI* Gene Arrangements , Cytosine Methylation and Expression, 4.
- Mergner, J., Frejno, M., List, M., Papacek, M., Chen, X., Chaudhary, A., ... Johannes, F. (2020). Mass-spectrometry-based draft of the Arabidopsis proteome. *Nature*, 579(March). <https://doi.org/10.1038/s41586-020-2094-2>
- Min, J., Zhang, X., Cheng, X., Grewal, S. I. S., & Xu, R. (2002). Structure of the SET domain histone lysine methyltransferase Clr4. *Nature Structural Biology*, 9(11). <https://doi.org/10.1038/nsb860>
- Stroud, H., Greenberg, M. V. C., Feng, S., Bernatavichute, Y. V., & Jacobsen, S. E. (2013). Comprehensive Analysis of Silencing Mutants Reveals Complex Regulation of the Arabidopsis Methylome. *Cell*, 152(1–2), 352–364. <https://doi.org/10.1016/j.cell.2012.10.054>
- Zhang, L., & Ma, H. (2012). Complex evolutionary history and diverse domain organization of SET proteins suggest divergent regulatory interactions. *New Phytologist*, 195, 248–263. <https://doi.org/10.1111/j.1469-8137.2012.04143.x>
- Zhang, X., Yang, Z., Khan, S. I., Horton, J. R., Tamaru, H., Selker, E. U., & Cheng, X. (2003). Structural Basis for the Product Specificity of Histone Lysine Methyltransferases. *Molecular Cell*, 12(1), 177–185. [https://doi.org/10.1016/s1097-2765\(03\)00224-7](https://doi.org/10.1016/s1097-2765(03)00224-7)

Chapter 4: Conclusions and Perspectives

Written by JK with editing from Dr. Judith Bender

Many studies have underlined the importance of SUVH4 in broad H3K9me throughout the *Arabidopsis thaliana* genome. My thesis work has focused on characterization of *Arabidopsis* SUVH5 and SUVH6 H3K9 methyltransferases and understanding molecular underpinnings of their locus-specific behavior. SUVH5 and SUVH6 belong to a clade of diversified SUVH proteins that share amino-terminal extension sequence features: the presence of amino-terminal conserved motif (ACM) sequences in an elongated disordered sequence context. I have shown in Chapter 2 that the ACM is a unique motif that contributes to both SUVH6 nuclear localization and activity at a preferred target. In Chapter 3, my work demonstrates that the SUVH6 and SUVH5 amino-terminal extensions contain residues critical for locus-specificity within the SUVH6 clade. My results show that the SUVH6 clade amino-terminal extensions regulate activity of SUVH6 clade carboxy-terminal domains but may not regulate the more distant SUVH4 carboxy-terminal domains.

Evolution of the SUVH6 clade

The SUVH6 clade is present in gymnosperms and ferns with fully sequenced genomes (Chapter 1) but is not present in the moss *Selaginella*. Ferns have much larger average genomes than mosses due to whole genome duplications, resulting in expansions of transposons and duplications of basal SUVH sequences. The pattern of SUVH6 clade evolution with large genome sizes lends support to the idea that SUVH6 clade proteins were diversified to

transcriptionally silence increased loads of transposons within genomes.

Expression analysis of dicots *Arabidopsis* and tomato and monocots corn and rice (Chapter 1 Figure 4, Figure 5) demonstrates that sperm have especially high expression of *SUVH6* clade genes. The importance of SUVH6 clade proteins in transcriptional silencing for larger genomes could be particularly important in gametophytes.

Because diversified SUVH proteins largely lack introns, SUVH clades are hypothesized to have diversified from a basal SUVH4 clade ancestral sequence through retrotransposition of a SUVH4 cDNA, (Baumbusch et al., 2001). I have characterized distinguishing features of the SUVH6 clade: presence of the ACM and extension of the amino-termini. The absence of the ACM from non-SUVH6 clade proteins and the lack of conservation of the SUVH6 clade amino-termini to other sequences has led us to question the origins of the SUVH6 amino-termini. BLAST searches using *Arabidopsis* SUVH5 and SUVH6 amino-terminal extensions as query sequences have not yielded insights to the origin of the SUVH6 clade amino-terminal extensions. In my motif-finder analysis I occasionally noticed diverged SRA and catalytic domain motifs in SUVH6 clade amino-terminal sequences. From this observation I propose that the SUVH6 clade could have originated as a tandem repeat duplication of SUVH4 cDNAs, with the sequences amino-terminal to the pre-SRA domain subsequently undergoing non-constrained evolution, diversifying from original domain functions.

Unlike the SRA motifs that were sporadically found in SUVH6 amino-terminal extensions, the ACM is a consistently conserved motif. However, I

noticed that the ACM shows similarity to the second alpha helix in the pre-SRA domain (Figure 1). In SUVH4 clade members and in SUVH6 clade members, the second pre-SRA alpha-helix sequence has notable similarities to the ACM sequence: enrichment of basic residues, several invariant non-polar residues, and enrichment for glycine residues at the carboxy-terminal end (Figure 1). However, the SUVH4 clade consensus and the SUVH6 clade consensus are substantially different from each other. Perhaps one or more duplications of the SRA cDNA sequence led to diversification and retention of the ACM in SUVH6 clade proteins. This also would provide a mechanistic explanation for the specificity of the ACM to SUVH6 clade proteins.

The ACM encodes a nuclear localization signal

Although the consensus ACM does not contain a classic nuclear localization signal (NLS), it is enriched for basic residues at the amino-terminal end (Chapter 2, Figure 1). I found that a *GFP-SUVH6* construct with the ACM deleted displays largely cytoplasmic localization (Chapter 2 Figure 4), indicating that one role of the SUVH6 ACM is to function in nuclear localization. However, in my *GFP-SUVH6* experiments there is residual nuclear signal in both Δ ACM and *K51A R52A* mutants, and these mutant proteins can promote slight remethylation at the chromosome two 25S rDNA repeats (Chapter 2 Figure 3 and Figure 4). Therefore, while the ACM clearly functions as the preferred NLS, perhaps other basic patches within the SUVH6 sequence are capable of mediating

inefficient nuclear localization. In addition, the *K51A R52A* mutant *SUVH6-GFP* fusion indicates that the KR residues are not required for chromatin localization because mitotic chromosomes show GFP enrichment after the nuclear envelope breaks down and the *K51A R52A* mutant protein has access to chromatin (Chapter 2 Supplemental Figure 8).

Unlike the consensus ACM, the *Arabidopsis* SUVH6 ACM encodes a cluster of four basic residues that constitutes a classic monopartite NLS. It is unclear if ACM sequences in other species without the four basic residues also mediate nuclear localization, with the ACM providing a new type of non-canonical NLS. This question can be tested with a similar approach to my analysis of *Arabidopsis* SUVH6, with GFP-SUVH fusion proteins stably expressed in transgenic *Arabidopsis* or using transient transfections of fluorescently-tagged SUVH constructs into *Nicotiana benthamiana* (tobacco) tissue. Comparing fluorescent protein localization patterns for wild type and ACM deletion or missense mutant pairs would indicate which ACMs are necessary for nuclear localization. The grape VvSUVH5a single ACM sequence would be a priority to analyze with this strategy because it is well conserved yet contains only two basic residues at the amino-terminal end (Chapter 2, Figure 1C). If the VvSUVH5a ACM proved to be necessary for nuclear localization, this result would indicate that the ACM can function as a non-canonical NLS. The fluorescent protein fusion approach can also be used to analyze the contributions of the different individual ACMs to nuclear localization in proteins that contain two or three ACMs.

To determine if other residues in the *Arabidopsis* SUVH6 ACM contribute to a non-canonical NLS or to the positive regulatory role observed for F61, further mutagenesis is required. In particular, ACM consensus sites A9, R11, P14, P15, and G18 show the highest conservation (Chapter 2 Figure 1B). Single amino acid mutagenesis could provide further evidence about the nature of these conserved residues and which part of ACM functionality they contribute to. The FPP sequence from residues 13-15 of the consensus ACM is a notably large and hydrophobic stretch that could perform hydrophobic stacking interactions.

The *Arabidopsis* SUVH4 NLS has not been experimentally determined. However, experiments with epitope-tagged domains of the rice OsSUVH4 support one or more NLS in the amino-terminus through pre-SRA regions of the protein, but not in the carboxy-terminal domains of the protein (Ding et al., 2007a, Ding et al., 2007b). From these observations, it could be that the pre-SRA domain contains the major SUVH4 NLS, which agrees with the similarities I found between the SUVH4 pre-SRA helix and the ACM sequences (Figure 1). An NLS in the SUVH4 pre-SRA domain also provides a potential mechanism for the minor nuclear localization observed in Δ ACM lines; these transgenes could use an inefficient and diverged pre-SRA helix as the NLS. Mutagenesis of the basic residues in the *Arabidopsis* SUVH4 pre-SRA helix will reveal whether this region plays a role in nuclear localization, providing more clues to the origin of the SUVH6 clade ACM. However, selection of residues for mutagenesis needs to be made carefully due to the known contacts that SUVH4 pre-SRA helices make with the SRA and SET domains (Du et al., 2014).

The ACM has a second role as a positive regulator of SUVH6 activity at a preferred target

The *Arabidopsis* SUVH6 *F61A* mutant shows normal nuclear localization but mediates only weak remethylation phenotypes at the preferred target *PAI1-PAI4* (Chapter 2). The *F61A* mutation could impair SUVH6 activity by interfering with an interaction either in *trans* with other factors or in *cis* with other SUVH6 residues. The histone deacetylase HDA6 was previously identified as a SUVH6 interacting protein, but the interaction interface was mapped to the SUVH6 carboxy-terminus. (Yu et al., 2017). In addition, I used a yeast two-hybrid approach with a SUVH6 amino-terminal extension bait to screen for interacting proteins but did not recover any relevant interacting proteins. However, a more comprehensive approach with immunoprecipitation followed by mass-spectrometry could identify additional SUVH6-interacting proteins, with ACM mutations including *F61A* allowing dissection of a role in ACM function. Similarly, RNA immunoprecipitation with SUVH6 wild type versus ACM mutants could be used to determine if RNA species are involved in ACM function.

My analysis demonstrates that the *F61A* mutant does not impair nuclear localization, gross protein stability, or general chromatin localization. Additionally, the *F61A* mutant localizes to chromocenters and mitotic chromosomes (Chapter 2), indicating that the mutation does not abrogate large scale chromatin localization. However, more analysis is required to determine if *F61A* chromatin binding is intact throughout the genome, particularly on

chromosome arms where the *PAII-PAI4* locus is. The *F6IA* mutant could demonstrate defects in DNA binding, catalytic activity, or processivity after methylation is catalyzed.

Crystal structures of SUVH4 and SUVH6 without amino-terminal extensions have revealed *cis* interactions in and between the carboxy-terminal domains (Du et al., 2014, Li et al., 2018). For both *Arabidopsis* SUVH4 and SUVH6, the pre-SRA second alpha helix makes important interactions with both the SET domain and to the bound DNA that has a flipped-out nucleotide (Du et al., 2012, Li et al., 2018). Given the similarities between the pre-SRA second helix and ACM (Figure 1), the ACM could also function in *cis* with the SET domain, at least transiently. Perhaps the SUVH6 clade ACM and pre-SRA helix each interact with the SET domain and/or with bound DNA in different enzyme conformations. Due to difficulties obtaining a crystal structure with the likely disordered amino-termini, NMR spectroscopy should be considered to identify *cis* interactions with the ACM.

Amino-terminal extensions are a conserved feature of the SUVH6 clade

I have described another defining feature of SUVH6 clade proteins: elongated amino-terminal extensions (Chapter 2 Figure 1). A key difference between SUVH4 clade proteins and SUVH6 clade proteins is amino-terminal length. However, with the exception of the ACM sequences, the SUVH6 clade amino-termini are largely diverged (Chapter 2, Supplemental Figure 1). In

seeking to determine if there are conserved secondary structure features that might function in the amino-termini, I discovered that SUVH amino-termini consistently show strong disorder predictions. The shorter amino-termini in SUVH4 clade proteins also show strong disorder predictions (Chapter 2, Supplemental Figure 3). These putatively disordered regions could function in a phase-separated chromatin compartmentalization mechanism.

To test the function of the SUVH6 putatively disordered sequences, I used a transgene add-back approach with deletion mutation *ΔLDR* that removes the majority of sequences from the region between the ACM and the pre-SRA domain. I found that the SUVH6 *ΔLDR* protein has at most a weak impairment of activity in remethylating the preferred target *PAII-PAI4* (Chapter 2 Figure 2). However, *PAII-PAI4* is a euchromatic target, so the effects of the LDR on SUVH6 activity at heterochromatic loci is unclear, particularly within chromocenters. I also found that the GFP-tagged wild type and *ΔLDR* mutant have similar localization to chromocenters, which have been hypothesized as phase separated nuclear compartments (Wang & Liu, 2019) (Chapter 2 Figure 4, Figure 5). These observations argue against a phase separation mechanism for the SUVH6 amino-terminal extension. However, because even wild type SUVH6 has weak activity across the variety of sequences concentrated in the chromocenter (see below), SUVH6 could have a poor phase-separation mechanism at chromocenters or has weak activity within certain phase-separated compartments.

The lack of impaired phenotypes for the *ALDR* mutant suggest that post-translational modifications of amino-terminal sequences are not a critical component of SUVH6 function at *PAI-PAI4*. For example, even though a recent proteomic analysis found that the SUVH6 residue serine 248 is phosphorylated (Mergner et al., 2020), removal of this site in the *ALDR* mutant does not result in obvious phenotypic consequences in our assays. However, SUVH5 contains two serine phosphorylation sites in its amino-terminal extension region (Mergner et al., 2020). Perhaps the overall level of phosphorylation in the amino-terminal extension enhances activity at some target regions such as the SUVH5 preferred target 25S rDNA repeats. Phosphorylation of residues in the disordered amino-terminal extension sequences could influence activity by altering the physical properties of the disordered region or by creating a regulatory signal. The SUVH5 amino-terminal extension is more complex than that of SUVH6 with two ACM sequences and a different disorder prediction topology, but a similar transgenic add-back approach could assess the contributions of each ACM, two phosphorylation sites, and disordered residues. By discovering that the 25S rDNA repeats are a preferred SUVH5 target with strong remethylation even from low expression transgenes (Chapter 3 Figure 2), I have developed an appropriate assay system for detailed protein analysis of SUVH5.

My initial results indicate that addition of the elongated SUVH6 amino-terminus to the SUVH4 carboxy-terminal DNA-binding and catalytic domains may not restrict the broad activity of SUVH4 (Chapter 3 Figure 6). An important caveat of this result is that the SUVH6 amino-terminus may be cleaved from the

transcript or protein in this swap fusion. However, the SUVH5 and SUVH6 amino-terminal extensions confer altered targeting preferences in the context of a SUVH5/SUVH6 carboxy-terminus (Chapter 3 Figure 5, Figure 7). These observations support a mechanism where the long amino-terminus has co-evolved with the carboxy-terminal domains in the SUVH6 clade to create a regulatory switch for H3K9 methyltransferase function. This regulatory switch could involve direct interactions between SUVH6 clade domains and/or involve interacting co-factors.

The SUVH6 clade amino-terminal extension contributes to locus-specific activity

In Chapter 3 I found that the SUVH5 and SUVH6 amino-terminal extensions can contribute to locus-specific preferred activity in the context of SUVH5 or SUVH6 carboxy-terminal domains. My results demonstrate that a swap of the SUVH5 amino-terminal extension on the SUVH6 carboxy-terminal domains confers increased remethylation at the 25S rDNA repeats. Conversely, a swap of the SUVH6 amino-terminal extension on the SUVH5 carboxy-terminal domains restricts remethylation at the 25S rDNA repeats.

SUVH5 and SUVH6 have differences in ACM composition and placement within the amino-terminal extension: SUVH5 has two ACMs, with the second more diverged and central to the amino-terminal extension, and SUVH6 has a single ACM at the far amino-terminus (Chapter 2, Figure 1). Do both

SUVH5 ACM sequences play regulatory roles? Analysis of transgenes with mutations to either ACM sequence can elucidate this significant difference between SUVH5 and SUVH6 amino-termini. The second SUVH5 ACM could be a nonfunctional relic, given the lack of conservation with the consensus ACM. Alternatively, the second ACM could switch off with the first ACM in different SUVH5 states. Though both SUVH5 and SUVH6 amino-termini are predicted to be disordered (Chapter 2), the disorder plots differ between them. SUVH5 and SUVH6 amino-termini could confer locus-specific preferences by contributing different biophysical properties to the carboxy-terminal domains. These differences could also provide varied phase-separation properties or activities within a phase separated compartment.

While the few known preferred loci for SUVH5 and SUVH6 have been instrumental in assaying transgenic protein features, determining where SUVH5 and SUVH6 each function can provide information about how features of the loci contribute to SUVH5/SUVH6 regulation. A comprehensive integration of SUVH5 versus SUVH6 preferred targets with other features including histone modifications, transcriptional activity, associated RNA species, and organization in nuclear domains could reveal patterns consistent among each group of targets. Chromatin immunoprecipitation-sequencing with antibodies to H3K9me1 and H3K9me2 for *suvh4 suvh5*, *suvh4 suvh6*, and *suvh4 suvh5 suvh6* mutants could identify many new preferred targets for endogenous SUVH5 in isolation and endogenous SUVH6 in isolation, leading to identification of informative patterns.

Determination of other SUVH6 preferred loci will be especially informative. To date, the *PAII-PAI4* transcribed inverted repeat is the only identified locus where SUVH6 displays highly efficient activity: SUVH6 alone in a *suvh4 suvh5* mutant is able to maintain H3K9me and 5meC at similar levels to wild type (Ebbs & Bender, 2006). The *PAII-PAI4* locus is transcribed across by RNA polymerase II from a heterologous distal promoter, but the central *PAI* duplicated sequences are modified by H3K9me and 5meC. Reflecting the hybrid externally-transcribed but internally-silenced nature of this locus, chromatin immunoprecipitation analysis shows enrichment for both the H3K4me mark associated with transcription and the H3K9me mark associated with silencing (Ebbs et al., 2005). In light of these observations, it could be that SUVH6 is most efficient at loci that are associated with RNA polymerase II and/or that contain some euchromatic histone modifications. Finding more SUVH6 preferred target loci will uncover to what extent transcription correlates with SUVH6 activation. Subsequent analysis can inform mechanisms of how the SUVH5 and SUVH6 clade amino-termini could impact locus-specificity.

Many H3K9me targets including *PAI2*, 5S rDNA repeats, and centromeric repeats are weak targets for both SUVH5 and SUVH6 compared to SUVH4 (Ebbs & Bender, 2006). For example, *PAI2* was not strongly remethylated in any of the SUVH5 and SUVH6 mutants, promoter variants, or domain swap transgenes that I tested (Chapter 2, Chapter 3). Furthermore, it is clear from my SUVH6 analysis that deleting amino-terminal extension sequences does not release SUVH6 to be more broadly active at *PAI2* or the 25S rDNA repeats (Chapter 2 Figure 2, Figure

3). These results argue against the view that the sole function of the amino-terminal extension is to restrict SUVH6 activity except at preferred targets. As an alternate approach to understanding the restricted activity of SUVH6 at *PAI2*, it would be interesting to test swap constructs of domains between SUVH6 and SUVH4, because SUVH4 can readily remethylate *PAI2* in a transgene add-back assay (Chapter 3 Figure 4). However, I determined that only *SUVH4* genomic constructs are effective at *PAI2* remethylation so it will be technically challenging to make SUVH4/SUVH6 swaps because sequence changes will need to be engineered in the context of the SUVH4 intron and exon structure, which could affect splicing efficiency.

Future characterization of preferred loci

In considering three-dimensional genomic context, *Arabidopsis thaliana* does not contain topologically associated domains but does contain A/B type compartments, which could underlie the observed differences between SUVH5 and SUVH6 preferred loci (Dogan & Liu, 2018). *PAI1-PAI4* is likely within an A compartment because RNA Polymerase II actively transcribes this locus. A comprehensive list of SUVH5 and SUVH6 preferred targets from the approaches suggested above combined with knowledge of A/B nuclear compartments might reveal a pattern for preferred targets. Beyond A/B compartments, the best characterized *Arabidopsis* three-dimensional nuclear structure thus far is the KNOT element of ten chromosome interactions that is enriched in transposons

(Grob et al., 2014). None of the sites of known locus-specificity occur within the KNOT chromosome sites, indicating that localization of SUVH5 or SUVH6 to the KNOT does not explain the locus-specificity described here.

Many of the sites assayed in this work occur within chromocenters, another key compartment in the *Arabidopsis* nuclear space. The chromosome two silenced 25S rDNA repeats colocalize with chromocenters, along with pericentromeric regions containing 180 bp repeats and a variety of transposable elements including SUVH5 preferred targets *Ta3* and *Mu1* (Fransz et al., 2002; Ebbs & Bender, 2006). Using GFP-SUVH fusion proteins, I determined that both SUVH5 and SUVH6 are localized to chromocenters (Chapter 2 Figure 4, Figure 5; Chapter 3 Figure 8). However, neither SUVH5 nor SUVH6 works efficiently at the 180 bp repeats, and SUVH6 also does not work efficiently at the 25S rDNA repeats or *Ta3* and *Mu1* transposons (Ebbs & Bender, 2006; Chapter 2 Figure 3). If target selectivity is controlled by nuclear compartmentalization, my results indicate that the compartments are organized at a finer level within the chromocenter than can be resolved by microscopy analysis of SUVH protein accumulation. Besides detecting SUVH5 at the chromocenters, in Chapter 3 I determined that overexpression of SUVH5 improves its activity at the euchromatic target *PAII-PAI4* (Chapter 3 Figure 2, Figure 3). Therefore, it could be that outside the target sequences assembled in the chromocenters, SUVH5 activity is mainly restricted by its low expression levels.

SUVH6 clade proteins in other plants

There is little known about the functions of SUVH4 and SUVH6 clade proteins in plants other than *Arabidopsis*. In particular, we do not know if locus-specificity is a conserved feature within the SUVH6 clade. Because the rice OsSUVH4 protein has already been characterized as a functional H3K9 methyltransferase (Ding et al., 2007), and because rice is well suited for genetic and genomic approaches, rice would be an excellent model monocot to determine the contributions of OsSUVH5a, OsSUVH5b, and OsSUVH5c in comparison to *Arabidopsis* SUVH6 clade proteins. A previous study has reduced OsSUVH5a transcript expression with RNA interference and found an increase in transcription for one probed transposon, indicating a reduction in silencing modifications (Qin et al., 2010). Of special interest is the function of the OsSUVH5c protein, which is a monocot-specific SUVH6 clade variant. OsSUVH5c contains two ACMs with diverged sequences and forms an outgroup with ZmSUVH5c to other SUVH5 and SUVH6 proteins.

Broader Perspectives

With changing environments and climate, we need to develop genetically modified crops that can withstand a harsher future. Understanding how H3K9 methyltransferases function at specific loci will be critical for preventing silencing of beneficial transgenes while still maintaining repeat and transposable element silencing. Transgene silencing is even a technical barrier in algae that could

provide new biofuels. For example, in the model chlorophytic algae *Chlamydomonas reinhardtii* nuclear transgenes are often susceptible to transcriptional silencing (Weisheit et al., 2009). The SUVH6 clade gives us the opportunity to elucidate a mechanism for regulating H3K9 methyltransferases to function preferentially at loci where silencing protects the genome. For example, my results suggest that SUVH5 functions preferentially at targets within the chromocenters whereas SUVH6 functions preferentially at the euchromatic target *PAII-PAI4*. Future structure/function analyses of SUVH5 and SUVH6 could reveal further ways to modify the preference for chromocenter transposons and repeats over euchromatic transgene insertions, allowing a means to create a better genetic background for transgenic crop plants and algae.

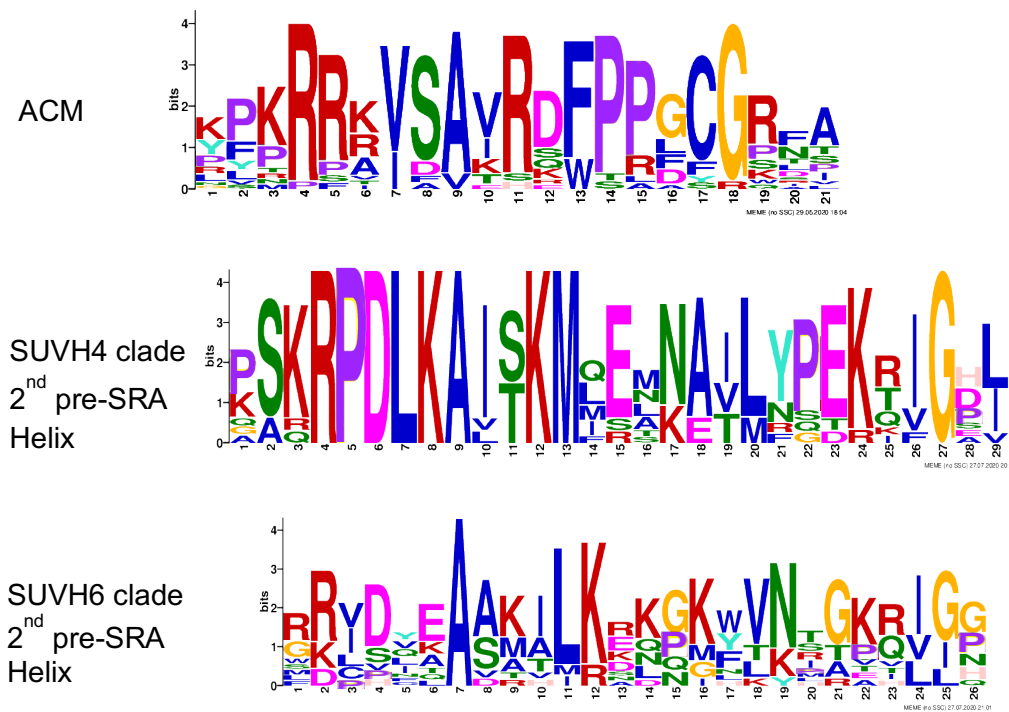


Figure 1: ACM consensus sequences resemble those of the SUVH4 consensus pre-SRA second helix. Consensus motifs generated with MEME on angiosperm SUVH4 and SUVH6 clade proteins used in Chapter 2 analyses. Residue letters scaled to frequency. Motifs were aligned for the highly conserved alanine.

References

- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., ... Noble, W. S. (2009). MEME SUITE : tools for motif discovery and searching. *Nucleic Acids Research*, 37(May), 202–208. <https://doi.org/10.1093/nar/gkp335>
- Baumbusch, L. O., Thorstensen, T., Krauss, V., Fischer, a, Naumann, K., Assalkhou, R., ... Aalen, R. B. (2001). The Arabidopsis thaliana genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. *Nucleic Acids Research*, 29(21), 4319–4333.
- Ding, B., Zhu, Y., Gao, J., Yu, Y., Cao, K., Shen, W., & Dong, A. (2007). Molecular characterization of three rice SET-domain proteins, 172, 1072–1078. <https://doi.org/10.1016/j.plantsci.2007.03.009>
- Ding, Y., Wang, X., Su, L., Zhai, J., Cao, S., Zhang, D., ... Cao, X. (2007). SDG714, a Histone H3K9 Methyltransferase, Is Involved in Tos17 DNA Methylation and Transposition in Rice. *The Plant Cell Online*, 19(1), 9–22. <https://doi.org/10.1105/tpc.106.048124>
- Doğan, E. S., & Liu, C. (2018). Three-dimensional chromatin packing and positioning of plant genomes. *Nature Plants*, 4(August), 521–529. <https://doi.org/10.1038/s41477-018-0199-5>
- Du, J., Johnson, L. M., Groth, M., Feng, S., Hale, C. J., Li, S., ... Jacobsen, S. E. (2014). Mechanism of DNA methylation-directed histone methylation by KRYPTONITE. *Molecular Cell*, 55(3), 495–504. <https://doi.org/10.1016/j.molcel.2014.06.009>
- Ebbs, M. L., Bartee, L., & Bender, J. (2005). H3 lysine 9 methylation is maintained on a transcribed inverted repeat by combined action of SUVH6 and SUVH4 methyltransferases. *Molecular and Cellular Biology*, 25(23), 10507–10515. <https://doi.org/10.1128/MCB.25.23.10507>
- Ebbs, M. L., & Bender, J. (2006). Locus-specific control of DNA methylation by the Arabidopsis SUVH5 histone methyltransferase. *The Plant Cell*, 18(5), 1166–1176. <https://doi.org/10.1105/tpc.106.041400.1>
- Fransz, P., Jong, J. H. De, Lysak, M., Castiglione, M. R., & Schubert, I. (2002). Interphase chromosomes in Arabidopsis are organized as well defined chromocenters from which euchromatin loops emanate.
- Grob, S., Schmid, M. W., & Grossniklaus, U. (2014). Article Hi-C Analysis in Arabidopsis Identifies the KNOT , a Structure with Similarities to the

- flamenco Locus of *Drosophila*. *Molecular Cell*, 55(5), 678–693.
<https://doi.org/10.1016/j.molcel.2014.07.009>
- Li, X., Harris, C. J., Zhong, Z., Chen, W., Liu, R., Jia, B., ... Du, J. (2018). Mechanistic insights into plant SUVH family H3K9 methyltransferases and their binding to context-biased non-CG DNA methylation. *Proceedings of the National Academy of Sciences*, 115(37), 201809841.
<https://doi.org/10.1073/pnas.1809841115>
- Mergner, J., Frejno, M., List, M., Papacek, M., Chen, X., Chaudhary, A., ... Johannes, F. (2020). Mass-spectrometry-based draft of the Arabidopsis proteome. *Nature*, 579(March). <https://doi.org/10.1038/s41586-020-2094-2>
- Qin, F. J., Sun, Q. W., Huang, L. M., Chen, X. S., & Zhou, D. X. (2010). Rice SUVH histone methyltransferase genes display specific functions in chromatin modification and retrotransposon repression. *Molecular Plant*, 3(4), 773–782. <https://doi.org/10.1093/mp/ssq030>
- Yu, C., Tai, R., Wang, S., Yang, P., Luo, M., Yang, S., ... Wu, K. (2017). HISTONE DEACETYLASE6 Acts in Concert with Histone Methyltransferases SUVH4 , SUVH5 , and SUVH6 to Regulate Transposon Silencing. *The Plant Cell*, 29(8), 1970–1983.
<https://doi.org/10.1105/tpc.16.00570>