

EPIGENETIC REGULATION OF THE DEFENSE GENE INDUCTION
IN *ARABIDOPSIS THALIANA* IN RESPONSE TO *PSEUDOMONAS*
SYRINGAE

by

Yogendra Bordiya, M.S.

A dissertation submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
with a Major in Aquatic Resources
December 2017

Committee Members:

Hong-Gu Kang, Chair

Nihal Dharmasiri

Sunethra Dharmasiri

Daniel F. Klessig

Ping He

COPYRIGHT

by

Yogendra Bordiya

2017

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

Fair Use

This work is protected by the Copyright Laws of the United States (Public Law 94-553, section 107). Consistent with fair use as defined in the Copyright Laws, brief quotations from this material are allowed with proper acknowledgement. Use of this material for financial gain without the author's express written permission is not allowed.

Duplication Permission

As the copyright holder of this work I, Yogendra Bordiya, authorize duplication of this work, in whole or in part, for educational or scholarly purposes only.

DEDICATION

To my grandfather, Mr. Bhagirath Bordiya from whom I learned the value of hard work, discipline, and conscientiousness. He converted barren lands into cultivable, fertile fields. He devoted his life to the improvement of his local community and society. He was a teacher, a farmer, a leader and a visionary.

ACKNOWLEDGEMENTS

This work would not be possible without Dr. Hong-Gu Kang. I am very grateful for the opportunity he provided me to work in his lab, for his vast knowledge, expert guidance, and patience while advising me. I have learned a lot since the time I started my doctoral research with him and appreciate his desire for perfection and highly analytical way of designing experiments and interpreting the data. All of his afore mentioned virtues have pushed me to become a better student and researcher. I would also like to thank my committee members Dr. Nihal Dharmasiri, Dr. Sunethra Dharmasiri, Dr. Daniel Klessig and Dr. Ping He for their valuable time and critical feedback on my research. I extend my gratitude to April Bonnard, Anthony Ledet, Kinsey Cozby, Nicole Beisel, Ji-Chul Nam and Dr. Sung Il Kim for their assistance during my experiments and feedback on my research. I would like to acknowledge the members of Dr. Dharmasiri's lab, especially Dr. Praveen Kumar Kathare and Thilanka Darshana Jayaveera for highly analytical discussions during weekly lab meeting which helped me hone my professional skills. I would like to thank our collaborator Dr. Habil Zare from the department of Computer Science at Texas State University and his student Gabriel Hurtado who performed the bioinformatics analysis on the RASL-seq data and Dr. David Rodriguez and Trina Guerra from the department of Biology for technical help with the Illumina-MiSeq. I express my appreciation for Dr. Benjamin Larman from the Johns Hopkins School of Medicine who graciously shared his RASL-seq procedure with us. The department of Biology's financial support through Doctoral Instructional assistantship for four years gave me my bread and

butter for which I am very grateful. Lastly, I am deeply grateful to the Bordiya and Taylor family, my fiancé Hallie Taylor and friends for supporting me through these five and half years of intense research

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	v
LIST OF TABLES	vii
LIST OF FIGURES	viii
ABSTRACT	x
CHAPTER	
I. PATHOGEN INFECTION AND MORC1 AFFECT CHROMATIN ACCESSIBILITY OF TRANSPOSABLE ELEMENTS AND EXPRESSION OF THEIR PROXIMAL GENES IN ARABIDOPSIS	1
II. ROLE OF RNA DIRECTED DNA METHYLATION MEDIATED TRANSCRIPTIONAL GENE SILENCING IN PLANT IMMUNITY	33
III. GLOBAL IDENTIFICATION OF DEFENSE GENES WITH RAPID INDUCTION KINETICS AND THEIR CHARACTERIZATION IN <i>DCL</i> MUTANTS	59
IV. MATERIALS AND METHODS	81
APPENDIX SECTION	91
REFERENCES	100

LIST OF TABLES

Table	Page
2.1. Epigenetic mutants crossed with <i>morc1/2</i> display visible phenotypes	43
3.1. RNA-seq sample composition to assess the induction dynamics of defense genes ...	67
3.2. RASL-seq sample detail	87
3.3. Sequence of the adapter, barcode and sequencing primers used in the RASL-seq.....	88
3.4. List of the criteria and number of genes used in the RASL-seq	88

LIST OF FIGURES

Figure	Page
1.1. DNase I hypersensitive sites are predominantly located in the promoter regions of genes and TEs.....	8
1.2. Pairwise comparisons among naïve, mock- or <i>Pst</i> -inoculated WT and MORC-deficient plants reveal differential DHSs (dDHSs).....	10
1.3. Pathogen infection and mutations in MORC family members alter chromatin accessibility in TEs	12
1.4. Effects of MORC1 and infection on DNase I accessibility of heterochromatic TEs	13
1.5. Quantitative PCR analysis of selected dDHSs identified by DNase-Seq	15
1.6. The degree of representation of genes neighboring TE-dDHSs	17
1.7. MORC1/2 influences the kinetics and/or amplitude of defense genes induced by <i>Pst</i>	19
1.8. MORC1 is physically associated with infection-induced TE-dDHSs	21
1.9. Silencing of <i>Pst</i> -induced TE-dDHSs proximal to a MORC1 binding site compromised induction of adjacent defense genes.....	25
1.10. Working model to explain the role of MORC1 in plant immunity and epigenetics.....	30
2.1. Canonical and non-canonical RdDM pathway of transcriptional gene silencing.....	42
2.2. Curly leaf phenotype of some of the high order mutants between <i>morc1/2</i> and additional epigenetic mutants	44
2.3. Expression of Copia TEs is induced in several epigenetic mutants and its induction was further enhanced by <i>Pst</i> infection.....	45
2.4. Release of TGS coincides with increase in TE activity immediately after pathogen infection.....	46

2.5. Some <i>dcl</i> mutants are more susceptible to avirulent <i>Pst</i> but more resistant to virulent <i>Pst</i>	49
2.6. Induction dynamics of <i>PR1</i> is affected in the <i>dcl</i> mutants.....	51
2.7. Induction dynamics of <i>PR2</i> is affected in the <i>dcl</i> mutants.....	52
2.8. Induction dynamics of <i>PR5</i> is affected in the <i>dcl</i> mutants.....	53
2.9. Arabidopsis DCL proteins are functionally redundant	55
2.10. Working model	57
3.1. RASL-seq procedure.....	63
3.2. RASL-seq library is compatible with Illumina sequencing platforms such as Hi-Seq and Mi-Seq	64
3.3. Clustering of defense genes displaying rapid induction dynamics in the RNA-seq data.....	68
3.4. Identification of 9 clusters of genes differentially expressed in Arabidopsis infected with avirulent and virulent <i>Pst</i>	69
3.5. <i>Pst</i> infection downregulates the expression of RdDM components including MORC1.....	70
3.6. Optimization of RASL procedure using RASL-PCR	72
3.7. High correlation of RASL-seq vs qRT-PCR and RASL-seq vs RNA-seq	73
3.8. RASL-seq showed that the expression kinetics of defense genes is altered in the <i>dcl</i> mutants	75
3.9. Basal level expression of a subset of defense genes is high in <i>dcl3-1</i>	76
3.10. Induction dynamics of some defense genes is compromised in <i>dcl1-7</i>	76

ABSTRACT

Plants and animals respond to an ever changing environment by making changes in the physiological level of various proteins and metabolites. This rapid physiological change in response to the environment is achieved by massive transcriptional reprogramming. Therefore, switching the large number of genes on and off at the right time in the right place requires highly sophisticated transcriptional regulation and is very important in mounting responses appropriate for the environmental change/stress. Recent studies suggest that epigenetics is one of the critical components in the regulation of transcription to help responding and adapting to environmental changes/stresses. The epigenetic regulation is achieved through the modification of chromatin structure, which is generally mediated by DNA/histone modifications, small RNAs (sRNAs), long non-coding RNA, and nucleosome positioning. In my dissertation, I have studied the role of epigenetic components which regulate defense responses in *Arabidopsis thaliana* in response to *Pseudomonas syringae*. To this end, I assessed biotic-stress-triggered changes in chromatin accessibility and characterized several epigenetic mutants in gene silencing in Arabidopsis. From these assessments, I particularly focused on testing the hypothesis that these epigenetic changes are important in the induction dynamics of defense genes triggered by infection. Intriguingly, I found that biotic-stress-triggered chromatin changes were frequently associated with transposable elements (TEs) proximal to defense genes, some

of which functioned as transcriptional enhancers. This observation suggested that a TE controlling mechanism(s) might be important in defense responses. Indeed, I found that more than a hundred TEs become transcriptionally induced under biotic stress, which merited further characterization of mutants involved in RNA-dependent DNA methylation (RdDM), the best characterized regulatory mechanism for TEs. I chose to characterize four *DCL* (dicer-like) genes that are important in the biogenesis of sRNAs, critical modulators for TEs and chromatin remodeling. Among these *dcl* mutants, *dcl1* displayed the most compromised resistance and induction of defense genes against avirulent *P. syringae*, suggesting that some sRNAs may be necessary for the rapid defense responses. In contrast, *dcl2* and *dcl3* showed marginally enhanced resistance and elevated expression of defense genes to the avirulent pathogen. In particular, *dcl2* and *dcl3* showed substantially increased expression of defense genes without pathogen challenges, suggesting that DCL2/3-generated sRNAs are important in suppressing defense genes. Note that the expression analysis of defense genes was performed using a novel targeted RNA-seq analysis known as RASL-seq (RNA-mediated oligonucleotide Annealing, Selection, and Ligation with next-generation sequencing). Selection of the defense genes used for the RASL-seq was chosen on the basis of my RNA-seq analysis, which identified rapidly induced genes at different time points in response to avirulent *P. syringae* relative to the virulent counterpart. In addition to altered defense gene induction in the mutants, I found that many RdDM genes including *MORC1* were transcriptionally suppressed as early as 6 hr post infection,

suggesting that RdDM components may play a role regulating the dynamics of defense responses.

CHAPTER I

PATHOGEN INFECTION AND MORC1 AFFECT CHROMATIN ACCESSIBILITY OF TRANSPOSABLE ELEMENTS AND EXPRESSION OF THEIR PROXIMAL GENES IN ARABIDOPSIS

Abstract

To assess MORC1's role in epigenetics in relation to plant immunity, genome-wide chromatin accessibility was compared between mock- or *Pseudomonas syringae* pv. *tomato* (*Pst*)-inoculated wild type (WT) Arabidopsis and/or the *morc1/2* double mutant. Most changes in chromatin accessibility, scored by DNase I hypersensitive sites (DHSs), were located in the promoters of genes and transposable elements (TEs). Comparisons between *morc1/2* and WT receiving the same treatment revealed differential DHSs (dDHSs) predominantly associated with heterochromatic TEs. By contrast, comparisons between mock- and *Pst*-inoculated plants from the same genotype showed dDHSs associated with biotic/abiotic stress-related genes; a smaller but significant population was in TEs. Interestingly, many defense genes, including *PR-1*, *PR-2* and *PR-5*, were proximal to *Pst*-induced, TE-associated dDHSs. A random subset of these defense genes showed moderately delayed/reduced expression in *Pst*-infected *morc1/2* as compared with WT. MORC1 was physically bound to chromatin in a *Pst* infection-responsive manner at sites dispersed throughout the genome. Notably, silencing of TE-associated dDHSs proximal to these infection-induced, MORC1-interacting sites led to significant suppression of *Pst*-induced transcription of adjacent defense genes, including *PR-1*. These results provide evidence that MORC1 is associated with TEs and suggest that a subset of these TEs may help regulate their proximal defense genes.

Introduction

Plants have evolved a variety of defense mechanisms to protect themselves from potentially pathogenic microorganisms. Upon infection, recognition of pathogen-associated molecular patterns (PAMPs) by extracellular surface receptors leads to the activation of PAMP-triggered immunity (PTI). While PTI is sufficient to prevent the further colonization of many microbes, evolutionary selection has led to the appearance of some pathogens containing effectors that suppress PTI. Whether such pathogens can effectively infect the host plant is then determined by whether the plant expresses a resistance (R) protein that recognizes one of these effectors (also termed avirulence [Avr] factors) and induces effector-triggered immunity (ETI; Jones & Dangl, 2006). Following the activation of either ETI or PTI, a variety of defense responses are induced, such as the accumulation of salicylic acid (SA) and the increased expression of certain *PR* genes (Vlot et al., 2009). In addition, the activation of plant immunity is associated with large-scale changes in gene expression (Vlot et al., 2009, Moore et al., 2011). This transcriptional reprogramming involves the highly coordinated action of myriad transcription factors and their associated proteins that function to recruit or modulate RNA polymerase II (RNAPII; Moore et al., 2011). Growing evidence suggests that chromatin modification/remodeling, which regulates the accessibility of DNA to the transcriptional machinery, as well as post-transcriptional regulation of defense-associated mRNAs, are essential for this phenomenon (Berr et al., 2012, Ma et al., 2011, Staiger et al., 2012).

Epigenetic gene regulation plays a critical role in cell differentiation and reprogramming. It also maintains genome integrity by silencing the expression of transposable elements (TEs) and other repeat sequences (Boyko & Kovalchuk, 2011, Zaratiegui et al., 2007).

Whether DNA is packaged as transcriptionally active euchromatin or transcriptionally silent heterochromatin is determined by the interplay between repressive epigenetic marks, such as DNA methylation and specific histone modifications, and chromatin remodeling complexes that reposition, evict, or alter the composition of nucleosomes (Boyko & Kovalchuk, 2008, Ma et al., 2011). Evidence supporting a role for epigenetic regulation of plant immune responses comes from the combined demonstrations that i) loss of histone modifying enzymes, including histone deacetylase (Kim et al., 2008b, Wang et al., 2010a), histone methyltransferases (Berr et al., 2011b, Palma et al., 2010, Alvarez-Venegas et al., 2007, De-La-Pena et al., 2012) or histone ubiquitin-ligase (Dhawan et al., 2009a), alters defense gene expression and resistance to bacterial and fungal pathogens, ii) SA treatment leads to altered histone methylation/acetylation in key defense genes, including *PR-1* (Mosher et al., 2006, Alvarez-Venegas et al., 2007), iii) mutations in several members of the SWI/SNF family of ATP-dependent chromatin remodeling factors, including Deficient in DNA methylation 1 (DDM1), Actin-related protein 6 (ARP6) and Photoperiod-independent Early Flowering 1 (PIE1), alter defense gene expression and/or influence plant immunity (Downen et al., 2012, March-Diaz et al., 2008, Li et al., 2010, Cheng et al., 2013), and iv) mutants lacking components in the RNA-directed DNA methylation (RdDM) pathway, which mediates epigenetic DNA methylation, exhibit altered expression of numerous pathogen-responsive genes and enhanced resistance to virulent, avirulent, and non-pathogenic strains of *Pst* (Downen et al., 2012, Yu et al., 2013).

TEs have long been considered transcriptionally inactive junk DNA (Probst & Almouzni, 2011, Lewin, 1986). However, several recent reports indicate that dynamic changes in TEs can occur, particularly in response to stress (Ito et al., 2011a, Grandbastien et al., 2005,

Tittel-Elmer et al., 2010). For example, the first comprehensive stress-induced DNA methylation map of the *Arabidopsis* genome indicated that TEs are differentially methylated in response to biotic stress or SA treatment (Downen et al., 2012). Moreover, demethylation of TEs following treatment with the bacterial PAMP flg22 correlated with the down-regulation of components involved in RdDM and the concomitant activation of TE expression (Yu et al., 2013). SA treatment also led to decreased methylation of some TEs, and this corresponded not only with their up-regulation, but also with increased expression of neighboring genes (Downen et al., 2012). In addition, a wide range of TEs, which are transcribed in response to stress with stress-responsive transcription, including *ONSEN* and *mPing*, have been implicated in inducing neighboring genes (Naito et al., 2009, Yasuda et al., 2013, Ito et al., 2011a, Makarevitch et al., 2015). Together, these observations suggest that TEs may play a role in triggering the transcriptional activation of proximal defense genes.

Efforts to elucidate the components required for resistance to turnip crinkle virus (TCV) in *Arabidopsis* led us to identify the *CRT1* (*Compromised Recognition of TCV*) gene (Kang et al., 2008). Sequence analysis revealed that *CRT1* contains the ATPase and S5 domains characteristic of Microrchidia (MORC) proteins (Kang et al., 2010). Additional members of the *Arabidopsis* CRT1 family have been identified by other groups and variously named MORC, DMS, or CRT/CRH (Brabbs et al., 2013, Lorkovic et al., 2012, Moissiard et al., 2012). To avoid further confusion, we use the nomenclature proposed in a previous report and refer to CRT1 and its family members as MORC1-7 (Langen et al., 2014). Through genetic and biochemical analyses, we and others demonstrated that MORC1 is required for multiple layers of immunity, including ETI and PTI, following inoculation by a wide range

of plant pathogens (Kang et al., 2008, Kang et al., 2010, Kang et al., 2012, Wang et al., 2011). In addition, MORC1 is one of very few proteins known to date that physically associates with a large number of immune components, including at least 12 R proteins and the PAMP recognition receptor FLS2 (Kang et al., 2008, Kang et al., 2010, Kang et al., 2012, Langen et al., 2014).

The demonstration that a small subpopulation of MORC1 is present in the nucleus, and that nuclear MORC1 levels increase following activation of PTI or ETI in *Arabidopsis* (Kang et al., 2012), raised the possibility that this protein also has a nuclear function during plant immunity. Consistent with this hypothesis, MORC proteins from a wide variety of eukaryotes and prokaryotes have been implicated in DNA recombination/repair and/or chromatin modification (Perry & Zhao, 2003, Iyer et al., 2008). Furthermore, MORC1 binds nucleic acids and exhibits Mn^{2+} -dependent endonuclease activity (Kang et al., 2012) as well as ATPase activity (Kang et al., 2008). Mouse MORC1, the first MORC protein identified, is required for meiotic nuclear division (Watson et al., 1998); more recently it was implicated in TE repression in the male germline (Pastor et al., 2014). Suggesting a similar role for *Arabidopsis* MORC proteins, MORC1 and/or MORC6 were identified in three independent screens for mutants defective for gene silencing (Moissiard et al., 2012, Lorkovic et al., 2012). The de-repression of silenced reporter genes and TEs in *morc1* and *morc6* mutants suggests that these proteins are required for epigenetic gene silencing (Lorkovic et al., 2012, Brabbs et al., 2013, Moissiard et al., 2012). However, it is currently unclear whether MORC1 and MORC6 fulfill their functions by modulating RdDM (Brabbs et al., 2013, Lorkovic et al., 2012) or by influencing chromatin structure (Moissiard et al., 2012).

In this study, we used both DNase-Seq (Hesselberth et al., 2009) and ChIP (Chromatin immunoprecipitation)-Seq (Furey, 2012) approaches to investigate the potential role MORC proteins play in regulating gene silencing and the mechanism through which this process impacts plant immunity. Analyses of DNase hypersensitive sites (DHSs) in mock- and *Pst*-inoculated wild type (WT) and *morc1/2* plants revealed that an enriched proportion of differential DHSs (dDHSs) are located in TEs. Strikingly, TE-associated dDHSs induced by *Pst* infection were identified near many defense-related genes. This finding, combined with the demonstration that i) MORC1 is a chromatin-interacting protein, ii) *Pst* infection increases MORC1 binding to *Pst*-induced dDHSs associated with genes and TEs, and iii) expression of many defense genes is delayed and/or reduced in *morc1/2* mutant plants, suggests that MORC proteins modulate plant immune responses by binding TEs and thereby influencing both their expression and that of proximal genes following pathogen infection.

Results

Pathogen infection did not dramatically alter the percentage of promoter-associated DHSs in either WT or MORC-deficient Arabidopsis

To characterize changes in the chromatin landscape in response to pathogen infection, we prepared and sequenced DNase-Seq libraries from three genotypes, i) WT, ii) *morc1-2 morc2-1* (*morc1/2*) (Kang et al., 2010), a double knock-out mutant lacking MORC1 and its closest homolog MORC2, and iii) *morc1-2 morc2-1 morc6-3* (*morc1/2/6*) (Moissiard et al., 2014), a triple knock-out mutant in which *morc6* was introduced into the *morc1/2* background. Plants from each genotype were subjected to four different treatments: no treatment (naïve), or inoculation with buffer (mock), virulent *Pst* (*Pst*), or avirulent *Pst* carrying *AvrRpt2* (*avrPst*). A total of 554 million reads were obtained. After discarding reads aligned to the Arabidopsis chloroplast genome, a total of 224 million reads were uniquely aligned to the Arabidopsis genome while 2.4 million reads were aligned with the *Pst* genome, suggesting very little contamination with bacterial DNA. The Pearson correlation coefficient between biological replicates for each combination ranged from 0.76 to 0.97; only four of the 36 libraries were lower than 0.85. Given that independently grown batches of plants were used for the biological triplicates to reduce false signals, these correlation coefficient values suggest high reproducibility of the data sets.

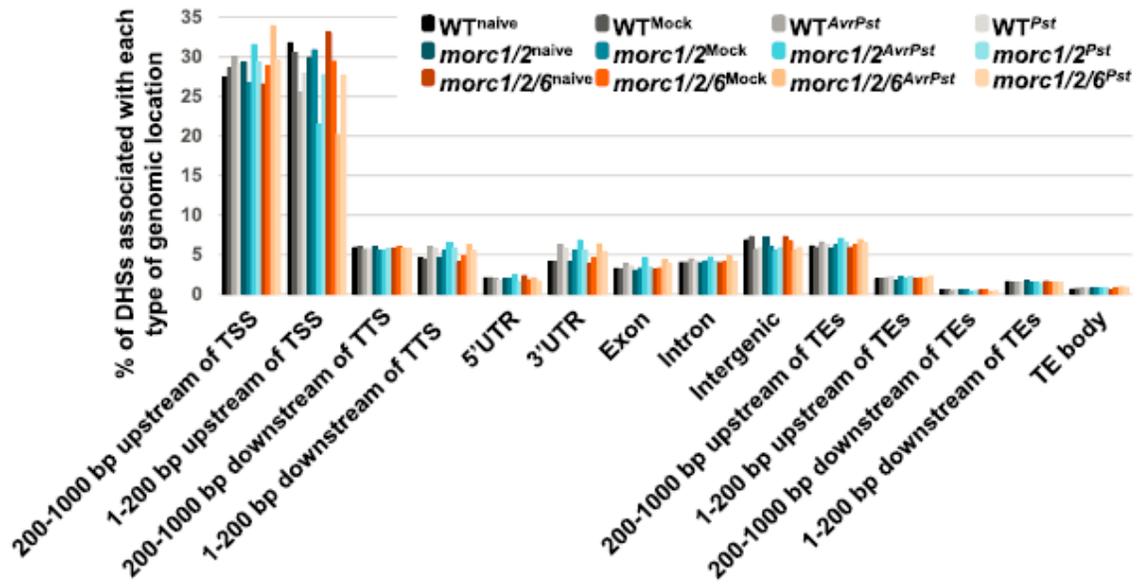


Fig. 1.1 DNase I hypersensitive sites are predominantly located in the promoter regions of genes and TEs. Genomic locations of DNase I hypersensitive sites (DHSs) relative to annotated protein-coding genes and TEs. The y-axis shows the percentage of DHSs in the indicated locations over the total number of DHSs associated with genes and/or TEs. TSS: transcription start site; TTS: transcription termination site; UTR: untranslated region.

A total of 29,450 DHSs were identified in the Arabidopsis genome (Fig. 1.1). As expected, the distribution of DHSs was biased toward euchromatin over heterochromatin, clearly displaying a strong correlation with the density of genes (Appendix 1.4). Among the DHSs, approximately 65% were located within 1 kb upstream of the transcription start sites (TSSs) or 1 kb downstream of the transcription termination site (TTS) of protein-coding genes, ~18% were associated with gene body, such as exons, introns, 5' and 3' UTRs, ~11% were located in sequences 1 kb upstream or downstream of TEs, ~1% were detected within TEs, and the remaining ~5% were in intergenic regions (Fig. 1.1A; Appendix 1.1). For protein-coding genes the majority of DHSs (64%) were detected within 1 kb upstream of their TSSs. Similarly, 78% of the DHSs associated with TEs were located in this region. These data suggest that the identified DHSs are heavily enriched in promoter regions, which are normally located within 1 kb upstream of the TSS. Pairwise comparisons among untreated

naïve plants and mock- or pathogen-inoculated WT, *morc1/2* or *morc1/2/6* mutants revealed no dramatic alterations in percentage of DHS at different genomic locations (Fig. 1.1).

Differential DHSs induced by *Pst* infection or loss of MORC family members were enriched in the promoters of TEs

To assess chromatin dynamics in response to pathogen infection and/or MORC1/2/6 deficiency, we performed pair-wise comparisons of the DHSs identified in the different plant backgrounds in the presence or absence of pathogen infection. The dDHSs identified by this process represent genomic sites at which the level of chromatin accessibility differs within the corresponding comparison. Pathogen infection induced significant changes in chromatin accessibility, since comparisons of DHSs between mock- and pathogen-infected plants within the WT, *morc1/2* or *morc1/2/6* backgrounds revealed hundreds of dDHSs (Fig. 1.2A; Appendix 1.2). By contrast, pairwise comparisons of DHSs in naïve vs mock or *avrPst* vs *Pst* identified 17 or fewer dDHSs (Fig. 1.2A; Appendix 1.2), indicating that mock inoculation does not substantially alter the DHS pattern present in untreated naïve plants, and that the DHSs induced by virulent and avirulent *Pst* are very similar.

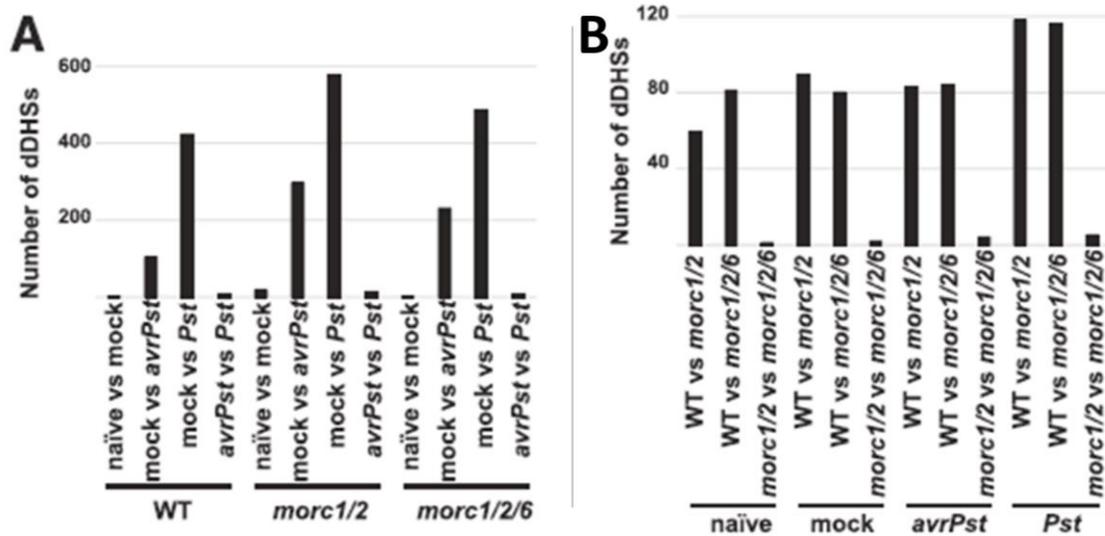


Fig. 1.2. Pairwise comparisons among naïve, mock- or *Pst*-inoculated WT and MORC-deficient plants reveal differential DHSs (dDHSs). **A**, The number of dDHSs detected in the indicated comparisons between different treatments are plotted for each genetic background. **B**, The number of dDHS from the indicated comparisons between different genetic backgrounds after a particular treatment are plotted.

To further investigate whether MORC proteins influence chromatin accessibility, we compared the DHSs identified in different genotypes subjected to the same treatment. Modest numbers of dDHSs, ranging from 59 to 118 (Fig. 1.2B; Appendix 1.2), were observed in WT vs *morc1/2* and WT vs *morc1/2/6* comparisons in which the plants were responding to the same treatment. By contrast, the DHS comparisons between *morc1/2* and *morc1/2/6* identified very few dDHSs, with a maximum of 5. This latter finding is in line with i) RNA-Seq analyses of *morc1/2* and *morc1/2/6*, which revealed very little difference between these lines (Moissiard et al., 2014), and ii) our findings that *morc1/2* and *morc1/2/6* exhibit comparable levels of enhanced susceptibility to virulent and avirulent *Pst* (Appendix 1.5). Given the small number of dDHSs observed in comparisons between i) naïve and mock, ii) *avrPst* and *Pst*, and iii) *morc1/2* and *morc1/2/6*, only analyses derived from combinations of mock- or *Pst*-inoculated WT plants (WT^{mock} and WT^{*Pst*}) and mock-

or *Pst*-inoculated *morc1/2* (*morc1/2*^{mock} and *morc1/2*^{*Pst*}) will be presented hereafter (Fig. 1.2).

Comparisons between the DHSs identified in mock- and pathogen-infected plants within the same genetic background (termed *Pst*-induced dDHSs) revealed a substantially larger number of dDHSs in genes than in TEs (Fig. 1.3A). By contrast, DHS comparisons between WT and *morc1/2* plants subjected to the same treatment (termed *morc1/2*-enhanced dDHSs) identified much lower numbers of dDHSs overall, with slightly more present in TEs than in genes. Interestingly, when the percentage of dDHSs associated with TEs (termed TE-dDHSs) was calculated for each pairwise comparison, a significantly greater number of TE-dDHSs was observed than would be statistically expected (indicated as a broken line in Fig. 1.3B). Approximately 21% of the *Pst*-induced dDHSs were associated with TEs in both WT and *morc1/2* backgrounds (Fig. 1.3B). Notably, over 60% of the *morc1/2*-enhanced dDHSs were associated with TEs, irrespective of infection. These striking results suggest that MORC family members participate in modulating the physical accessibility of TE-associated sequences.

To gain insights into the putative biological functions of the gene-associated dDHSs, their gene annotations were analyzed using the TAIR database (Fig. 1.3C). Of the *Pst*-induced dDHSs, a significant number were associated with ‘response to stress’ or ‘response to biotic/abiotic stress genes’ (see black and dark gray bars in Fig. 1.3C). Due to the low number of the *morc1/2*-enhanced dDHSs (43 or fewer), little significance was found in their gene annotations (Fig. 1.3C).

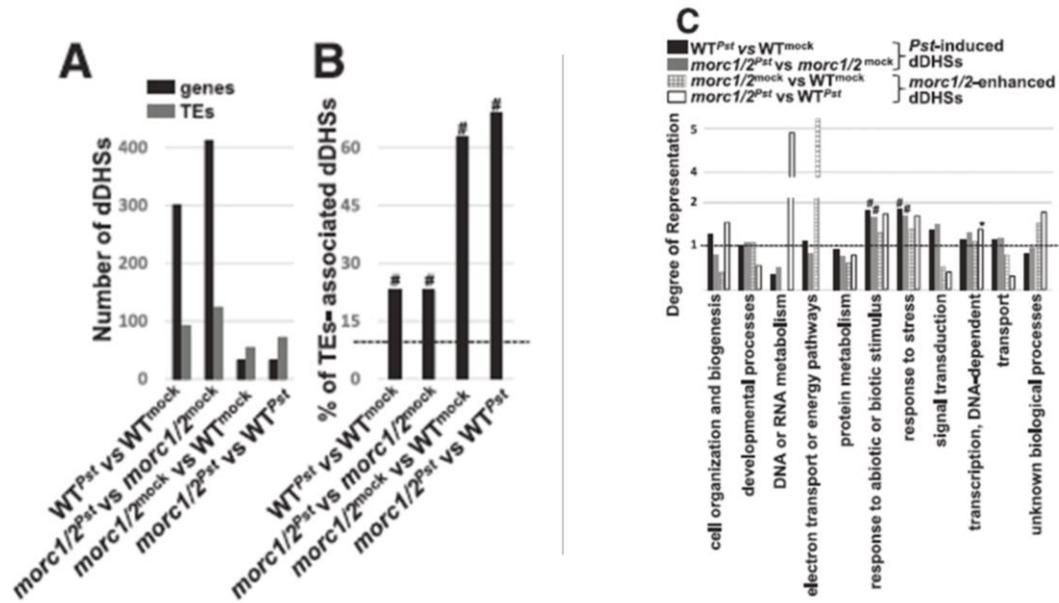


Fig. 1.3. Pathogen infection and mutations in MORC family members alter chromatin accessibility in TEs. **A**, The number of dDHSs present in the indicated pairwise comparisons were plotted. Genes and TEs are presented as black and grey bars, respectively. **B**, Percentage of TE-dDHSs. Red line indicates the percentage (11.9%) of all TE-DHSs based on the total 29,450 DHSs. **C**, Excluding TEs, the degree of representation in each biological process category was calculated for the dDHSs from each pairwise comparison based on annotation in the TAIR database. Biological process categories are presented and their representation in the Arabidopsis genome is calculated at 1, denoted by a broken black line. Statistical significance was determined using a one sample χ^2 test (B) and the Benjamini-Hochberg procedure (C): * $P < 0.05$; # $P < 0.01$.

Members of the MORC family modulated the DNase I accessibility of heterochromatic TEs, while *Pst* infection altered the accessibility of TEs distributed throughout the genome.

TEs are predominantly localized in heterochromatin; these heterochromatic TEs are subject to transcriptional gene silencing primarily via RdDM (Matzke & Moshier, 2014). By contrast, the silencing of euchromatic TEs, which also involves DNA methylation, appears to be mediated by a distinct mechanism (Zemach et al., 2013). For instance, *ddm1* mutant plants mainly lose repression of heterochromatic, but not euchromatic, TEs. To assess the location of the TE-dDHSs identified in the four comparisons, we plotted their genomic distribution against the relative density of genes and TEs (Fig. 1.4).

Interestingly, the *morc1/2*-enhanced TE-dDHSs were preferentially associated with heterochromatin (horizontal tracks 3 and 4 in Fig. 1.4), whereas the *Pst*-induced TE-dDHSs were more evenly distributed across the genome (tracks 5 and 6 in Fig. 1.4). There is little, if any, overlap between the TE- or gene-associated dDHSs induced by *Pst* inoculation and those induced by the loss of MORC family members.

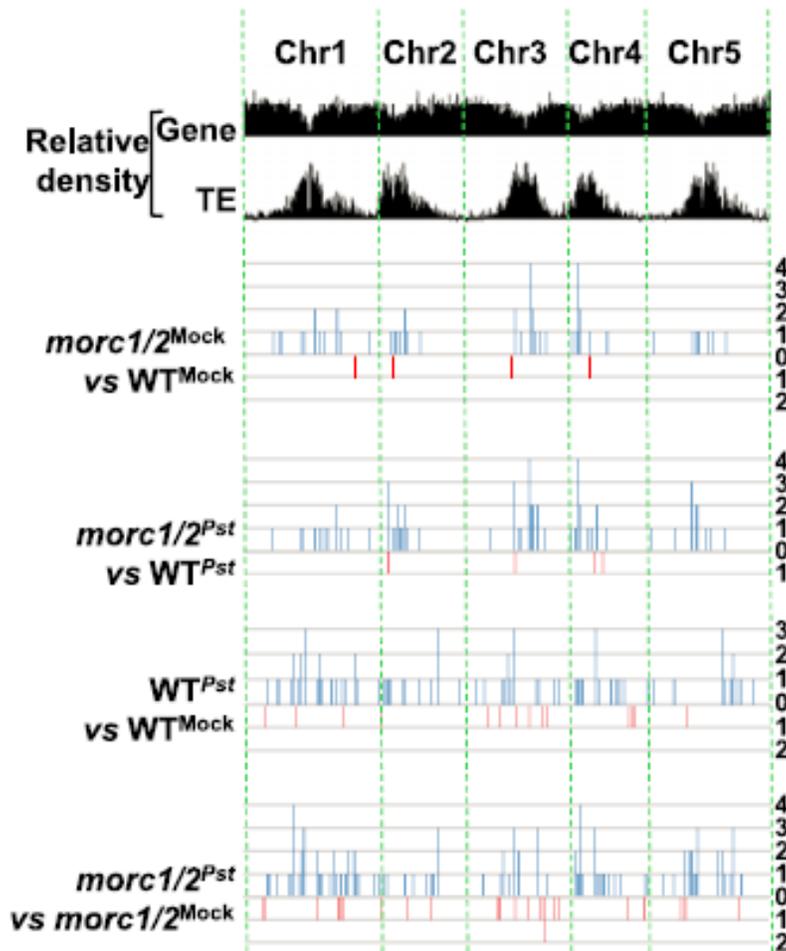


Fig. 1.4. Effects of MORC1 and infection on DNase I accessibility of heterochromatic TEs. The relative densities of genes and TEs are presented in the top two tracks and the genomic distribution of TE-associated dDHSs found in the four pairwise comparisons are shown in the lower four tracks. The y-axis indicates the number of dDHSs in a 100 kb window. For each comparison, the DHSs enhanced in the plants listed first are presented in blue, while the DHSs enhanced in the plant listed second are presented in red. Dotted green lines indicate the borders between chromosomes.

DNase I-qPCR analysis confirmed the chromatin accessibility of select TE-dDHSs identified by DNase-Seq

Nine dDHSs associated with TEs either in heterochromatin or in euchromatin were randomly selected and analyzed using DNase I-quantitative PCR (qPCR) to further assess the reproducibility of DNase-Seq dataset. The *Pst*-inducible TE-dDHS AP8539 was additionally chosen as it mapped to a site proximal to the well-known defense gene, *PR-1* (Fig. 1.5A). This TE-dDHS spans a large region that encompasses two dDHSs: one resides in the promoter of a TE approximately 2 kb upstream of the *PR-1* gene (designated as AP8539b), and the other is located within the *PR-1* promoter (designated as AP8539a). Thus, we monitored chromatin accessibility at both sites. Using DNA prepared from DNase I-treated nuclei, qPCR was performed for the selected TE-dDHSs. Note that while DNase-Seq identifies the ends of DNAs that are cut with DNase I, DNase I-qPCR amplifies DNAs that are not disrupted by DNase I. Consequently, less amplification via qPCR corresponds to increased genome accessibility. Comparisons between the DNase-Seq reads and DNase I-qPCR analyses for *morc1/2*-enhanced TE-dDHSs (Fig. 1.5B) and for *Pst*-induced TE-dDHSs (Fig. 1.5C) indicated that the results obtained from both techniques were consistent. The only exception was AP4254, which exhibited greater accessibility in *morc1/2* plants than in WT plants following DNase-Seq, but only a little difference between these plants following DNase I-qPCR.

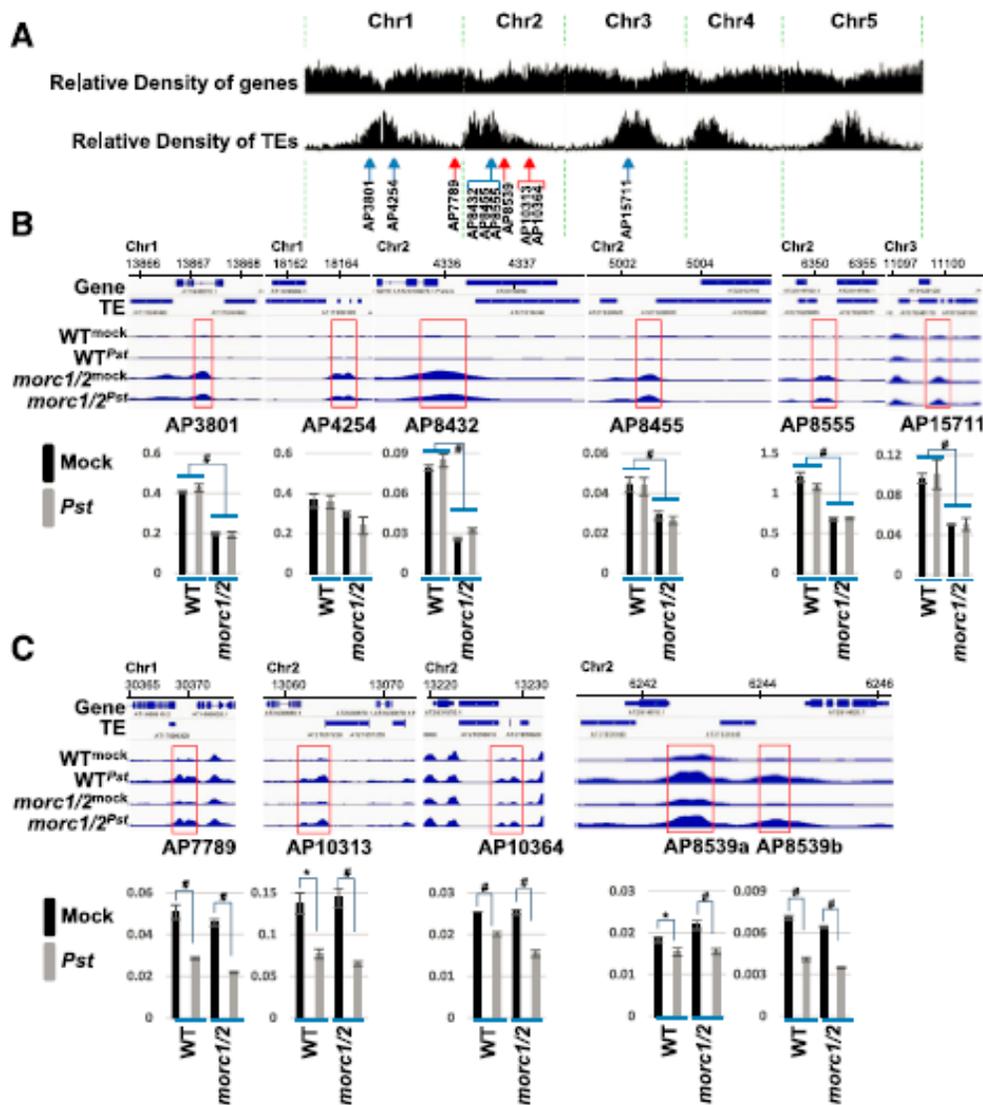


Fig. 1.5. Quantitative PCR analysis of selected dDHSs identified by DNase-Seq. **A**, Genomic locations of nine randomly selected TE-associated dDHSs, as well as TE-associated dDHSs near the defense gene *PR-1* (AP8539). dDHSs located in heterochromatic and euchromatic positions are indicated by blue and red arrows, respectively. A dotted green vertical line denotes the border between chromosomes. **B and C**, Schematic presentation of Arabidopsis genes and TEs with their corresponding genome coordinates in kb, as well as DNase-Seq read densities of TE-associated dDHSs (indicated as red boxes) in WT^{mock}, WT^{Pst}, *morc1/2*^{mock}, and *morc1/2*^{Pst}, and qPCR analysis. **B**, Selected *morc1/2*-induced dDHSs. **C**, Selected *Pst* infection-induced dDHSs. The relative amount of each dDHS was determined through qPCR analysis using *Tip41* as a reference gene (represented in the y-axis). Note that DNase I-qPCR amplifies DNA that is not disrupted by DNase I; thus, lower levels of amplification indicate increased genome accessibility. Two repeats were performed for each dDHS. Statistical significance was determined using a student t test: *P<0.05; #P<0.01.

Defense genes were overrepresented in the genomic regions proximal to *Pst*-induced TE-dDHSs and their induction by *Pst* was delayed and/or reduced in *morc1/2*

Decreased methylation of some TEs in response to SA was shown to increase expression of neighboring genes (Downen et al., 2012), suggesting that derepression of TEs might trigger the transcriptional activation of proximal genes. Therefore, annotated genes within 5 kb of TE-dDHSs were analyzed for their functional association (Fig. 1.6). Of the *Pst*-induced TE-dDHSs, a significant number were associated with genes involved in ‘response to abiotic or biotic stimulus’ or ‘response to stress’ (Fig. 1.6). A significant number of the *morc1/2*-enhanced TE-dDHSs were proximal to genes that are annotated as ‘unknown biological processes’.

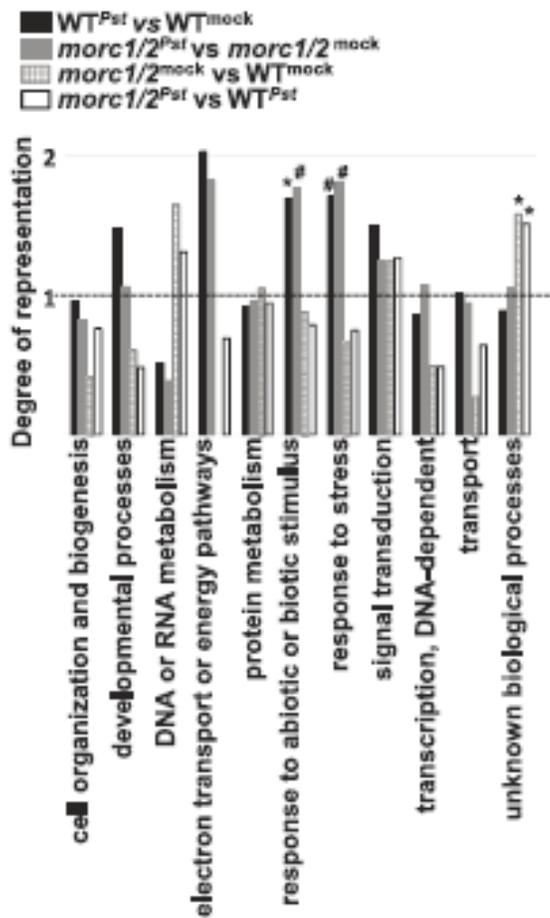


Fig. 1.6. The degree of representation of genes neighboring TE-dDHSs. Based on the TAIR annotation, biological process categories of the genes located within 5 kb of TE-dDHSs from each pairwise comparison were analyzed. The degree of representation of each category was plotted on the y-axis, as compared to their overall representation from all the genes in the Arabidopsis genome, which was set at 1 and denoted by a broken line. Statistical significance was determined using the Benjamini-Hochberg procedure: * $P < 0.05$; # $P < 0.01$.

Of these TE-neighboring genes, those annotated as ‘response to abiotic or biotic stimulus’ included several well-known defense genes, such as *PR-1* and *PR-2*. In addition, we selected 12 more defense genes that neighbor *Pst*-induced TE-dDHSs and assessed their transcript levels at several time points after mock or virulent *Pst* infection. All of the genes analyzed displayed strong induction in response to infection, although their induction kinetics varied (Fig. 1.7). Although our analyses suggested that MORC proteins primarily associate with heterochromatic TEs, the previous demonstration that a wide range of TEs

are upregulated in *morc1* and/or *morc6* (Moissiard et al., 2012), led us to test whether the defense genes neighboring *Pst*-induced TE-dDHSs exhibit altered expression in *morc1/2*. Notably, most of the 14 defense genes displayed various degrees of delayed and/or reduced transcriptional induction in *Pst*-inoculated *morc1/2* vs. WT plants, suggesting that MORC1/2 promotes their expression. Note that for four of the 14 genes (*At1g33960*, *At1g80820*, *At2g20010*, and *At1g44350*), the *Pst*-induced expression kinetics were not significantly different between *morc1/2* and WT plants.

PR-5 is another extensively used marker gene for defense signaling. In contrast to *PR-1* and *PR-2*, which have proximal *Pst*-induced TE-dDHSs, *PR-5* has a *Pst*-induced dDHS within its own promoter, and this location does not correlate with any annotated TEs. Nonetheless, *PR-5* showed compromised induction by *Pst* in *morc1/2* as compared with WT (Fig. 1.7B). This observation prompted us to further assess whether the dDHS present in the *PR-5* promoter is associated with a cryptic TE. Remarkably, four truncated putative TE sequences were observed at this dDHS (Appendix 1.6). Together, these observations raise an intriguing possibility that TE-associated genomic regions help regulate gene expression in response to biotic stresses.

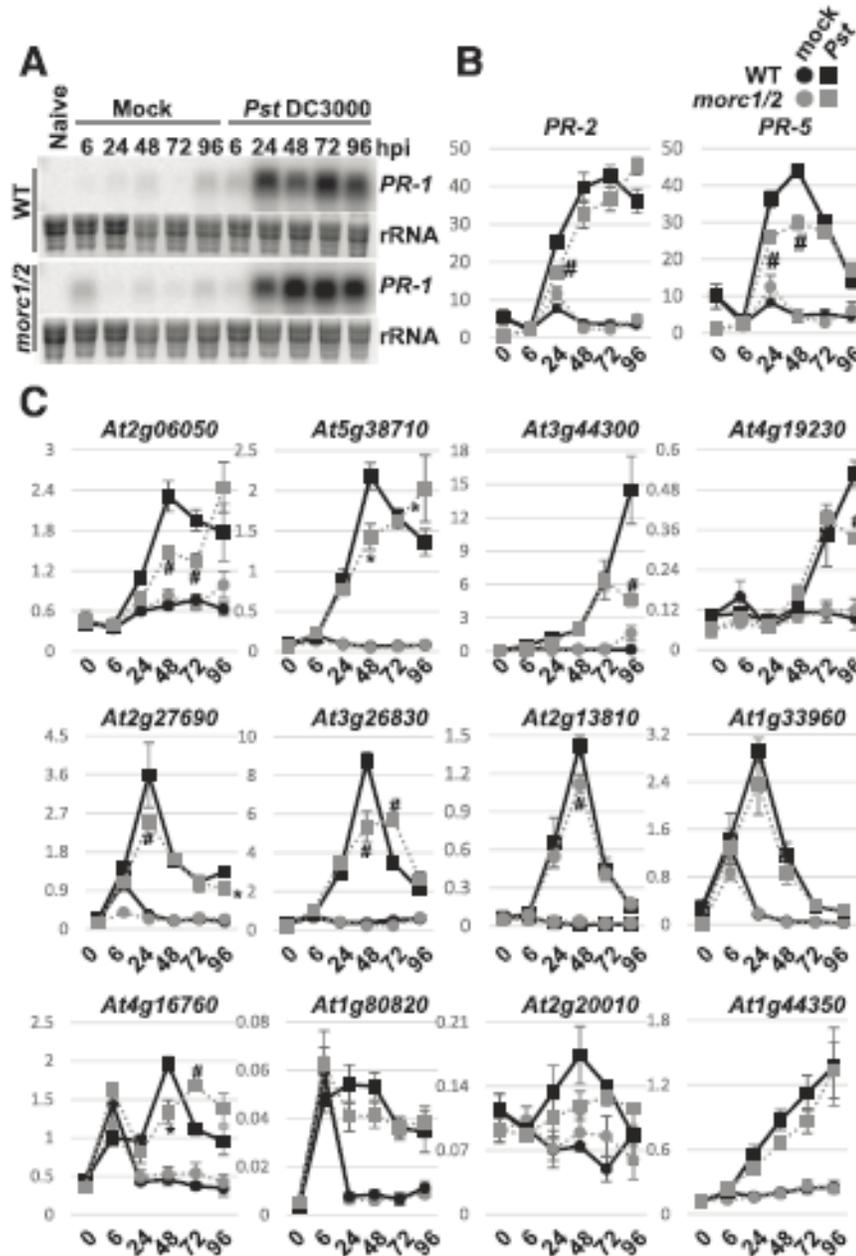


Fig. 1.7. MORC1/2 influences the kinetics and/or amplitude of defense genes induced by *Pst*. A, *PR-1*, B, *PR-2*, *PR-5* and C, selected defense genes that neighbor TE-associated dDHS were analyzed. Leaves were inoculated with buffer (mock) or 10^6 cfu/ml of *Pst*. Untreated (naïve) control leaves were harvested at 0 hours post inoculation (hpi). RNAs prepared from WT and *morc1/2* plants at the indicated time points after inoculation were resolved in a 1.5% agarose gel followed by northern blot analysis with a probe for *PR-1* (for the panel A) and were used for qRT-PCR analysis with primers specific for indicated genes (for the panels B and C). rRNA was used as a loading control for northern blot (A). The *TIP41-like* gene was used as a reference gene for qRT-PCR and the mean \pm SE ($n \geq 6$) including minimum of two biological replicates is presented (B and C). Statistical significance from WT was determined using t-test: * $P < 0.05$; # $P < 0.01$.

***Pst* infection changed the profile of MORC1-TE interaction**

Although we previously showed that MORC1 binds DNA with little sequence specificity *in vitro* (Kang et al., 2012), it is plausible that some of the *morc1/2*-enhanced TE-dDHSs are due to the loss of protection by bound MORC proteins. To address this possibility, chromatin immuno-precipitation followed by Illumina-based DNA sequencing (ChIP-Seq) was used to assess the interaction between Myc-tagged MORC1 and chromatin. These analyses were performed using an anti-Myc antibody and mock- or *Pst*-inoculated plants from a transgenic *Arabidopsis morc1/2* line carrying a Myc-tagged *MORC1* transgene expressed under its own promoter (*Myc-gMORC1*). This transgene was previously shown to complement *morc1/2* (Kang et al., 2012). To control for non-specific background binding, ChIP-Seq was performed in parallel on mock- and *Pst*-inoculated WT plants. Note that each sample was prepared in three independent biological triplicates.

Analysis of the ChIP-Seq peaks from mock- and *Pst*-inoculated *Myc-gMORC1* plants (the 3rd and 4th tracks, Fig. 1.8A) revealed that MORC1 binds sites distributed throughout the genome, although it shows a strong preference for heterochromatin. To find genomic regions exhibiting altered levels of MORC binding after pathogen infection, the intensity of ChIP-Seq peaks from mock- and *Pst*-inoculated plants was compared. Differential ChIP-Seq peaks (dChIP-peaks) exhibiting increased (*Pst*-induced, blue lines in the 5th track, Fig. 1.8A) or decreased (*Pst*-suppressed, red lines in the 5th track, Fig. 1.8A) intensity after *Pst* infection as compared to mock infection were not primarily associated with heterochromatic sites, but instead were dispersed throughout the genome (Fig. 1.8A). Given that *Pst*-induced and *morc1/2*-enhanced dDHSs are enriched in TEs (Fig. 1.3), we assessed whether the genomic regions exhibiting altered interaction with MORC1 after

pathogen infection correspond to TEs. Indeed, over 70% of the dChIP- peaks which showed either increased or decreased MORC1 binding were TE-associated (Appendix 1.7). Together, these results suggest that MORC1 is a chromatin-interacting protein that displays a significant preference for TEs and exhibits altered binding affinity after pathogen infection. Interestingly, the MORC1-associated peaks identified by ChIP-Seq were relatively broad, averaging around 8.5kb. This contrasts with the typically narrow peaks associated with transcription factors, and suggests that MORC1 exhibits little sequence preference but instead may recognize higher-order chromatin structure.

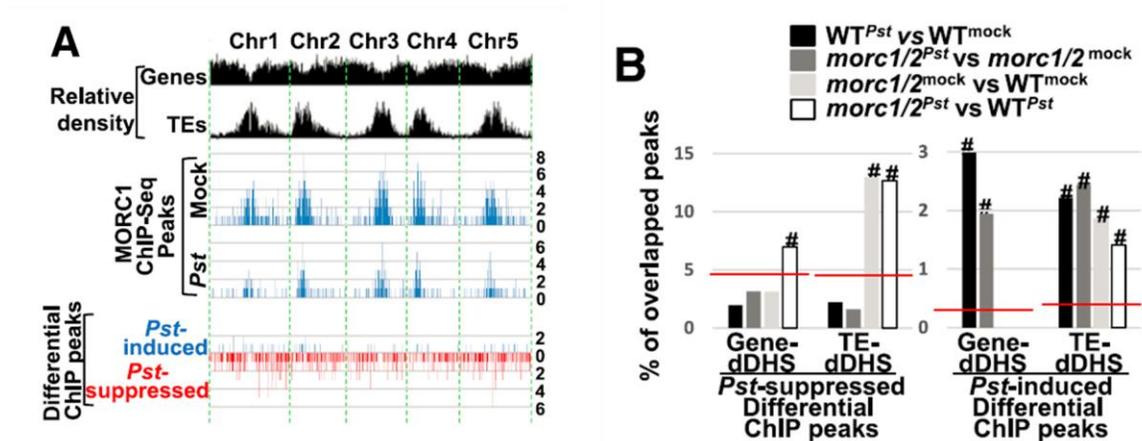


Fig. 1.8. MORC1 is physically associated with infection-induced TE-dDHSs.

Myc-gMORC1 plants and WT were mock-or *Pst*-infected (5×10^5 cfu/ml) for 1 day. ChIP was performed using an anti-Myc antibody and the recovered DNAs were sequenced using Illumina Hi-Seq 2500. SICER (Zang et al., 2009) was used to identify peaks that are associated with MORC1 in three independent biological replicates; WT was used as a background control. **A**, The relative densities of genes and TEs are presented in the top two tracks. The genome position of ChIP-Seq peaks ($p < 0.01$) in Mock- and *Pst*-infected *Myc-gMORC1* plants as compared to the WT background control are presented in the 3rd and 4th track. The y-axis indicates the number of ChIP peaks in a 100 kb window. ChIP-Seq peaks ($FDR < 0.001$) where levels are substantially different in mock- vs *Pst*-inoculated plants (termed Differential ChIP peaks) are presented in the 5th track. **B**, Percentage of dDHS peaks overlapping with ChIP-peaks were calculated to show association of MORC1 with dDHSs. The four pairwise comparisons presented in fig. 1.5 were analyzed separately. To better display TE-association with MORC1, gene- and TE-dDHSs were separately analyzed. A background level of overlap between ChIP and dDHS peaks was calculated using total gene- ($n=24,340$) and TE-DHSs ($n=3,520$). Statistically significant overrepresentation was determined between indicated groups and all DHSs using one sample χ^2 test: # $P < 0.01$.

Whether the MORC1-chromatin interaction sites overlap the *morc1/2*-enhanced and/or *Pst*-induced dDHSs was then assessed. All of the gene- and TE-associated dDHSs were checked for overlaps with the dChIP- peaks; however, dDHSs located in the intergenic regions were excluded. The background level of overlap (denoted as a red line in Fig. 1.8B) was determined by calculating the percentage of total gene- or TE-dDHSs in each group that overlap with *Pst*-suppressed or *Pst*-induced dChIP-peaks. Interestingly, dChIP- peaks that were reduced after *Pst* infection correlated most strongly with *morc1/2*-enhanced TE-dDHSs, although a lower, but statistically significant percentage overlapped with gene-associated dDHSs detected in *morc1/2* vs. WT plants after *Pst* infection. In comparison, the dChIP- peaks that were enhanced by *Pst* infection showed a specific overlap with *Pst*-induced, but not *morc1/2*-enhanced dDHSs located in genes. A low, but statistically significant overlap between *Pst*-induced dChIP- peaks and both *Pst*-induced and *morc1/2*-enhanced TE-dDHSs also was observed. Given the connection between MORC1 and TEs, we assessed whether cryptic TEs are present in the gene-dDHSs that overlap chromatin sites displaying increased MORC1 binding after *Pst* infection. Interestingly, cryptic TEs were identified in all of these gene-dDHSs, suggesting that MORC1 interacts with these previously unidentified TEs. In summary, our findings suggest that *Pst* infection leads to reduced MORC1/2 binding at dDHSs that are preferentially associated with heterochromatic TEs while in contrast infection enhances MORC1/2 binding at *Pst*-induced dDHSs located in a small population of euchromatic and heterochromatic TEs as well as in genes, likely via unannotated cryptic TEs.

Silencing of *Pst*-induced TE-dDHSs proximal to a MORC1 binding region compromised induction of adjacent defense genes

Although MORC1/2 physically associates with some *Pst*-induced TE-dDHSs (Fig. 1.8B), none of the TE-dDHSs neighboring the defense genes monitored in fig. 1.7 directly overlap *Pst*-induced dChIP-peaks. However, a majority of these TE-dDHSs are within 250 kb of a *Pst*-induced dChIP peak. Based on these results, we hypothesized that, following *Pst*-induced binding of MORC1 at adjacent sites, these TE-dDHSs are local enhancers that up-regulate the expression of their neighboring defense genes. Suppression of these enhancers by RNAi-mediated silencing would therefore be expected to interfere with proximal gene induction. To test this hypothesis, transgenic *Arabidopsis* lines expressing a hairpin construct (Wesley et al., 2001) targeting each of seven TE-dDHSs from fig. 1.7 were generated. Five of these TE-dDHSs were located within 250 kb of one or more infection-induced dChIP peak; the other two dDHSs, which were not proximal to a dChIP peak, were chosen as controls.

Silencing of the targeted regions was assessed via *McrBC*-qPCR, which involves qPCR of DNA digested with or without *McrBC*, an enzyme that specifically cuts methylated DNA. This quantitative analysis to test whether DNA methylation is induced by a hairpin RNA verified that all seven lines have significant DNA methylation in the intended regions (Fig. 1.9A). Silencing the putative TE-associated enhancers proximal to a *Pst*-induced dChIP-peak significantly reduced the induction of four of the five neighboring defense genes, including *PR-1*, *At2g06050*, *At4g19230* and *At2g27690*, but did not impact *At3g44300* (Fig. 1.9B). By contrast, silencing the TE-dDHSs lacking an adjacent *Pst*-induced dChIP peak had little impact on the induction of defense genes *At2g19230* or *At4g16760* (Fig.

1.9B). These results support the possibility that these MORC1-associated TEs are local enhancers that induce adjacent defense gene expression.

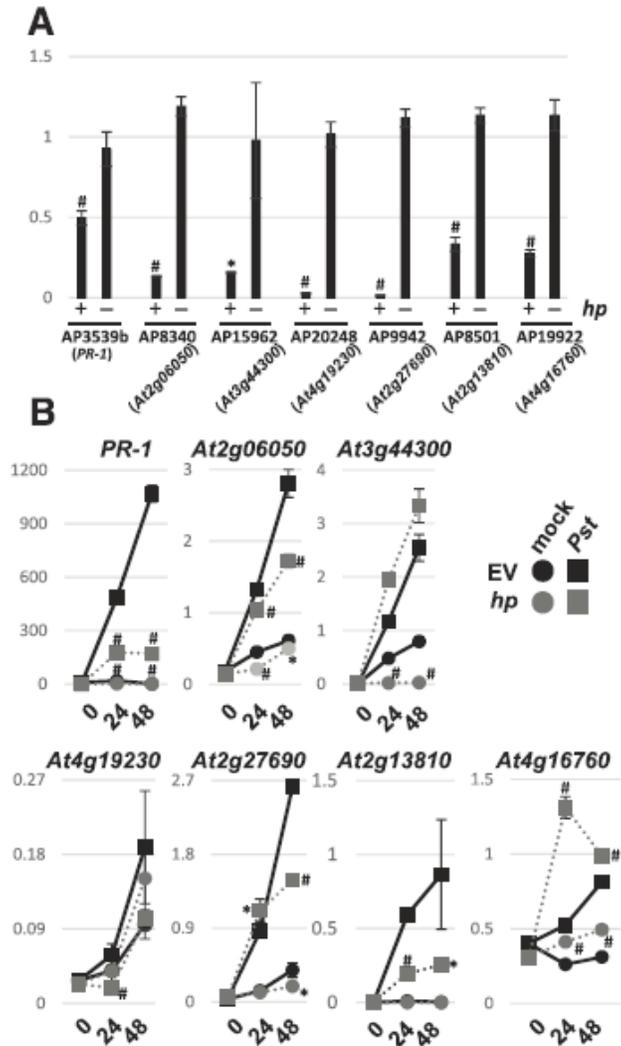


Fig. 1.9. Silencing of *Pst*-induced TE-dDHSS proximal to a MORC1 binding site compromised induction of adjacent defense genes. **A**, RNAi-mediated silencing of selected TE-dDHSSs in transgenic Arabidopsis plants carrying a hairpin construct (*hp*) was verified via *McrBC*-qPCR; WT plants was used as a control. The relative amount of *McrBC*-digested DNA in comparison to undigested DNA was analyzed in three replicates and presented in the y-axis. Note that higher levels of amplification indicate decreased DNA methylation. Several independent lines were generated for each hairpin construct. **B**, *Pst*-induced expression of defense genes proximal to silenced TE-associated dDHSSs was analyzed via qRT-PCR. Each panel corresponds to a representative transgenic line (*hp*) in which the TE-dDHSS adjacent to the indicated defense gene has been silenced by a hairpin construct; an empty vector (EV) transformed line was used as a control. Leaves were inoculated with buffer (mock) or 10^6 cfu/ml of *Pst*. Untreated (naïve) control leaves were harvested at 0 hpi. RNAs prepared from WT and *morc1/2* plants at the indicated time points after inoculation were used for qRT-PCR analysis with primers specific for indicated genes. The *TIP41-like* gene was used as a reference gene for qRT-PCR and the mean \pm SE ($n \geq 3$). A minimum of one more independent line was tested and found to have an expression pattern comparable to those presented here. Statistical significance from WT was determined using t-test: * $P < 0.05$; # $P < 0.01$.

Discussion

To address the relationship between the roles MORC proteins play in gene silencing and plant immunity, we mapped the genomic location of DHSs in mock- and pathogen-inoculated WT and MORC-deficient plants. In a previous study, ~45% of the DHSs in the Arabidopsis genome were mapped to putative promoter regions (Zhang et al., 2012). Underscoring the correlation between DHSs and transcriptionally active chromatin, these sites were depleted of nucleosomes and were tightly associated with RNAPII binding sites. Furthermore, the prevalence of DHSs was dramatically reduced in transcriptionally silent peri-centromeric regions, which contain highly methylated DNA. Similar to these findings, the majority of DHSs we identified were located within 1 kb upstream of the TSS of protein-coding genes. Analysis of the TE-DHSs also revealed that the majority also were located within 1 kb of the TSS. Neither *Pst* infection nor loss of MORC family members substantially altered the genomic distribution of DHSs. Given that DNase I sensitivity is influenced by DNA methylation levels (Zhang et al., 2012) and that genome-wide DNA methylation levels are not appreciably altered in *morc1* or *morc6* mutants as compared with WT plants (Moissiard et al., 2012) or in *Pst*- vs. mock-inoculated Arabidopsis (Downen et al., 2012), our DHS results are consistent with these methylation studies.

Although global DHS distribution was not affected by *Pst* inoculation or the loss of MORC family members, pairwise comparisons of the DHSs detected in mock- or *Pst*-inoculated WT or *morc1/2* plants revealed notable differences. The number of *morc1/2*-enhanced dDHSs was much lower than that induced by *Pst* infection (Fig. 1.3). Of these *morc1/2*-enhanced dDHSs, fewer than 40% were located in genes. Instead, *morc1/2*-enhanced dDHSs were highly enriched in TEs, particularly those located in heterochromatic regions

(Fig. 1.4). Analysis of the genomic regions identified by ChIP-Seq confirmed that MORC1 preferentially binds heterochromatic and TE-associated regions (Fig. 1.8A; Appendix 1.7). Furthermore, since *morc1/2*-enhanced TE-dDHSs overlap the genomic regions corresponding to *Pst*-suppressed dChIP- peaks at a rate substantially greater than that of the expected background, *Pst* infection appears to lead to lowered MORC1/2 binding at heterochromatic TEs (Fig. 1.8B). It is interesting to note that a majority of the TEs transcriptionally activated in *morc1*, *morc6* and *morc1/6* (Moissiard et al., 2012) are physically located close (less than 160kb) to *morc1/2*-enhanced TE-dDHSs. Together, these results are consistent with the reported association of MORC1 and MORC6 with heterochromatin (Moissiard et al., 2012) and the derepression of multiple families of TEs and endogenous genes preferentially associated with heterochromatin in *morc1*, *morc2* and/or *morc6* mutants (Brabbs et al., 2013, Moissiard et al., 2012, Moissiard et al., 2014).

Given the considerable interaction between MORC1 and heterochromatin (Fig. 1.8A; Appendix 1.7), it seems surprising that dramatic differences in chromatin accessibility, as determined by the number of dDHSs, were not observed in comparisons between *morc1/2* and WT plants receiving the same mock or *Pst* treatment (Fig. 1.2 and Fig. 1.3). One possible explanation for the discrepancy in these findings is that the development of DHSs in the *morc1/2* background may be partially suppressed by one or more functionally redundant members of the MORC family. Indeed, MORC3, the next closest homolog of MORC1 after MORC2, appears to be functionally redundant with MORC1 to some degree, since it restored TCV coat protein-induced cell death in *morc1-1* plants expressing an inducible coat protein transgene (Kang et al., 2008). Unfortunately, we cannot test whether MORC3 also suppresses DHS development in the *morc1/2* background, because the

morc3-1 knock-out mutation (SALK_000009) is lethal in the homozygous state (Kang et al., 2010).

An alternative possibility is that MORC1/2 does not directly bind DNA, but instead influences chromatin accessibility by directly or indirectly interacting with other proteins. For example, MORC family members may influence DNase I hypersensitivity in heterochromatic regions by interacting with proteins involved in RdDM. Indeed, the combined observations that i) derepression of reporter genes in *morc6* mutants correlates with a decrease in their DNA and histone methylation levels (Brabbs et al., 2013, Lorkovic et al., 2012), and ii) MORC1, MORC2, and/or MORC6 interact with several proteins involved in the RdDM pathway, including DMS3 (Defective in Meristem Silencing 3; Lorkovic et al., 2012) and the SET domain-containing proteins SUVH9 and SUVH2 (Liu et al., 2014), suggest that MORC family members influence RdDM, which in turn could affect DNase I sensitivity. The relationship between MORC proteins and RdDM, however, is currently unclear. Notably, older *morc6* plants develop stochastic, cell-autonomous silencing of a GFP reporter gene that is consistently expressed in younger plants (Brabbs et al., 2013). Since this silencing of GFP expression was associated with the reappearance of DNA methylation in the reporter gene sequence, it was proposed that MORC6 promotes, but is not obligately required for RdDM (Brabbs et al., 2013). Consistent with this proposition, analyses of *morc1* and *morc6* mutants failed to detect a correlation between changes in genome-wide methylation levels and the activation of a silenced reporter gene (Moissiard et al., 2014, Moissiard et al., 2012). Instead, since peri-centromeric heterochromatin was decondensed in these mutants, it was proposed that MORC1 and MORC6 enforce the higher order compaction of methylated, silenced chromatin; such a

function would presumably also modulate the DNase I sensitivity of these sequences. The very broad MORC1-associated peaks identified by our ChIP-Seq analysis also support this possibility.

In comparison to the *morc1/2*-enhanced dDHSs, those induced by *Pst* infection were primarily located in protein-encoding genes, particularly those associated with biotic/abiotic stresses. They also were enriched in TEs; these TE-dDHSs were spread throughout the genome and shared little overlap with those enhanced by *morc1/2* (Fig. 1.4). Strikingly, the 5' flanking regions of a wide range of defense genes, including *PR-1*, *PR-2* and *PR-5*, contained *Pst*-induced TE-dDHSs. Since the expression kinetics of these *PR* genes, as well as 12 additional, randomly selected defense genes, was generally modestly delayed and/or weaker in *Pst*-inoculated *morc1/2* as compared to WT plants (Fig. 1.7), MORC1/2 appears to promote their expression. Consistent with this possibility, MORC1 association with *Pst*-induced TE-dDHSs was enhanced after *Pst* infection. Note that all of the *Pst*-induced gene dDHSs were found to be associated with unannotated TEs, essentially making them TE-dDHSs (Fig. 1.8B). It is interesting to note that infection-induced transcription of *PR-1* and other defense genes was generally suppressed by RNAi-silencing of the neighboring (although in some cases distant) TE-dDHS, but only if the TE-dDHS was adjacent to a *Pst*-induced dChIP-peak. Given that enhancers can be located as far as 1 Mb from the genes they regulate, our silencing results suggest that i) at least some *Pst*-induced TE-dDHSs serve as enhancers of neighboring defense genes and ii) these putative enhancers are activated by *Pst*-induced MORC1/2 binding to sites that, while nearby, do not necessarily overlap with the dDHS. This latter conclusion may explain why the overlap frequency between dDHSs and ChIP peaks is relatively low (Fig. 1.8B).

MORC1 in plant immunity and Epigenetics

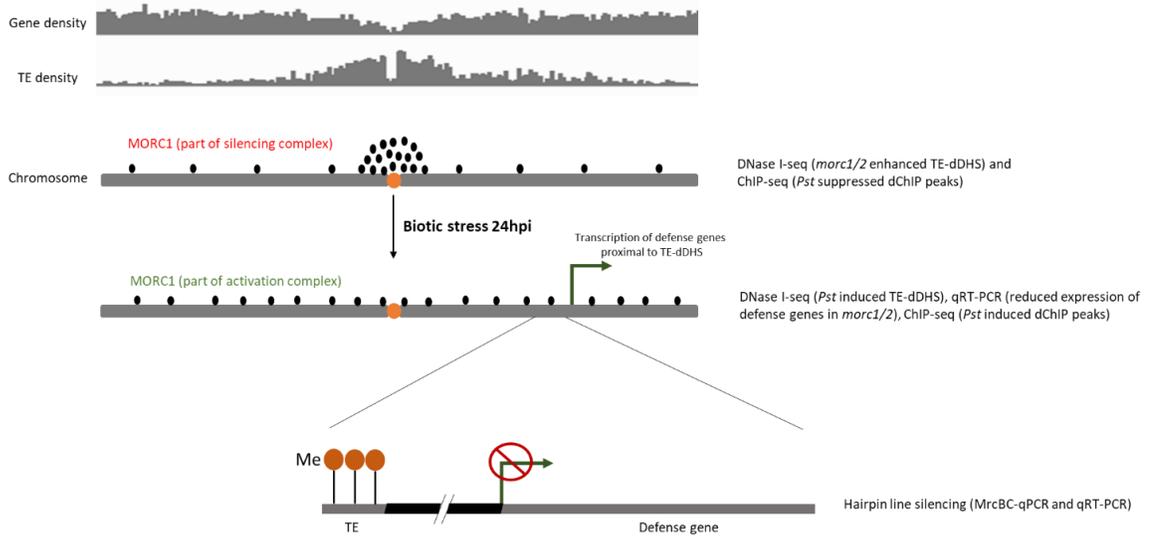


Fig. 1.10. Working model to explain the role of MORC1 in plant immunity and epigenetics. Under normal conditions, MORC1 is localized primarily in the heterochromatin (suggested by *morc1/2* enhanced TE-dDHSs and ChIP-seq peaks in the heterochromatin). Immediately after pathogen infection the expression MORC1 is downregulated along with the other components of the RdDM (discussed in Chapter 2 and 3). 24 hours post infection its expression resumed to the normal level. However, MORC1 localization is changed from primarily in heterochromatin before to primarily in euchromatin after (suggested by significantly high overlap between *Pst* induced dChIP peaks and *morc1/2* enhanced TE-dDHS). Following the biotic stress, MORC1 perhaps becomes a component of the gene activating complex indicated by reduced expression of the *Pst* induced TE-dDHS neighboring defense genes in *morc1/2* (it is tempting to speculate that some other components of RdDM might do the same suggesting molecular machineries in the cell share the multitasking proteins such as chromatin remodelers to perform repression and activation). The transcriptional gene silencing of *Pst* induced TE-dDHS proximal to the defense genes using the hairpin line suppressed the expression of neighboring defense genes shown by qRT-PCR on the hairpin line suggesting that MORC1/2 indeed regulates the expression of defense genes by changing chromatin accessibility of the TEs.

Analyses of the genomic regions exhibiting differentially increased or decreased MORC1 binding following *Pst* infection indicated that they were strongly associated with TEs dispersed over the entire genome (Fig. 1.8A; Appendix 1.7). This finding, combined with the discovery that cryptic TEs are present in most of the gene-associated dDHSs exhibiting increased MORC1 binding after *Pst* infection, suggests that MORC1 may regulate gene expression by binding the super structure associated with TEs or TE-like sequences. A long list of studies also have suggested a link between TEs and gene regulation. In humans, TEs

and repeat DNAs (Feschotte, 2008) are associated with cell-specific transcription (Thurman et al., 2012). Furthermore, a substantial portion of regulatory elements in the promoters of human and plants genes are derived from TEs or pseudo/former TEs (Jordan et al., 2003, Martienssen et al., 1990, Girard & Freeling, 1999). In Arabidopsis, TE density is enriched in the ~2kb upstream of the TSS and near the 3' end of genes, whose induction by SA or *Pst* treatment is associated with the appearance of differentially methylated regions (Downen et al., 2012). Analysis of these *Pst* or SA-inducible genes containing differentially methylated regions revealed that many exhibit functions associated with plant immunity. A recent study also indicated that DNA demethylation of TEs is important for the activation of some defense genes (Le et al., 2014). Our finding that *Pst*-induced TE-dDHSs are associated with a wide range of defense genes further argues that TEs are important regulatory elements for controlling transcription, and that their modulation by RdDM components, including MORC1, plays an important role in the activation of defense responses after pathogen attack.

In summary, we demonstrate that *Pst* infection primarily suppresses MORC1 binding at dDHSs associated with heterochromatic TEs, but enhances its binding at infection-induced dDHSs in genes and TEs. These results, combined with MORC1's previously demonstrated involvement in heterochromatin condensation and gene silencing (Brabbs et al., 2013, Lorkovic et al., 2012, Moissiard et al., 2012), and our finding that defense gene expression is attenuated in the *morc1/2* mutant, suggest that MORC1 plays important roles in both gene silencing and gene induction. We propose that MORC1 mediates these divergent effects via its interaction with different chromatin-binding proteins. In this scenario, the *Pst*-induced loss of MORC1 at heterochromatic TEs would disrupt a complex

involved in gene silencing, thereby leading to activation of TE expression after pathogen infection. By contrast, the *Pst*-induced addition of MORC1 to a protein complex present at TEs would temporarily relieve silencing, thereby promoting robust expression of proximal genes. In support of this model, growing evidence suggests that epigenetic regulation is often context specific (Sarris et al., 2014). For instance, enhancer of Zeste 2 homolog (EZH2), which is a core enzymatic subunit of the polycomb repressive complex 2 (PRC2) in human cells, exerts opposing effects on gene expression. As a part of PRC2, EZH2 is involved in silencing a wide range of genes; however, when EZH2 is post-translationally modified, it functions as a co-activator for androgen receptor-target genes (Xu et al., 2012). Further characterization of MORC1's interacting proteins, including epigenetic factors such as DMS3 (Lorkovic et al., 2012), the SUVH9, and SUVH2 (Liu et al., 2014), will likely provide important insights into the mechanisms through which MORC1 impacts gene silencing, defense gene induction, and plant immunity.

CHAPTER II

ROLE OF RNA DIRECTED DNA METHYLATION MEDIATED TRANSCRIPTIONAL GENE SILENCING IN PLANT IMMUNITY

Abstract

Gene expression is mainly regulated at the transcriptional and post transcriptional level. RdDM (RNA directed DNA Methylation) is a plant specific transcriptional gene silencing (TGS) pathway which leads to the methylation of DNA *via* sRNAs (small RNAs). The DNA methylation is recognized by histone modifying enzymes which add a repressive mark on chromatin. TGS mediated by RdDM (RdDM-TGS) mostly targets and suppresses TEs and repeats in facultative heterochromatin. Given the role of RdDM in suppression of TEs, and the observation that biotic- stress-triggered alteration in chromatin accessibility was highly associated with TEs, we hypothesized that the de-repression of TEs and its re-repression is important in the induction dynamics of immune responses in plants. Indeed, TEs, mostly known as transcriptionally inactive, were induced by infection with *Pst*. Furthermore, in several RdDM mutants combined with *morc1/2*, I found that the bacterial infection significantly induced a *Copia* TE, suggesting that RdDM and stress are closely associated. These findings prompted a systematic assessment of RdDM-TGS components. I chose *DCL* (dicer-like) genes for my further study as their functions are very well characterized in sRNA biogenesis, important regulators for TEs and chromatin remodeling. Among the *dcl* mutants, *dcl1* displayed the most compromised resistance and induction of defense genes against avirulent *Pst*, suggesting that some sRNAs may be necessary for the rapid defense responses. In contrast, *dcl2* and *dcl3* showed marginally enhanced resistance and elevated expression of defense genes to the virulent pathogen. In particular, *dcl2* and

dcl3 showed elevated expression of defense genes without pathogen challenges, suggesting that DCL2/3-generated sRNAs are important in suppressing defense genes.

Introduction

DNA methylation can be achieved in three different sequence contexts. Methylation of CG and CHG (H is any nucleotide except G) is catalyzed by the enzyme DNA Methyltransferase1 (MET1) and Chromomethylase3 (CMT3), respectively. CHH methylation on the other hand is mediated by Chromomethylase2 (CMT2) and Domains rearranged methyltransferase2 (DRM2) (Law & Jacobsen, 2010). MET1 and CMT3 are required for the maintenance of DNA methylation when the DNA is replicated for cell division, whereas DRM2 is required for *de novo* DNA methylation. *De novo* DNA methylation is activated when a transposable element (TE) is freshly transposed to a new location in the genome (Bender, 2004). Consequently, the genomic area carrying TEs and repeats is highly methylated. Because TEs can be mobile, DNA methylation which is believed to suppress the mobility is indispensable for the genome stability (Jones, 2000). Biotic stress suppresses DNA methylation (hypomethylation) in the genome including the regulatory sequences of the defense related genes and TEs (Deleris et al., 2016) (Downen et al., 2012). This hypomethylation in response to stress has been shown to be actively achieved by a DNA demethylase enzyme, Repressor of Silencing (ROS1) (Yu et al., 2013). This epigenetic change was proposed to be a part of immune responses from host plants, which is perhaps an important player in the transcriptional activation of resistance genes including Resistance Methylated Gene 1 (RMG1) (Yu et al., 2013).

The DNA wraps around an octamer of histone proteins, two copies of each H2A, H2B, H3 and H4. This histone and DNA complex is a structural unit of a eukaryotic chromosome, a nucleosome. A number of chemical modifications like methylation, acetylation, phosphorylation and ubiquitination on the tails of H3 and H4 are known to have an

significant effect on the chromatin structure, DNA methylation and ultimately gene transcription (Karlic et al., 2010). While histone methylation can either increase or decrease transcription, depending on which subunit is modified, the acetylation, phosphorylation and ubiquitination are generally known to increase transcription (Sridhar et al., 2007, Zhang et al., 2007). H3K27 and H3K9 dimethylation have been reported to repress RNA transcription while H3K9 trimethylation activates it (Zhang et al., 2007). Histone modifications are directly implicated in immunity, including priming of the defense responses (Conrath, 2011). Loss of histone modifying enzymes, including histone deacetylase (Kim et al., 2008a, Wang et al., 2010b), histone methyltransferases (Berr et al., 2011a, Dangl et al., 2010, Alvarez-Venegas, 2007, De-La-PeÑA et al., 2012) and histone ubiquitin ligase (Dhawan et al., 2009b), alters resistance to *Pst* as well as two fungal pathogens, and changes the expression of defense genes. The crucial role of histone modification in defense signaling is further underscored by the observation that lysine (K) residues in H3, such as H3K4, H3K9, and H3K14, are differentially decorated with methylation and/or acetylation in key defense genes, including *PR1* in response to salicylic acid (Alvarez-Venegas, 2007).

The chromatin architecture can be changed by varying the distribution density of nucleosomes. This change is primarily performed to allow the access of transcriptional machinery to transcriptionally active sites (Pique-Regi et al., 2011). Thus, densely packed nucleosomes are the characteristics of transcriptionally silent regions. Using this characteristics, chromatin accessibility analyses have discovered transcriptionally active regions and their regulatory sites (Thurman et al., 2012, Bordiya et al., 2016). Chromatin remodeling complexes such as SWI/SNF bind to the promoter and terminators to regulate

the transcription of mRNAs as well as non-coding RNAs (Archacki et al., 2017). Apart from affecting the distribution of nucleosomes, chromatin remodeling involves replacing the canonical histones with their variants. For example, replacement of H2A with the H2A.Z variant occurs under biotic and abiotic stress, which is proposed to have an important regulatory role in the stress responses in Arabidopsis (March-Diaz et al., 2008, Sura et al., 2017).

Protein coding sequences constitute a very minor portion of eukaryotic genomes. For instance, the coding sequences make up mere 2% of the human genome (Elgar & Vavouri, 2008). Furthermore, most transcriptionally active RNAs never get converted into proteins, but instead function as non-coding RNAs. Surprisingly, growing evidence indicates that non-coding RNAs play crucial roles in a wide range of cellular processes, including chromosomal silencing, growth and developmental, transcriptional regulation, and stress responses (Liu et al., 2012). An elegant study by Tsai et al. showed that, early in embryonic development, histones have two kinds of modifications, an activating mark such as H3K4 methylation and a suppressing mark such as H3K27 methylation (Miao-Chih Tsai, 2010). This study reported that the HOTAIR lncRNA (long non-coding RNA) acts as a scaffold and interacts with two different enzymes; the 5' and 3' end of the HOTAIR lncRNA respectively interacts with PRC2 (polycomb repressive complex 2) and KDM1 (H3K4 Lysine demethylase). Combined action of these enzymes, one removing H3K4 methylation and another adding H3K27 methylation mark, triggers strong silencing, which generally leads to a tightly packed chromatin. In plants, transcription of non-coding RNAs is executed by RNA polymerase II as well as plant-specific RNA polymerase IV and V, critical components for gene silencing in plants (Haag & Pikaard, 2011). The RdDM

pathway uses RNAs transcribed by Pol IV and Pol V and converts them into small RNAs (sRNAs) using Dicer like proteins, which range 21-24 nt in size. The small RNAs use the sequence complementarity to either transcriptionally silence (Transcriptional Gene Silencing; TGS) or post transcriptionally silence target genes (Post Transcriptional Gene Silencing; PTGS) (Matzke & Mosher, 2014). This silencing process is initially reported as an anti-viral response to target invading viral DNAs/RNAs (Goldbach et al., 2003). A number of microRNAs (miRNAs) have also been shown to contribute to the plant immunity by regulating the expression of innate immune receptors and also by targeting the expression of negative immune regulator (Li et al., 2012a, Katiyar-Agarwal & Jin, 2010).

Significant improvement of DNA sequencing technology and its analysis tools have revolutionized genome sciences. Through the improved sequencing, a big surprise is that eukaryote genomes are full of sequence repeats, TEs, and non-coding RNAs (Initiative, 2000). Arabidopsis genome is not an exception (Initiative, 2000). TEs are kept transcriptionally silent by almost all the components discussed above (DNA methylation, repressive histone modifications, chromatin remodeling, and sRNAs). An elegant study have recently characterized the mobilome of Arabidopsis and shown a considerable number of active TEs that are capable of transposition (Quadrana et al., 2016). Accumulating evidence begins raising a possibility that active TEs may play an important role in transcriptional regulatory mechanisms and even evolution of transcriptional network in eukaryotes (Fedoroff, 2012, Ito et al., 2011b, Grandbastien, 1998, Barsh et al., 2010). I reported along with others that a number of defense related genes have TEs in the

promoter region and may regulate the gene expression under biotic stress (Bordiya et al., 2016, Downen et al., 2012), supporting the role of TEs in transcriptional regulation.

Since the first discovery of the RNAi (RNA interference) on anthocyanin accumulation in *Petunia* (Napoli et al., 1990), a number of studies have found that sRNAs not only post transcriptionally silence genes but also transcriptionally silence genes via DNA methylation (Morris et al., 2004). RdDM is a plant specific epigenetic pathway which leads to the methylation of DNA via sRNAs. In general, the RdDM process is known to involve the following steps. RNA pol IV recognizes and transcribes chromatin regions decorated with H3K9 methylation. This recognition is mediated by RNA Pol IV interacting protein SHH1 (SAWADEE HOMEODOMAIN HOMOLOGUE 1) (Matzke & Mosher, 2014, Law et al., 2013). RNA Pol IV transcribed RNAs are then converted into dsRNAs by RDR2 (RNA dependent RNA polymerase) and cut into small 24 nt sRNAs by DCL3 (Dicer-like 3). These siRNAs are loaded onto AGO4 and imported into the nucleus in which complementary sequences are methylated using Pol V-transcribed RNAs complexed with DRM2 (Domain rearranged methyltransferase2) (Zhang & Zhu, 2011). RdDM mainly targets TEs and repeat sequences to maintain the facultative heterochromatin (Matzke et al., 2015). The stable and reliable silencing of the heterochromatin, however, is not a function of RdDM. Rather, it is mediated by the chromatin remodeler DDM1 (Decrease in DNA Methylation1) and MET1 (Methyltransferase 1) (Ito & Kakutani, 2014), which explains why RdDM-associated mutants usually do not show a severe developmental phenotype (Matzke et al., 2015).

Components in the biogenesis of sRNAs and RISC (RNA Induced Silencing Complex) are well characterized for their role in PTGS and TGS (Bologna & Voinnet, 2014). These

epigenetic components directly or indirectly affect antiviral and antibacterial responses in plants and animals (Pelaez & Sanchez, 2013). For example, RdDM components such as Ago4 and RNA Pol IV and V have been shown to be directly involved in plant immune responses. Lopez et al. showed that, using ChIP-PCR analysis, that active chromatin marks such as H3K4me3 and H3K9ac are enriched at *PR1* locus in *nrpd* and *nrpe* mutants, which make these mutants resistant to *Pst* compared to WT (Lopez et al., 2011). In contrast, *ago4* was reported to display enhanced susceptibility to *Pst* (Agorio & Vera, 2007). Zemach et al. reported that RdDM pathway also controls the methylation of euchromatic short TEs (Zemach et al., 2013). Interestingly, my study presented in Chapter 1 also suggests that TEs which are found to be associated with infection-induced dDHS are present throughout the chromosome (Bordiya et al., 2016). Since many defense genes have TEs in their proximal regions (Feschotte, 2008), it is possible therefore that these TEs near defense genes are epigenetically controlled by RdDM.

A number of studies have shown that CMT3, a methyltransferase in the RdDM, recognizes the H3K9me1 through its chromodomain and methylates the CHG sites. This CHG methylation is recognized by the Kryptonite H3K9 methyltransferase (Kyp; also known as SUVH4) which methylates H3K9 (Han et al., 2015). The CMT3 and Kyp protein therefore form a positive reinforcing loop of repressive epigenetic mark. It is also known that SNF2 chromatin remodeler protein FRG1 and -2 are required for the RdDM (Groth et al., 2014). Given that DNA methylation is often correlated with condensed chromatin (John S. Choy, 2010) and hypomethylation is triggered during stress (Yu et al., 2013), we hypothesized that de-repression of RdDM-TGS and its re-repression is important in plant immune responses. To test this hypothesis, we investigated the role of Dicer like (DCL) proteins,

one of the major components of the RdDM, in plant immune responses, with emphasis on the induction dynamics of defense genes. Although, out of four DCLs in the Arabidopsis genome, only DCL3 is shown to be involved in the canonical RdDM, various non-canonical RdDM pathways exist in Arabidopsis which may use DCLs other than DCL3 (Fig. 2.1) (Panda et al., 2016, Matzke et al., 2015). Thus, I decided to include all the *DCL* genes in my following studies to better understand their role in immune responses in plants.

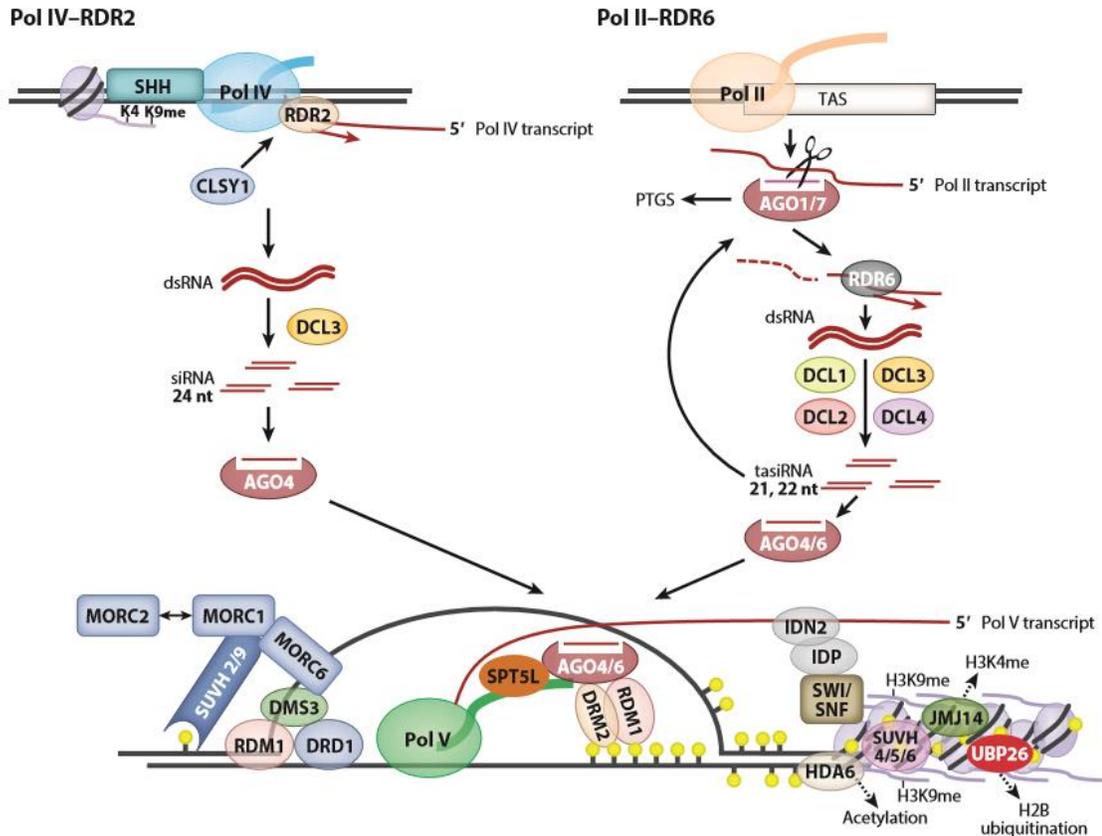


Fig. 2.1. Canonical and non-canonical RdDM pathway of transcriptional gene silencing (TGS). Image adapted from Matzke et al. (Matzke et al., 2015). Top left side of the figure depicts the canonical RdDM. RNA pol IV transcribes the genomic region having methylated H3K9 and unmethylated H3K4 which can be recognized by RNA Pol IV interacting protein SHH1 (SAWADEE HOMEODOMAIN HOMOLOGUE 1). This transcript is converted into dsRNA by RDR2 (RNA dependent RNA polymerase) which is then cut into small 24nt siRNA by DCL3 (Dicer-like 3). In the non-canonical RdDM shown on the right side, RDR6 is used to convert the ssRNA into dsRNA and various Dicer like protein are involved in dicing the dsRNA into siRNAs. These siRNAs are then loaded onto AGO4 and imported to the nucleus where complementary sequences are methylated at CHG sites using Pol V transcripts as scaffold by CMT3 and DRM2 (Domain rearranged methyltransferase2), a de novo DNA methyltransferase. CHG methylation is recognized by the Kryptonite6 (kyp6 aka SUVH4) which transfers the methyl group on the H3K9 (H3K9me1). CMT3, through its chromodomain, recognizes the H3K9 methylation and adds the methylation group at CHG sites forming a positive reinforcing loop of repressive epigenetic mark.

Results

MORC1 genetically interacts with a wide range of RdDM components

I assessed the genetic interaction of MORC1 with other epigenetic factors to further understand how MORC1 modulates both immunity and epigenetics. To this end, I crossed *morc1/2* with various epigenetic mutants (Table 2.1), mostly in the RdDM pathway.

Table 2.1. Epigenetic mutants crossed with *morc1/2* display visible phenotypes. High order mutants have been generated in which *morc1/2* was introduced to the indicated mutant. Progeny phenotypes different from those of the parents are indicated. Otherwise, it is indicated as ‘parental’.

Mutants	Known/putative function	Phenotype with <i>morc1/2</i>
<i>ago6</i>	RNA-induced silencing complex	Moderately curled leaves
<i>dcl3</i>	RNA-induced silencing complex	Curly leaves
<i>atx2</i>	Histone methylation-H3K4	Parental
<i>kyp6</i>	Histone methylation-H3K9	Curly leaves
<i>jmj24</i>	Histone demethylase	Parental
<i>cmt3</i>	DNA methylation (maintenance)	Curly leaves
<i>drm1/2</i>	DNA methylation (<i>de novo</i>)	Curly leaves
<i>drd1</i>	Chromatin remodeling factor (SWI/SNF2)	Curly leaves
<i>arp/suf</i>	Chromatin remodeling factor (SWR1)	Parental
<i>ddm1-2</i>	Chromatin remodeling factor (SNF2)	Curly leaves
<i>met1-1</i>	DNA methylation (maintenance)	Lethal
<i>clf28</i>	Histone methylation H3K27	Parental

Majority of epigenetic mutants tested (Table 2.1) displayed non-parental phenotypes when combined with *morc1/2*, suggesting their genetic interaction. These non-parental phenotypes are mostly curly leaves (Fig. 2.2). Based on this genetic interaction, expression of TEs were then checked since RdDM is the main mechanism suppressing TEs. Northern analysis using a *Copia* (*Copia-Romaniat5-AT1TE43225*, At1g35735) TE as a probe was performed for the mutants listed in Table 2.1. Interestingly, several lines with high-ordered mutant backgrounds displayed elevated expression of *Copia* (the top panel in Fig. 2.3). Expression of *Copia* was further enhanced when the plants were infected with *Pst*

especially in the case of *morc1/2* combined with *dcl3-1*, *drm1/2* and *ddm1*. This suggests that RdDM and chromatin remodeling factors play an important role in regulating TEs under biotic stress in plants (Katiyar-Agarwal & Jin, 2010). However, the underlying mechanism how coordinating epigenetic elements with plant immunity and why TEs become transcriptionally active is currently unclear.

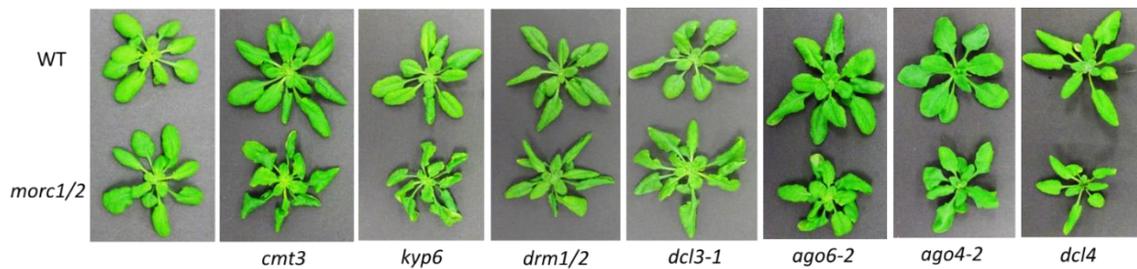


Fig. 2.2. Curly leaf phenotype of some of the high order mutants between *morc1/2* and additional epigenetic mutants. *morc1/2* was crossed with various components of the RdDM pathway. Images shown are the single mutants in Col-0 ecotype and mutants in combination with *morc1/2*.

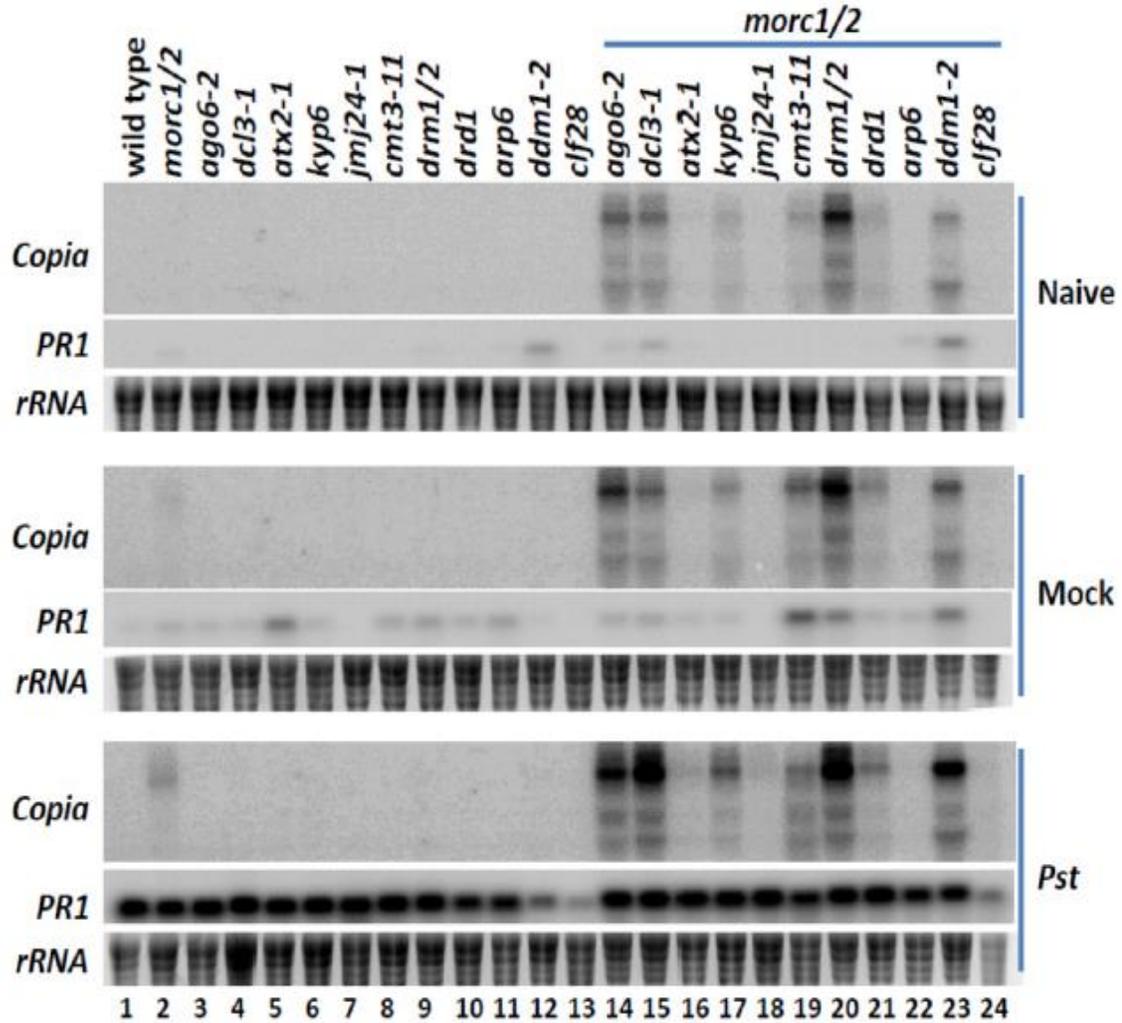


Fig. 2.3. Expression of *Copia* TEs is induced in several epigenetic mutants and its induction was further enhanced by *Pst* infection. RNA levels for a *Copia* TE were assessed in the leaves of 4 week-old naïve, mock- or *Pst*-infected (10^6 cfu/ml) WT plants and the indicated mutants at 24 hpi. The no treatment samples (Naïve) were collected at 0 hpi. rRNA was used as a loading control. RNA was resolved in 1.5% agarose gel followed by northern blot analysis. *Copia* (*Copia-Romaniat5-AT1TE43225*, At1g35735) and *PR1* were used as the probe. A single gel and membrane were used for the analysis.

Induced expression of TEs in response to pathogen infection is widespread in *Arabidopsis*

We analyzed publicly available RNA-seq data (Howard et al., 2013) to check the effect of pathogen infection on the expression of TEs. Interestingly, a wide range of TEs were transcriptionally induced at 1 and 6 hpi of *Pst* and their expression were reduced back lower

than their normal level by 12hpi (Fig. 2.4), suggesting that suppression of TE expression is temporarily lifted under biotic stress. In the Chapter 3, I used RNA-seq to analyze infection-induced *Arabidopsis* transcriptome, several RdDM-TGS components were shown to be down-regulated after infection with *Pst*. The downregulation of RdDM-TGS components coinciding with transcriptional induction of TEs under *Pst* infection argues the possibility that de-repression of TGS and its re-repression may be an important player in regulating TEs and perhaps their neighboring defense genes in plants.

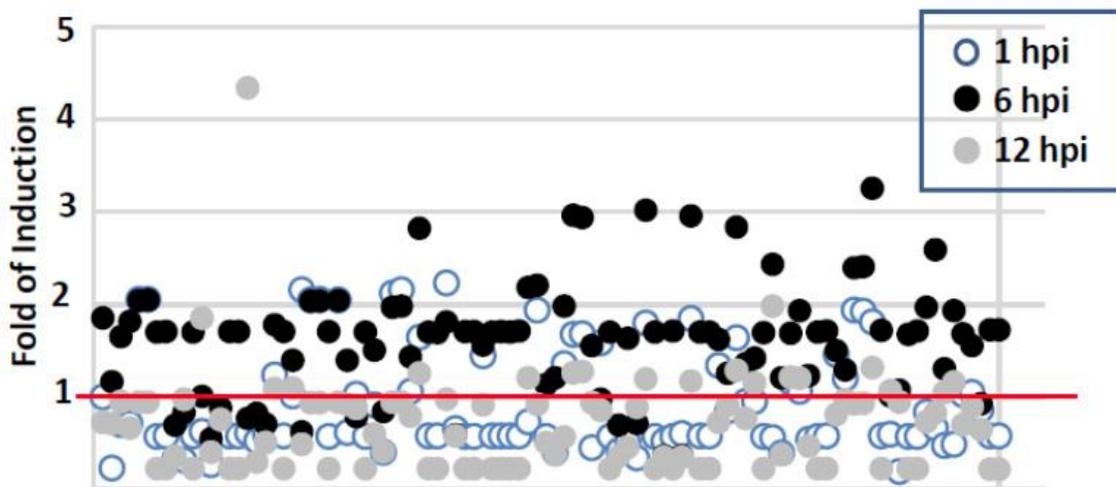


Fig. 2.4. Release of TGS coincides with increase in TE activity immediately after pathogen infection. We analyzed publicly available RNA-seq data and found that expression of TEs becomes dynamic after pathogen infection. RNA expression of TEs was increased at 1 and 6 hpi (white and black dots respectively) and subside by 12hpi (grey dots).

Role of DCL (Dicer like) proteins, a major component of the RdDM, in the defense responses in plants

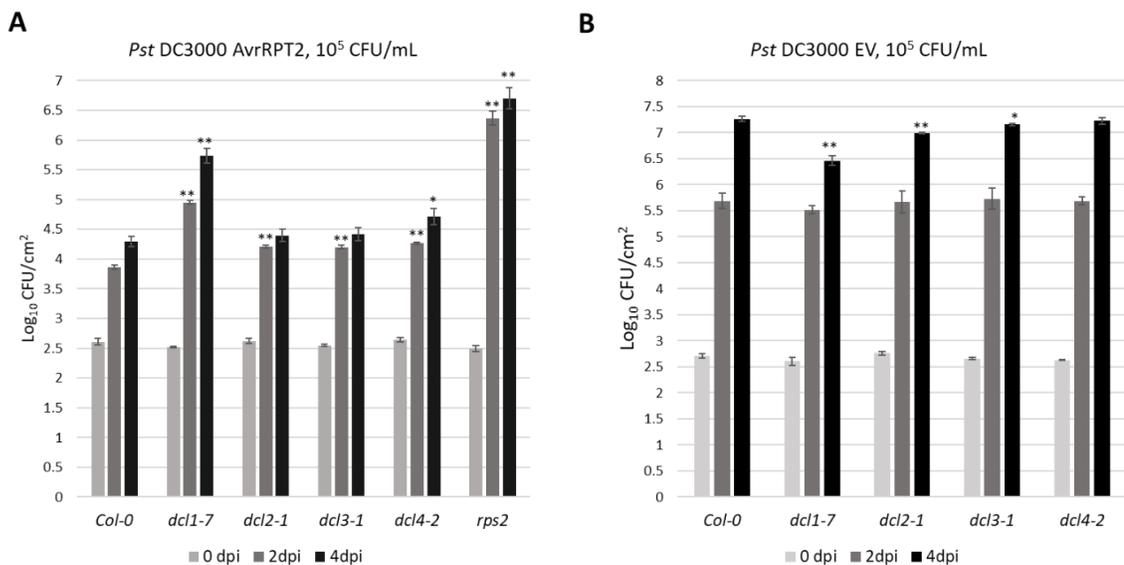
DCLs are the main enzymes for generating sRNAs from longer double dsRNAs in plants (Borges & Martienssen, 2015). There are four DCLs in *Arabidopsis thaliana*, DCL1, DCL2, DCL3 and DCL4. These DCL proteins possess the RNase-III domain which dices long dsRNA into sRNAs (Bologna & Voinnet, 2014). sRNAs generated by DCL proteins have been shown to have diverse functions in different plant-pathogen interactions (Huang

et al., 2016). DCL2 and DCL4 are involved in resistance against virus by targeting/cleaving viral DNA/RNAs (Qin et al., 2017). The role of DCLs in plant immunity, however, is not limited to the antiviral responses (Pelaez & Sanchez, 2013, Wang et al., 2017b). Several miRNAs which are processed by DCL proteins have been identified as an important regulator of antibacterial immunity in plants. For example, miRNA393 is induced upon bacterial infection and it downregulates the expression of TIR1, an Auxin receptor, which negatively regulates plant immunity (Zhang et al., 2011, Dharmasiri et al., 2005). A more recent study reported that miRNA miR863-3p regulates the timing and amplitude of defense responses against *Pst* both positively and negatively (Niu et al., 2016). miR863-3p, silences negative defense regulators, ARLPK1 (atypical receptor-like pseudokinase1) and ARLPK2, in the early stage of infection and silences a positive defense regulator SERRATE in the later stages. miRNAs and siRNA are also known to regulate the expression of plant NB-LRR R-gene family and thus preventing auto immune response in the plants (Zhai et al., 2011). Given the multiple roles of sRNAs in defense responses in host plants, it is not surprising to find that non-viral pathogens have evolved mechanisms to suppress the host immune-related RNA silencing pathway. The bacterial effector protein AvrPtoB is reported to suppress the expression of miRNA393a and miRNA393b in Arabidopsis (Lionel Navarro, 2008). Fungal pathogens deliver small RNAs as well as their own RNA silencing suppressors to compromise plant defense responses by interfering host RNAi pathways (Weiberg et al., 2013) (Qiao et al., 2013). These growing examples therefore highlight the importance of RNA silencing pathway in plant resistance.

To assess whether DCL proteins, important enzymes in RNA silencing pathway, are involved in plant immunity against *Pst*, I performed a bacterial growth assay on individual

DCL mutants. Most difference in resistance was observed from *dcl1-7* when challenged with avirulent *Pst*. In both times observed at 2 and 4 days post infiltration (dpi), *dcl1-7* supported more than 10-fold growth of avirulent *Pst* relative to WT, suggesting significant compromise in the defense responses. (Fig. 2.5A and 2.5C). The other *dcl* mutants, however, showed a marginal but consistent increase in *Pst* growth. This observation is consistent with a previous report suggesting a positive regulatory role of DCL4 in defense responses to avirulent *Pst* through the production of a novel class of small RNA known as lsiRNAs (Katiyar-Agarwal et al., 2007).

DCL mutants, however, showed small increase in resistance against virulent *Pst*, especially *dcl1-7* (Fig. 2.5B). This contrasting resistance phenotype in the *dcl* to avirulent vs virulent *Pst* suggests that sRNA biogenesis may be important in resistance requiring fast induction of defense genes while be not as important when defense responses is slowed by virulent pathogens.



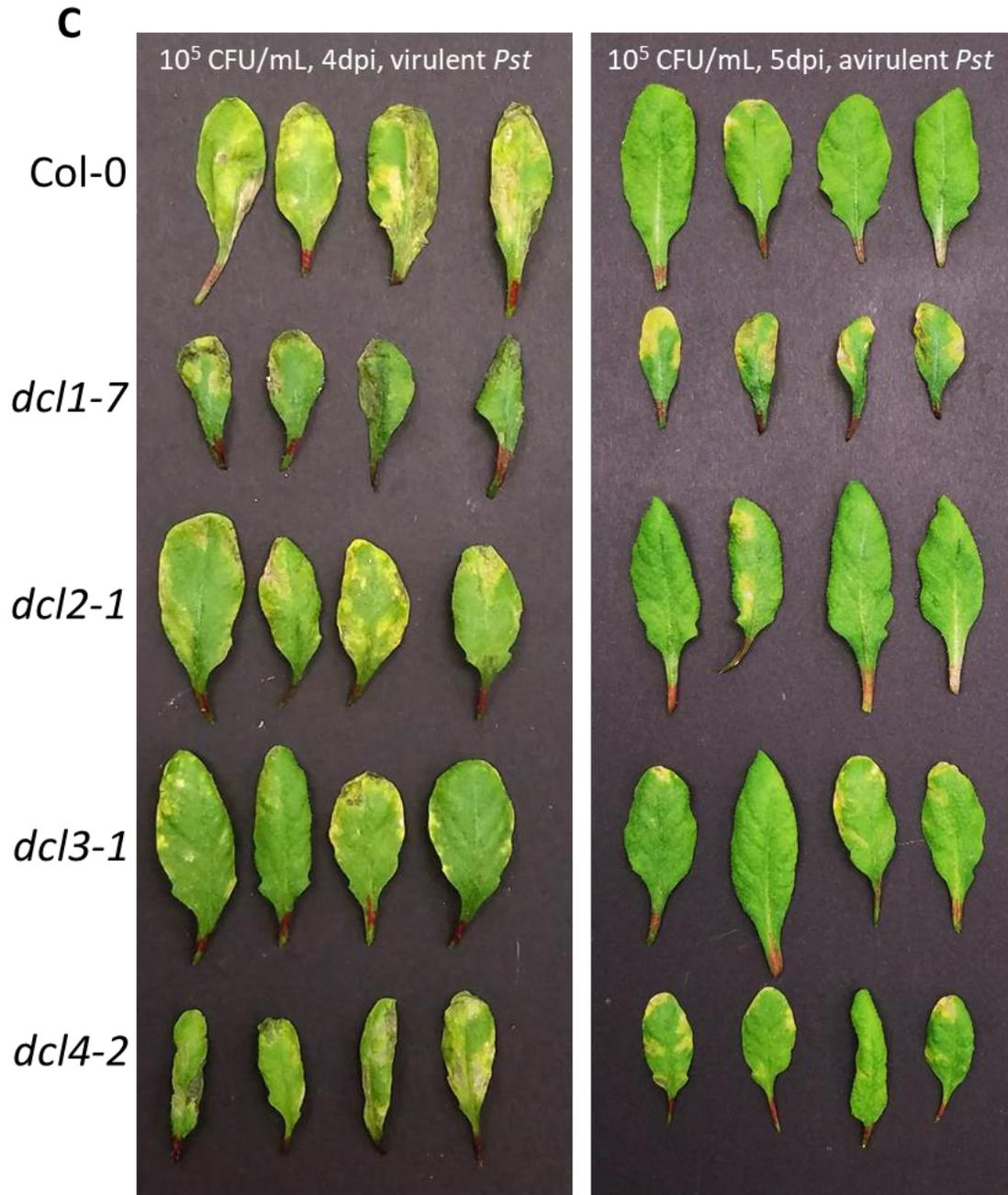


Fig. 2.5. Some *dcl* mutants are more susceptible to avirulent *Pst* but more resistant to virulent *Pst*. Bacterial growth was analyzed in WT and DCL mutant plants. 3.5 weeks old plants were infiltrated with 10⁵ cfu/mL avirulent (A) and virulent (B) *Pst*. bacterial population was analyzed at 0, 2 and 4 dpi. *rps2* was used as a control. Error bars indicate standard deviation. Statistical difference from the wild type is indicated; *P < 0.05 and **P < 0.01 (Student's t-test). C, Phenotype of WT and the *dcl* mutants at four or five dpi infected with virulent (left panel) and avirulent *Pst* (right panel).

Induction dynamics of *PR* gene is affected in the *dcl* mutants

To further characterize antibacterial defense responses in the *dcl* mutants, I checked the expression of *PR* genes. Consistent with my resistance analysis (Fig. 2.5), expression of *PR1* in *dcl1-7* was particularly compromised at 6 hpi when challenged with avirulent *Pst* (Fig. 2.6B and 2.6E). On the other hand, *dcl2-1* and *dcl3-1* showed heightened expression of the *PR* genes at the basal level prior to pathogen challenge (Fig. 2.6C, 2.7C, and 2.8C). This enhanced expression may be due to derepressed RdDM-TGS which may suppress defense genes in the absence of infection. Expression of *PR* genes in *dcl2-1* and *dcl3-1* stayed elevated during pathogen infection as compared with WT (Fig. 2.6, 2.7 and 2.8). In the case of *PR2* and *PR5*, the expression of these genes is lower at 6 hpi in *dcl1-7* relative to WT in response to avirulent *Pst* infection while it is higher at 24 hpi. Consistent with the *PR1* expression pattern, both *PR2* and *PR5* showed lower expression in *dcl1-7* relative to WT under virulent *Pst* infection.

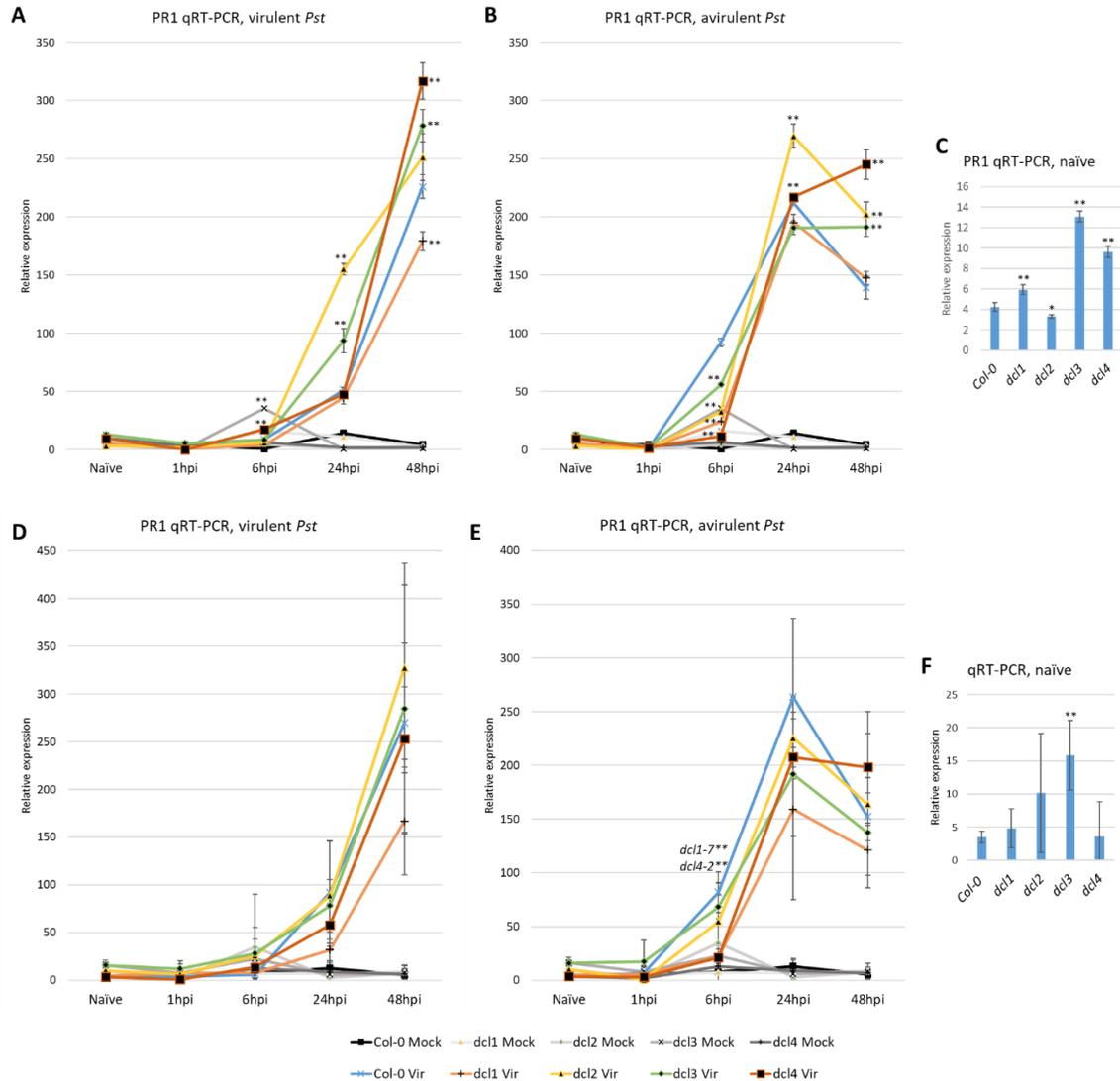


Fig. 2.6. Induction dynamics of *PR1* is affected in the *dcl* mutants. Expression analysis of *PR1* using qRT-PCR was performed at 0, 1, 6, 24 and 48 hpi with virulent and avirulent *Pst*. **A and B**, Leaves were inoculated with buffer (mock) or 10^6 cfu/ml of virulent (**A**) and avirulent (**B**) *Pst*. Untreated (naïve) control leaves were harvested at 0 hpi. RNA prepared from WT and the *dcl* mutant plants at the indicated time points after inoculation were converted into cDNA and were used for qRT-PCR analysis with the *PR1* specific primers. **C**, Basal expression of *PR1* in naïve plants. **D-F**, Combined representation of three biological replicates including the ones in (**A**), (**B**) and (**C**). The *TIP41-like* gene was used as a reference gene for qRT-PCR. Error bars indicate the standard deviation. Statistical significance from WT was determined using t-test: * $P < 0.05$; ** $P < 0.01$.

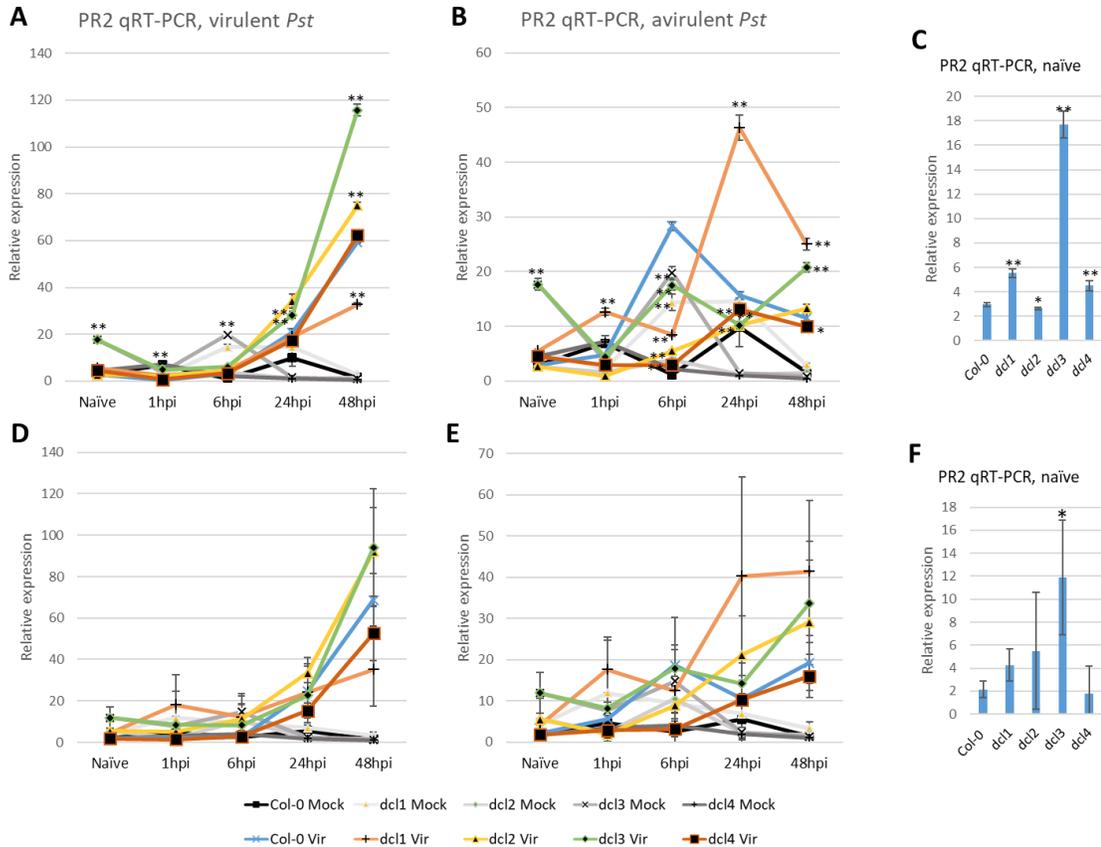


Fig. 2.7. Induction dynamics of *PR2* is affected in the *dcl* mutants. Expression analysis of *PR2* using qRT-PCR was performed at 0, 1, 6, 24 and 48 hpi with virulent and avirulent *Pst*. **A and B**, Leaves were inoculated with buffer (mock) or 10^6 cfu/ml of virulent (**A**) and avirulent (**B**) *Pst*. Untreated (naïve) control leaves were harvested at 0 hpi. RNA prepared from WT and the *dcl* mutant plants at the indicated time points after inoculation were converted into cDNA and were used for qRT-PCR analysis with the *PR2* specific primers. **C**, Basal expression of *PR2* in naïve plants. **D-F**, Combined representation of three biological replicates including the ones in (**A**), (**B**) and (**C**). The *TIP41-like* gene was used as a reference gene for qRT-PCR. Error bars indicate the standard deviation. Statistical significance from WT was determined using t-test: * $P < 0.05$; ** $P < 0.01$.

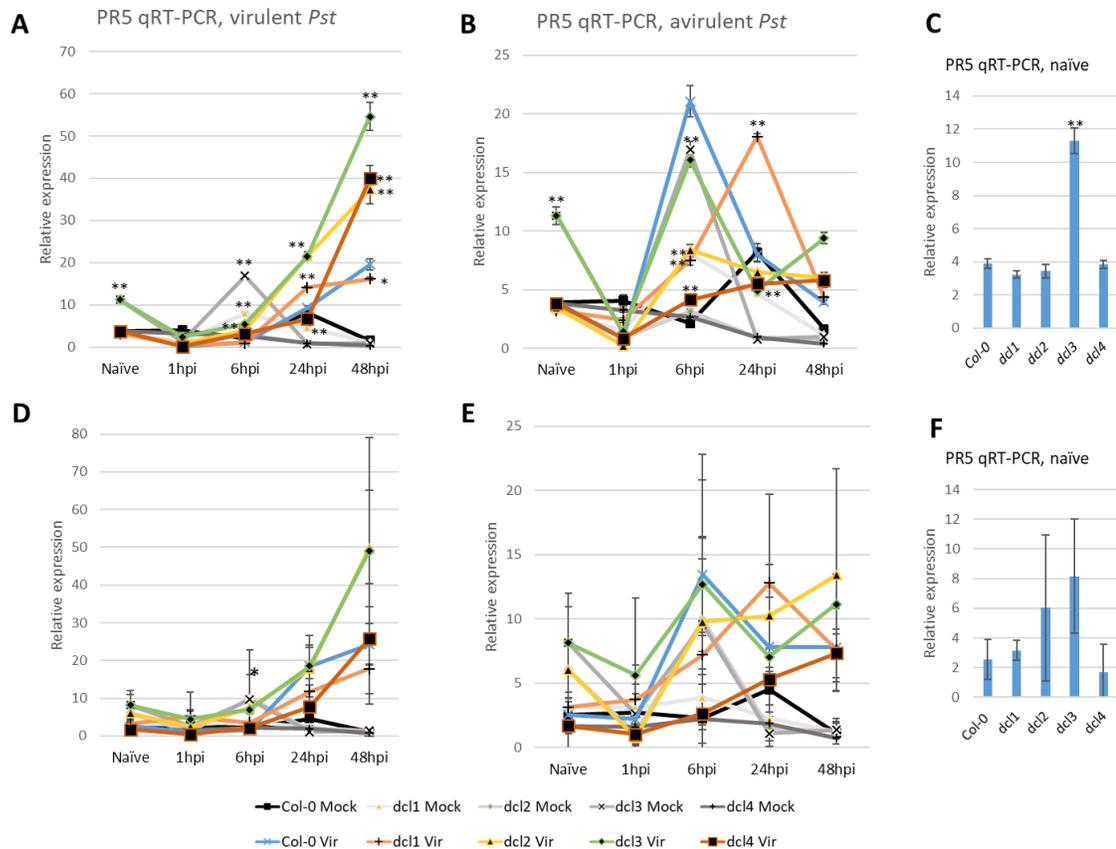


Fig. 2.8. Induction dynamics of *PR5* is affected in the *dcl* mutants. Expression analysis of *PR5* using qRT-PCR was performed at 0, 1, 6, 24 and 48 hpi with virulent and avirulent *Pst*. **A and B**, Leaves were inoculated with buffer (mock) or 10^6 cfu/ml of virulent (**A**) and avirulent (**B**) *Pst*. Untreated (naïve) control leaves were harvested at 0 hpi. RNA prepared from WT and the *dcl* mutant plants at the indicated time points after inoculation were converted into cDNA and were used for qRT-PCR analysis with the *PR5* specific primers. **C**, Basal expression of *PR5* in naïve plants. **D-F**, Combined representation of three biological replicates including the ones in (**A**), (**B**) and (**C**). The *TIP41-like* gene was used as a reference gene for qRT-PCR. Error bars indicate the standard deviation. Statistical significance from WT was determined using t-test: * $P < 0.05$; ** $P < 0.01$.

DCL2, DCL3 and DCL4 show functional redundancy

Although the individual mutants *dcl2-1*, *dcl3-1* and *dcl4-2* showed quite marginal defense responses, this minor change may be due to functional redundancy. A gene could be induced in a mutant background in which its functionally related/redundant gene is knocked down. To assess this possibility, I analyzed the expression of the *DCL* genes in all the *dcl* mutant backgrounds. Indeed, the expression level of *DCL2* in *dcl3-1* increased

significantly and that of *DCL3* was also enhanced in *dcl2* (Fig. 2.9A), suggesting that DCL2 and DCL3 may complement each other. Although the transcript level of *DCL4* did not change in any of the individual mutant backgrounds, it was significantly increased in the *dcl2-1/dcl3-1* double mutant background (Fig. 2.9B), highlighting a complicated functional relationship among DCL2, DCL3 and DCL4. *DCL1* was substantially induced in *dcl1-7* (Fig. 2.9A). *dcl1-7* carrying a point mutation and this increased expression may be a feedback loop to compensate for the defective protein.

To find out if this change in the transcript level leads to alteration in sRNA biogenesis, I examined the size ratio of sRNAs in all *dcl* mutants since each DCL predominantly produces sRNAs in a specific size(s). It is widely accepted that DCL1 and DCL4 dice long dsRNA into 21 nt sRNA and DCL2 and DCL3 into 22 and 24 nt sRNA, respectively (Bologna & Voinnet, 2014). I used publicly available data (Kristin D. Kasschau, 2007) and found a notable change in the population of the 20-22 nt sRNAs in *dcl3-1* and an increased ratio of 23 nt sRNA in *dcl1-7* (Fig. 2.9C). These results together with the transcriptional changes of *DCL* genes in each *dcl* mutant suggest complicated functional overlap/redundancy among the DCL proteins in Arabidopsis. Therefore, to better understand the role of these proteins in bacterial resistance, higher-order *dcl* mutants are currently being analyzed.

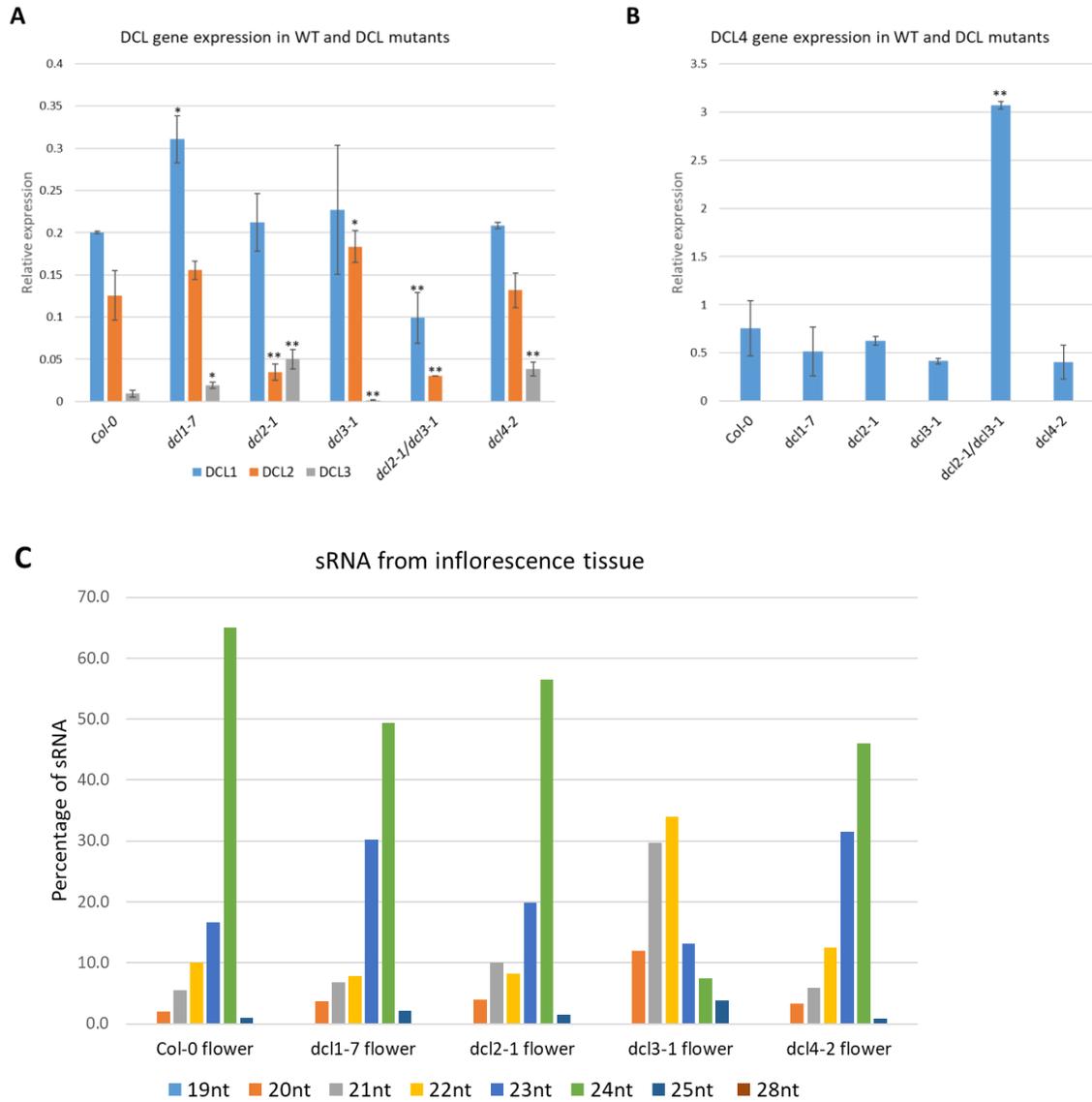


Fig. 2.9. Arabidopsis DCL proteins are functionally redundant. **A**, Expression analysis of the *DCL1*, *DCL2* and *DCL3* gene under untreated (Naive) condition in WT, *dcl1-7*, *dcl2-1*, *dcl3-1*, *dcl2-1/dcl3-1* and *dcl4-2* using qRT-PCR. **B**, Expression analysis of the *DCL4* gene expression under untreated (Naive) condition in WT, *dcl1-7*, *dcl2-1*, *dcl3-1*, *dcl2-1/dcl3-1* and *dcl4-2* using qRT-PCR. Statistical significance from WT was determined using Student's t-test: * $P < 0.05$; ** $P < 0.01$. **C**, Analysis of the publicly available data showing the percentage distribution of various sRNA size in WT, *dcl1-7*, *dcl2-1*, *dcl3-1* and *dcl4-2*. Note that the sRNAs were prepared from inflorescence.

Discussion

Synergistic increase in expression of the *Copia* TE (Fig. 2.3) and development of curly leaf phenotypes when *morc1/2* is crossed with various RdDM mutants (Fig. 2.2) suggest that MORC1 and tested RdDM components function together in RdDM-TGS. In addition, MORC1 has been shown to be a part of the DDR chromatin remodeling complex of the RdDM pathway (Fig. 2.1) (Lorković et al., 2012), raising a possibility that MORC1 may function as a RdDM-associated chromatin remodeling protein. Therefore, it will be interesting to find the composition of the MORC1-associated complex involving in activation of TEs and defense genes in response to biotic stress in the future.

Numerous studies have reported induction of the TE expression under stress (Wang et al., 2017a) (Seidl & Thomma, 2017). However, the exact role of stress-responsive TEs is currently unknown. As shown in the Chapter 1, silencing TEs led to compromised expression of neighboring defense genes, suggesting that these TEs function as a transcriptional enhancer. RdDM has been shown to control the expression of both TEs and genes (Au et al., 2017, Rowley et al., 2017). It is interesting to test therefore if the induction dynamics of RdDM-TGS is correlated with the strength of defense response (i.e., incompatible vs compatible defense responses). In the following chapter, I attempted to address this possibility by analyzing a large number of defense genes in the *dcl* mutants.

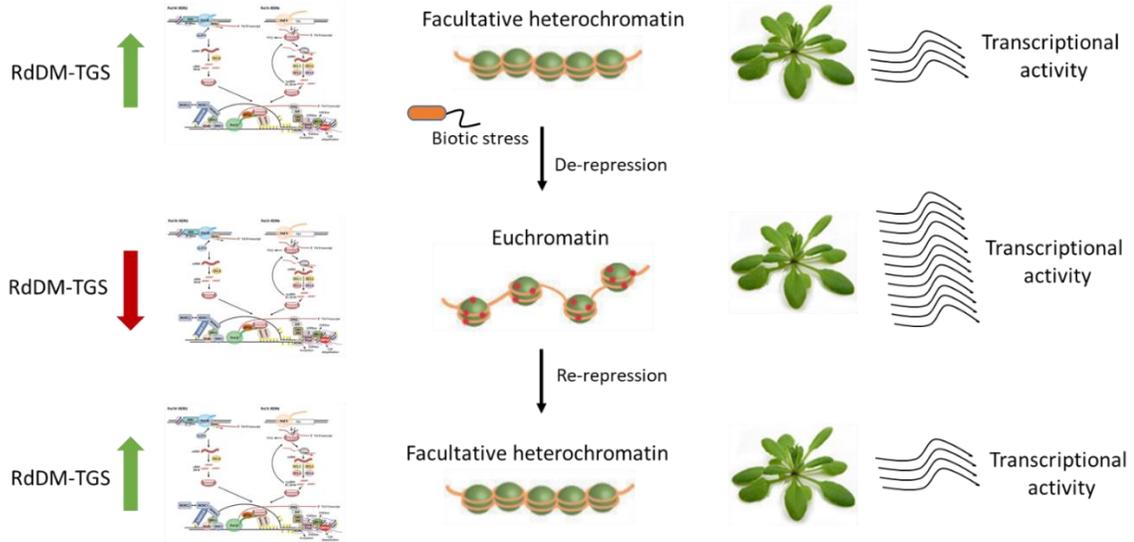


Fig. 2.10. Working model. Under normal growth conditions, the targets of RdDM-TGS are TEs, repeat sequences and genes distributed throughout the genome to maintain facultative heterochromatin. Under biotic stress at early time points, RdDM-TGS is released which jump-starts the transcriptional activity of defense genes in the genome. In the absence of an RdDM component such as DCL3, TGS is constitutively released under normal condition and plants have a heightened basal level expression of *PR* genes and misregulated kinetics of defense genes (Fig. 2.6C, 2.7C, 2.8C), which lead to altered resistance phenotypes in Arabidopsis (Fig. 2.5). By 24 hpi, WT plants re-repress the induced transcriptional activity. In contrast, RdDM defective mutants such as *dcl3* fail to perform the re-repression, resulting in the high transcriptional level of defense genes for an extended time (Fig. 2.6, 2.7 and fig. 2.8).

In this chapter, I showed that a wide range of TEs become transcriptionally active after *Pst* infection. To further characterize the role of RdDM-TGS, a pathway which controls the activity of these TEs, in plant immunity, all the *DCL* genes which are critical in sRNA biogenesis were characterized in resistance and expression of *PR* genes to *Pst* infection. Interestingly, transcriptional expression of *PR* gene was notably elevated in *dcl3-1* before and after infection with virulent *Pst*, raising a possibility that release of RdDM-TGS due to the *dcl3-1* mutation led to derepressed defense genes, and that a tight regulation of defense genes may require the RdDM-TGS pathway. In addition, *dcl1-7* displayed significant compromise in resistance to avirulent *Pst* and delayed/reduced induction of *PR* genes, suggesting that DCL1 may be involved directly or indirectly in regulating the induction of defense genes. Together, my findings suggest that compromised RdDM-TGS

leads to altered defense gene dynamics because RdDM-TGS may regulate timely de-repression and re-repression of defense genes following pathogen infection. However, given that a small subset of defense genes were tested in this chapter, the findings may be difficult to be generalized. Therefore, in the following chapter, I established a list of genes that become highly dynamic in transcription under avirulent and virulent *Pst* infection, and further tested hundreds of defense genes in the *dcl* mutants to gain insights into the role of DCLs in defense gene induction dynamics.

CHAPTER III

GLOBAL IDENTIFICATION OF DEFENSE GENES WITH RAPID INDUCTION KINETICS AND THEIR CHARACTERIZATION IN *DCL* MUTANTS

Abstract

Appropriate timing and magnitude of defense gene induction is critical in plant resistance. In previous chapters, several epigenetic mutants displayed compromised induction dynamics of a few selected defense genes. To further study the expression dynamics of defense genes and its potential dependence on epigenetic components, I performed transcriptomic analysis to identify defense genes with rapid induction kinetics and assessed a wide range of defense genes in the *dcl* mutants. A conventional mRNA-seq was performed on Arabidopsis WT plants infected with avirulent and virulent *Pst* for 1, 6, 24 and 48 hours. To assess the dynamics of mRNA transcriptome in response to bacterial infection, a recently revised bioinformatic tool, maSigPro, was used for clustering defense genes based on their induction dynamics. Once candidate defense genes were identified and clustered, I turned to a targeted RNA quantitation approach to examine a large number of RNA samples collected from the *dcl* mutants as well as WT plants that were challenged with avirulent and virulent *Pst* at five time points. RASL-seq (RNA-mediated oligonucleotide Annealing, Selection, and Ligation with next-generation sequencing) was developed several years ago as a targeted RNA analysis and recently improved to eliminate a well-known low signal-to-noise ratio issue. I confirmed that this revised RASL-seq approach was quantitatively reliable relative to RNA-seq and qRT-PCR analysis performed on the same RNA samples. Correlation analysis showed that the *R*-squared value for RASL-seq vs RNA-seq and RASL-seq vs qRT-PCR exceeded 0.95, suggesting that RASL-

seq is acceptable for RNA quantitation analysis. Using the RASL-seq, I found that some early and late induced defense genes displayed significantly lower expression in *dcl1-7* infected avirulent *Pst* at 6 hpi relative to WT. In addition, *dcl2-1* and *dcl3-1* showed heightened basal expression of defense genes prior to infection as compared with the WT. These findings suggest that the induction dynamics of defense genes is affected in the *dcl* mutants, which therefore support the possibility that epigenetic components play an important role in regulating the defense genes.

Introduction

Quantifying the spatiotemporal expression of genes under various environmental stimuli is one of the most informative experimental procedures to understand biology at the molecular level. The RNA blotting, also known as northern blotting, through which expression of one gene can be checked at a time in a qualitative manner, has been used for decades. More recently, the system-wide tools such as microarray and RNA-seq have been developed which have been used to get the quantitative measure of essentially most of the genes, if not all, in an organism. As RNA-seq depending on next-generation DNA sequencing technology becomes popular, , there is a growing need for multiplexing targeted transcriptome to accommodate a large number of samples (Metzker, 2010, Mardis, 2013, Feng et al., 2015).

Two multiplexing options for a targeted RNA analysis are currently available although they are not widely adopted. The first utilizes a probe carrying multiple fluorophores that are tagged with targeted oligonucleotides (Geiss et al., 2008, Reis et al., 2011). This tool, named NanoString nCounter (NanoString), is capable of analyzing hundreds of genes at a higher sensitivity than that of microarray. It relies on a highly sensitive CCD (charge-coupled device) camera to detect the fluorophore probes, which enables the analysis of a large number of samples. However, because NanoString involves an advanced CCD camera and the fluorophore-tagged oligonucleotides, the overall cost is very high; the base machine (NanoString Sprint) alone is currently sold at \$150k. The second tool is RASL (RNA-mediated oligonucleotide Annealing, Selection, and Ligation with next-generation sequencing)-seq (Yeakley et al., 2002, Li et al., 2012b). While this tool was developed earlier than NanoString, it had a considerable background issue, which limits its utility in

analyzing low-expressing genes. Recently, a modification in RASL-seq was shown to overcome this shortcoming. Original RASL-seq used a DNA ligase to anneal two DNA oligonucleotides that were bound to target RNA. In the improved RASL-seq, an RNA ligase is used instead. This modification is achieved by adding two ribo-nucleotides at the end of one of oligonucleotides, which substantially enhances the efficacy of oligonucleotide annealing at more than 100 folds, and therefore, significantly reduces the background and enhances the sensitivity (Larman et al., 2014).

To prepare the library for RASL-seq, as shown in fig. 3.1, a pair of 20 base-long oligonucleotides is designed complementary to a target mRNA. One of the oligonucleotides has phosphate at the 5' end (donor probe) and the another has two ribonucleotides at the 3' end (acceptor probe). After mRNA and oligonucleotides are annealed and washed (to remove un-annealed oligonucleotides and RNA), Rnl2 (T4 dsRNA ligase2) is used to ligate the oligonucleotides. Following ligation, multiplexing barcode primers are used in PCR which only amplifies the ligated pairs. Note that an 8 nt barcode sequence was added to both barcoding primers. All the samples can be pooled together after PCR and run on an agarose gel for band isolation of the library followed by quantification and sequencing.

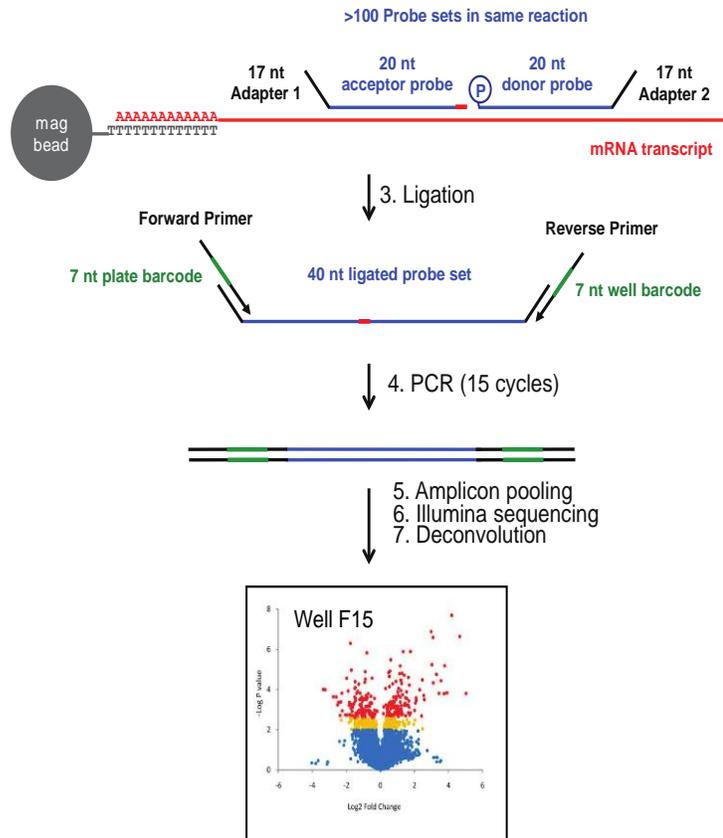


Fig. 3.1. RASL-seq procedure (adapted from Larman et al 2014, *Nucleic Acid Research*). mRNA is enriched using biotinylated oligo-dT from total RNA and the acceptor probes (with ribonucleotides on 3' end) and the donor probes (with phosphate group on the 5' end) are annealed. After the probe annealing and washing of the unannealed probes and RNA, a double stranded T4 RNA ligase2 (Rnl2) is used for ligating the annealed probes. After ligation, barcoding PCR is performed for multiplexing to accommodate a large number of samples. Library DNA is then mixed, run on an agarose gel followed by purification of the expected size band and the Illumina next-gen sequencing.

The library structure of RASL-seq is shown in the fig. 3.2. P5 and P7 sequences allow the library DNA to bind to the Illumina flow cell. With the dual index sequencing capability, up to 900 samples can be multiplexed with mere 30 P5 and 30 P7 multiplexing barcode primers. The length of the insert sequence is 40 base pairs (20bp from the donor probe, and 20bp from the acceptor probe). Two adapter sequences on the either side of the insert allows binding of “Read1”, “Read2” and “Index1” primers for sequencing. The RASL-seq library uses same “Read2” and “Index1” primers as standard Illumina HiSeq and MiSeq

Results

Identification of highly dynamic differentially expressed genes, using RNA-seq, in *Arabidopsis* infected with *Pseudomonas*

Successful defense responses often rely on their rapidity and robustness (Tao et al., 2003). While the dynamics of defense gene induction has been known to be important, the underlying mechanism contributing massive gene induction under biotic stress is still unclear. As I have characterized some of epigenetics components that are known to be involved in defense responses, it became apparent that epigenetic changes associated with chromatin play a role in induction dynamics of defense genes as I demonstrated with MORC1 family in *Arabidopsis* (Bordiya et al., 2016). Unfortunately, to date, there is no solid *Arabidopsis* study done to investigate the dynamics of defense gene induction at the system level. Howard et al (2013) has performed an mRNA transcriptome study using avirulent and virulent *Pst*. However, this study suffers significant statistical issues because it was only done in duplicates and each replicate shows high variability. Recently, another transcriptome report with a much greater set-up using *Pst* was published (Lewis et al., 2015). Unfortunately, this study only utilized virulent *Pst* and mainly investigated MAMP-triggered immunity. Thus, I decided to set up another transcriptome experiment with an aim to examine the induction kinetics of defense genes in response to avirulent as well as virulent *Pst*.

In my transcriptome analysis, *Arabidopsis* WT plants were infected with avirulent and virulent *Pst* for 1, 6, 24 and 48 hrs. This time course was used to target early as well as late responsive defense genes (Table3.1); mock and no-infection (naïve) controls were also included. To have statistical strength significant enough for multi-time point kinetic

analysis, three biological replicates were prepared for all the samples generated. Over 250 million reads were obtained through a conventional mRNA sequencing protocol via an Illumina sequencing (see the next Chapter for more detail in the procedure).

Most commonly used bioinformatics tools for an mRNA-seq to date include edgeR (Robinson et al., 2010) and DEseq (Anders & Huber, 2010). These early methods were developed to accommodate experiments with low replication because next-gen sequencing several years back was very expensive. Recently, another bioinformatics tool, maSigPro, was revised to accommodate time-course approaches with high replication, three or more times (Nueda et al., 2014). Thus, we utilized this time-course friendly bioinformatics tool to process my sequence reads. The maSigPro analysis revealed that 9 clusters display distinct dynamics of total 1061 genes (Fig. 3.3). To better visualize the trend of each cluster, we used the Pigengene package, recently developed in our collaborator's lab (Foroushani et al., 2017), which summarizes each cluster into one value (i.e., an eigengene; Fig. 3.4). Typical Early and late responsive defense genes belonged to cluster 3 and 6, respectively. In addition, although cluster 4 genes respond to mock treatment, gene in this category also show the induction kinetics as rapid as cluster 3. Cluster 9 is another interesting group showing faster induction in response to virulent *Pst*, relative to the avirulent counterpart. The other clusters, although displaying clear difference as compared with the mock control, mostly show reduction or marginal induction. Therefore, for the subsequent targeted RNA-seq approach, I mainly chose genes from the clusters 3, 4, 6 and 9 based on their rapid induction dynamics.

Table 3.1. RNA-seq sample composition to assess the induction dynamics of defense genes. RNA-seq was performed on WT, which were infected with mock, virulent and avirulent *Pst* at 1, 6, 24 and 48 hpi. Untreated WT was included (Naive) as a control. Each biological batch therefore consists of 13 samples, which was replicated three times.

Treatment	Hours post infiltration (hpi)					Row total
	0 hpi	1hpi	6hpi	24hpi	48hpi	
Naïve (N) No treatment	1					1
Mock (M)		1	1	1	1	4
Virulent <i>Pst</i> (V)		1	1	1	1	4
Avirulent <i>Pst</i> (A)		1	1	1	1	4
Total number of samples per biological replicate						13

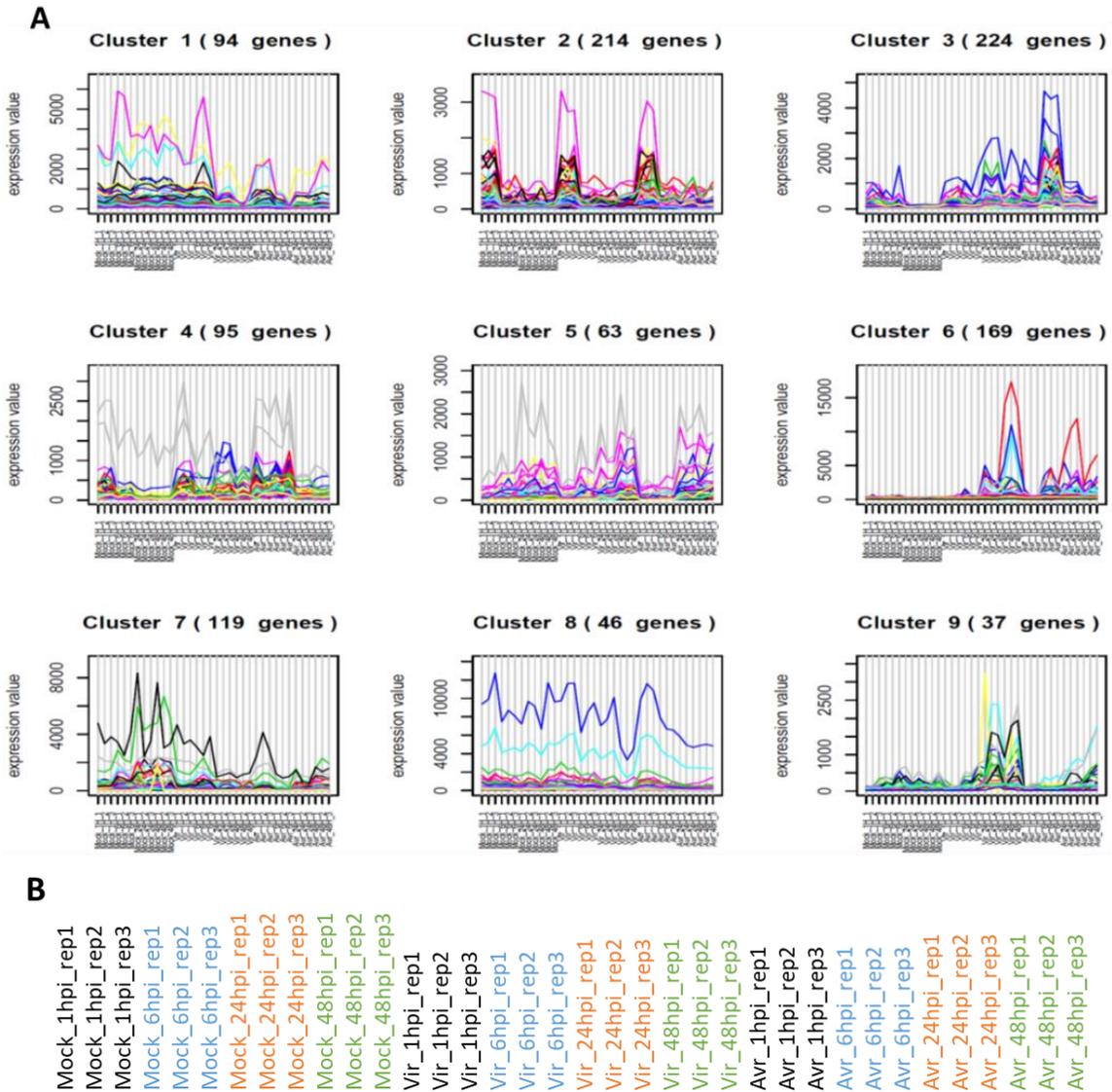


Fig. 3.3. Clustering of defense genes displaying rapid induction dynamics in the RNA-seq data. **A**, RNA-seq was performed on WT. Depending on the induction pattern after a particular time and/or treatment, genes were divided into 9 different clusters by using the maSigPro R package. The y-axis is the RPKM value and the x-axis shows time and treatment of three biological replicates. The order of the samples on x-axis is shown (Mock 1, 6, 24 and 48 hpi; virulent *Pst* 1, 6, 24 and 48 hpi; avirulent *Pst* 1, 6, 24 and 48 hpi). **B**, The left to right label of the x-axis in the graphs from panel A.

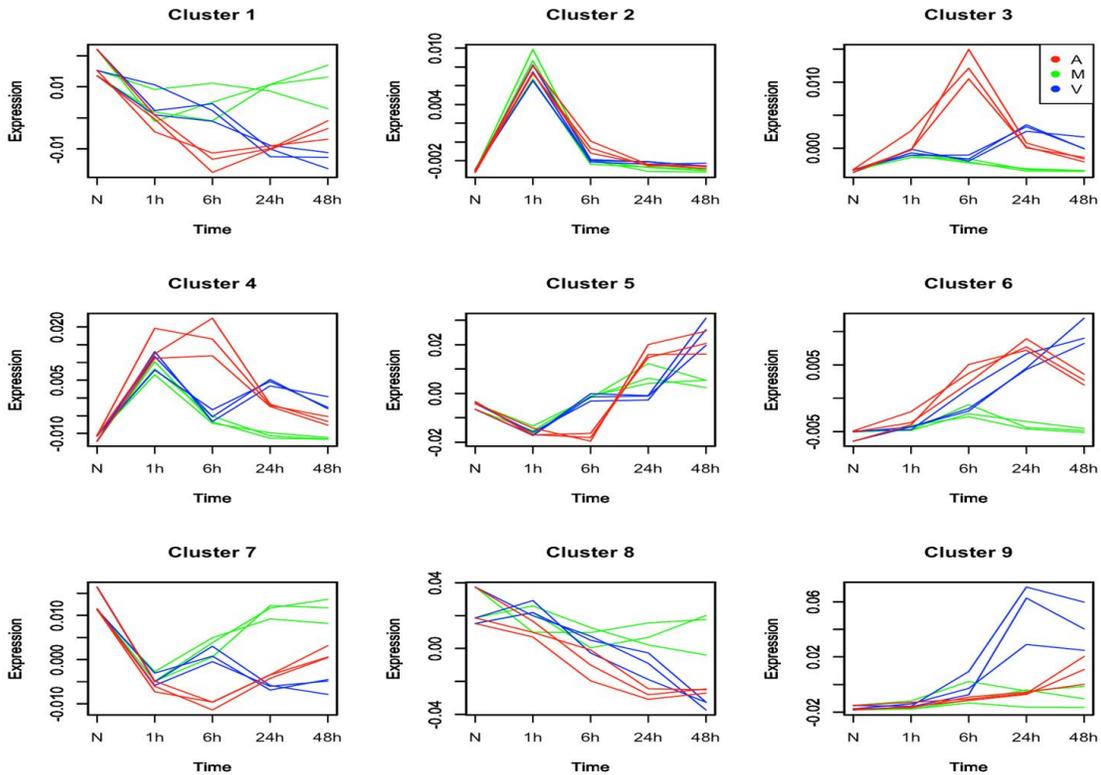


Fig. 3.4. Identification of 9 clusters of genes differentially expressed in *Arabidopsis* infected with avirulent and virulent *Pst*. To better visualize the trend of each cluster from Fig. 3.3, Eigengene analyses was performed. Green, blue and red color lines indicate mock, virulent *Pst* and avirulent *Pst* treatment respectively. Each line indicates one biological replicate.

Expression of RdDM components, including MORC1, was downregulated upon *Pst* infection

Expression of the major components in the RdDM pathway is reduced after treatment with flg22 a PAMP, suggesting that suppression of RdDM-TGS after pathogen infection may induce demethylation. This sequence of events in turn may play a role in the activation of neighboring defense genes such as RMG1 (Yu et al., 2013). My RNA-seq data revealed a similar transcriptional suppression of the RdDM pathway (Fig. 3.5). This suppression was greater when plants are challenged with avirulent *Pst* as compared with the virulent counterpart, suggesting that the strength of defense response may be correlated with the magnitude of RdDM-TGS suppression. This reduction in the expression of RdDM

components after pathogen infection coincides with the increase in the TEs activity as shown in Chapter 2 (Fig. 2.4) suggesting that RdDM-TGS, TE activity and biotic stress are closely associated.

	N	M.1	M.6	M.24	M.48	V.1	V.6	V.24	V.48	A.1	A.6	A.24	A.48
NRPE1	1.01	1.16	1.47	1.83	1.31	0.6	1.43	0.66	1.36	0.73	0.71	1.49	1.91
NRPD1A	0.82	0.47	1.32	0.87	0.85	0.5	0.85	0.87	1	0.45	0.85	1.09	0.97
AGO6	0.48	0.24	0.52	0.4	0.47	0.42	0.39	0.12	0.01	0.18	0.22	0.4	0.13
AGO4	5.79	2.57	4.27	4.75	4.93	2.25	3.47	1.02	1.26	1.68	0.71	1.88	3
DCL3	0.37	0.64	0.45	0.83	0.56	0.4	0.37	0.31	0.52	0.29	0.22	0.42	0.68
IDN2	3.97	2.12	4.56	4.64	3.85	3.04	5.38	2.38	3.07	1.35	1.6	3.46	4.22
MORC1	8.41	5.82	7.88	8.34	7.76	6.11	6.69	6.43	8.32	5.06	3.97	8.06	8.88

Fig. 3.5. *Pst* infection downregulates the expression of RdDM components including MORC1. Expression value (RPKM) for various components of the RdDM pathway in response to avirulent and virulent *Pst*. Green to red color gradient represents low to high expression respectively. The numbers indicate RPKM value. The column name consists of the following codes: N=Naïve, M=Mock, V=Virulent *Pst* and A= Avirulent *Pst*; 1, 6, 24 and 48 indicate hpi (e.g. V.48 = virulent *Pst* infection for 48 hours).

RASL-PCR is highly quantitative

RASL-seq is a massive parallel analysis that quantitates a large number of targeted RNAs (Fig. 3.1). A biotin-tagged oligo dT primer further provides the selection of RNAs since this step removes most unbound oligonucleotides. These oligonucleotides that were purified through the process are then subject to PCR for the sample barcoding. There are a few significant advantages of RASL-seq. First, there is no need for cDNA synthesis and additional library preparation before the sequencing, which save time and simplify the procedure. T4 DNA ligase was used in the original RASL-seq to join DNA oligonucleotides on an RNA template. This enzyme, however, showed poor performance (Larman et al., 2014). In contrast, Rnl2, a dsRNA ligase, was at least a hundred times more efficient than the DNA ligase (Larman et al., 2014). Interestingly, Rnl2 ligates both DNA and RNA if a minimum of two ribonucleotides are present between two joining

oligonucleotides. Therefore, two ribonucleotides were added at the 3' end of acceptor. This significant improvement appeared to transform the RASL-seq as the arguably best parallel transcriptomic tool. Therefore, I decided to employ this tool to track some of the defense genes, identified through my RNA-seq, in the *dcl* mutants.

To ensure that RASL-seq worked as reported, I have run a pilot RASL experiment which combines RASL and PCR (RASL-PCR; Fig. 3.6). Several pairs of oligonucleotides for housekeeping genes in Arabidopsis were prepared; the pair of primers has a phosphate at the 5' end and two ribonucleotides at the 3' end, respectively. This pair was successfully annealed and amplified only in the presence of RNA and Rnl2 as shown in fig. 3.6A. To further assess the quantitation characteristics of RASL, the mixed amount of Arabidopsis and Maize RNAs was assessed for the RASL-PCR experiment. The targeted sequences by oligonucleotide probes are absent in Maize and were rarely amplified (the last lane in Fig. 3.6C). This suggests that RASL-PCR is specifically dependent upon its target. In addition, the amplification was highly proportional to the relative amount of Arabidopsis RNA (Fig. 3.6B), indicating that RASL was performed in a highly quantitative manner.

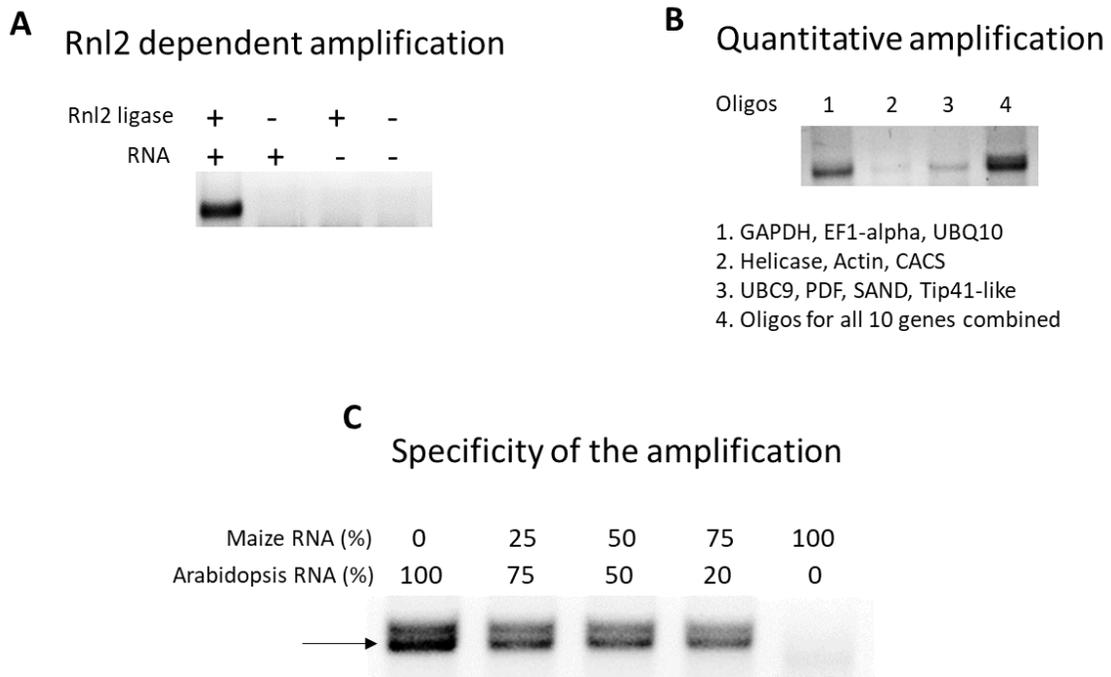


Fig. 3.6. Optimization of RASL procedure using RASL-PCR. **A**, After annealing, selection, and ligation, PCR was run using barcode primers. The band in the first lane shows that Rnl2 dependent amplification was achieved. **B**, To check whether i) the amount of probe (10nM) used in RASL provide the visible difference between highly and lowly expressed genes and that ii) quantitative difference can be seen if the probes specific to many genes are used, semi-quantitative RASL-PCR was performed. Note that for lanes 1 through 3, the number of transcripts decrease left to right, based on their known expression values. Lane 4 contains all the probes.. **C**, To check the specificity of RASL procedure, probes specific to the Arabidopsis genome which do not have complementary sequence in the Maize genome were used and RASL-PCR was performed with various percentages of the Maize and Arabidopsis RNA.

Quantitation data from a full-scale RASL-seq was comparable with those of a conventional mRNA-seq

After my semi quantitative experiment suggested that the improved RASL-seq shows a strong quantitative correlation (Fig. 3.6), I increased the testing capacity of RASL-seq to 227 of genes; note that for the full-scale RASL-seq, WT as well as four *dcl* mutants were examined. As RASL-seq is yet to be widely adopted, it is critical to demonstrate that this RNA analysis tool is as quantitative as RNA-seq. To this end, I used the same RNA set that was processed in my earlier RNA-seq approach (Table 3.1) and compared the outcomes between these two RNA analysis as shown in fig. 3.7B; only *PRI* was shown for

brevity. R-squared value for *PR1* between these RNA analyses is 0.99, suggesting that a wide range of the expression values are highly correlative. I performed correlation analysis between the qRT-PCR tested samples (Table 3.2; Materials and Methods) of the *dcl* mutants which were used for the RASL-seq and found R-squared value more than 0.95 for *PR1* (Fig. 3.7A). In summary, the full-scale RASL-seq is highly quantitative, which is consistent with the earlier report (Larman et al., 2014).

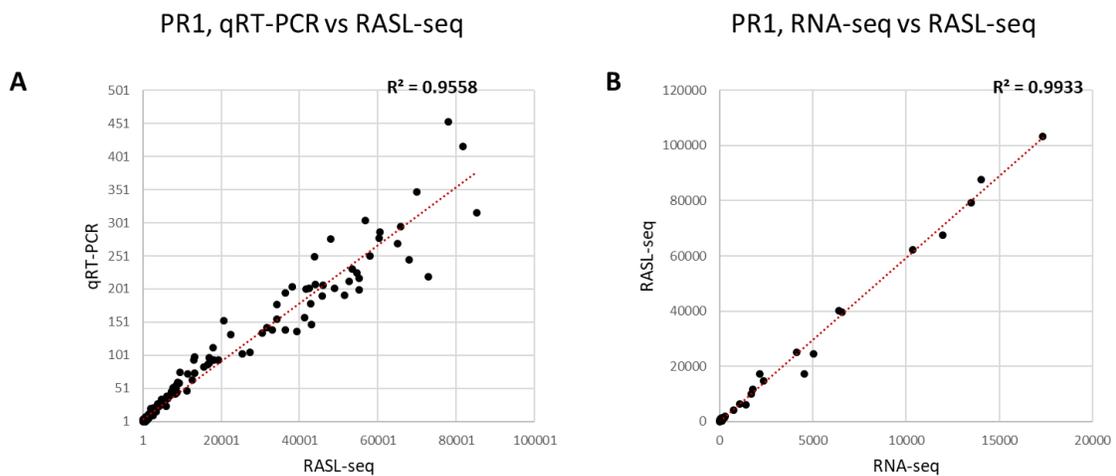


Fig. 3.7. High correlation of RASL-seq vs qRT-PCR and RASL-seq vs RNA-seq. A, Correlation analysis of *PR1* between qRT-PCR and RASL-seq from the same RNA set. **B,** Correlation analysis of *PR1* between RNA-seq and RASL-seq from the same RNA set.

RASL-seq analysis revealed the differential expression profile of defense genes in the *dcl* mutants

Once the Illumina reads were separated based on the barcode sequences at each end, they sequence identity was determined by using a conventional alignment program, Bowtie (Langmead & Salzberg, 2012) against a reference sequence containing all the target genes. We closely followed the RASL-seq informatics procedure as described in (Larman et al., 2014). Relative expression of target genes are then analyzed by comparing the reads

between the target gene and an average of 7 housekeeping genes that were shown to display consistent and constitutive expression (Czechowski et al., 2005).

The RASL-seq analysis revealed that the expression of many defense genes in *dcl1-7* at 6 hpi with avirulent *Pst* were substantially low relative to WT (Fig. 3.8). This reduced expression in *dcl1-7* is consistent with compromised induction of *PR* genes and the susceptibility of *dcl1-7* to avirulent *Pst* (Fig. 2.5A); this compromise induction occurred more to early defense genes (cluster 3) than late defense genes (cluster 6). A list of selected defense genes showing the difference between WT and *dcl1* at 6 hpi with avirulent *Pst* is shown in fig. 3.10. In addition, expression of many genes in cluster 3 were also transcriptionally elevated in *dcl2-1* and *dcl3-1* in response to avirulent *Pst* infection (Fig. 3.8). Some of the cluster 3 genes in *dcl4* were marginally induced at the basal level. A list of selected defense genes showing the difference between WT and *dcl3* at their basal level is shown in fig. 3.9. As discussed in the previous chapter, it has been proposed that an sRNA may regulate the basal expression of NB-LRR genes to keep plants from expressing defense genes constitutively (Shivaprasad et al., 2012, Fei et al., 2016). To check this possibility, we analyzed two *R* genes, *RMG1* and *SNCI*, that were shown to be transcriptionally induced by infection (Yu et al., 2013, Li et al., 2010) and found that their expression level is high in the *dcl* mutants as compared to WT (Appendix 3.2). Consistent with this observation, *dcl4-2* was shown to have high expression of *R* genes (Appendix 3.2) since the phasi-RNAs (phased, secondary, siRNA) which control the *R* gene expression are processed by DCL4 (Fei et al., 2013).

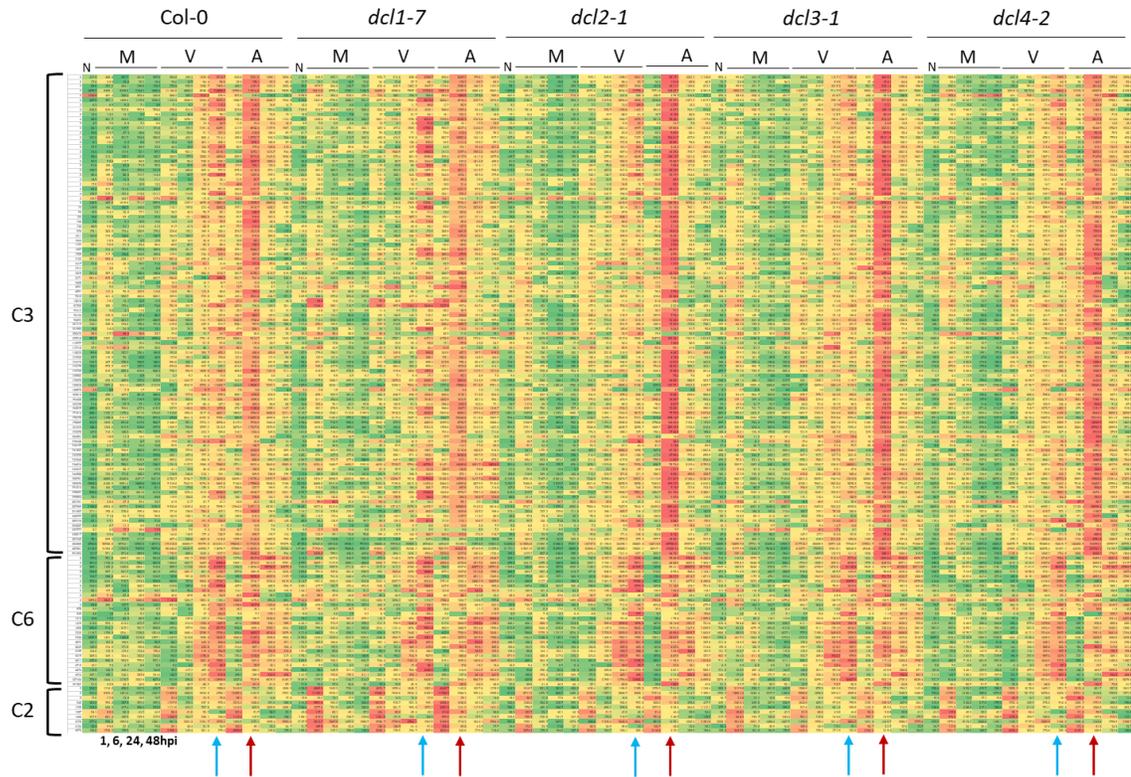


Fig. 3.8. RASL-seq showed that the expression kinetics of defense genes is altered in the *dcl* mutants. RASL-seq raw read count was normalized using the read count of 8 house-keeping genes. The figure shows mean normalized RASL-seq (of three biological replicates) read count of 103 cluster 3 genes, 28 cluster 6 genes and 10 cluster 2 genes. Cluster 3 and cluster 6 genes are specifically induced by *Pst* infection (as shown by the blue and red arrows for virulent *Pst* at 48 hpi, and avirulent *Pst* at 6 hpi). Note that genes in cluster 3 are early responsive genes (peaking at 6 hpi with avirulent *Pst*) while genes in cluster 6 are late responsive genes (peaking at 24 hpi with avirulent *Pst*). Each row represents one gene and each column represents the time and treatment. Green to red color gradient indicates low to high expression, respectively. Each treatment was indicated on the top together with its genetic background; N=Naïve, M=Mock, V= virulent *Pst* and A= avirulent *Pst*. Except of the Naïve treatment, each treatment has four time points; 1, 6, 24 and 48 hpi. These time points were shown at the first occurrence for the brevity of figure. The red arrow indicates avirulent *Pst* treatment at 6hpi and blue arrow indicates virulent *Pst* treatment at 48hpi, respectively.

Gene_ID	D3.N/C.N	D3.M.1/C.M.1	D3.M.6/C.M.6	D3.M.24/C.M.24	D3.M.48/C.M.48	D3.V.1/C.V.1	D3.V.6/C.V.6	D3.V.24/C.V.24	D3.V.48/C.V.48	D3.A.1/C.A.1	D3.A.6/C.A.6	D3.A.24/C.A.24	D3.A.48/C.A.48	P-value D3N vs CN
AT2G14610	6.55	1.63	2.67	0.49	1.40	1.87	3.83	1.10	1.01	2.40	0.80	0.68	0.87	0.0002
AT3G28510	1.84	0.61	1.39	0.75	0.91	1.41	2.35	1.35	0.82	1.07	1.03	0.88	0.70	0.0037
AT1G32640	2.33	2.88	1.31	0.95	1.02	1.25	0.97	0.92	1.17	0.96	1.14	1.04	1.09	0.0126
AT3G57260	6.24	1.43	4.01	0.38	0.90	2.55	2.25	0.75	1.31	1.36	0.71	1.16	1.29	0.0253
AT1G75040	4.97	1.09	3.70	0.23	0.99	3.18	3.44	0.92	1.65	2.48	0.69	0.73	1.10	0.0301
AT2G04040	1.79	2.27	1.17	0.87	0.93	1.04	0.91	0.92	0.72	0.72	1.19	0.78	0.99	0.0346
AT3G28540	2.32	0.93	1.40	0.82	1.02	1.30	1.56	1.16	0.94	1.10	0.99	0.95	0.82	0.0384
AT1G02930	2.87	1.52	1.06	1.10	1.04	1.15	0.90	1.11	0.67	0.73	1.04	0.77	1.22	0.0562
AT5G20900	1.22	1.64	1.05	1.02	0.95	0.95	0.95	1.00	0.98	1.01	0.80	0.97	0.98	0.0566
AT1G70700	2.27	2.80	1.32	0.97	1.25	0.79	0.99	0.85	0.90	0.79	1.04	1.30	1.03	0.0599
AT1G21310	3.95	1.66	1.75	0.53	0.92	1.37	1.34	0.97	1.06	1.02	1.17	0.84	0.93	0.0667
AT1G45145	2.00	2.45	1.69	1.04	1.05	1.18	1.18	1.04	0.73	0.75	1.16	1.05	0.81	0.0678
AT1G55210	1.51	1.00	1.54	0.71	1.07	1.19	1.08	0.89	1.17	1.14	0.90	1.11	1.24	0.0701
AT3G07390	3.66	1.35	1.88	0.65	1.21	1.18	3.24	1.50	0.72	0.91	1.49	0.45	1.39	0.0815
AT4G39950	1.90	1.28	1.32	0.86	1.04	0.86	0.93	0.72	0.56	1.09	1.39	0.69	1.14	0.0821
AT3G28600	3.66	2.21	1.64	1.40	1.73	1.05	0.88	1.08	0.71	0.96	1.20	0.75	1.21	0.0826
AT1G02920	2.58	1.65	1.78	1.01	1.14	1.56	1.51	1.94	0.96	0.85	1.53	0.82	1.38	0.0845
AT5G24200	2.51	0.91	1.51	0.64	0.94	1.61	1.21	0.92	0.85	1.05	1.30	0.75	0.95	0.0897
AT2G30770	3.27	2.26	1.71	1.14	1.14	0.77	1.03	1.06	0.66	0.74	1.27	0.80	1.12	0.0903
AT1G51760	2.48	2.44	1.33	1.33	0.93	0.94	0.95	0.83	0.67	0.69	1.00	0.96	0.92	0.0913
AT4G21830	1.27	1.11	1.23	0.82	1.02	0.97	1.18	0.87	0.66	0.78	1.53	1.15	1.08	0.0952
AT5G13320	1.71	1.01	1.03	0.54	0.62	0.83	1.23	1.13	0.65	0.83	1.63	0.76	1.00	0.0955
AT3G13610	3.23	0.70	2.27	1.30	0.84	1.43	2.26	1.32	0.67	1.32	1.05	0.55	1.15	0.0993

Fig. 3.9. Basal level expression of a subset of defense genes is high in *dcl3-1*. Many defense genes showed high basal level expression in *dcl3-1*. Numbers indicate the normalized expression ratio between *dcl3-1* and WT. The column name consists of the following codes: N= naïve, M= mock, V= virulent *Pst*, A= avirulent *Pst*; D3= *dcl3-1*, C= WT(Col-0); 1, 6, 24 and 48 refers to hours post infiltration (e.g. D3.V.6 = *dcl3-1* treated with virulent *Pst* for 6 hours). Student's T-test P-value between *dcl3-1* naïve vs WT naïve comparison is shown on the right. Green to red color gradient indicate low to high expression, respectively.

Gene_ID	D1.N/C.N	D1.M.1/C.M.1	D1.M.6/C.M.6	D1.M.24/C.M.24	D1.M.48/C.M.48	D1.V.1/C.V.1	D1.V.6/C.V.6	D1.V.24/C.V.24	D1.V.48/C.V.48	D1.A.1/C.A.1	D1.A.6/C.A.6	D1.A.24/C.A.24	D1.A.48/C.A.48	P-value CA6 vs D1A6
AT1G02920	1.15	0.48	0.66	0.53	0.51	0.24	0.39	0.30	0.24	0.13	0.11	0.13	0.20	0.0030
AT1G02930	0.99	0.36	0.47	0.60	0.56	0.24	0.45	0.27	0.42	0.25	0.24	0.27	0.37	0.0621
AT2G14610	1.96	0.64	1.05	0.65	2.00	1.46	1.27	0.46	0.58	0.89	0.26	0.69	0.83	0.0178
AT3G28510	1.18	0.52	0.82	0.86	1.11	1.13	0.93	0.47	0.96	0.77	0.28	0.78	1.00	0.0449
AT3G28540	1.05	0.78	0.88	1.10	1.50	1.43	0.83	0.71	1.14	1.15	0.32	0.99	1.09	0.0029
AT2G26400	1.55	0.95	0.74	0.85	0.62	1.72	0.81	0.35	0.91	1.37	0.43	0.71	0.99	0.0163
AT2G13810	1.63	1.85	1.03	0.84	1.19	1.02	1.13	0.31	0.61	0.97	0.44	0.66	0.62	0.0831
AT1G33960	1.01	0.72	0.96	0.77	0.68	0.69	0.50	0.43	1.03	0.70	0.44	0.91	0.87	0.0175
AT2G30550	1.63	0.91	1.75	1.51	0.85	1.02	1.10	0.95	1.49	1.25	0.46	1.49	1.51	0.0498
AT5G01900	1.24	1.60	0.92	0.77	0.88	1.09	0.67	0.63	0.96	1.44	0.49	1.09	0.89	0.0233
AT5G26340	1.73	1.66	1.60	0.86	1.45	0.88	1.57	0.63	0.87	0.63	0.49	0.61	0.99	0.0632
AT5G55450	0.87	0.78	0.73	0.78	0.68	1.41	0.91	0.63	0.65	1.08	0.54	1.08	0.88	0.0469
AT3G03470	0.52	0.39	1.19	1.75	1.11	0.63	1.10	0.63	1.45	1.49	0.43	0.54	1.46	0.0114
AT1G21250	0.83	0.56	0.95	0.68	0.71	1.12	0.82	0.74	0.66	0.77	0.55	0.94	1.06	0.0073
AT1G21310	1.75	0.94	0.96	0.39	0.92	1.06	1.08	0.49	0.86	0.48	0.56	0.39	0.81	0.0361
AT5G24200	1.56	0.57	0.85	1.03	1.13	1.63	0.66	0.62	1.21	0.73	0.57	0.99	1.24	0.0124
AT3G25882	0.83	0.75	1.03	0.85	0.56	0.74	0.77	0.74	1.13	0.74	0.58	0.97	0.95	0.0411
AT2G19190	1.00	1.02	0.99	0.68	0.54	0.75	0.79	0.74	1.11	0.75	0.59	1.08	1.22	0.0172
AT2G04450	1.17	0.76	1.06	0.87	0.84	0.89	0.95	0.59	0.91	0.85	0.60	0.73	0.93	0.0034
AT5G22570	1.25	0.87	0.96	0.93	0.78	1.01	0.86	0.75	1.10	0.86	0.60	1.10	0.94	0.0179
AT3G18250	1.63	1.21	1.18	1.55	1.85	1.36	1.04	0.69	1.30	0.76	0.62	1.02	1.51	0.0271
AT2G04430	1.06	0.83	1.18	0.59	0.52	0.71	0.92	0.69	0.77	0.80	0.63	1.07	1.33	0.0578
AT5G44568	1.06	0.76	0.92	1.03	0.88	0.87	0.87	0.59	0.62	0.88	0.65	0.71	0.91	0.0198
AT2G18660	0.89	1.27	0.80	0.95	1.22	2.32	1.24	1.03	0.98	0.82	0.66	0.95	1.11	0.0893

Fig. 3.10. Induction dynamics of some defense genes is compromised in *dcl1-7*. RASL-seq revealed that a considerable number of defense gene expression in *dcl1-7* especially at 6 hours after avirulent *Pst* infection was compromised. Numbers indicate the normalized expression ratio between *dcl1-7* and WT. The column name consists of the following codes: N= naïve, M= mock, V= virulent *Pst*, A= avirulent *Pst*; D1= *dcl1-7*, C= WT (Col-0); 1, 6, 24 and 48 refers to hours post infiltration (e.g. D1.A.6 = *dcl1-7* treated with avirulent *Pst* for 6 hours). The P-value of D1.A.6 vs C.A.6 comparison is shown on the right (Student's T-test). Green to red color gradient shows low to high expression respectively.

Discussion

Sequencing whole genomes, especially when dealing with a large sample size, remains prohibitively expensive. Meyer et al. reported that ultra-deep amplicon sequencing (up to 49,000X) of suspected target region in the genome led to the finding of a relapse-specific rare cancer causing mutation in the *NT5C2* gene (Meyer et al., 2013), justifying, in some case, the necessity of a large number of samples. Targeted sequencing approaches like RASL-seq have been, therefore, sought to process a large number of samples at reasonable expense. In addition, targeted sequencing approaches can be performed with a low amount of material (Chung et al., 2016). My optimization data indeed suggested that RASL-seq is highly quantitative as a gene such as *PRI* displaying a large range of expression values with a modest input RNA (Fig. 3-7). However, for this assay to be highly quantitative, I found the consistency in primer binding efficiency across samples is critical. For example, the T_m and length of RASL-probes are important in the assay consistency. Due to a high number of probes in RASL-seq (288 genes in this study with 288 acceptor and 288 donor probes), partial complementarity among RASL probes also influences the efficiency of RASL. Thus, to accommodate a wide range of expression for target genes and to make sure that RASL probes are not a limiting factor, 10 nM of each probe (about 60 billion molecules) with no partially complementing sequence in the genome was used during the RASL seq library preparation.

Release of RdDM-TGS occurs in response to *Pst* infection, which was evident by induction of TEs (Fig. 2.4) and downregulation of RdDM-TGS components such as *DCL3* and *MORC1* (Fig. 3.5). In contrast, expression of *DCL1* in WT was shown to be upregulated upon pathogen infection and proposed to positively regulate defense responses through

PTGS of the negative regulators such as *TIR1*, *AFB2*, and *AFB3* (Pelaez & Sanchez, 2013). This observation is consistent with compromised resistance to avirulent *Pst* in *dcl1-7* (Fig. 2.5). On the other hand, enhanced resistance in *dcl1-7* following virulent *Pst* infection seemed to be accompanied by heightened expression of many defense genes at 48 hpi as indicated by blue arrow in Fig. 3.8. Together, this observation suggests that reduction in DCL1 leads to contrasting outcomes depending on the strength of defense responses (Fig. 2.5). DCL1 is reported to be a target of the bacterial silencing suppressor AvrPto (Lionel Navarro, 2008). Virulent *Pst* could therefore hijack the host silencing machinery to suppress host defense responses via suppressing DCL1 as well as other silencing components (Fig. 2.5) although it remains to be found the existence of the additional targets.

Another intriguing possibility is that DCL1 is involved in saRNA (small activating RNA) biogenesis. It has been reported in a mammalian system that small RNAs are not only involved in gene silencing but in gene activation (Li, 2017). Decrease in the expression of some defense genes after avirulent *Pst* infection in *dcl1-7* (Fig. 3.8 and 3.10) may therefore be due to compromised DCL1-mediated production of saRNAs. This is supported by our preliminary analysis showing that many defense genes in cluster 3 and cluster 6 have strong small RNA inducibility (Appendix 3.1) compared to the genes from the other clusters. Note that this analysis is based on an sRNA dataset provided by Dr. Klessig, which was produced from WT infected with *Pst* at 4 hpi. Interestingly, the size of these sRNAs induced by *Pst* was beyond the conventional range (21-24 nt), which is consistent with the earlier saRNA reports (Zhang & Zhang, 2017). However, I did not find strong correlation between compromised induction in *dcl1* and the presence of *Pst*-inducible saRNAs, suggesting that

DCL1-mediated regulation may involve additional factors. We are currently performing additional experiment(s) to support the possible presence of saRNAs under biotic stress.

In contrast to *dcl1*, *dcl2* and *dcl3* showed enhanced expression of cluster 3 defense genes at 6 hpi with avirulent *Pst* (indicated by red arrow in Fig. 3.8). This observation, along with the observation that basal expression of several defense genes is high in *dcl3* (Fig. 3.9), suggests that DCL3 may be involved in the TGS of a subset of defense genes. Interestingly, although the expression of defense genes is higher in *dcl2* and *dcl3* relative to WT when infected with avirulent as well as virulent *Pst*, both *dcl2* and *dcl3* are mildly susceptible to avirulent *Pst* but resistance to virulent *Pst* (Fig. 2.5A). Modestly heightened resistance phenotype of *dcl2* and *dcl3* to virulent *Pst* (Fig. 2.5B) is likely due to the heightened expression of defense genes. However, it is unclear why same heighten expression of defense genes in *dcl2* and *dcl3* resulted in mild susceptibility to avirulent *Pst*.

In this chapter, I optimized the RASL-seq procedure for defense genes in Arabidopsis and confirmed the reliability of this RNA quantitation tool by comparing it with RNA-seq as well as qRT-PCR analysis. Assessing a large number of defense genes in the *dcl* mutants at the multiple time points revealed that all *dcl* mutants display altered dynamics of defense gene induction to various degrees and fashions. *dcl1* which showed significant reduction in resistance to avirulent *Pst* also displayed compromised induction of defense genes. This reduced defense response may be due to the lower activities of positive defense regulators which requires DCL1. Alternatively, DCL1-generated sRNAs may function positively; this possibility is currently being examined. By contrast, *dcl3* showed enhanced basal expression of defense genes as observed in the previous chapter, suggesting that DCL3-generated sRNAs suppresses the expression of defense genes when no stress is present. It

is also notable that a considerable number of defense genes were not significantly changed in the defense gene induction, suggesting that there are multiple factors regulating the dynamics of defense genes. Together with the genome accessibility study uncovering TEs as transcriptional enhancers, my findings with the *dcl* mutants demonstrate that these epigenetic components play an important role in the induction dynamics of defense genes.

CHAPTER IV

MATERIALS AND METHODS

Plant material, bacterial infiltration and analysis of anti-bacterial resistance in Arabidopsis

Plants were grown in soil at 22°C, 60% relative humidity, and a 16 hr light period. Three and half week old plants were syringe-infiltrated with *Pst* (5×10^5 cfu/mL) or 10mM MgCl₂ (mock) and treated leaves were harvested 1 dpi for DNase-Seq, DNase-qPCR and ChIP-Seq described below or at the indicated times for RNA analysis. To analyze anti-bacterial resistance in Arabidopsis, leaves of Arabidopsis plants were infiltrated with 10^5 cfu/mL of *Pst* with or without *AvrRpt2* in 10 mM MgCl₂ using a needleless syringe. Infected leaves were harvested at the given time points after the infiltration, then used for bacterial titer determination as previously described (Kang et al., 2008).

Preparation for DNase-Seq and DNase-qPCR

Around 4.5 g of leaf tissue was harvested in a 50 mL tube and incubated in diethyl ether for 3 min followed by washing three times with cold water. The tissue was homogenized in 5 ml of the homogenization buffer at 4 °C using a T10 Ultra-Turrax homogenizer (IKA). Nuclei were enriched as previously described (Manzara & Gruissem, 1995). In addition, the nucleus pellet was washed five times with the homogenization buffer to remove chloroplasts and *Pst*, followed by a Percoll gradient purification as described (Henfrey & Slater, 1988).

Prepared nuclei were subject to DNase I digestion as described (Hesselberth et al., 2009) with the following modification: After concentration of nuclei was calculated using

hemocytometer under the microscope, 8×10^5 nuclei was incubated with one unit of DNase I (Roche) at 37 °C for 10 min which gave DNA cleavage comparable to that in the yeast DNase-Seq experiment (Hesselberth et al., 2009). DNA was size-fractionated using a sucrose gradient as described (Hesselberth et al., 2009) to obtain DNA ranging from 100-700 bp. DNA was further purified using a Qiagen PCR purification kit. DNA concentration was determined using PicoGreen (Life Technologies) and DNA quality was checked on a 1% agarose gel stained with SYBR Green (Life Technologies) and scanned with a Typhoon Trio imager (GE Healthcare).

Construction and sequencing of DNase I libraries

Each combination of genotype and treatment was prepared in biological triplicates, resulting in a total of 36 independent libraries. DNase I libraries were prepared using Genomic DNA Sample Prep kit (Illumina) according to the manufacturer's protocol. The libraries were sequenced on an Illumina HiSeq 2500 system at Cornell University Life Sciences Core Laboratories Center with single-end, 100 bp mode.

DNase-Seq data and gene ontology analysis

DNase-Seq reads were first aligned to Arabidopsis chloroplast genome using Bowtie (Langmead et al., 2009), allowing up to two mismatches and those aligned were discarded. The resulting reads were then aligned to Arabidopsis genome (TAIR10 release) using Bowtie allowing up to two mismatches and only the best hits were kept. Only reads uniquely mapped (having one single best hit) to the genome were used for further analysis. The uniquely mapped reads from biological replicates were combined and their mapping information was then fed to F-seq (Boyle et al., 2008) to identify DHSs with default

parameters. The identified DHSs from different samples that overlapped were merged into one DHS. If the peak position was covered by more than one type of genomic features, the annotation was determined by a predefined order. The order of genomic features was 1) TSS upstream, 2) TE upstream, 3) UTR, 4) exon or intron, 5) TE, 6) TTS downstream, 7) TE downstream, and 8) Intergenic. Then for each DNase-Seq library, the number of reads mapped to each of the identified DHSs was counted and normalized to reads per million mapped reads (RPM). The raw count information was fed to DESeq (Anders & Huber, 2010) to identify differentially accessible DHSs upon infection by *Pst* in WT and *MORC*-deficient mutants with a cutoff of corrected *p* value < 0.05 and fold change > 2.

To evaluate enrichment in the Gene Ontology categories, known or deduced biological functions of the genes associated with dDHSs were annotated with TAIR (Berardini et al., 2004) as of January 2015. Broad biological categories that are ‘other biological processes’, ‘other cellular processes’ and ‘other metabolic processes’. The raw P values were calculated using the hypergeometric distribution, which were then adjusted for multiple testings using the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995).

ChIP

ChIP was performed as described (Tai et al., 2005) with the following changes: 4.5 g of leaf tissue/sample was crosslinked for 60 min and nuclei were prepared as described above except that the Percoll gradient step was skipped. Magnetic beads conjugated with Protein A/G (Thermo Fisher Scientific) and α -Myc (Abcam) antibodies were used to immunoprecipitate Myc-tagged MORC1 and its interacting chromatin. DNA was purified using Qiagen PCR purification kit. ChIP-Seq libraries were generated using NEBNext ChIP-Seq Library Prep Reagent kit (New England Biolabs) according to the manufacturer’s

instructions. The libraries were sequenced using an Illumina HiSeq 2500 system at the Genomic Sequencing and Analysis Facility of The University of Texas at Austin.

ChIP-Seq data analysis

ChIP-Seq raw reads were first processed to remove adapter and low quality sequences using Trimmomatic (Bolger et al., 2014). For each library, identical reads were then collapsed into a single unique read. The resulting ChIP-Seq reads were aligned to the Arabidopsis genome (TAIR10 release) using Bowtie (Langmead et al., 2009) allowing up to one mismatch, and only the best hits were kept and only reads uniquely mapped (having one single best hit) to the genome were kept. The alignments from different biological replicates were combined using SAMtools (Li et al., 2009), then converted into Browser Extensible Data (BED) format using BEDtools (Quinlan & Hall, 2010). The read mapping coordinates in BED format were fed into SICER (Zang et al., 2009; redundancy threshold at 1, window size at 600, fragment size at 150, effective genome fraction at 0.96, and gap size at 1800.) to identify wide peaks from the two pairs of libraries i) *Myc-gMORCI*^{mock} and WT^{mock} 2) *Myc-gMORCI*^{Pst} and WT^{Pst}. SICER were run at the following parameters: redundancy threshold at 1, window size at 600, fragment size at 150, effective genome fraction at 0.96, and gap size at 1800. SICER first identified significant peaks in each of the two pairs and then merge the two sets of peaks. Then for each merged peak, its level in *Pst* was compared with that in mock to determine the significance of changes. Significantly differential peaks were identified when the fold change of peak levels between *Pst* and mock should be larger than 1.1 and the FDR (false discovery rate) should be less than 0.001.

Then for each ChIP-Seq peak, the number of mapped reads in each sample was counted and normalized to reads per kilobase per million mapped reads (RPKM). Raw counts were then fed to edgeR (Robinson et al., 2010) to identify differential peaks between WT and *Myc-gMORC1* plants under mock- or *Pst*-inoculation. ChIP-Seq peaks with fold changes of peak levels greater than 1.1 and p-value less than 0.01 were identified as significant and MORC1-associated.

qPCR

qPCR was used to quantify the DNA templates prepared from DNase I and ChIP experiments. Maxima SYBR Green qPCR Master Mix (Thermo Fisher Scientific) was used with initial incubation at 50 °C for 2 min and at 95 °C for 10 min followed by 40 cycles of 95 °C for 25 sec and 60 °C for 1 min. Primer set used for qPCR was listed in Appendix 1.3. Level of target DNA was calculated from the difference of threshold cycle (Ct) values between reference and target gene (Schmittgen & Livak, 2008). The *TIP41-like* gene (Czechowski et al., 2005) was used as the reference for DNase I-qPCR.

Northern blot and qRT-PCR analysis

RNA was extracted using TRIzol (Life Technologies) following the manufacturer's protocol and quantified by NanoDrop 1000 (Thermo Fisher Scientific). Northern analysis was performed with the *PR-I* probe as described (Kang & Klessig, 2005)^[103]. (Kang et al., 2010)^{Kang et al., 2010} with the ^[103] with the primers listed in Appendix 1.3 (Czechowski et al., 2005)

McrBC-qPCR

Overnight digestion of 500 ng of genomic DNA was performed with and without *McrBC* as suggested by the manufacturer (NEB). The digested sample was diluted 10-fold and used for qPCR which is described above.

DNA constructs for complementation

Constructs complementing *morc1-1* were previously described (Kang et al., 2008). *NheI*-digested BAC DNA F23E13 carrying genomic MORC3 was cloned into pART27. This construct was used for the complementation of *morc1-1*.

RNA isolation for RASL-seq

Total RNA from the leaf tissue was extracted using PureLink RNA mini kit from Ambion and 1 micro gram was used for the RASL-seq library preparation without any further treatment. The sample information used for RASL-seq is shown in table 3.2.

Table 3.2. RASL-seq sample detail. The number of samples per biological replicate comes from table 3.1. Cohort 1 is an independently prepared batch of Col-0 wild type for RNA-seq (table3.1). Cohort 1 was used to perform the correlation analysis between RNA-seq and RASL-seq (Fig. 3.3). Cohort 2 which consists of 195 samples (39X5) was used for the transcript analysis in DCL mutants. Before performing RASL-seq, PR gene qRT-PCR was run on these samples to check the samples quality and have data for the correlation analysis between qRT-PCR and RASL-seq (Fig. 3.3).

Genotype		Samples per biological replicate	Number of biological replicates	Samples per genotype
Cohort 1	Col-0 wild type (RNA-seq)	13	3	39
Cohort 2	Col-0 wild type (qRT-PCR tested)	13	3	39
	<i>dcl1-7</i> (qRT-PCR tested)	13	3	39
	<i>dcl2-1</i> (qRT-PCR tested)	13	3	39
	<i>dcl3-1</i> (qRT-PCR tested)	13	3	39
	<i>dcl4-2</i> (qRT-PCR tested)	13	3	39
Grand total (Number of samples)				234

Probe design for RASL-seq

The NCBI primer BLAST tool was used for the probe design. Minimum, optimum and maximum Tm was set at 60, 68 and 85 respectively. The GC percentage allowed was set at 30 and 70 minimum and maximum respectively. The maximum length of the primer designed using this tool is 36 bases, however, since for the RASL-seq 40 bases are required, 4 bases were manually added towards 3' end of the sequence. The 40 bases were split in half and used as donor (with 5' phosphate) and acceptor (with two 3' ribose) probes. To be able to use the PCR product library ready for sequencing, 17 base adapter sequence was added to each probe making the total length of each probe 37 bases. The adapter sequence and the barcode primer sequence information was obtained from (Credle et al., 2017) and is also described in table 3.3.

Table 3.3. Sequence of the adapter, barcode and sequencing primers used in the RASL-seq

Name	5'-3'
Donor probe adapter	AGATCGGAAGAGCACAC
Acceptor probe adapter	GGAGCTGTCGTTCACTC
P5 barcode primer	AATGATACGGCGACCACCGAGATCTACACBBBBBBBACACTCTTCCGATCTGGAGCTGTCGTTCACTC
P7 barcode primer	CAAGCAGAAGACGGCATAACGAGATBBBBBBBGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Custom_read1 sequencing primer	ACACTCTTCCGATCTGGAGCTGTCGTTCACTC
Read2 sequencing primer	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Index1 sequencing primer	GATCGGAAGAGCACACGTCTGAACTCCAGTCAC

RASL-seq probe mix preparation

All the probes are mixed together in deionized water to a final concentration of 10nM. 10ml of this mix was prepared and stored as 0.5ml aliquots in 1.5ml Eppendorf tubes at -20°C. For high reproducibility, these probe mix aliquots were thawed and used for all the experiments. In this study, a probe mix containing oligo pairs from 288 genes (table 3.4) was prepared.

Table 3.4. List of the criteria and number of genes used in the RASL-seq. Def_ann= defense annotated genes, RdDM= genes involved in the RdDM pathway, TEs= transposable elements, Hormone= genes involved in the Auxin, Ethylene and Jasmonic acid response.

Criteria	Number of genes
Cluster 2 genes from RNA-seq data (C2)	10
Cluster 3 genes from RNA-seq data (C3)	103
Cluster 4 genes from RNA-seq data (C4)	10
Cluster 6 genes from RNA-seq data (C6)	28
Cluster 9 genes from RNA-seq data (C9)	10
Def_ann	15
RdDM	20
TEs	20
House keeping genes	9
Pathogen induced TEs	27
Hormone	36
Total	288

Preparation of oligo-dT coated streptavidin beads

For one RASL reaction, 3ul MagnaBind streptavidin bead slurry (Thermo Scientific; 21344) was washed three times in 6ul 1x B&W buffer (5 mM Tris-HCl (pH 7.5), 0.5 mM EDTA, 1 M NaCl) followed by two times washing with solution A (DEPC-treated 0.1 M NaOH, DEPC-treated 0.05 M NaCl) and twice with the solution B (DEPC-treated 0.1 M NaCl). Beads were then resuspended in 9.6ul 1x B&W buffer followed by the addition of 0.4ul biotinylated oligo dT probe (Promega; PR-Z5261, 50pmol/ μ L, to get final concentration of 5uM). This mix was incubated at a shaking incubator for 30min at room temperature to allow for the binding of biotinylated oligo dT with streptavidin beads. After 30min, the beads were washed twice with 1x B&W and once with 4x SSC (0.6M NaCl, 60mM Sodium Citrate pH7.0) to remove unbound oligo dT. Finally, beads were resuspended in 10ul 4x SSC.

Annealing, Selection and Ligation

For annealing and selection (40ul per RASL reaction), in PCR tubes/plate, 1ug total RNA in 20ul volume was mixed with 10ul biotinylated oligo-dT-streptavidin coated beads (in 4XSSC prepared in the step above) and 10ul 10nM probe mix. All three components were mixed well by pipetting and incubated in thermo cycler at 70°C for 10 min followed by incubation at 45°C for 30min. After the incubation, thermo cycler was set up to hold the temperature at 30°C. To minimize the background annealing, the delay between addition of probes and beads and moving tubes to 70°C was avoided.

After annealing, the beads were washed twice with 50ul washing buffer (20 mM Tris-HCl pH 7.5, 0.1 M NaCl) and once with 20ul 1x Rnl2 (T4 dsRNA Ligase2, NEB; M0239L)

buffer by re-suspending the beads completely (washing is important to remove un-annealed probes and RNA thus reducing the background ligation). The beads were then resuspended in 1x ligase buffer containing 5U of Rnl2. After mixing well by pipetting the solution was incubated at 37°C for 30 min. After ligation, the tubes were placed on the magnet to remove all the supernatant. Finally, the beads were resuspended in 10ul deionized water.

PCR (Barcoding) and band isolation

From 10ul resuspended beads in the step above, 2ul was used as template for the PCR reaction. 15 cycles of PCR were run using P5 and P7 barcode primers (Table 3.3) and Herculase II polymerase (Agilent Technologies; NC9390548). After PCR, all the reactions/libraries were pooled (2ul each) together into one tube and ran on a 1.5% agarose gel to band isolate the size of interest (expected size 176bp). The library was quantified using TapeStation 2200 (Agilent).

Sequencing

12pmol library was used for sequencing on Illumina MiSeq V3 flow cell. The de-multiplexing was done by the Illumina BaseSpace based on dual index information provided to the sequencer in the sample sheet.

APPENDIX SECTION

Appendix 1.1. Number of DNase I hypersensitive sites (DHSs) relative to genes and TEs.

Region	No. DHSs
1-200 upstream of TE	740
201-1000 upstream of TE	1,992
TE Body	263
1-200 downstream of TE	138
201-1000 downstream of TE	386
1-200 upstream of TSS	8,017
201-1000 upstream of TSS	7,657
5'-UTR	535
Exon	1,418
Intron	1,445
3'-UTR	1,933
1-200 downstream of TTS	1,721
200-1000 downstream of TTS	1,614
Intergenic	1,591
Total	29,450

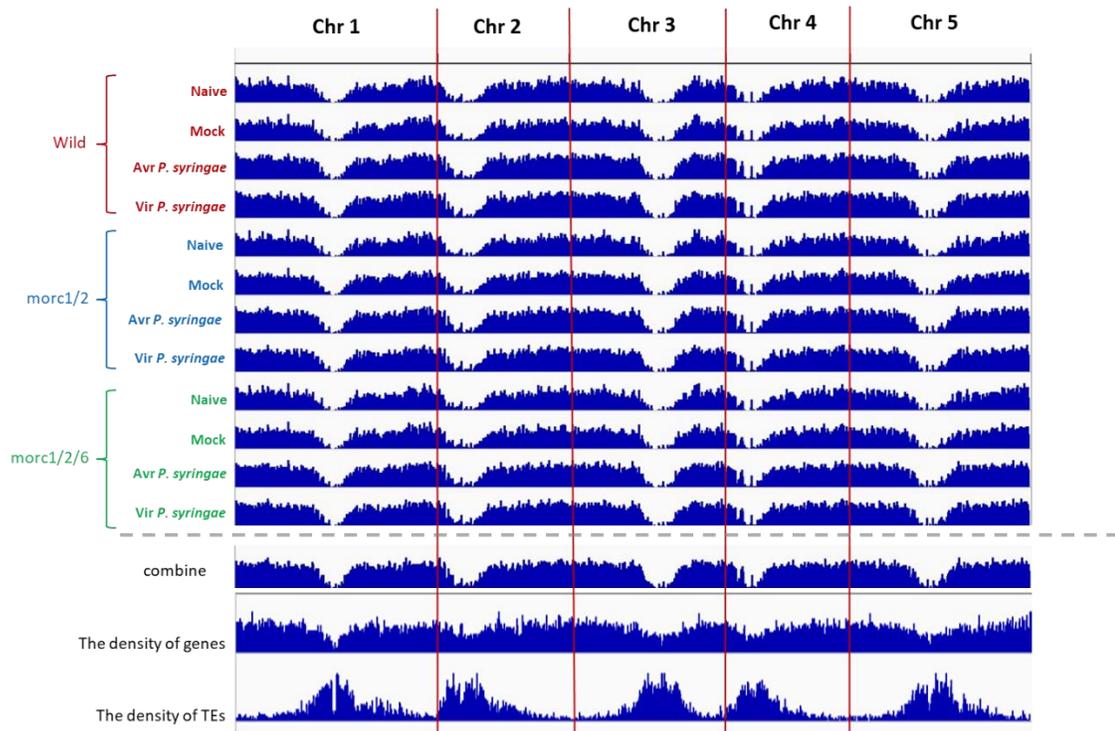
Appendix 1.2. Summary of differential DNase I hypersensitive sites (dDHSs)

Comparison	No. of dDHSs
WT ^{naive} vs WT ^{mock}	2
WT ^{mock} vs WT ^{AvrPst}	102
WT ^{mock} vs WT ^{Pst}	423
WT ^{AvrPst} vs WT ^{Pst}	7
morc1/2 ^{naive} vs morc1/2 ^{mock}	17
morc1/2 ^{mock} vs morc1/2 ^{AvrPst}	296
morc1/2 ^{mock} vs morc1/2 ^{Pst}	577
morc1/2 ^{AvrPst} vs morc1/2 ^{Pst}	12
morc1/2/6 ^{naive} vs morc1/2/6 ^{mock}	2
morc1/2/6 ^{mock} vs morc1/2/6 ^{AvrPst}	231
morc1/2/6 ^{mock} vs morc1/2/6 ^{Pst}	486
morc1/2/6 ^{AvrPst} vs morc1/2/6 ^{Pst}	8
WT ^{naive} vs morc1/2 ^{naive}	59
WT ^{naive} vs morc1/2/6 ^{naive}	81
morc1/2 ^{naive} vs morc1/2/6 ^{naive}	1
WT ^{mock} vs morc1/2 ^{mock}	89
WT ^{mock} vs morc1/2/6 ^{mock}	80
morc1/2 ^{mock} vs morc1/2/6 ^{mock}	2
WT ^{AvrPst} vs morc1/2 ^{AvrPst}	83
WT ^{AvrPst} vs morc1/2/6 ^{AvrPst}	84
morc1/2 ^{AvrPst} vs morc1/2/6 ^{AvrPst}	1
WT ^{Pst} vs morc1/2 ^{Pst}	118
WT ^{Pst} vs morc1/2/6 ^{Pst}	116
morc1/2 ^{Pst} vs morc1/2/6 ^{Pst}	5

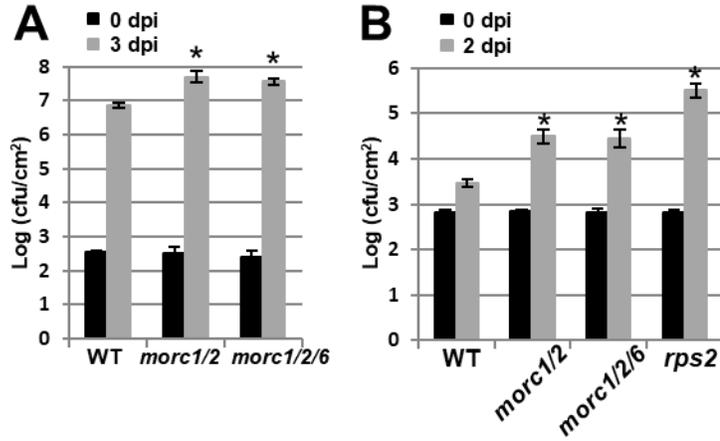
Appendix 1.3. List of primers used in qPCR.

DNase I-qPCR primers		qRT PCR primers		McrBC qPCR primers	
Primer name	Sequence (5'-3')	Primer name	Sequence (5'-3')	Primer name	Sequence (5'-3')
AP8539a-F	tcatgtcattcagttgtttgt	Tip41-Like-R-q	ggatacccttcgagatagagac	At2g06050_qPCR-F	gcctcttcatgagtaacagaa
AP8539a-R	actgtccgatacgattttct	Tip41-Like-F-q	gcgattttggctgagagttgat	At2g06050_qPCR-R	cccattgtgagaacaaatagaga
AP8539b-F	accttttcaacaatctattaacaa	RMG1-R-q	ctaagcggctgaactcactcc	At3g44300_qPCR-F	atgcatgagttggtttctct
AP8539b-R	aatctaatagaaaatgccatctgg	RMG1-F-q	cagttaccagaagaagggctaaagc	At3g44300_qPCR-R	cgccgacataccttcttattcac
AP10313-F	atagagaattgagaaggtgaaa	PR1-R_QRT	ccaccattgttacacctcaattt	At2g14610_qPCR-F	aggtaaacaacgtttcaatatttcaaa
AP10313-R	ttctaattctaaaatgccaatga	PR1-F_QRT	aaaacttagctgggtagcgg	At2g14610_qPCR-R	tttcaaaaggagacaactatc
AP7789-F	ttgatagaaaaccacttggaat	PR2-R_QRT	tgtaagagccacaactgctcc	At4g19230_qPCR-F	tactcaactcagacaaactatg
AP7789-R	aagttgtgattgtgagtatta	PR2-F_QRT	atcaaggagcttagcctcac	At4g19230_qPCR-R	ctgcttctcatgatatacaagaaaa
AP10364-F	gaaagaaaagtgtgaaaatgtcg	PR5-R_QRT	gaagcactggagtaattc	At2g27690_qPCR-F	agctttattccttaattttactacgat
AP10364-R	ttttgatagaaggcagattg	PR5-F_QRT	ctcttctcgtgttcatcac	At2g27690_qPCR-R	ttcgacaagccaaaataaaaatg
AP8555-F	ttttctattcaacacctccaca	AT1G33960_qRT_F	ggcaatgagcagagatgatggaga	At4g16760_qPCR-F	gtcagattagctagctagctgtggaag
AP8555-R	gaatttggggtaggggtgtt	AT1G33960_qRT_R	cttcattcagacagcatcacg	At4g16760_qPCR-R	tggtaggtaaataaaatagagcaaac
AP8455-F	acaatcacatgtaccattgatcca	AT1G44350_qRT_F	gtatgagactagtaggcttatta	At2g13810_qPCR-F	ttctaatagaacaacattgcagataaa
AP8455-R	atccttattctctctaccg	AT1G44350_qRT_R	ttccattcaactgcttctctg	At2g13810_qPCR-R	aaaccgacgtgcgaaatcctatat
AP3801-F	tgtttgctactttctcattttgta	AT1G80820_qRT_F	tgaccgacgatcccagacaaa		
AP3801-R	ccaaaactgacagattacattatt	AT1G80820_qRT_R	atggcttgagtgctcaggttaggg		
AP15711-F	gaaaaatcatgggctgacatg	AT2G06050_qRT_F	gatccgaggggtcccacat		
AP15711-R	ttgtgtttctctagaagtgta	AT2G06050_qRT_R	ctccattaggttgatacactgc		
AP19209-F	gaaaagataaacaaccacaaaa	AT2G13810_qRT_F	gatatttctcgtctcagctactc		
AP19209-R	caaaaacatgaaatcaagaacaa	AT2G13810_qRT_R	acattttgttactcttggttttt		
AP4254-F	acatggaaacccaacaaaaatca	AT2G20010_qRT_F	cagcaagaggtatggaatccgaga		
AP4254-R	ttttcttactaccaccacaacc	AT2G20010_qRT_R	cgcgagccgaggatga		
AP8432-F	aagagttggataggttgagaaaa	AT2G27690_qRT_F	tcccgaacaaccgagacact		
AP8432-R	cgttaaagcatgaagtgcacaaat	AT2G27690_qRT_R	tggtcagttagagaacaagactc		
		AT3G26830_qRT_F	caaaggaatgatctcggaca		
		AT3G26830_qRT_R	tggttagatcttctctttgatt		
		AT3G44300_qRT_F	ctttgtacgccaaggcattg		
		AT3G44300_qRT_R	gatcagggaatctttacgaaggc		
		AT4G16760_qRT_F	tgcaattacaggtgctcgattcc		
		AT4G16760_qRT_R	aaagcttccagtacaacatcaggg		
		AT4G19230_qRT_F	ttcatcaagattcgaggtggc		
		AT4G19230_qRT_R	cgctcgtcgtccaacaat		
		AT5G38710_qRT_F	actcgggtaactagggcaagaa		
		AT5G38710_qRT_R	gcagttatcaaccggccgtatg		

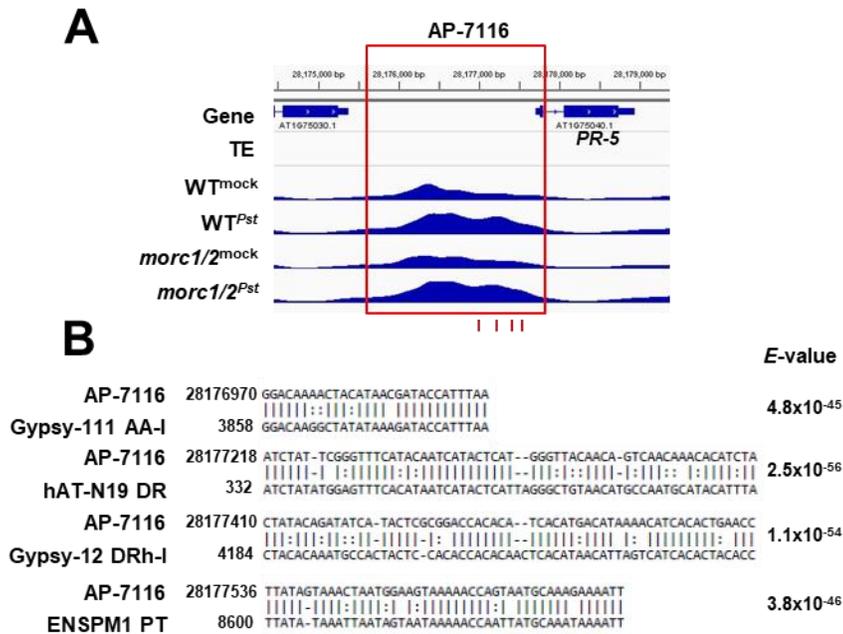
Appendix 1.4. Distribution of the DNase I hypersensitive sites (DHSs) across the Arabidopsis genome.



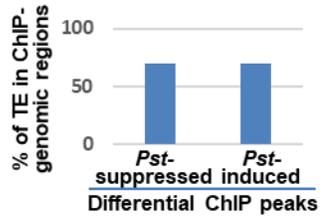
Appendix 1.5. Susceptibility to avirulent and virulent *Pst* in *morc1/2/6* and *morc1/2* mutants. Bacterial growth in the leaves of WT, *morc1/2* and *morc1/2/6* plants was measured after inoculation with virulent *Pst* (10^5 cfu/ml) for 3 days (a) and avirulent *Pst* carrying *avrRpt2* (10^5 cfu/ml) for 2 days (b). Data are the mean \pm SD (n = 4). Statistical significance was determined using a student t test: $P < 0.05$.



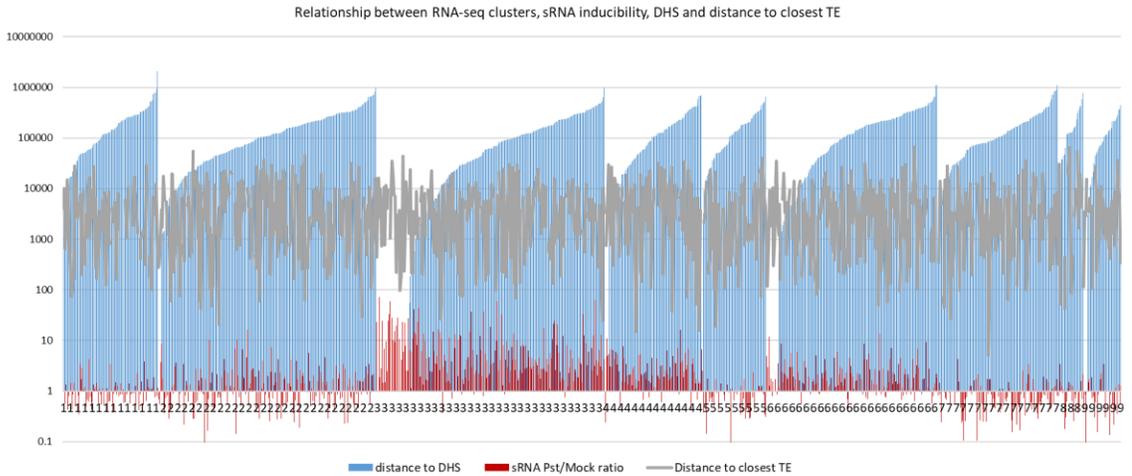
Appendix 1.6. The infection-induced dDHS, AP-7116, in the *PR-5* promoter contains TE-like nucleotide sequences. The AP-7116 sequence (2,409 bp) highlighted by a red box was BLASTed against Repbase (<http://www.girinst.org/replibase/index.html>; total 109,617,951 bp in size) to identify putative TE-like sequences using default BLASTN parameters. **(A)** Genes and TEs are presented in the top two tracks. DNase-Seq read densities of the dDHS AP-7116 in WT^{mock}, WT^{Pst}, *morc1/2*^{mock}, and *morc1/2*^{Pst} are presented in the next four tracks. Note that the presented region does not contain an annotated TE as per the TAIR database. The position of sequences homologous to TEs is indicated by red lines. **(B)** Aligned sequences between AP-7116 and the TEs indicated in **(A)** are presented. The *E*-value is indicated, following the sequence alignment.



Appendix 1.7. MORC1-interacting genomic regions are predominantly associated with TEs.
All genomic regions representing the differential ChIP-Seq peaks as shown in Fig. 8A were analyzed for sequences associated with genes or TEs. The y-axis indicates the percentage of TEs.



Appendix 3.1. Relationship between RNA-seq clusters, sRNA inducibility after *Pst* infection, DHS and distance to the nearest TE. X-axis left to right is all 1038 genes from the RNA-seq clustering analysis from cluster 1 to cluster 9. Genes which show robust response to the pathogen infection (cluster 3, and 6) as shown in RNA-seq data in fig. 3.5, also have very high inducibility of the sRNA (shown as *Pst*/mock ratio in the red lines). While, cluster 2 genes which respond to the touch or genes from cluster 1,4, 5, 7 and 8 which respond to the mock treatments as well, do not show consistent high inducibility of the sRNA. Note that there is no correlation between the sRNA inducibility and distance to DNase I hypersensitive sites (DHS, shown by the blue line). And, there is no correlation between the sRNA inducibility and distance to the nearest TE (shown in grey line) supporting our claim that these sRNAs are involved in enhancing the defense response via RNAa.



Appendix 3.2. High expression of SNC1 and RMG1 R-genes in *dcl3-1* and *dcl4-2*. Using RASL-seq we found that SNC1 (Suppressor of *npr1-1*, Constitutive 1), one of the TIR-NB-LRR type R gene, and RMG1 (Resistance Methylated Gene 1), also a TIR-NB-LRR type R gene display high level of expression in *dcl3-1* and *dcl4-2* mutants. Green to red color gradient represents low to high expression respectively. The numbers indicate RASL-seq normalized expression value. N=Naïve, M=Mock, V=*Pst* DC3000 and A= *Pst* DC3000 AvrRpt2 treatment. 1, 6, 24 and 48 indicate hours post infiltration (hpi).

	N	M1	M6	M24	M48	V1	V6	V24	V48	A1	A6	A24	A48	
SNC1	WT	36.5	33.0	45.2	38.5	49.5	37.5	25.3	53.9	78.4	32.5	79.6	67.3	72.6
	<i>dcl1-7</i>	49.1	56.2	44.2	45.7	41.9	52.3	36.5	66.8	86.8	45.7	81.7	82.9	66.6
	<i>dcl2-1</i>	46.0	29.6	48.6	45.9	46.6	46.6	44.3	78.9	85.5	22.0	94.5	57.7	56.2
	<i>dcl3-1</i>	36.7	51.8	45.4	40.6	44.3	50.7	32.8	73.6	79.7	35.1	106.2	78.7	49.3
	<i>dcl4-2</i>	42.9	44.5	32.5	46.0	45.1	24.6	35.4	76.2	80.6	35.4	99.6	77.9	60.2
RMG1	WT	43.3	56.0	33.9	41.0	40.2	48.8	27.3	49.5	52.9	68.7	110.5	34.1	24.3
	<i>dcl1-7</i>	51.9	42.2	40.4	45.0	41.4	44.5	35.4	38.4	72.2	48.4	148.0	52.6	40.7
	<i>dcl2-1</i>	45.1	63.3	36.0	48.5	36.5	55.3	33.1	50.5	47.7	51.1	137.4	35.5	29.8
	<i>dcl3-1</i>	37.1	40.8	35.5	41.9	36.4	47.8	41.1	53.7	54.1	65.1	184.0	40.1	38.6
	<i>dcl4-2</i>	43.3	61.1	33.3	38.8	40.9	62.7	33.3	39.1	47.6	73.4	190.3	43.5	37.8

REFERENCES

- Agorio A, Vera P, 2007. ARGONAUTE4 is required for resistance to *Pseudomonas syringae* in *Arabidopsis*. *Plant Cell* 19, 3778-90.
- Alvarez-Venegas, 2007. Epigenetic control of a transcription factor at the cross section of two antagonistic pathways. *Epigenetics : official journal of the DNA Methylation Society* 106.
- Alvarez-Venegas R, Abdallat AA, Guo M, Alfano JR, Avramova Z, 2007. Epigenetic control of a transcription factor at the cross section of two antagonistic pathways. *Epigenetics : official journal of the DNA Methylation Society* 2, 106-13.
- Anders S, Huber W, 2010. Differential expression analysis for sequence count data. *Genome Biol* 11, R106.
- Archacki R, Yatusovich R, Buszewicz D, *et al.*, 2017. Arabidopsis SWI/SNF chromatin remodeling complex binds both promoters and terminators to regulate gene expression. *Nucleic Acids Res* 45, 3116-29.
- Au PCK, Dennis ES, Wang MB, 2017. Analysis of Argonaute 4-Associated Long Non-Coding RNA in *Arabidopsis thaliana* Sheds Novel Insights into Gene Regulation through RNA-Directed DNA Methylation. *Genes (Basel)* 8.
- Barsh GS, Tittel-Elmer M, Bucher E, *et al.*, 2010. Stress-Induced Activation of Heterochromatic Transcription. *PLoS Genetics* 6, e1001175.
- Bender J, 2004. Chromatin-based silencing mechanisms. *Curr Opin Plant Biol* 7, 521-6.
- Benjamini Y, Hochberg Y, 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological* 57, 289-300.
- Berardini TZ, Mundodi S, Reiser L, *et al.*, 2004. Functional annotation of the *Arabidopsis* genome using controlled vocabularies. *Plant Physiol* 135, 745-55.
- Berr A, Menard R, Heitz T, Shen WH, 2012. Chromatin modification and remodelling: a regulatory landscape for the control of *Arabidopsis* defence responses upon pathogen attack. *Cell Microbiol* 14, 829-39.
- Berr A, Shafiq S, Shen W-H, 2011a. Histone modifications in transcriptional activation during plant development. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1809, 567-76.
- Berr A, Shafiq S, Shen WH, 2011b. Histone modifications in transcriptional activation during plant development. *Biochim Biophys Acta* 1809, 567-76.

- Bolger AM, Lohse M, Usadel B, 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114-20.
- Bologna NG, Voinnet O, 2014. The diversity, biogenesis, and activities of endogenous silencing small RNAs in Arabidopsis. *Annu Rev Plant Biol* 65, 473-503.
- Bordiya Y, Zheng Y, Nam JC, *et al.*, 2016. Pathogen Infection and MORC Proteins Affect Chromatin Accessibility of Transposable Elements and Expression of Their Proximal Genes in Arabidopsis. *Mol Plant Microbe Interact* 29, 674-87.
- Borges F, Martienssen RA, 2015. The expanding world of small RNAs in plants. *Nat Rev Mol Cell Biol* 16, 727-41.
- Boyko A, Kovalchuk I, 2008. Epigenetic control of plant stress response. *Environ Mol Mutagen* 49, 61-72.
- Boyko A, Kovalchuk I, 2011. Genetic and epigenetic effects of plant-pathogen interactions: an evolutionary perspective. *Mol Plant* 4, 1014-23.
- Boyle AP, Davis S, Shulha HP, *et al.*, 2008. High-resolution mapping and characterization of open chromatin across the genome. *Cell* 132, 311-22.
- Brabbs TR, He Z, Hogg K, *et al.*, 2013. The stochastic silencing phenotype of Arabidopsis morc6 mutants reveals a role in efficient RNA-directed DNA methylation. *Plant J* 75, 836-46.
- Cheng C, Gao X, Feng B, Sheen J, Shan L, He P, 2013. Plant immune response to pathogens differs with changing temperatures. *Nat Commun* 4, 2530.
- Chung J, Son DS, Jeon HJ, *et al.*, 2016. The minimal amount of starting DNA for Agilent's hybrid capture-based targeted massively parallel sequencing. *Sci Rep* 6, 26732.
- Conrath U, 2011. Molecular aspects of defence priming. *Trends Plant Sci* 16, 524-31.
- Credle JJ, Itoh CY, Yuan T, *et al.*, 2017. Multiplexed analysis of fixed tissue RNA using Ligation in situ Hybridization. *Nucleic Acids Res* 45, e128.
- Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible WR, 2005. Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. *Plant Physiol* 139, 5-17.
- Dangl JL, Palma K, Thorgrimsen S, *et al.*, 2010. Autoimmunity in Arabidopsis acd11 Is Mediated by Epigenetic Regulation of an Immune Receptor. *PLoS Pathogens* 6, e1001137.

- De-La-Pena C, Rangel-Cano A, Alvarez-Venegas R, 2012. Regulation of disease-responsive genes mediated by epigenetic factors: interaction of Arabidopsis-Pseudomonas. *Mol Plant Pathol* 13, 388-98.
- De-La-Peña C, Rangel-Cano A, Alvarez-Venegas R, 2012. Regulation of disease-responsive genes mediated by epigenetic factors: interaction of Arabidopsis-Pseudomonas. *Molecular Plant Pathology* 13, 388-98.
- Deleris A, Halter T, Navarro L, 2016. DNA Methylation and Demethylation in Plant Immunity. *Annu Rev Phytopathol* 54, 579-603.
- Dharmasiri N, Dharmasiri S, Weijers D, *et al.*, 2005. Plant development is regulated by a family of auxin receptor F box proteins. *Dev Cell* 9, 109-19.
- Dhawan R, Luo H, Foerster AM, *et al.*, 2009a. HISTONE MONOUBIQUITINATION1 interacts with a subunit of the mediator complex and regulates defense against necrotrophic fungal pathogens in Arabidopsis. *Plant Cell* 21, 1000-19.
- Dhawan R, Luo H, Foerster AM, *et al.*, 2009b. HISTONE MONOUBIQUITINATION1 Interacts with a Subunit of the Mediator Complex and Regulates Defense against Necrotrophic Fungal Pathogens in Arabidopsis. *The Plant Cell Online* 21, 1000-19.
- Downen RH, Pelizzola M, Schmitz RJ, *et al.*, 2012. Widespread dynamic DNA methylation in response to biotic stress. *Proc Natl Acad Sci U S A* 109, E2183-91.
- Elgar G, Vavouri T, 2008. Tuning in to the signals: noncoding sequence conservation in vertebrate genomes. *Trends Genet* 24, 344-52.
- Fedoroff NV, 2012. Transposable Elements, Epigenetics, and Genome Evolution. *Science* 338, 758-67.
- Fei Q, Xia R, Meyers BC, 2013. Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks. *Plant Cell* 25, 2400-15.
- Fei Q, Zhang Y, Xia R, Meyers BC, 2016. Small RNAs Add Zing to the Zig-Zag-Zig Model of Plant Defenses. *Mol Plant Microbe Interact* 29, 165-9.
- Feng Y, Zhang Y, Ying C, Wang D, Du C, 2015. Nanopore-based fourth-generation DNA sequencing technology. *Genomics, proteomics & bioinformatics / Beijing Genomics Institute* 13, 4-16.
- Feschotte C, 2008. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet* 9, 397-405.

- Foroushani A, Agrahari R, Docking R, *et al.*, 2017. Large-scale gene network analysis reveals the significance of extracellular matrix pathway and homeobox genes in acute myeloid leukemia: an introduction to the Pigengene package and its applications. *BMC medical genomics* 10, 16.
- Furey TS, 2012. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat Rev Genet* 13, 840-52.
- Geiss GK, Bumgarner RE, Birditt B, *et al.*, 2008. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol* 26, 317-25.
- Girard L, Freeling M, 1999. Regulatory changes as a consequence of transposon insertion. *Developmental Genetics* 25, 291-6.
- Goldbach R, Bucher E, Prins M, 2003. Resistance mechanisms to plant viruses: an overview. *Virus Research* 92, 207-12.
- Grandbastien M-A, 1998. Activation of plant retrotransposons under stress conditions. *Trends in Plant Science* 3, 181-7.
- Grandbastien MA, Audeon C, Bonnivard E, *et al.*, 2005. Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae. *Cytogenet Genome Res* 110, 229-41.
- Groth M, Stroud H, Feng S, *et al.*, 2014. SNF2 chromatin remodeler-family proteins FRG1 and -2 are required for RNA-directed DNA methylation. *Proc Natl Acad Sci U S A* 111, 17666-71.
- Haag JR, Pikaard CS, 2011. Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nat Rev Mol Cell Biol* 12, 483-92.
- Han SK, Wu MF, Cui S, Wagner D, 2015. Roles and activities of chromatin remodeling ATPases in plants. *Plant J* 83, 62-77.
- Henfrey RD, Slater RJ, 1988. Isolation of plant nuclei. *Methods Mol Biol* 4, 447-52.
- Hesselberth JR, Chen X, Zhang Z, *et al.*, 2009. Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nat Methods* 6, 283-9.
- Howard BE, Hu Q, Babaoglu AC, *et al.*, 2013. High-throughput RNA sequencing of pseudomonas-infected Arabidopsis reveals hidden transcriptome complexity and novel splice variants. *PLoS One* 8, e74183.
- Huang J, Yang M, Lu L, Zhang X, 2016. Diverse Functions of Small RNAs in Different Plant-Pathogen Communications. *Frontiers in Microbiology* 7.
- Initiative TaG, 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* 408, 796-815.

- Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, Paszkowski J, 2011a. An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature* 472, 115-9.
- Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, Paszkowski J, 2011b. An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature* 472, 115-9.
- Ito H, Kakutani T, 2014. Control of transposable elements in *Arabidopsis thaliana*. *Chromosome Res* 22, 217-23.
- Iyer LM, Abhiman S, Aravind L, 2008. MutL homologs in restriction-modification systems and the origin of eukaryotic MORC ATPases. *Biology direct* 3, 8.
- John S. Choy SW, Ju Yeon Lee, Song Tan, Steven Chu, Tae-Hee Lee, 2010. DNA Methylation Increases Nucleosome Compaction and Rigidity. *J. AM. CHEM. SOC.* 132, 1782-3.
- Jones JD, Dangl JL, 2006. The plant immune system. *Nature* 444, 323-9.
- Jones KDRaPA, 2000. DNA methylation: past, present and future directions. *Carcinogenesis Oxford* 21, 461-7.
- Jordan IK, Rogozin IB, Glazko GV, Koonin EV, 2003. Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends in genetics : TIG* 19, 68-72.
- Kang HG, Hyong WC, Von Einem S, *et al.*, 2012. CRT1 is a nuclear-translocated MORC endonuclease that participates in multiple levels of plant immunity. *Nat Commun* 3, 1297.
- Kang HG, Klessig DF, 2005. Salicylic acid-inducible *Arabidopsis* CK2-like activity phosphorylates TGA2. *Plant Mol Biol* 57, 541-57.
- Kang HG, Kuhl JC, Kachroo P, Klessig DF, 2008. CRT1, an *Arabidopsis* ATPase that interacts with diverse resistance proteins and modulates disease resistance to turnip crinkle virus. *Cell Host and Microbe* 3, 48-57.
- Kang HG, Oh CS, Sato M, *et al.*, 2010. Endosome-associated CRT1 functions early in resistance gene-mediated defense signaling in *Arabidopsis* and tobacco. *Plant Cell* 22, 918-36.
- Karlic R, Chung HR, Lasserre J, Vlahovicek K, Vingron M, 2010. Histone modification levels are predictive for gene expression. *Proc Natl Acad Sci U S A* 107, 2926-31.

- Katiyar-Agarwal S, Gao S, Vivian-Smith A, Jin H, 2007. A novel class of bacteria-induced small RNAs in Arabidopsis. *Genes Dev* 21, 3123-34.
- Katiyar-Agarwal S, Jin H, 2010. Role of Small RNAs in Host-Microbe Interactions. *Annual Review of Phytopathology* 48, 225-46.
- Kim KC, Lai Z, Fan B, Chen Z, 2008a. Arabidopsis WRKY38 and WRKY62 Transcription Factors Interact with Histone Deacetylase 19 in Basal Defense. *The Plant Cell Online* 20, 2357-71.
- Kim KC, Lai Z, Fan B, Chen Z, 2008b. Arabidopsis WRKY38 and WRKY62 transcription factors interact with histone deacetylase 19 in basal defense. *Plant Cell* 20, 2357-71.
- Kristin D. Kasschau NF, Elisabeth J. Chapman, Christopher M. Sullivan, Jason S. Cumbie, Scott A. Givan, James C. Carrington, 2007. Genome-Wide Profiling and Analysis of Arabidopsis siRNAs. *PLoS Biology* 5, 479-93.
- Langen G, Von Einem S, Koch A, *et al.*, 2014. The compromised recognition of turnip crinkle virus1 subfamily of microrchidia ATPases regulates disease resistance in barley to biotrophic and necrotrophic pathogens. *Plant Physiol* 164, 866-78.
- Langmead B, Salzberg SL, 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357-9.
- Langmead B, Trapnell C, Pop M, Salzberg SL, 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25.
- Larman HB, Scott ER, Wogan M, Oliveira G, Torkamani A, Schultz PG, 2014. Sensitive, multiplex and direct quantification of RNA sequences using a modified RASL assay. *Nucleic Acids Res* 42, 9146-57.
- Law JA, Du J, Hale CJ, *et al.*, 2013. Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature* 498, 385-9.
- Law JA, Jacobsen SE, 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics* 11, 204-20.
- Le TN, Schumann U, Smith NA, *et al.*, 2014. DNA demethylases target promoter transposable elements to positively regulate stress responsive genes in Arabidopsis. *Genome Biol* 15, 458.
- Lewin R, 1986. "Computer genome" is full of junk DNA. *Science* 232, 577-8.
- Lewis LA, Polanski K, De Torres-Zabala M, *et al.*, 2015. Transcriptional Dynamics Driving MAMP-Triggered Immunity and Pathogen Effector-Mediated Immunosuppression in Arabidopsis Leaves Following Infection with *Pseudomonas syringae* pv tomato DC3000. *Plant Cell* 27, 3038-64.

- Li F, Pignatta D, Bendix C, *et al.*, 2012a. MicroRNA regulation of plant innate immune receptors. *Proc Natl Acad Sci U S A* 109, 1790-5.
- Li H, Handsaker B, Wysoker A, *et al.*, 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-9.
- Li H, Qiu J, Fu XD, 2012b. RASL-seq for massively parallel and quantitative analysis of gene expression. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* Chapter 4, Unit 4 13 1-9.
- Li LC, 2017. Small RNA-Guided Transcriptional Gene Activation (RNAa) in Mammalian Cells. *Adv Exp Med Biol* 983, 1-20.
- Li Y, Tessaro MJ, Li X, Zhang Y, 2010. Regulation of the expression of plant resistance gene SNC1 by a protein with a conserved BAT2 domain. *Plant Physiol* 153, 1425-34.
- Lionel Navarro FJ, Kinya Nomura, Sheng Yang He, Olivier Voinnet, 2008. Suppression of the MicroRNA Pathway by Bacterial Effector Proteins. *Science* 321, 964-7.
- Liu J, Jung C, Xu J, *et al.*, 2012. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in Arabidopsis. *Plant Cell* 24, 4333-45.
- Liu ZW, Shao CR, Zhang CJ, *et al.*, 2014. The SET domain proteins SUVH2 and SUVH9 are required for Pol V occupancy at RNA-directed DNA methylation loci. *PLoS Genet* 10, e1003948.
- Lopez A, Ramirez V, Garcia-Andrade J, Flors V, Vera P, 2011. The RNA silencing enzyme RNA polymerase v is required for plant immunity. *PLoS Genet* 7, e1002434.
- Lorkovic ZJ, Naumann U, Matzke AJ, Matzke M, 2012. Involvement of a GHKL ATPase in RNA-directed DNA methylation in Arabidopsis thaliana. *Curr Biol* 22, 933-8.
- Lorković Zdravko j, Naumann U, Matzke Antonius jM, Matzke M, 2012. Involvement of a GHKL ATPase in RNA-Directed DNA Methylation in Arabidopsis thaliana. *Current Biology* 22, 933-8.
- Ma KW, Flores C, Ma W, 2011. Chromatin configuration as a battlefield in plant-bacteria interactions. *Plant Physiol* 157, 535-43.
- Makarevitch I, Waters AJ, West PT, *et al.*, 2015. Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLoS Genet* 11, e1004915.
- Manzara T, Grissem W, 1995. Identification of promoter sequences that interact with DNA-binding proteins. *Methods in Plant Molecular Biology. A Laboratory Course Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY*, 233-60.

- March-Diaz R, Garcia-Dominguez M, Lozano-Juste J, Leon J, Florencio FJ, Reyes JC, 2008. Histone H2A.Z and homologues of components of the SWR1 complex are required to control immunity in Arabidopsis. *Plant J* 53, 475-87.
- Mardis ER, 2013. Next-generation sequencing platforms. *Annu Rev Anal Chem (Palo Alto Calif)* 6, 287-303.
- Martienssen R, Barkan A, Taylor WC, Freeling M, 1990. Somatic heritable switches in the DNA modification of Mu transposable elements monitored with a suppressible mutant in maize. *Genes Dev* 4, 331-43.
- Matzke MA, Kanno T, Matzke AJ, 2015. RNA-Directed DNA Methylation: The Evolution of a Complex Epigenetic Pathway in Flowering Plants. *Annu Rev Plant Biol* 66, 243-67.
- Matzke MA, Mosher RA, 2014. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat Rev Genet* 15, 394-408.
- Metzker ML, 2010. Sequencing technologies - the next generation. *Nature reviews. Genetics* 11, 31-46.
- Meyer JA, Wang J, Hogan LE, *et al.*, 2013. Relapse-specific mutations in NT5C2 in childhood acute lymphoblastic leukemia. *Nat Genet* 45, 290-4.
- Miao-Chih Tsai OM, Yue Wan, Nima Mosammaparast, Jordon K. Wang, Fei Lan, Yang Shi, Eran Segal, Howard Y. Chang, 2010. Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. *Science* 329, 689-93.
- Moissiard G, Bischof S, Husmann D, *et al.*, 2014. Transcriptional gene silencing by Arabidopsis microorchidia homologues involves the formation of heteromers. *Proc Natl Acad Sci U S A* 111, 7474-9.
- Moissiard G, Cokus SJ, Cary J, *et al.*, 2012. MORC family ATPases required for heterochromatin condensation and gene silencing. *Science* 336, 1448-51.
- Moore JW, Loake GJ, Spoel SH, 2011. Transcription dynamics in plant immunity. *Plant Cell* 23, 2809-20.
- Morris KV, Chan SW-L, Jacobsen SE, Looney DJ, 2004. Small Interfering RNA-Induced Transcriptional Gene Silencing in Human Cells. *Science* 305, 1289-92.
- Mosher RA, Durrant WE, Wang D, Song J, Dong X, 2006. A comprehensive structure-function analysis of Arabidopsis SNI1 defines essential regions and transcriptional repressor activity. *Plant Cell* 18, 1750-65.

- Naito K, Zhang F, Tsukiyama T, *et al.*, 2009. Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature* 461, 1130-4.
- Napoli C, Lemieux C, Jorgensen R, 1990. Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. *The Plant cell* 2, 279-89.
- Niu D, Lii YE, Chellappan P, *et al.*, 2016. miRNA863-3p sequentially targets negative immune regulator ARLPKs and positive regulator SERRATE upon bacterial infection. *Nat Commun* 7, 11324.
- Nueda MJ, Tarazona S, Conesa A, 2014. Next maSigPro: updating maSigPro bioconductor package for RNA-seq time series. *Bioinformatics* 30, 2598-602.
- Palma K, Thorgrimsen S, Malinovsky FG, *et al.*, 2010. Autoimmunity in Arabidopsis acd11 is mediated by epigenetic regulation of an immune receptor. *PLoS Pathog* 6, e1001137.
- Panda K, Ji L, Neumann DA, Daron J, Schmitz RJ, Slotkin RK, 2016. Full-length autonomous transposable elements are preferentially targeted by expression-dependent forms of RNA-directed DNA methylation. *Genome Biol* 17, 170.
- Pastor WA, Stroud H, Nee K, *et al.*, 2014. MORC1 represses transposable elements in the mouse male germline. *Nat Commun* 5, 5795.
- Pelaez P, Sanchez F, 2013. Small RNAs in plant defense responses during viral and bacterial interactions: similarities and differences. *Frontiers in plant science* 4, 343.
- Perry J, Zhao Y, 2003. The CW domain, a structural module shared amongst vertebrates, vertebrate-infecting parasites and higher plants. *Trends Biochem Sci* 28, 576-80.
- Pique-Regi R, Degner JF, Pai AA, Gaffney DJ, Gilad Y, Pritchard JK, 2011. Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res* 21, 447-55.
- Probst AV, Almouzni G, 2011. Heterochromatin establishment in the context of genome-wide epigenetic reprogramming. *Trends in genetics : TIG* 27, 177-85.
- Qiao Y, Liu L, Xiong Q, *et al.*, 2013. Oomycete pathogens encode RNA silencing suppressors. *Nat Genet* 45, 330-3.
- Qin C, Li B, Fan Y, *et al.*, 2017. Roles of Dicer-Like Proteins 2 and 4 in Intra- and Intercellular Antiviral Silencing. *Plant Physiol* 174, 1067-81.
- Quadrana L, Bortolini Silveira A, Mayhew GF, *et al.*, 2016. The Arabidopsis thaliana mobilome and its impact at the species level. *Elife* 5.

- Quinlan AR, Hall IM, 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841-2.
- Reis PP, Waldron L, Goswami RS, *et al.*, 2011. mRNA transcript quantification in archival samples using multiplexed, color-coded probes. *BMC Biotechnol* 11, 46.
- Robinson MD, McCarthy DJ, Smyth GK, 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139-40.
- Rowley MJ, Rothi MH, Bohmdorfer G, Kucinski J, Wierzbicki AT, 2017. Long-range control of gene expression via RNA-directed DNA methylation. *PLoS Genet* 13, e1006749.
- Sarris M, Nikolaou K, Talianidis I, 2014. Context-specific regulation of cancer epigenomes by histone and transcription factor methylation. *Oncogene* 33, 1207-17.
- Schmittgen TD, Livak KJ, 2008. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* 3, 1101-8.
- Seidl MF, Thomma B, 2017. Transposable Elements Direct The Coevolution between Plants and Microbes. *Trends Genet.*
- Shivaprasad PV, Chen HM, Patel K, Bond DM, Santos BA, Baulcombe DC, 2012. A microRNA superfamily regulates nucleotide binding site-leucine-rich repeats and other mRNAs. *Plant Cell* 24, 859-74.
- Sridhar VV, Kapoor A, Zhang K, *et al.*, 2007. Control of DNA methylation and heterochromatic silencing by histone H2B deubiquitination. *Nature* 447, 735-8.
- Staiger D, Korneli C, Lummer M, Navarro L, 2012. Emerging role for RNA-based regulation in plant immunity. *New Phytol.*
- Sura W, Kabza M, Karlowski WM, *et al.*, 2017. Dual Role of the Histone Variant H2A.Z in Transcriptional Regulation of Stress-Response Genes. *Plant Cell* 29, 791-807.
- Tai HH, Tai GC, Beardmore T, 2005. Dynamic histone acetylation of late embryonic genes during seed germination. *Plant Mol Biol* 59, 909-25.
- Tao Y, Xie Z, Chen W, *et al.*, 2003. Quantitative nature of Arabidopsis responses during compatible and incompatible interactions with the bacterial pathogen *Pseudomonas syringae*. *Plant Cell* 15, 317-30.
- Thurman RE, Rynes E, Humbert R, *et al.*, 2012. The accessible chromatin landscape of the human genome. *Nature* 489, 75-82.
- Tittel-Elmer M, Bucher E, Broger L, Mathieu O, Paszkowski J, Vaillant I, 2010. Stress-induced activation of heterochromatic transcription. *PLoS Genet* 6, e1001175.

- Vlot AC, Dempsey DA, Klessig DF, 2009. Salicylic Acid, a multifaceted hormone to combat disease. *Annu Rev Phytopathol* 47, 177-206.
- Wang C, Gao F, Wu J, Dai J, Wei C, Li Y, 2010a. Arabidopsis putative deacetylase AtSRT2 regulates basal defense by suppressing PAD4, EDS5 and SID2 expression. *Plant Cell Physiol* 51, 1291-9.
- Wang C, Gao F, Wu J, Dai J, Wei C, Li Y, 2010b. Arabidopsis Putative Deacetylase AtSRT2 Regulates Basal Defense by Suppressing PAD4, EDS5 and SID2 Expression. *Plant and Cell Physiology* 51, 1291-9.
- Wang D, Qu Z, Yang L, *et al.*, 2017a. Transposable elements (TEs) contribute to stress-related long intergenic noncoding RNAs in plants. *Plant J* 90, 133-46.
- Wang H, Seo JK, Gao S, Cui X, Jin H, 2017b. Silencing of AtRAP, a target gene of a bacteria-induced small RNA, triggers antibacterial defense responses through activation of LSU2 and down-regulation of GLK1. *New Phytol* 215, 1144-55.
- Wang W, Barnaby JY, Tada Y, *et al.*, 2011. Timing of plant immune responses by a central circadian regulator. *Nature* 470, 110-4.
- Watson ML, Zinn AR, Inoue N, *et al.*, 1998. Identification of morc (microorchidia), a mutation that results in arrest of spermatogenesis at an early meiotic stage in the mouse. *Proc Natl Acad Sci U S A* 95, 14361-6.
- Weiberg A, Wang M, Lin FM, *et al.*, 2013. Fungal small RNAs suppress plant immunity by hijacking host RNA interference pathways. *Science* 342, 118-23.
- Wesley SV, Helliwell CA, Smith NA, *et al.*, 2001. Construct design for efficient, effective and high-throughput gene silencing in plants. *Plant Journal* 27, 581-90.
- Xu K, Wu ZJ, Groner AC, *et al.*, 2012. EZH2 oncogenic activity in castration-resistant prostate cancer cells is Polycomb-independent. *Science* 338, 1465-9.
- Yasuda K, Ito M, Sugita T, *et al.*, 2013. Utilization of transposable element as a novel genetic tool for modification of the stress response in rice. *Mol Breed* 32, 505-16.
- Yeakley JM, Fan JB, Doucet D, *et al.*, 2002. Profiling alternative splicing on fiber-optic arrays. *Nat Biotechnol* 20, 353-8.
- Yu A, Lepere G, Jay F, *et al.*, 2013. Dynamics and biological relevance of DNA demethylation in Arabidopsis antibacterial defense. *Proc Natl Acad Sci U S A* 110, 2389-94.

- Zang C, Schones DE, Zeng C, Cui K, Zhao K, Peng W, 2009. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* 25, 1952-8.
- Zaratiegui M, Irvine DV, Martienssen RA, 2007. Noncoding RNAs and gene silencing. *Cell* 128, 763-76.
- Zemach A, Kim MY, Hsieh PH, *et al.*, 2013. The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* 153, 193-205.
- Zhai J, Jeong DH, De Paoli E, *et al.*, 2011. MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. *Genes Dev* 25, 2540-53.
- Zhang H, Zhu J-K, 2011. RNA-directed DNA methylation. *Current Opinion in Plant Biology* 14, 142-7.
- Zhang K, Sridhar VV, Zhu J, Kapoor A, Zhu JK, 2007. Distinctive core histone post-translational modification patterns in Arabidopsis thaliana. *PLoS One* 2, e1210.
- Zhang W, Zhang T, Wu Y, Jiang J, 2012. Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in Arabidopsis. *Plant Cell* 24, 2719-31.
- Zhang X, Zhao H, Gao S, *et al.*, 2011. Arabidopsis Argonaute 2 regulates innate immunity via miRNA393(*)-mediated silencing of a Golgi-localized SNARE gene, MEMB12. *Mol Cell* 42, 356-66.
- Zhang Y, Zhang H, 2017. RNAa Induced by TATA Box-Targeting MicroRNAs. *Adv Exp Med Biol* 983, 91-111.