creative commons
COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

A DISSERTATION FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Epigenetic Regulation of Developmental Process and Morphological Divergence in *Arabidopsis thaliana* and *Brassica rapa* Subspecies

애기장대의 발달 과정 및 배추의 아종 내 형태학적 다양성에 대한 후성유전학적 조절 연구

AUGUST, 2022

JOO YOUNG LIM

MAJOR IN HORTICULTURAL SCIENCE AND BIOTECHNOLOGY

DEPARTMENT OF AGRICULTURE, FORESTRY AND BIORESOURCES

COLLEGE OF AGRICULTURE AND LIFE SCIENCES

THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

Epigenetic Regulation of Developmental Process and Morphological Divergence in *Arabidopsis thaliana* and *Brassica rapa* Subspecies

UNDER THE DIRECTION OF DR. JIN HOE HUH
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF
SEOUL NATIONAL UNIVERSITY

BY
JOO YOUNG LIM

MAJOR IN HORTICULTURAL SCIENCE AND BIOTECHNOLOGY
DEPARTMENT OF AGRICULTURE, FORESTRY AND BIORESOURCES

AUGUST, 2022

APPROVED AS A QUALIFIED DISSERTATION OF JOO YOUNG LIM
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
BY THE COMMITTEE MEMBERS

CHAIRMAN	**Doil Choi, Ph.D.**
VICE-CHAIRMAN	**Jin Hoe Huh, Ph.D.**
MEMBER	**Byoung-Cheorl Kang, Ph.D.**
MEMBER	**Yeonhee Choi, Ph.D.**
MEMBER	**Jinju Han, Ph.D.**

Epigenetic Regulation of Developmental Process and Morphological Divergence in *Arabidopsis thaliana* and *Brassica rapa* Subspecies

JOO YOUNG LIM

Department of Agriculture, Forestry and Bioresources

The Graduate School of Seoul National University

ABSTRACT

Epigenetic mechanisms play crucial roles in diverse biological processes such as cell differentiation and responses to developmental and environmental cues. Epigenetic modifications including DNA methylation and histone modifications can alter gene expression through modulating chromatin structure without DNA sequence changes. Plants can perceive and respond to internal or external signals, where gene expression may be epigenetically regulated. Moreover, some epigenetic variations may be stably inherited and conceivably contribute to important phenotypic traits. However,

the detailed mechanisms of epigenetic factor-mediated developmental and environmental plasticity and their evolutionary significance associated with traits remain to be explored in plants. In this study, I investigated environmental and developmental responses regulated by DNA demethylases in Arabidopsis. DNA methylation can be actively removed by the DEMETER (DME) family of DNA glycosylases, together with REPRESSOR OF SILENCING 1 (ROS1), DEMETER-LIKE 2 (DML2) and DML3. Two *ros1* mutant lines were hypersensitive to abscisic acid (ABA). Downregulation of ABA-inducible genes was accompanied by DNA hypermethylation at their promoter regions in *ros1*, indicating ROS1-dependent DNA demethylation is required for transcriptional activation in ABA-mediated drought and osmotic stress responses. I further extended this research to examine the combined function of DNA demethylases in reproductive development. I observed several developmental defects such as delayed growth and flowering, aberrant floral development and short unfertilized siliques in the *dme ros1 dml2 dml3* (*drdd*) quadruple homozygous mutant. Gene expression is significantly reduced in *drdd* compared to WT, rather than in *dme* and *rdd*, accompanied with an increase in DNA methylation levels. These findings suggest that DNA demethylases act redundantly for the proper regulation of genes during reproductive development. In addition, I analyzed the morphological

divergence between the two subspecies Chinese cabbage (*Brassica rapa* subsp. *pekinensis*) and turnip (*B. rapa* subsp. *rapa*) despite high genetic similarity. Comprehensive analysis of transcriptome and epigenome revealed that accessible chromatin regions (ACRs) were associated with the expression dynamics, histone H3 lysine 27 acetylation enrichment and the depletion of DNA methylation. The distant ACRs of the two subspecies were highly conserved but displayed divergent chromatin accessibility with differential enrichment of transcription factor motifs. These results indicate that subspecies-specific divergence of distal enhancers might be responsible for morphotype diversification. Taken together, this study will broaden the understanding of regulatory mechanisms of DNA demethylation in response to environmental and developmental cues and provide insights into distal enhancer-derived subspecies diversification during evolution.

Keywords: Arabidopsis, DNA demethylase, Chinese cabbage, turnip, enhancer

Student number: 2017-34703

CONTENTS

LIST OF TABLES

CHAPTER 1

CHAPTER 2

CHAPTER 3

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

LIST OF ABBREVIATION

5mC	5-methylcytosine
ABA	Abscisic acid
ABCB4	ATP-BINDING CASSETTE B4
ACR	Accessible chromatin region
ATAC-seq	Assay for transposase-accessible chromatin sequencing
BEH3	BRASSINAZOLE RESISTANT1 HOMOLOG 3
BEH4	BRASSINAZOLE RESISTANT1 HOMOLOG 4
BER	Base excision repair
BES1	BRASSINOSTEROID INSENSITIVE1-EMS-SUPPRESSOR 1
BZR1	BRASSINAZOLE RESISTANT1
CaMV	Cauliflower mosaic virus
ChIP-seq	Chromatin immunoprecipitation sequencing
CMT3	CHROMOMETHYLASE 3
CSLA11	CELLULOSE SYNTHASE LIKE A11
dACR	Distant accessible chromatin region
DAS	Day-after-sowing
DEG	Differentially expressed gene
DMC	Differentially methylated cytosine
DME	DEMETER
DML2	DEMETER-LIKE 2
DML3	DEMETER-LIKE 3
DMR	Differentially methylated region
DNMT	DNA methyltransferase

drdd	dme ros1 dml2 dml3
DRM2	DOMAINS REARRANGED METHYLTRANSFERASE 2
EPF2	EPIDERMAL PATTERNING FACTOR 2
EXPA5	EXPANSIN A5
EXPA7	EXPANSIN A7
FIS2	FERTILIZATION INDEPENDENT SEED 2
FWA	FLOWERING WAGENINGEN
gACR	Genic accessible chromatin region
gbM	Gene body methylation
H3K27ac	Histone H3 lysine 27 acetylation
H3K27me3	Histone H3 lysine 27 trimethylation
H3K4me1	Histone H3 lysine 4 monomethylation
H3K9me2	Histone H3 lysine 9 dimethylation
JKD	JACKDAW
LHL3	LONESOME HIGHWAY LIKE 3
LINE	Long interspersed nuclear element
LTR	Long terminal repeat
MEA	MEDEA
MET1	METHYLTRANSFERASE 1
MGP	MAGPIE
Nam	Nicotinamide
NIC3	NICOTINAMIDASE 3
NUC	NUTCRACKER
pACR	Proximal accessible chromatin region
PILS5	PIN-LIKES 5
PRK4	POLLEN RECEPTOR LIKE KINASE 4

QTL	Quantitative trait loci
rdd	ros1 dml2 dml3
RdDM	RNA-directed DNA methylation
RHD2	ROOT HAIR DEFECTIVE 2
RNA-seq	RNA sequencing
ROS1	REPRESSOR OF SILENCING 1
RPT2	ROOT PHOTOTROPISM 2
RVE1	REVEILLE1
SCR	SCARECROW
SHR	SHORTROOT
SINE	Short interspersed nuclear element
siRNA	Small interfering RNA
SNP	Single nucleotide polymorphism
STARR-seq	Self-transcribing active regulatory region sequencing
TDG	Thymine DNA glycosylase
TE	Transposable element
TET	Ten-eleven translocation
TF	Transcription factor
TMAC2	TWO OR MORE ABRES-CONTAINING GENE 2
WT	Wild type

GENERAL INTRODUCTION

DNA methylation

In higher eukaryotes, DNA methylation plays essential roles in diverse biological processes such as gene imprinting, transposon silencing and X chromosome inactivation (Huh et al., 2008; Law and Jacobsen, 2010; Smith and Meissner, 2013). DNA methylation is achieved by DNA methyltransferases (DNMTs) catalyzing the addition of a methyl group to the C5 position of cytosine (5-methylcytosine; 5mC). DNA methylation in mammals is primarily found in the symmetric CG context, whereas DNA methylation in plants occurs at cytosines in all sequence contexts: CG, CHG and CHH (H = A, T or C) (Law and Jacobsen, 2010).

In mammals, DNA methylation is established by the DNMT3 family of *de novo* methyltransferases and maintained through the activity of DNMT1 (Law and Jacobsen, 2010; Wu and Zhang, 2010). During DNA replication, ubiquitin-like plant homeodomain and RING finger domain 1 specifically binds to hemimethylated DNA and recruits DNMT1 to copy the methylation pattern onto the daughter strand (Law and Jacobsen, 2010; Smith and Meissner, 2013; Wu and Zhang, 2010). In plants, DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2), a homolog of

DNMT3, catalyzes the establishment of DNA methylation via the RNA-directed DNA methylation (RdDM) pathway (Matzke and Mosher, 2014). RNA polymerase IV initiates the biogenesis of 24-nucleotide small interfering RNAs (siRNAs) that direct DRM2-mediated *de novo* methylation on RNA polymerase V-transcribed loci (Matzke and Mosher, 2014; Zhang et al., 2018). Once established, maintenance of DNA methylation requires different DNA methyltransferases acting on each sequent context. Symmetric CG and CHG methylation are maintained by DNA METHYLTRANSFERASE 1 (MET1) and CHROMOMETHYLASE 3 (CMT3), respectively, and DRM2 and CMT2 are responsible for maintaining asymmetric CHH methylation (Law and Jacobsen, 2010; Zhang et al., 2018).

DNA demethylation

DNA demethylation is a reverse process of DNA methylation, which can take place through passive or active mechanisms (Law and Jacobsen, 2010). Passive DNA demethylation is a process by which 5mC is replaced with cytosine in a replication-dependent manner when maintenance DNA methyltransferases such as DNMT1 and MET1 are downregulated or inactivated. In contrast, active DNA demethylation refers to the enzymatic removal of 5mC in a replication-independent manner (Wu and Zhang, 2010).

Different types of active demethylation machineries are employed in mammals and plants. In mammals, the ten-eleven translocation (TET) family proteins catalyze the oxidation of 5mC to form 5-hydroxymethylcytosine (5hmC), then oxidized to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Ito et al., 2011; Tahiliani et al., 2009). These intermediates are excised by the thymine DNA glycosylase (TDG) and further replaced with unmethylated cytosine via the base excision repair (BER) pathway (He et al., 2011; Kohli and Zhang, 2013; Wu and Zhang, 2017). On the contrary to the conversion of 5mC to other bases in mammals, direct removal of 5mC is achieved by specific DNA glycosylases in plants. As bifunctional DNA glycosylases/lyases, the DEMETER (DME) family of proteins excise 5mC and cleave the sugar-phosphate backbone, creating an abasic site, by β- and δ-elimination processes (Choi et al., 2002; Gehring et al., 2006; Gong et al., 2002; Penterman et al., 2007). The gap is processed to provide 3′-OH for subsequent polymerization via the BER pathway (Lee et al., 2014; Martinez-Macias et al., 2012).

Four DNA glycosylase family members are found in Arabidopsis – DME, REPRESSOR OF SILENCING 1 (ROS1), DEMETER-LIKE 2 (DML2) and DML3 (Penterman et al., 2007). *DME* is primarily expressed in the central cell of the female gametophyte, which is important for the

establishment of gene imprinting in the endosperm, whereby a maternal *dme* allele results in seed abortion (Choi et al., 2002; Gehring, 2013; Huh et al., 2008). DME activates the maternal allele-specific expression of *MEDEA* (*MEA*), *FLOWERING WAGENINGEN* (*FWA*) and *FERTILIZATION INDEPENDENT SEED 2* (*FIS2*) by removing DNA methylation (Choi et al., 2002; Jullien et al., 2006; Kinoshita et al., 2004). In contrast to *DME*, *ROS1*, *DML2* and *DML3* are broadly expressed in vegetative tissues. ROS1 is necessary for the inhibition of transcriptional silencing of transgenes and endogenous genes, and along with DML2 and DML3, prevents the spreading of DNA methylation at genomic regions (Gong et al., 2002; Lister et al., 2008; Penterman et al., 2007; Qian et al., 2012; Stroud et al., 2013). Notably, ROS1 controls the expression of *EPIDERMAL PATTERNING FACTOR 2* (*EPF2*) for stomatal development and stress-responsive genes for *Fusarium oxysporum* resistance by antagonizing RdDM at transposable elements (Le et al., 2014; Tang et al., 2016; Yamamuro et al., 2014). Moreover, recent studies attempted to generate the *dme ros1 dml2 dml3* quadruple mutant by central cell-specific *DME* complementation, indicating the biological function of DNA demethylases in flowering and biotic stress response (Williams et al., 2021; Zeng et al., 2021).

Enhancers in transcriptional regulation

Enhancers are *cis*-regulatory DNA elements that act independently of the distance, location or orientation to their target genes (Li et al., 2016; Long et al., 2016). Generally, the assembly of transcription complex at enhancers is initiated by pioneer transcription factors that bind to DNA and generate open chromatin, followed by the recruitment of coactivators such as histone acetyltransferases and chromatin remodelers to increase the accessibility of chromatin (Li et al., 2016; Shlyueva et al., 2014). Then, other transcription factors that interact with components of the Mediator complex bind to the DNA to recruit RNA polymerase II, thereby leading to augmentation of basal transcription levels (Malik and Roeder, 2010; Ong and Corces, 2011). Importantly, chromatin looping brings enhancers in close proximity to their cognate promoters, established by cohesin and the Mediator complex (Ong and Corces, 2011; Schoenfelder and Fraser, 2019). Such looping allows long-range interactions between distal enhancers and their cognate genes, even up to 1 Mb away (Lettice et al., 2003).

Enhancers play crucial roles in cell identity control, embryonic development, diseases and evolution by precise spatiotemporal regulation of gene expression (Ong and Corces, 2011; Shlyueva et al., 2014). As multiple enhancers with diverse spatiotemporal activities regulate the expression of

their target genes, enhancers are activated in response to external or internal cues (Shlyueva et al., 2014). In mammals, active enhancers are typically associated with high chromatin accessibility, enrichment of histone H3 lysine 4 monomethylation (H3K4me1) and H3K27 acetylation (H3K27ac), non-coding enhancer RNA transcription and low DNA methylation levels (Andersson et al., 2014; Kouzarides, 2007; Rada-Iglesias et al., 2011). In contrast, inactive enhancers display low chromatin accessibility and enrichment of H3K4me1 and H3K27 trimethylation (H3K27me3) (Rada-Iglesias et al., 2011).

Identification of enhancers in plants

As enhancers can activate transcription in a distance- and orientation-independent manner, enhancer discovery has remained challenging, especially in plants. High-throughput sequencing technologies have enabled the identification of enhancer chromatin features in the human and other animal genomes, which leads to the discovery of tens of thousands of enhancer candidates (Andersson et al., 2014; Kvon et al., 2014). However, in plants, genome-wide chromatin features of enhancers and their spatiotemporal control in development remain unclear. By assaying chromatin accessibility in a high-throughput manner, together with histone

mark enrichment and DNA methylation analysis, DNase-hypersensitive sites or transposase-accessible chromatin regions were predicted as enhancer candidates in the Arabidopsis, rice and maize genomes (Oka et al., 2017; Ricci et al., 2019; Zhang et al., 2012; Zhu et al., 2015). Recently, a massive parallel reporter assay, self-transcribing active regulatory region (STARR)-seq was used to directly examine enhancer activity in rice and maize (Ricci et al., 2019; Sun et al., 2019). Nevertheless, in tobacco, when enhancer candidates were located in the 3'-untranslated regions under the control of a minimal promoter as in the mammalian STARR-seq system, they failed to exhibit enhancer activities due to mRNA degradation, indicating the limitation of species-specific tolerances to mRNA decay for plant STARR-seq (Jores et al., 2020).

Morphological divergence of the Brassicaceae family

The Brassicaceae (Cruciferae) family comprises 338 genera and 3,709 species, with agronomical and economical importance as oilseeds, condiments, fodder or vegetable crops (Warwick et al., 2006). These crops include *Brassica rapa* (Chinese cabbage, pak choi and turnip), *Brassica oleracea* (broccoli, cabbage and cauliflower), *Brassica napus* (rapeseed), *Raphanus sativus* (radish), *Armracia rusticana* (horseradish) and *Sinapis alba*

(white mustard) (Gómez-Campo, 1980). After a whole genome triplication event that took place approximately 9-15 million years ago, *Brassica* genomes have experienced diversification as well as extensive gene fractionation. Domestication and breeding of the *Brassica* crops also facilitated highly diverse morphotypes involving leafy heads, enlarged organs (roots, stems and inflorescences) and extensive axillary branching (Cheng et al., 2016; Cheng et al., 2014; Wang et al., 2011). In addition to the evolution of many distinct phenotypic traits, similar morphotypes are likely to be selected between species, indicating convergent crop domestication, exemplified by the leafy heads of Chinese cabbage (*B. rapa*) and cabbage (*B. oleracea*) and enlarged roots or stems in turnip (*B. rapa*), kohlrabi (*B. oleracea*), swede (*B. napus*) and tuberous mustard (*B. juncea*) (Cheng et al., 2016; Cheng et al., 2014).

This study focused on the role of DNA demethylases in abscisic acid response and reproductive development in Arabidopsis. This thesis work also includes the comparative analysis of genome and epigenome landscapes in *Brassica rapa* subspecies and addresses the following three topics:

Chapter 1: Plant-specific DNA demethylation in response to abscisic acid in Arabidopsis

Chapter 2: Regulation of reproductive development by DNA demethylases in Arabidopsis

Chapter 3: Comparative analysis of genome and epigenome landscapes in *Brassica rapa* subspecies

REFERENCES

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al. (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455-461.

Cheng, F., Sun, R., Hou, X., Zheng, H., Zhang, F., Zhang, Y., Liu, B., Liang, J., Zhuang, M., Liu, Y., et al. (2016). Subgenome parallel selection is associated with morphotype diversification and convergent crop domestication in *Brassica rapa* and *Brassica oleracea*. Nat. Genet. *48*, 1218-1224.

Cheng, F., Wu, J., and Wang, X. (2014). Genome triplication drove the diversification of *Brassica* plants. Hortic. Res. *1*, 14024.

Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J.J., Goldberg, R.B., Jacobsen, S.E., and Fischer, R.L. (2002). DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*. Cell *110*, 33-42.

Gehring, M. (2013). Genomic imprinting: insights from plants. Annu. Rev. Genet. *47*, 187-208.

Gehring, M., Huh, J.H., Hsieh, T.F., Penterman, J., Choi, Y., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (2006). DEMETER DNA glycosylase establishes *MEDEA* polycomb gene self-imprinting by allele-specific demethylation. Cell *124*, 495-506.

Gómez-Campo, C. (1980). Morphology and morpho-taxonomy of the tribe Brassiceae. Brassica crops and wild allies.[I]. 3-31.

Gong, Z., Morales-Ruiz, T., Ariza, R.R., Roldan-Arjona, T., David, L.,

and Zhu, J.K. (2002). *ROS1*, a repressor of transcriptional gene silencing in *Arabidopsis*, encodes a DNA glycosylase/lyase. Cell *111*, 803-814.

He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science *333*, 1303-1307.

Huh, J.H., Bauer, M.J., Hsieh, T.F., and Fischer, R.L. (2008). Cellular programming of plant gene imprinting. Cell *132*, 735-744.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science *333*, 1300-1303.

Jores, T., Tonnies, J., Dorrity, M.W., Cuperus, J.T., Fields, S., and Queitsch, C. (2020). Identification of plant enhancers and their constituent elements by STARR-seq in tobacco leaves. Plant Cell *32*, 2120-2131.

Jullien, P.E., Kinoshita, T., Ohad, N., and Berger, F. (2006). Maintenance of DNA methylation during the *Arabidopsis* life cycle is essential for parental imprinting. Plant Cell *18*, 1360-1372.

Kinoshita, T., Miura, A., Choi, Y., Kinoshita, Y., Cao, X., Jacobsen, S.E., Fischer, R.L., and Kakutani, T. (2004). One-way control of *FWA* imprinting in *Arabidopsis* endosperm by DNA methylation. Science *303*, 521-523.

Kohli, R.M., and Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. Nature *502*, 472-479.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Kvon, E.Z., Kazmar, T., Stampfel, G., Yanez-Cuna, J.O., Pagani, M.,

Schernhuber, K., Dickson, B.J., and Stark, A. (2014). Genome-scale functional characterization of *Drosophila* developmental enhancers *in vivo*. Nature *512*, 91-95.

Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. *11*, 204-220.

Le, T.N., Schumann, U., Smith, N.A., Tiwari, S., Au, P.C., Zhu, Q.H., Taylor, J.M., Kazan, K., Llewellyn, D.J., Zhang, R., et al. (2014). DNA demethylases target promoter transposable elements to positively regulate stress responsive genes in *Arabidopsis*. Genome Biol. *15*, 458.

Lee, J., Jang, H., Shin, H., Choi, W.L., Mok, Y.G., and Huh, J.H. (2014). AP endonucleases process 5-methylcytosine excision intermediates during active DNA demethylation in *Arabidopsis*. Nucleic Acids Res. *42*, 11408-11418.

Lettice, L.A., Heaney, S.J., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D., Elgar, G., Hill, R.E., and de Graaff, E. (2003). A long-range *Shh* enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum. Mol. Genet. *12*, 1725-1735.

Li, W., Notani, D., and Rosenfeld, M.G. (2016). Enhancers as non-coding RNA transcription units: recent insights and future perspectives. Nat. Rev. Genet. *17*, 207-223.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008). Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. Cell *133*, 523-536.

Long, H.K., Prescott, S.L., and Wysocka, J. (2016). Ever-changing landscapes: transcriptional enhancers in development and evolution. Cell *167*, 1170-1187.

Malik, S., and Roeder, R.G. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. Nat. Rev. Genet. *11*, 761-772.

Martinez-Macias, M.I., Qian, W., Miki, D., Pontes, O., Liu, Y., Tang, K., Liu, R., Morales-Ruiz, T., Ariza, R.R., Roldan-Arjona, T., and Zhu, J.K. (2012). A DNA 3' phosphatase functions in active DNA demethylation in *Arabidopsis*. Mol. Cell. *45*, 357-370.

Matzke, M.A., and Mosher, R.A. (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat. Rev. Genet. *15*, 394-408.

Oka, R., Zicola, J., Weber, B., Anderson, S.N., Hodgman, C., Gent, J.I., Wesselink, J.J., Springer, N.M., Hoefsloot, H.C.J., Turck, F., and Stam, M. (2017). Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize. Genome Biol. *18*, 137.

Ong, C.T., and Corces, V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat. Rev. Genet. *12*, 283-293.

Penterman, J., Zilberman, D., Huh, J.H., Ballinger, T., Henikoff, S., and Fischer, R.L. (2007). DNA demethylation in the *Arabidopsis* genome. Proc. Natl. Acad. Sci. U. S. A. *104*, 6752-6757.

Qian, W., Miki, D., Zhang, H., Liu, Y., Zhang, X., Tang, K., Kan, Y., La, H., Li, X., Li, S., et al. (2012). A histone acetyltransferase regulates active DNA demethylation in *Arabidopsis*. Science *336*, 1445-1448.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279-283.

Ricci, W.A., Lu, Z., Ji, L., Marand, A.P., Ethridge, C.L., Murphy, N.G., Noshay, J.M., Galli, M., Mejia-Guerra, M.K., Colome-Tatche, M., et al. (2019). Widespread long-range cis-regulatory elements in the maize genome. Nat. Plants *5*, 1237-1249.

Schoenfelder, S., and Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. Nat. Rev. Genet. *20*, 437-455.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nat. Rev. Genet. *15*, 272-286.

Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nat. Rev. Genet. *14*, 204-220.

Stroud, H., Greenberg, M.V., Feng, S., Bernatavichute, Y.V., and Jacobsen, S.E. (2013). Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. Cell *152*, 352-364.

Sun, J., He, N., Niu, L., Huang, Y., Shen, W., Zhang, Y., Li, L., and Hou, C. (2019). Global quantitative mapping of enhancers in rice by STARR-seq. Genomics Proteomics Bioinformatics *17*, 140-153.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., and Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930-935.

Tang, K., Lang, Z., Zhang, H., and Zhu, J.K. (2016). The DNA demethylase ROS1 targets genomic regions with distinct chromatin modifications. Nat. Plants *2*, 16169.

Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., Bai, Y., Mun, J.H., Bancroft, I., Cheng, F., et al. (2011). The genome of the mesopolyploid

crop species *Brassica rapa*. Nat. Genet. *43*, 1035-1039.

Warwick, S.I., Francis, A., and Al-Shehbaz, I.A. (2006). Brassicaceae: Species checklist and database on CD-Rom. Plant Syst. Evol. *259*, 249-258.

Williams, B.P., Bechen, L.L., Pohlmann, D.A., and Gehring, M. (2021). Somatic DNA demethylation generates tissue-specific methylation states and impacts flowering time. Plant Cell *34*, 1189-1206.

Wu, S.C., and Zhang, Y. (2010). Active DNA demethylation: many roads lead to Rome. Nat. Rev. Mol. Cell Biol. *11*, 607-620.

Wu, X., and Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. Nat. Rev. Genet. *18*, 517-534.

Yamamuro, C., Miki, D., Zheng, Z., Ma, J., Wang, J., Yang, Z., Dong, J., and Zhu, J.K. (2014). Overproduction of stomatal lineage cells in *Arabidopsis* mutants defective in active DNA demethylation. Nat. Commun. *5*, 4062.

Zeng, W., Huang, H., Lin, X., Zhu, C., Kosami, K.I., Huang, C., Zhang, H., Duan, C.G., Zhu, J.K., and Miki, D. (2021). Roles of DEMETER in regulating DNA methylation in vegetative tissues and pathogen resistance. J. Integr. Plant Biol. *63*, 691-706.

Zhang, H., Lang, Z., and Zhu, J.K. (2018). Dynamics and function of DNA methylation in plants. Nat. Rev. Mol. Cell Biol. *19*, 489-506.

Zhang, W., Wu, Y., Schnable, J.C., Zeng, Z., Freeling, M., Crawford, G.E., and Jiang, J. (2012). High-resolution mapping of open chromatin in the rice genome. Genome Res. *22*, 151-162.

Zhu, B., Zhang, W., Zhang, T., Liu, B., and Jiang, J. (2015). Genome-wide prediction and validation of intergenic enhancers in Arabidopsis using open chromatin signatures. Plant Cell *27*, 2415-2426.

CHAPTER 1

Plant-Specific DNA Demethylation in Response to Abscisic Acid

The research described in this chapter was published in *Plant Physiology*, https://doi.org/10.1104/pp.18.01471

ABSTRACT

In higher eukaryotes, DNA methylation plays pivotal roles in various biological processes including the response to environmental stress. In Arabidopsis, active DNA demethylation can be achieved by the activity of REPRESSOR OF SILENCING 1 (ROS1) that directly excises 5-methylcytosine from DNA. Abscisic acid (ABA) is a key phytohormone that is upregulated and accumulated under osmotic stress conditions. The involvement of DNA methylation in the ABA response has been reported, but the underlying mechanisms remain elusive. Here I report that *ros1* mutants exhibit hypersensitivity to ABA during early seedling development and root growth. In *ros1* mutants, some ABA-inducible genes were down-regulated, and more than 60 percent of their proximal regions became hypermethylated, demonstrating that ROS1-mediated DNA demethylation is required for the expression of a subset of ABA-inducible genes. Remarkably, *NICOTINAMIDASE 3* (*NIC3*) that encodes an enzyme that catalyzes the deamination of nicotinamide in the NAD^+ salvage pathway is linked to the loss of ABA responsiveness and DNA hypermethylation in *ros1*. The *nic3* mutants are hypersensitive to ABA, and moreover, ectopic expression of *NIC3* in *ros1* mutants restores normal ABA response. These findings indicate

that *NIC3* is inducible to ABA but requires ROS1-dependent DNA demethylation at its promoter as a prerequisite to transcriptional activation.

INTRODUCTION

Plants are sessile organisms that are constantly exposed to diverse harsh environmental conditions such as a deficit or an excess of water, salt, light and temperature. Plants perceive and respond to such abiotic stresses through signal transduction pathways, thereby regulating the expression of stress-responsive genes (Kinoshita and Seki, 2014; Mirouze and Paszkowski, 2011; Yamaguchi-Shinozaki and Shinozaki, 2006; Zhu, 2016). Abscisic acid (ABA) is a major plant hormone crucial for growth and development involving seed dormancy, stomatal aperture and environmental stress responses. In response to osmotic stress, ABA is accumulated in plant cells, then accompanied by transcriptional activation of stress-responsive genes via the ABA signaling pathway (Finkelstein, 2013; Mittler and Blumwald, 2015; Nambara and Marion-Poll, 2005; Yoshida et al., 2015).

Epigenetic modifications such as DNA methylation, histone modification and noncoding RNAs induce mitotically or meiotically heritable changes in gene expression by the alteration of chromatin structure without DNA sequence changes. Under environmental stress conditions, plants perceive and memorize stress-induced gene expression changes through the

mitotic or meiotic transmission of epigenetic states (Chinnusamy and Zhu, 2009; Ganguly et al., 2017; Heard and Martienssen, 2014; Mirouze and Paszkowski, 2011; Wibowo et al., 2016). Histone modifications that involve acetylation, methylation, phosphorylation and ubiquitination on specific residues of the histone N-terminal tails have been reported to regulate ABA, drought and salt stress responses (Chen et al., 2010; Chinnusamy and Zhu, 2009; Kinoshita and Seki, 2014). Particularly, Arabidopsis mutants defective in histone deacetylase activity displayed hypersensitivity to ABA and salt, whereas the transgenic plants overexpressing the histone deacetylase were insensitive to ABA and tolerant to drought and salt stresses (Cui et al., 2013; Sridha and Wu, 2006).

DNA methylation, which refers to the addition of a methyl group to the C5 position of cytosine (5-methylcytosine; 5mC), is responsible for diverse developmental processes including cell differentiation, gene imprinting, genome stability and X chromosome inactivation (Law and Jacobsen, 2010; Smith and Meissner, 2013; Wu and Zhang, 2014; Zhang and Zhu, 2012). Although DNA methylation in mammals predominantly takes place in the symmetric CG context, DNA methylation in plants occurs at cytosine bases in all sequence contexts: CG, CHG and CHH (H = A, T or C) (Law and Jacobsen, 2010). Plant *de novo* methylation is established by DOMAINS

REARRANGED METHYLTRANSFERASE 2 (DRM2) via the RNA-directed DNA methylation (RdDM) pathway (Matzke and Mosher, 2014). Once established, DNA methylation is maintained by three DNA methyltransferases, depending on different sequent contexts: DNA METHYLTRANSFERASE 1 (MET1), CHROMOMETHYLASE 3 (CMT3) and DRM2 at CG, CHG, and CHH sequences, respectively (Law and Jacobsen, 2010; Zhang et al., 2018). In contrast to the conversion of 5mC to other bases in mammals, active DNA demethylation in plants begins with direct excision of 5mC by DNA glycosylases, leading to unmethylated cytosine via the base excision repair pathway (Wu and Zhang, 2014; Zhu, 2009). Arabidopsis genome contains four DNA glycosylases - REPRESSOR OF SILENCING 1 (ROS1), DEMETER (DME), DEMETER-LIKE 2 (DML2), and DML3 (Choi et al., 2002; Penterman et al., 2007). All the DME/ROS1 family proteins can directly remove 5mC from DNA and then cleave the sugar-phosphate backbone by successive β- and δ-elimination reactions *in vitro* (Agius et al., 2006; Gehring et al., 2006; Morales-Ruiz et al., 2006; Ortega-Galisteo et al., 2008; Penterman *et al.*, 2007). Nevertheless, they play distinct roles in plant growth and development. *DME* is primarily expressed in the central cell of the female gametophyte, thereby responsible for gene imprinting and seed development (Choi et al., 2002; Gehring, 2013;

Huh et al., 2008). On the other hand, ROS1, DML2, and DML3 are expressed in the vegetative tissues with their biological functions uncharacterized (Penterman et al., 2007). ROS1 was isolated in a genetic mutant screen for transcriptional silencing of the ABA- and stress-responsive *RD29A::LUC* transgene and the endogenous *RD29A* gene (Gong et al., 2002). In addition to inhibition of transcriptional silencing, ROS1 prevents the spreading of DNA methylation at genomic regions, together with DML2 and DML3 (Gong et al., 2002; Lister et al., 2008; Penterman et al., 2007; Qian et al., 2012; Stroud et al., 2013).

In response to stress, DNA methylation alters the expression of stress-responsive genes. Genome-wide methylome analyses revealed that biotic stress-induced differentially methylated regions were highly associated with differentially expressed genes which were also upregulated in *met1* and *drm1 drm2 cmt3* (*ddc*) mutants, leading to enhanced tolerance to the bacterial pathogen (Dowen et al., 2012). DNA methylation changes at transposable elements (TEs) in response to abiotic stress, followed by alterations of nearby gene expression, were partially transmitted through mitosis in rice (Secco et al., 2015). Moreover, distinct regions that were susceptible to hyperosmotic stress-triggered DNA methylation changes were reinforced by repetitive exposure to stress and could be inherited to the offspring through the female

germline (Wibowo et al., 2016). Particularly, ROS1 controlled the expression of *EPIDERMAL PATTERNING FACTOR 2* (*EPF2*) in stomatal development in ABA response and stress-responsive genes required for *Fusarium oxysporum* resistance by counteracting the action of RdDM at TEs in the promoter regions (Le et al., 2014; Yamamuro et al., 2014).

In this study, the relationship between DNA methylation and ABA-mediated stress response was investigated. Here I showed that *ros1* mutants exhibited ABA hypersensitivity for early seedling development and root growth. In an attempt to understand the role of ROS1 in transcriptional regulation of ABA-responsive genes associated with DNA methylation changes, the transcriptome and DNA methylome were performed in the wild type (WT) and *ros1* mutants. Approximately 80 percent of differentially expressed genes (DEGs) were down-regulated in *ros1-4*, and more than 60 percent of the ABA-inducible genes were proximal to hypermethylated regions in *ros1* mutants. These results indicate that decreased expression of a subset of ABA-inducible genes in *ros1* mutants, associated with excessive DNA methylation, causes ABA hypersensitive phenotypes. Most notable among them is *NICOTINAMIDSE 3* (*NIC3*) which encodes a catalytic enzyme that converts nicotinamide (Nam) into nicotinic acid, as many other enzymes in the NAD^+ salvage pathway are known to be involved in stress

responses. In *ros1* mutants, TEs in the promoter region of *NIC3* were hypermethylated compared to the WT, with a decrease in ABA-inducibility. The *nic3* mutants were hypersensitive to ABA, and ABA hypersensitive phenotypes of *ros1* mutants were restored by ectopic expression of *NIC3*. Taken together, these findings suggest that ROS1-mediated active DNA demethylation maintains the active state of *NIC3* expression in response to ABA.

MATERIALS AND METHODS

Plant materials

Seeds *Arabidopsis thaliana* ecotype Columbia (Col-0) was used as the WT control in this study. The *ros1-3* seeds were provided (Penterman et al., 2007), and the homozygous T-DNA insertion lines *ros1-4* (SALK_045303), *nic3-1* (SALK_034040), and *nic3-2* (SALK_107343) were obtained from Arabidopsis Biological Resource Center. Seeds were sterilized and stratified at 4°C for 3 days in the dark. They were sown on Murashige and Skoog (MS) medium, then grown in a growth chamber at 22°C under 16 h of fluorescent light at 30 ± 10 μmol m^{-2} s^{-1}.

Cotyledon greening and root elongation assays

For sensitivity screening, cotyledon greening and root growth were analyzed on MS medium containing different concentrations of ABA or Nam. Seedlings with developed green cotyledons were counted every 4 days. For root elongation assay, 4-day-old seedlings grown on normal MS medium were transferred to MS medium supplemented with ABA or Nam. Seedlings were allowed to grow vertically for additional days, and then the primary root length was measured.

RNA sequencing (RNA-seq) and analysis

Total RNA was extracted using the RNeasy Plant Mini kit (Qiagen) and treated with RQ1 RNase-free DNase (Promega) to remove any genomic DNA contaminants. RNA-seq libraries were constructed according to the previous report (Zhong et al., 2011). There are three biological replicates per genotype. RNA-seq was performed on the Illumina HiSeq 2500 sequencing system.

Quantitative RT-PCR (RT-qPCR)

The complementary DNA synthesis was performed using the QuantiTech Reverse Transcription kit (Qiagen). Quantitative RT-PCR was conducted using the Roter-Gene Q (Qiagen) with SYBR green Q-master mix (Genet Bio). *UBIQUITIN10* was used as a control for the normalization of the relative transcript level of target genes. Primers used in RT-qPCR are listed in Table 1-1.

Whole genome bisulfite sequencing (BS-seq) and analysis

A total of 5 μg of genomic DNA was used to generate BS-seq libraries. The library preparation and sequencing were performed on Illumina HiSeq 2000. An average of approximately 43.4 M paired-end reads (2×101 bp) was obtained from each library. Low-quality sequences ($q < 20$) were trimmed

manually, and the trimmed reads were mapped to the TAIR 10 genome using bwa-meth (v0.09; (Pedersen et al., 2014)) and Bis-SNP (v. 0.82.2; (Liu et al., 2012)), with the parameters "–map-q 60 -T BisulfiteGenotyper -C CG,1 -C CHG,1 -C CHH,1 -out_modes EMIT_ALL_CYTOSINES". Only cytosine sites with 4× coverage were used for subsequent analysis. DMCs and DMRs were identified as described previously (Huang et al., 2013). DMRs were finally identified based on the regions with a length ≥ 100 bp, ≥ 5 DMCs, and the mean methylation difference ≥ 0.3 for CG, ≥ 0.15 for CHG, or ≥ 0.1 for CHH.

Local bisulfite sequencing

A total of 1 μg of genomic DNA was treated with sodium bisulfite using the Epitect Bisulfite Kit (Qiagen). Each region-of-interest was amplified from the bisulfite-treated DNA, and the individual PCR products were sequenced in more than quintuplicate. The sequences were analyzed and visualized using CyMATE (Hetzl et al., 2007).

Plant expression vector construction and transformation

To generate the *35S::NIC3* construct, the *NIC3* coding sequence was amplified from complementary DNA and cloned into the pGHX vector under

the control of cauliflower mosaic virus (CaMV) 35S promoter (Fujita et al., 2012). The verified construct was transformed into the Agrobacterium tumefaciens strain GV3101, and then transgenic Arabidopsis plants in the *ros1-4* background were generated by the Agrobacterium-mediated floral dip method. T1 plants were selected on MS medium containing 30 mg L^{-1} hygromycin with 25 mg L^{-1} cefotaxime. The verified T2 lines were used for further analyses. Oligos are listed in Table 1-1.

Table 1-1. List of primers.

Name	Sequence (5'→3')	Purpose
NIC3_qRT304f	CAAGTAACCGGACCAGACG	RT-qPCR
NIC3_qRT370r	GGAGGCGGGTTTTGTTAAA	RT-qPCR
ROS1_qRT3710f	TCTCCTGCAACAGCATCA	RT-qPCR
ROS1_qRT3847r	CATGATCCGCAAACACCT	RT-qPCR
UBQ10_561f	CGTTGACTGGGAAAACTATCACT	RT-qPCR
UBQ10_637r	GTCCTGGATCTTGGCTTTCA	RT-qPCR
ros1-3_AKFf	TGGAAGGGATCCGTCGTGGATTCT	Genotyping
ros1-3_AKFr	CCCGCGACTCTTGATTGTTTCAGCAACTT	Genotyping
ros1-4_SALKf	CCAGTTAAGGACAGAACACCG	Genotyping
ros1-4_SALKr	TCGTCTTTCGATCAAATCCAC	Genotyping
nic3_SALKf	CAGCATGACTTAGATGCTTTTAGC	Genotyping
nic3_SALKr	TACTGGGAACGGAAACGTATG	Genotyping
JL-202_AKF	CATTTTATAATAACGCTGCGGACATCTAC	Genotyping
LBb1.3_SALK	ATTTTGCCGATTTCGGAAC	Genotyping
NIC3_XbaIf	TTCTAGAATGGCTTCCTCATCAACG	Cloning
NIC3_XhoIr	TCTCGAGGTTACCGAGTAGACTTCGC	Cloning
ASA1pro_BSf	AATAGTAAGAATAATAGGAAGAGATATGTT	Local bisulfite sequencing
ASA1pro_BSr	CTCAATTTAATCTATAAACGAAACC	Local bisulfite sequencing
NIC3pro_BSf	T AGAGATATAATAATATGGTGAAGATAATATTG	Local bisulfite sequencing
NIC3pro_BSr	ATCTCACTTTCATATCTCATACTTTCTC	Local bisulfite sequencing
NIC3pro_BS2f	GATAGTGTTTTAATATAAATTAAAAATAGAGAA	Local bisulfite sequencing
NIC3pro_BS2r	CAATATTATCTTCACCATATTATTATATCTCTA	Local bisulfite sequencing
NIC3pro_BS3f	GTGTAAAGTAATATTGTTTYTTTATTATAGAATT	Local bisulfite sequencing
NIC3pro_BS3r	CTAAACCCTAATATATATTTTTCATATATATATC	Local bisulfite sequencing

RESULTS

The *ros1* mutants exhibit hypersensitivity to ABA

To determine ABA sensitivity, seed germination, early seedling development and root elongation assays have been widely used to identify genes involved in ABA signaling and ABA-mediated stress responses (Finkelstein, 2013; Yoshida et al., 2015; Zhu, 2016). Genetic screens were carried out on Arabidopsis T-DNA insertion mutants defective in DNA methylation and demethylation with regard to ABA responsiveness to examine the effect of DNA methylation on transcriptional control of ABA-responsive genes. Two *ros1* mutant lines *ros1-3* and *ros1-4* displayed hypersensitive phenotypes to ABA during early seeding establishment and root growth (Figure 1-1). Without ABA treatment, there was no significant difference between the WT and *ros1* mutants (Figures 1-1C to 1F). Upon ABA treatment, the WT successfully developed early seedlings and exhibited only slightly reduced primary root growth, in accordance with previous reports (Finkelstein, 2013; Yoshida et al., 2015). However, *ros1* mutants failed to establish early seedlings and showed a significant decrease in root

elongation in an ABA concentration-dependent manner (Figures 1-1C to 1F). These results indicated that ROS1 plays an important role in ABA response.

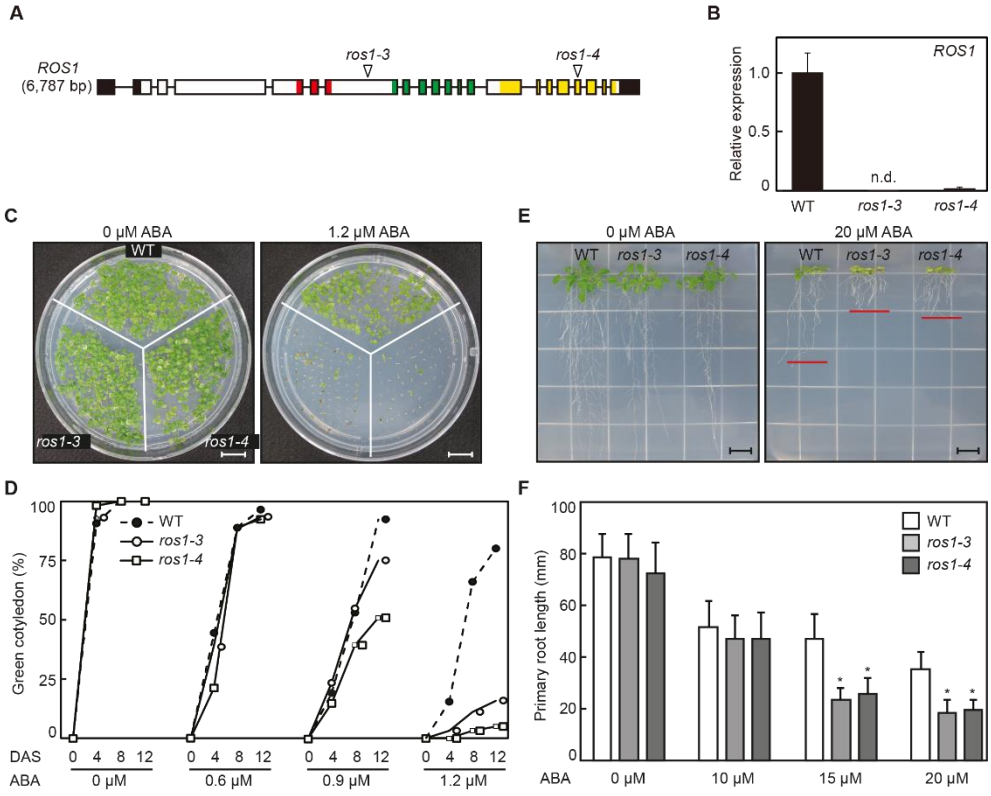

Figure 1-1. ABA hypersensitive phenotypes of *ros1* mutants.

(A) Schematic representation of the *ROS1* locus. Lines indicate introns, while closed and open boxes indicate untranslated regions and coding sequences, respectively. The relative locations of T-DNA insertions (*ros1-3* and *ros1-4*) are presented. Red, green and yellow regions indicate the catalytic essential domains. (B) Expression levels of *ROS1* in WT and two *ros1* lines as determined by RT-qPCR. Error bars indicate ± SD (n = 3). n.d.: not detected. (C) Early seeding establishment of the WT and *ros1* at 12 day-after-sowing (DAS) on MS medium in the absence or presence of ABA (D) Quantitative analysis of green cotyledon emergence at different ABA concentrations every 4 days. Data points are the percentage of green cotyledons against total seeds

on MS medium ($n > 80$). (E) Root elongation at 19 DAS on MS medium in the absence or presence of ABA. Red lines indicate average root lengths for each genotype. (F) Quantitative analysis of root growth at 19 DAS with different concentrations of ABA. Error bars indicate ± SD ($n = 4$). Asterisks indicate significant differences from WT (Student's *t*-test with Bonferroni-Holm correction; * $P < 0.05$). Scale bar = 10 mm.

The *ros1* mutation alters gene expression patterns upon ABA treatment

ROS1 represses transcriptional silencing of some endogenous genes such as *RD29A, EPF2*, and stress-responsive genes by direct removal of 5mCs (Gong et al., 2002; Le et al., 2014; Yamamuro et al., 2014). Based on the role of ROS1 in transcriptional regulation, it was hypothesized that some genes crucial for ABA response may be dysregulated in *ros1*, leading to increased sensitivity to ABA. RNA-seq analysis was conducted on the WT and *ros1* mutants with the criteria of a twofold change cutoff and a 5% false discovery rate. A total of 75 and 89 DEGs between the WT and *ros1-4* were identified in the absence and presence of ABA, respectively, among which nearly 80 percent of DEGs were down-regulated in *ros1-4* compared to the WT (Figures 1-2A and 2B). The *ros1-3* mutation also triggered a total of 113 and 140 DEGs in the absence and presence of ABA, and approximately 80 percent and 60 percent of them were down-regulated, respectively. As *EPF2* was down-regulated in *ros1-4*, consistent with the previous study (Yamamuro et al., 2014), but not in *ros1-3*, I focused on the transcriptome data of *ros1-4*.

To understand gene expression dynamics in response to ABA in the *ros1* background, hierarchical clustering analysis was performed. A total of 116 DEGs between the WT and *ros1-4* were grouped into six clusters (Figure 1-

2C). Cluster 1 (*n* = 12) and Cluster 2 (*n* = 9) are composed of the genes up-regulated by the *ros1-4* mutation in the ABA-dependent and -independent manners, respectively. Cluster 3 (*n* = 19) includes the genes showing ABA-inducible expression in the WT but not in *ros1* (Table 1-2). Cluster 4 (*n* = 38) represents the genes down-regulated in *ros1*, independent of ABA treatment. Cluster 5 (*n* = 31) involves the genes down-regulated by ABA in the WT but not in *ros1*. *EPF2* was also found in cluster 5. Cluster 6 (*n* = 7) includes the genes whose expression was repressed by ABA both in the WT and *ros1-4* mutant. As the candidate genes conferring ABA hypersensitivity in the *ros1* mutants might result from the loss of ABA inducibility, genes in cluster 3 were subjected to in-depth analysis.

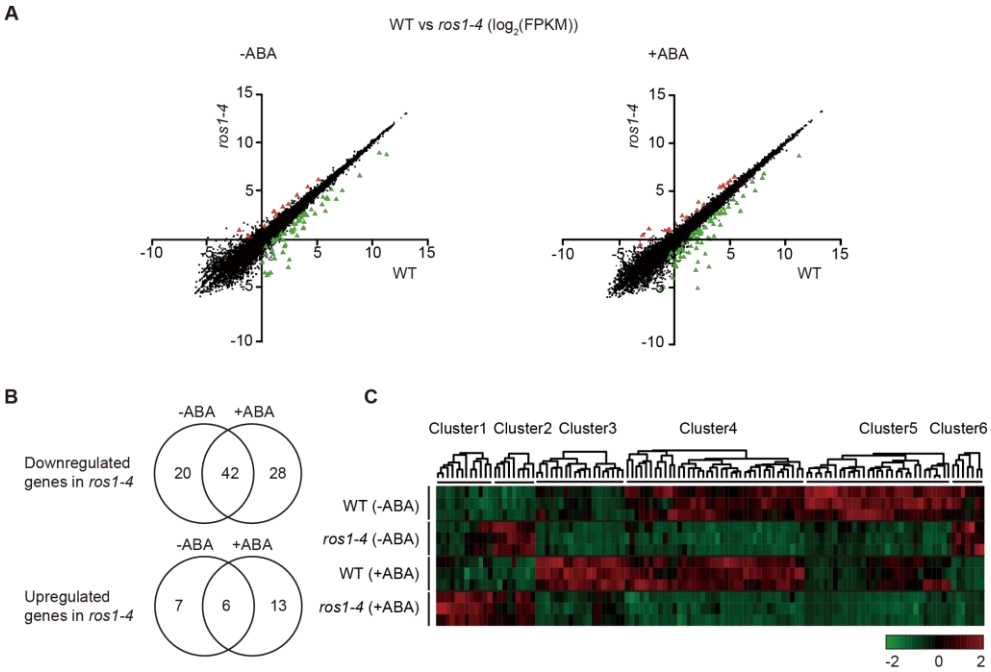

Figure 1-2. Transcriptome changes caused by the *ros1-4* mutation in the absence or presence of ABA.

(A) Scatter plot of gene expression levels in the WT and *ros1-4* as determined by RNA-seq. The mRNA expression level was calculated as $\log_2$-scaled fragments per kilobase of exon per million mapped fragments (FPKM). Red, green, and black dots indicate up-regulated genes, down-regulated genes, and non-DEGs, respectively. (B) The Venn diagram of down-regulated and up-regulated genes in *ros1-4*. (C) Hierarchical clustering of all 116 DEGs in *ros1-4* categorized into six clusters.

Table 1-2. List of a subset of genes in cluster 3.

Gene ID[a]	Symbol[a]	WT				*ros1-4*				Description[a]
		FPKM		FC	FDR	FPKM		FC	FDR	
		-ABA	+ABA			-ABA	+ABA			
AT1G58270[b]	ZW9	43.91	156.78	1.84	1.85E-55	12.36	33.48	1.44	3.31.E-24	TRAF-like family
AT2G02120[b]	LCR70	42.51	242.56	2.51	2.15E-28	33.13	82.88	1.32	4.32.E-08	Scorpion toxin-like knottin superfamily
AT5G22860[b]	-	7.40	12.24	0.73	4.61E-06	2.98	4.81	0.69	4.92.E-03	Ser carboxypeptidase S28 family
AT2G18193[b]	-	6.38	10.25	0.68	3.64E-02	3.27	3.53	0.11	1	P-loop containing nucleoside triphosphate hydrolases superfamily
AT5G23220[b]	NIC3	4.54	29.06	2.67	1.00E-31	5.72	10.17	0.83	3.38.E-03	Nicotinamidase 3
AT5G25120[b]	CYP71B11	2.56	5.11	1.00	4.14E-04	0.96	1.48	0.62	0.30	Cytochrome P450, family 71, subfamily B, polypeptide 11
AT1G43590[b]	-	1.40	2.30	0.72	0.62	0	0	0	1	Transposable element gene
AT2G01580	-	1.40	3.00	1.10	2.26E-02	0.63	1.01	0.67	0.70	Unknown
AT2G06002	-	1.10	2.69	1.28	0.11	0	0	0	1	Other RNA
AT1G52990[b]	-	0.35	1.72	2.25	4.71E-04	0.24	0.42	0.79	1	Thioredoxin family
AT5G08250[b]	-	0.30	1.00	1.72	4.40E-03	0.06	0.16	1.29	0.69	Cytochrome P450 superfamily
AT2G26750	-	0.17	1.08	2.59	1.40E-03	0.10	0.20	1.01	0.92	a/b-Hydrolases superfamily
AT5G28520	-	0.13	49.27	8.44	3.05E-128	0.07	13.63	7.34	1.81.E-59	Man-binding lectin superfamily

[a] Information is adopted from TAIR10 (http://arabidopsis.org/).

[b]Gene proximal to the *ros1*-hyper DMRs within 2 kb.

(–), no data; FC, log2-scaled fold change; FDR, false discovery rate; FPKM, fragments per kilobase of transcript per million mapped reads.

The *ros1* mutation contributes to DNA hypermethylation

To investigate genome-wide changes in DNA methylation patterns in *ros1* mutants, BS-seq analysis was performed. Differentially methylated cytosines (DMCs) and differentially methylated regions (DMRs) in *ros1-3* and *ros1-4* compared with the WT were identified. Both in *ros1-3* and *ros1-4*, consistent with the previous reports (Huang et al., 2013; Tang et al., 2016), the numbers of hypermethylated DMCs (hyper-DMCs) and hyper-DMRs were significantly higher than those of hypomethylated (hypo-) DMCs and DMRs, respectively (Figures 1-3C and 3D). Out of 171,550 common hyper-DMCs in *ros1-3* and *ros1-4* (*ros1* hyper-DMCs), CG hypermethylation predominantly occurred in *ros1* mutants (Figure 1-3A). Approximately 70% of the *ros1* hyper-DMCs were found in the CG context, 15 percent of which occurred in the CHG and CHH contexts, respectively (Figure 1-3A). I also identified 4,044 common hyper-DMRs in *ros1-3* and *ros1-4* (*ros1* hyper-DMRs; Figure 1-3B). The *ros1* hyper-DMRs in each sequence context significantly overlapped each other, and the number of CG and CHG DMRs was higher than that of CHH DMRs (Figure 1-3B). In accordance with the observation that 70% of the CG DMRs overlapped with CHG DMRs, CG and CHG DMRs displayed higher CHG and CG DNA methylation levels, respectively (Figures 1-3B and 3C). Hence, these results revealed that DNA

methylation levels of *ros1* mutants are substantially increased compared with the WT, a high portion of which occur in CG and CHG contexts. Furthermore, the distribution of *ros1* DMCs and DMRs across the genome indicated that the *ros1* hyper-DMCs and DMRs were found at high density on chromosome arms but low density in pericentromeric regions (Figures 1-3D). Consistent with the previous reports (Penterman et al., 2007; Qian et al., 2012), these observations demonstrated that ROS1 may act on discrete loci in euchromatin regions in the Arabidopsis genome.

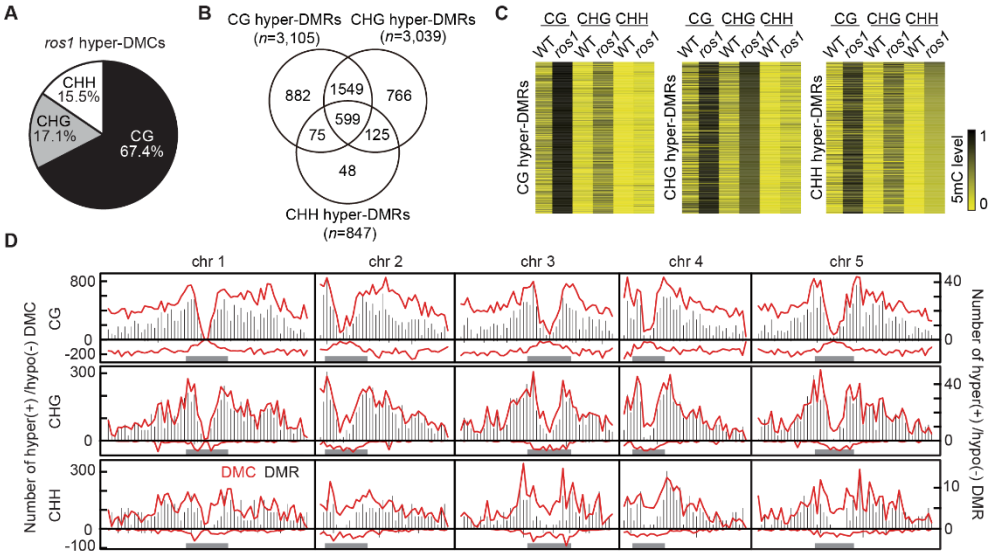

Figure 1-3. Genome-wide DNA methylation changes in *ros1* mutants.

(A) Fraction of the *ros1* hyper-DMCs as determined by BS-seq. (B) Overlap of the *ros1* hyper-DMRs in each sequence context. (C) Heat maps of DNA methylation levels for the *ros1* hyper-DMRs. Each horizontal line indicates the 5mC level of single DMR. (D) Distribution of the *ros1* DMCs and DMRs along the Arabidopsis chromosomes. The red lines and black bars are, respectively, the numbers of DMCs and DMRs in a bin (500 kb). The positive and negative numbers represent the counts of hyper- and hypo-DMCs/DMRs, respectively. Gray bars indicate pericentromeric regions.

Down-regulation of ABA-inducible genes in *ros1* is associated with an increase in DNA methylation levels

To examine the relationship between the *ros1*–down-regulated genes and the *ros1*-hyper DMRs, the number of genes proximal to the *ros1*-hyper DMRs within 2 kb were counted. The *ros1*-hyper DMRs were located in proximity to 5,692 genes (18.3% of Arabidopsis coding genes), among which 52 genes showed decreased expression in *ros1* compared with the WT. Interestingly, 63.2% (12 of 19) of the genes in cluster 3 were proximal to the *ros1*-hyper DMRs (Table 1-2). A subset of genes in cluster 3 exhibited ABA inducibility in the WT, but the induction was abolished in *ros1* (Figure 1-4A). Moreover, in the absence of ABA, *AT2G02120* and *NIC3* showed similar expression levels in the WT and *ros1* mutants, but in the presence of ABA, they were significantly down-regulated in *ros1* (Figure 1-4A). The *ros1*-hyper DMRs were located in the upstream regions of *AT2G02120* and *NIC3* and in the downstream regions of *AT1G58270* and *AT2G43670*, respectively, demonstrating that DNA methylation levels of the proximal regions were increased in *ros1* compared with the WT (Figure 1-4B). Therefore, these results indicate that loss of ABA-inducibility of the genes in cluster 3 highly correlates with DNA hypermethylation, resulting in ABA hypersensitivity of the *ros1* mutants.

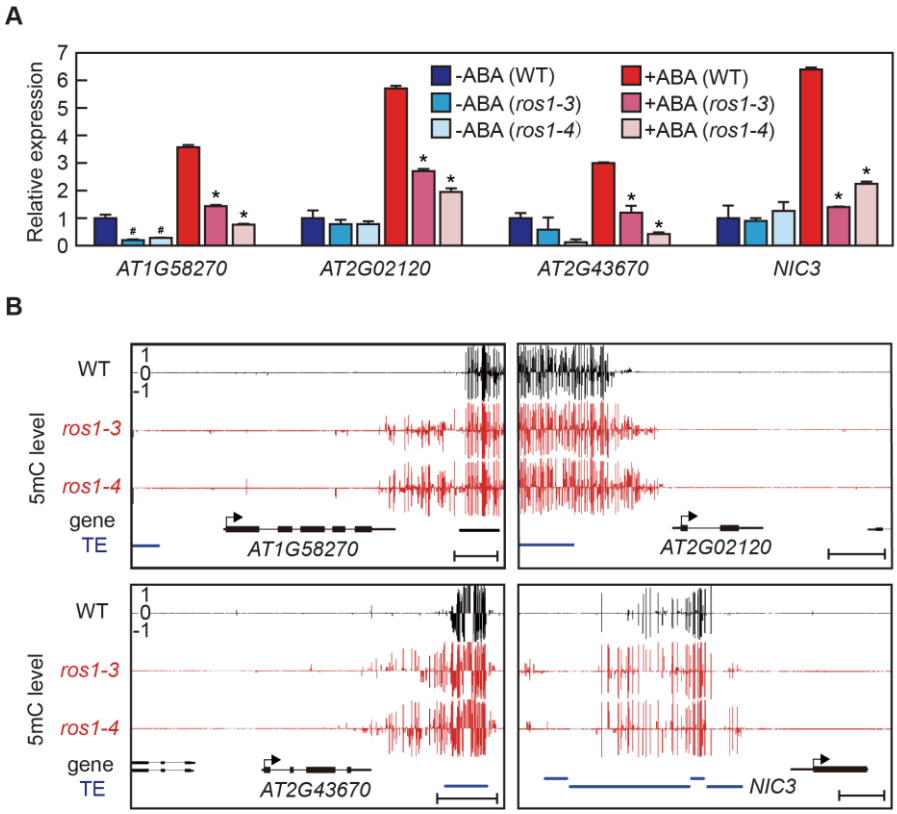

Figure 1-4. Association of decreased expression with DNA hypermethylation in the cluster 3 genes.

(A) The $\log_2$-scaled fold change in gene expression of the representative genes in cluster 3. Error bars indicate ± SD (n = 3). Hashes and asterisks indicate significant differences from the WT in the absence and presence of ABA, respectively (Student's t-test with Bonferroni-Holm correction; *P < 0.05). (B) Genome browser views of the DNA methylation levels at the representative genes in cluster 3 and their proximal genomic regions. The left side is the upstream region of the locus. Positive and negative bars indicate 5mC levels of single cytosine on the Watson (+1) and Crick (-1) strands, respectively. Guide bar = 500 bp.

DNA hypermethylation at the *NIC3* promoter is accompanied with decreased *NIC3* expression

The RNA-seq and BS-seq analysis revealed that several ABA-inducible genes were down-regulated in *ros1* accompanied by excessive DNA methylation. Among the candidate genes in cluster 3, I focused on *NIC3* which encodes an enzyme that converts Nam to nicotinic acid in the NAD^+ salvage pathway (Figures 1-4A and 4B). Yeast nicotinamidase Pnc1 is known as an activator of Silent information regulator 2 (Sir2), a NAD^+-dependent histone deacetylase crucial for transcriptional silencing at the rDNA locus in response to calorie restriction, heat and salt stresses (Anderson et al., 2003; Gallo et al., 2004). In Arabidopsis, three nicotinamidase genes *NIC1*, *NIC2*, and *NIC3* are designated as the homologs of yeast *PNC1* (Hunt et al., 2007). The *nic1* and *nic2* mutants exhibited hypersensitivity to ABA, salt stress and Nam, indicating that nicotinamidases are important for ABA-mediated osmotic stress responses (Hunt et al., 2007; Wang and Pichersky, 2007). Despite its high sequence similarity to *NIC2*, the biological roles of NIC3 are poorly understood. The RT-qPCR analysis revealed that there was no significant difference in *NIC3* expression between the WT and *ros1* mutants in the absence of ABA (Figure 1-5A). In contrast, in the presence of ABA, *NIC3* expression was significantly induced in WT but to a

less extent in *ros1* (Figure 1-5A). In addition, tissue-specific expression analysis showed that *NIC3* expression was most abundant in the roots, followed by imbibed seeds, flower buds and siliques, and relatively low in rosette leaves and cauline leaves (Figure 1-5B).

DNA methylation of the upstream region of *NIC3* promoter was further examined for its functional implication in the down-regulation of ABA-inducible expression of *NIC3* in *ros1* mutants. The upstream region of the *NIC3* promoter contains four TEs, two of which harbor the *ros1*-hyper DMRs (Figure 1-5C). The Local bisulfite sequencing analysis revealed that the first DMR (DMR I) at the RC/Helitron family of TE (AT5TE28305) showed increased DNA methylation levels for all sequence contexts, and the second DMR (DMR II) at the DNA/MuDR family of TE (AT5TE28295) exhibited an increase of DNA methylation only in CG context (Figure 1-5D). These findings strongly indicate that excessive DNA methylation of TEs at the upstream *NIC3* promoter contributes to the suppression of its expression and ABA hypersensitive phenotypes in *ros1* mutants.

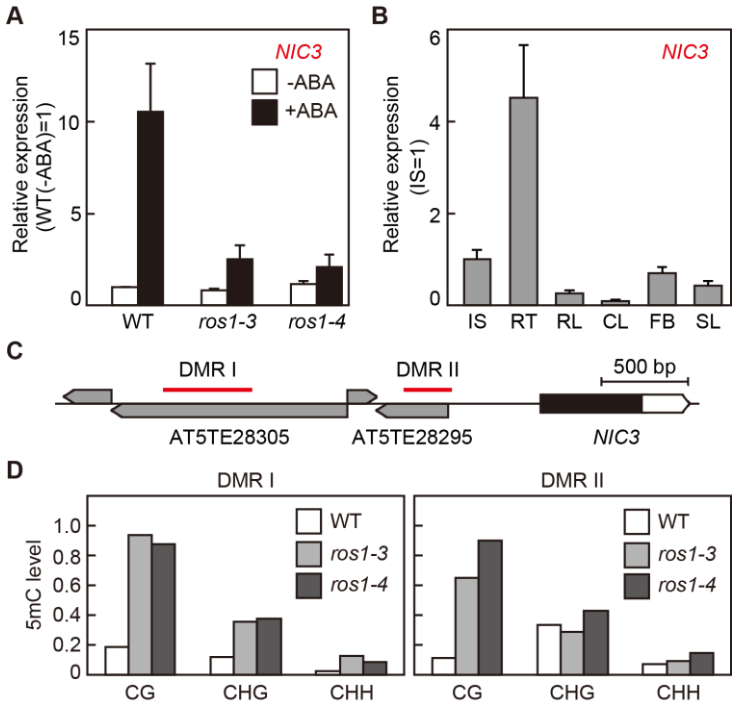

Figure 1-5. Suppression of ABA-inducible *NIC3* expression by DNA hypermethylation at the upstream region of the *NIC3* promoter in *ros1*.

(A) Relative expression levels of *NIC3* in the WT and *ros1* in the absence and presence of ABA as determined by RT-qPCR. Error bars indicate ± SD (*n* = 3). (B) Relative expression levels of *NIC3* in 24-h imbibed seeds (IS), roots (RT), rosette leaves (RL), cauline leaves (CL), flower buds (FB), and young siliques (SL) of 5-week-old Arabidopsis plants by RT-qPCR analysis. Error bars indicate ± SD (*n* = 3). (C) Schematic diagram of the *NIC3* locus. Lines indicate intergenic regions, whereas white and black boxes indicate untranslated regions and coding regions, respectively. Relative locations of the promoter TEs (gray boxes) and DMRs (red bars) are presented. (D) DNA methylation levels at the DMR I and DMR II in the WT and *ros1* as determined by local bisulfite sequencing.

Ectopic expression of *NIC3* mitigates ABA hypersensitivity of *ros1*

To further examine whether *NIC3* expression is necessary for ABA responses, two T-DNA insertion mutant lines *nic3-1* and *nic3-2* were subjected to ABA treatment in early seedling establishment and primary root growth (Figures 1-6A and 6B). During early seedling development, *nic3* mutants exhibited increased sensitivity to ABA in comparison with the WT, albeit their phenotypes were weaker than *ros1* (Figures 1-6C and 6D). Furthermore, primary root growth of *nic3* mutants was suppressed upon ABA treatment, similar to *ros1* (Figures 1-6E and 6F), suggesting that the down-regulation of *NIC3* in *ros1* mutants may be responsible for ABA hypersensitive phenotypes.

I next investigated whether ectopic expression of *NIC3* in *ros1* could restore the normal growth in response to ABA. Two independent lines of T2 transgenic plants were generated in the *ros1-4* background by overexpression of *NIC3* under the control of the constitutive CaMV 35S promoter (*NIC3*ox-1 and *NIC3*ox-2; Figures 1-7A and 7B). In the presence of ABA, *NIC3*ox lines restored normal early seeding development and root growth in a manner similar to the WT (Figures 1-7C to 7F). These observations demonstrate that ectopic expression of *NIC3* was able to mitigate the hypersensitivity to ABA

during early seedling establishment and root elongation, suggesting that ROS1-mediated DNA demethylation maintains the transcriptionally active state of *NIC3* in response to ABA.

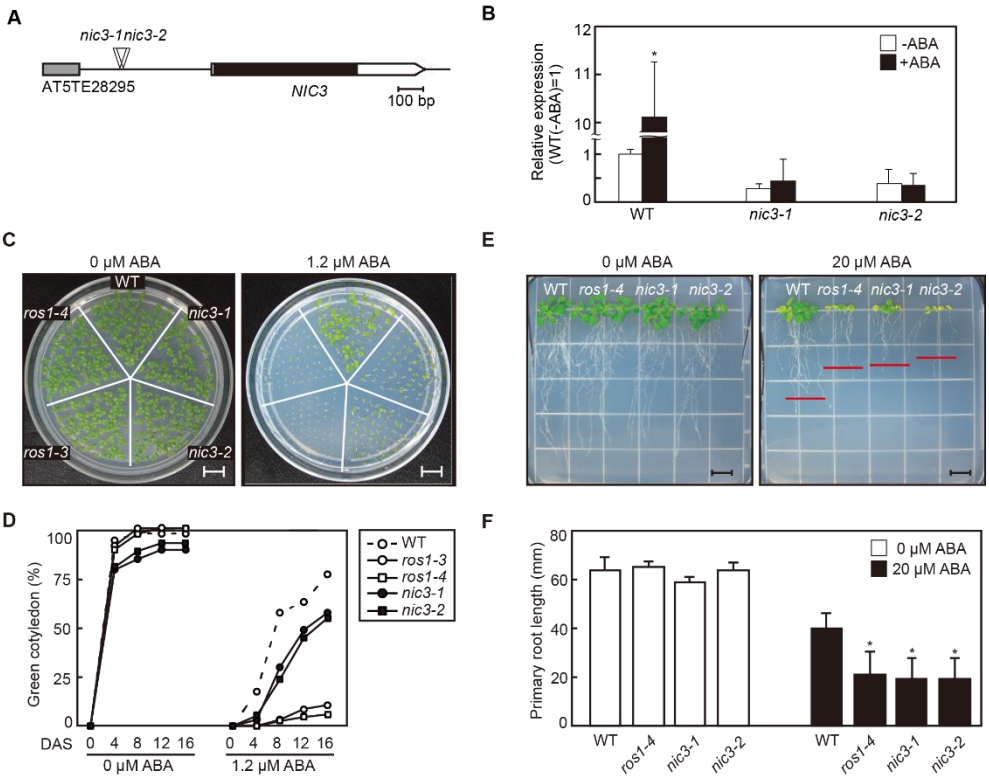

Figure 1-6. ABA hypersensitive phenotypes of *nic3* mutants.

(A) Schematic representation of the *NIC3* locus. Lines indicate intergenic regions while closed and open boxes indicate untranslated regions and coding sequences, respectively. The relative locations of T-DNA insertions (*nic3-1* and *nic3-2*) and the promoter TEs (gray boxes) are presented. (B) Expression levels of *NIC3* in the WT and *nic3* as determined by RT-qPCR. Error bars indicate ± SD (n = 3). The asterisk indicates a significant difference between the absence and presence of ABA (Student's t-test with Bonferroni-Holm correction; *P < 0.01). (C) Early seedling establishment of the WT, *ros1*, and *nic3* at 12 DAS on MS medium in the absence or presence of ABA. (D) Quantitative analysis of green cotyledon emergence every 4 days. Data points

are the percentage of green cotyledons against total seeds on MS medium (n > 70). (E) Root elongation at 19 DAS on MS medium in the absence or presence of ABA. Red lines indicate average root lengths for each genotype. (F) Quantitative analysis of root growth at 19 DAS. Error bars indicate ± SD (n = 4). Asterisks indicate significant differences from the WT (Student's t test with Bonferroni-Holm correction; *P < 0.05). Scale bar = 10 mm.

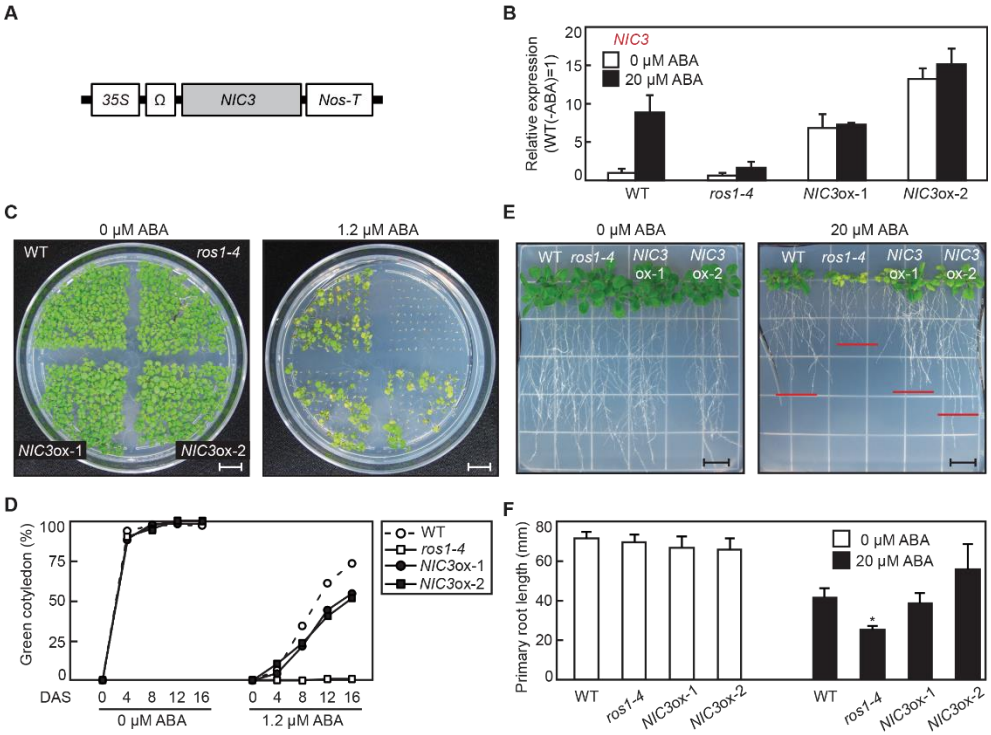

Figure 1-7. Restored normal growth by ectopic expression of *NIC3* in *ros1* upon ABA treatment.

(A) Schematic diagram of constitutive *NIC3* expression by the CaMV 35S promoter. (B) Relative expression of *NIC3* in the WT, *ros1-4*, and two *NIC3*ox lines in the *ros1-4* background (*NIC3*ox-1,2). Error bars indicate ± SD (n = 3). (C) Early seedling establishment at 15 or 17 DAS on MS medium in the absence or presence of ABA, respectively. (D) Quantitative analysis of green cotyledon emergence every 4 days. Data points are the percentage of green cotyledons against total seeds on MS medium (n > 70). (E) Root elongation at 18 or 23 DAS on MS medium in the absence or presence of ABA, respectively. Red lines indicate average root lengths for each genotype. (F) Quantitative analysis of root growth at 19 DAS. Error bars

indicate ± SD (n = 4). Asterisks indicate significant differences from the WT (Student's t-test with Bonferroni-Holm correction; *P < 0.05). Scale bar = 10 mm.

DISCUSSION

DNA methylation at TEs is generally associated with transcriptional repression of proximal genes (Diez et al., 2014; Lisch, 2013; Slotkin and Martienssen, 2007). A member of the Arabidopsis DNA demethylase family, ROS1 has been proposed to target specific genomic regions rather than act on random regions (Tang et al., 2016; Zhu et al., 2007). Previous reports revealed that ROS1-initiated DNA demethylation is responsible for the expression of *EPF2* and stress-responsive genes for stomatal development and *F. oxysporum* tolerance, respectively, by preventing DNA methylation spreading from TEs to nearby genes (Le et al., 2014; Yamamuro *et al.*, 2014). Interestingly, this study indicates that the ABA-inducible expression of *NIC3* requires ROS1-mediated DNA demethylation at the TEs in its promoter region, suggesting the role of NAD^+ biosynthesis and Nam turnover in ABA responses.

ROS1 maintains the transcriptionally active states of the ABA- and stress-responsive *RD29A::LUC* transgene and the endogenous *RD29A* gene, indicating the function of ROS1 in ABA-mediated stress responses (Gong et al., 2002). Without ABA treatment, the *ros1* mutants did not show any developmental defects (Figure 1-1), in accordance with the previous report

(Gong *et al.*, 2002). However, upon ABA treatment, the *ros1* mutants exhibited increased sensitivity to ABA during early seedling development and root growth in an ABA-concentration dependent manner, demonstrating that ROS1-initiated active DNA demethylation is crucial for ABA responses (Figure 1-1). Albeit the detailed mechanism by which ROS1 perceives and copes with environmental stresses including ABA hormone and osmotic stress remains to be elucidated, ROS1-dependent DNA demethylation may transmit environmental cues to the genome, thereby leading to prompt responsiveness to the plant cell.

The transcriptome and methylome analysis indicated that a majority of DEGs were down-regulated in *ros1-4* compared with the WT, associated with DNA hypermethylation in nearby regions (Figures 1-2 and 1-3). Interestingly, more than 60% of the cluster 3 genes that lost ABA inducibility in *ros1* mutants were found to be proximal to the *ros1* hyper-DMRs, suggesting a close relationship between transcriptional silencing and excessive DNA methylation (Table 1-2). In particular, the *nic3* mutants were hypersensitive to ABA (Figure 1-5), and the TEs in the upstream region of the *NIC3* promoter (AT5TE28305 and AT5TE28295) displayed increased DNA methylation levels in the *ros1* mutants, thus resulting in a decrease in *NIC3* expression (Figure 1-6). Taken these together, ROS1 is suggested to prevent

DNA methylation spreading from nearby TEs to the *NIC3* promoter, ensuring the ABA-inducible expression of *NIC3*. However, considering the numerous *ros1*-hyper DMRs, I cannot rule out the possibility that ROS1 activates the factor(s) involved in ABA-dependent induction of *NIC3*, rather than its direct activation. Furthermore, among the total of 1,489 genes up-regulated by ABA treatment in the WT, most of them (1,203 of 1,489) still showed normal ABA inducible expression in *ros1-4*. Approximately 84% (1,009 of 1,203) of these genes whose expression levels were increased upon ABA treatment both in the WT and *ros1-4* were not located proximal to the *ros1* hyper-DMRs. The normal ABA induction of well-known ABA-inducible genes such as *RD29B* and *MYB DOMAIN PROTEIN 41* in *ros1* mutants without nearby hyper-DMRs also suggested that ABA inducibility can be controlled in a ROS1-independent manner.

Nicotinamidase has the catalytic activity that converts Nam into nicotinic acid in the NAD^+ salvage pathway. Many enzymes in the NAD^+ salvage pathway including poly(ADP-ribose) polymerase and nicotinate/nicotinamide mononucleotide adenyltransferase are reported to be involved in stress responses (De Block et al., 2005; Hashida et al., 2010; Vanderauwera et al., 2007). This study suggests that the dynamic regulation of Nam and NAD^+ might contribute to ABA-dependent developmental

processes. Further examination of the link between the NAD^+ salvage pathway and the ABA response will uncover the connection between plant stress hormone signaling and NAD^+ metabolism. In addition, a homolog of NIC3 in yeast, Pnc1, is an activator of NAD^+-dependent histone deacetylase Sir2 and functions to lower the cellular Nam levels, which leads to lifespan extension in response to calorie restriction, heat, and hyperosmotic stresses (Anderson et al., 2003; Gallo et al., 2004). It remains unclear whether NIC3 plays a regulatory role similar to Pnc1 in the regulation of SIR2 or SIR2-like factors in plant ABA responses. By contrast with yeasts that do not employ DNA methylation for epigenetic control, plants may engage ROS1-dependent DNA demethylation for ABA responses, allowing rapid adaptation to environmental changes in a rather dynamic manner.

REFERENCES

Agius, F., Kapoor, A., and Zhu, J.K. (2006). Role of the *Arabidopsis* DNA glycosylase/lyase ROS1 in active DNA demethylation. Proc. Natl. Acad. Sci. U. S. A. *103*, 11796-11801.

Anderson, R.M., Bitterman, K.J., Wood, J.G., Medvedik, O., and Sinclair, D.A. (2003). Nicotinamide and *PNC1* govern lifespan extension by calorie restriction in *Saccharomyces cerevisiae*. Nature *423*, 181-185.

Chen, L.T., Luo, M., Wang, Y.Y., and Wu, K. (2010). Involvement of *Arabidopsis* histone deacetylase HDA6 in ABA and salt stress response. J. Exp. Bot. *61*, 3345-3353.

Chinnusamy, V., and Zhu, J.K. (2009). Epigenetic regulation of stress responses in plants. Curr. Opin. Plant Biol. *12*, 133-139.

Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J.J., Goldberg, R.B., Jacobsen, S.E., and Fischer, R.L. (2002). DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*. Cell *110*, 33-42.

Cui, X., Jin, P., Cui, X., Gu, L., Lu, Z., Xue, Y., Wei, L., Qi, J., Song, X., Luo, M., et al. (2013). Control of transposon activity by a histone H3K4 demethylase in rice. Proc. Natl. Acad. Sci. U. S. A. *110*, 1953-1958.

De Block, M., Verduyn, C., De Brouwer, D., and Cornelissen, M. (2005). Poly(ADP-ribose) polymerase in plants affects energy homeostasis, cell death and stress tolerance. Plant J. *41*, 95-106.

Diez, C.M., Roessler, K., and Gaut, B.S. (2014). Epigenetics and plant genome evolution. Curr. Opin. Plant Biol. *18*, 1-8.

Dowen, R.H., Pelizzola, M., Schmitz, R.J., Lister, R., Dowen, J.M., Nery,

J.R., Dixon, J.E., and Ecker, J.R. (2012). Widespread dynamic DNA methylation in response to biotic stress. Proc. Natl. Acad. Sci. U. S. A. *109*, E2183-2191.

Finkelstein, R. (2013). Abscisic acid synthesis and response. Arabidopsis Book *11*, e0166.

Fujita, M., Fujita, Y., Iuchi, S., Yamada, K., Kobayashi, Y., Urano, K., Kobayashi, M., Yamaguchi-Shinozaki, K., and Shinozaki, K. (2012). Natural variation in a polyamine transporter determines paraquat tolerance in *Arabidopsis*. Proc. Natl. Acad. Sci. U. S. A. *109*, 6343-6347.

Gallo, C.M., Smith, D.L., Jr., and Smith, J.S. (2004). Nicotinamide clearance by Pnc1 directly regulates Sir2-mediated silencing and longevity. Mol. Cell Biol. *24*, 1301-1312.

Ganguly, D.R., Crisp, P.A., Eichten, S.R., and Pogson, B.J. (2017). The Arabidopsis DNA methylome is stable under transgenerational drought stress. Plant Physiol. *175*, 1893-1912.

Gehring, M. (2013). Genomic imprinting: insights from plants. Annu. Rev. Genet. *47*, 187-208.

Gehring, M., Huh, J.H., Hsieh, T.F., Penterman, J., Choi, Y., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (2006). DEMETER DNA glycosylase establishes *MEDEA* polycomb gene self-imprinting by allele-specific demethylation. Cell *124*, 495-506.

Gong, Z., Morales-Ruiz, T., Ariza, R.R., Roldan-Arjona, T., David, L., and Zhu, J.K. (2002). *ROS1*, a repressor of transcriptional gene silencing in *Arabidopsis*, encodes a DNA glycosylase/lyase. Cell *111*, 803-814.

Hashida, S.N., Itami, T., Takahashi, H., Takahara, K., Nagano, M., Kawai-Yamada, M., Shoji, K., Goto, F., Yoshihara, T., and Uchimiya,

H. (2010). Nicotinate/nicotinamide mononucleotide adenyltransferase-mediated regulation of NAD biosynthesis protects guard cells from reactive oxygen species in ABA-mediated stomatal movement in *Arabidopsis*. J. Exp. Bot. *61*, 3813-3825.

Heard, E., and Martienssen, R.A. (2014). Transgenerational epigenetic inheritance: myths and mechanisms. Cell *157*, 95-109.

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

Huang, C.F., Miki, D., Tang, K., Zhou, H.R., Zheng, Z., Chen, W., Ma, Z.Y., Yang, L., Zhang, H., Liu, R., et al. (2013). A pre-mRNA-splicing factor is required for RNA-directed DNA methylation in *Arabidopsis*. PLoS Genet. *9*, e1003779.

Huh, J.H., Bauer, M.J., Hsieh, T.F., and Fischer, R.L. (2008). Cellular programming of plant gene imprinting. Cell *132*, 735-744.

Hunt, L., Holdsworth, M.J., and Gray, J.E. (2007). Nicotinamidase activity is important for germination. Plant J. *51*, 341-351.

Kinoshita, T., and Seki, M. (2014). Epigenetic memory for stress response and adaptation in plants. Plant Cell Physiol. *55*, 1859-1863.

Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. *11*, 204-220.

Le, T.N., Schumann, U., Smith, N.A., Tiwari, S., Au, P.C., Zhu, Q.H., Taylor, J.M., Kazan, K., Llewellyn, D.J., Zhang, R., et al. (2014). DNA demethylases target promoter transposable elements to positively regulate stress responsive genes in *Arabidopsis*. Genome Biol. *15*, 458.

Lisch, D. (2013). How important are transposons for plant evolution? Nat.

Rev. Genet. *14*, 49-61.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008). Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. Cell *133*, 523-536.

Liu, Y., Siegmund, K.D., Laird, P.W., and Berman, B.P. (2012). Bis-SNP: combined DNA methylation and SNP calling for Bisulfite-seq data. Genome Biol. *13*, 1-14.

Matzke, M.A., and Mosher, R.A. (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat. Rev. Genet. *15*, 394-408.

Mirouze, M., and Paszkowski, J. (2011). Epigenetic contribution to stress adaptation in plants. Curr. Opin. Plant Biol. *14*, 267-274.

Mittler, R., and Blumwald, E. (2015). The roles of ROS and ABA in systemic acquired acclimation. Plant Cell *27*, 64-70.

Morales-Ruiz, T., Ortega-Galisteo, A.P., Ponferrada-Marin, M.I., Martinez-Macias, M.I., Ariza, R.R., and Roldan-Arjona, T. (2006). *DEMETER* and *REPRESSOR OF SILENCING 1* encode 5-methylcytosine DNA glycosylases. Proc. Natl. Acad. Sci. U. S. A. *103*, 6853-6858.

Nambara, E., and Marion-Poll, A. (2005). Abscisic acid biosynthesis and catabolism. Annu. Rev. Plant. Biol. *56*, 165-185.

Ortega-Galisteo, A.P., Morales-Ruiz, T., Ariza, R.R., and Roldan-Arjona, T. (2008). Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks. Plant Mol. Biol. *67*, 671-681.

Pedersen, B.S., Eyring, K., De, S., Yang, I.V., and Schwartz, D.A. (2014). Fast and accurate alignment of long bisulfite-seq reads. arXiv preprint

arXiv:1401.1129.

Penterman, J., Zilberman, D., Huh, J.H., Ballinger, T., Henikoff, S., and Fischer, R.L. (2007). DNA demethylation in the *Arabidopsis* genome. Proc. Natl. Acad. Sci. U. S. A. *104*, 6752-6757.

Qian, W., Miki, D., Zhang, H., Liu, Y., Zhang, X., Tang, K., Kan, Y., La, H., Li, X., Li, S., et al. (2012). A histone acetyltransferase regulates active DNA demethylation in *Arabidopsis*. Science *336*, 1445-1448.

Secco, D., Wang, C., Shou, H., Schultz, M.D., Chiarenza, S., Nussaume, L., Ecker, J.R., Whelan, J., and Lister, R. (2015). Stress induced gene expression drives transient DNA methylation changes at adjacent repetitive elements. Elife *4*, e09343.

Slotkin, R.K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. Nat. Rev. Genet. *8*, 272-285.

Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nat. Rev. Genet. *14*, 204-220.

Sridha, S., and Wu, K. (2006). Identification of *AtHD2C* as a novel regulator of abscisic acid responses in Arabidopsis. Plant J. *46*, 124-133.

Stroud, H., Greenberg, M.V., Feng, S., Bernatavichute, Y.V., and Jacobsen, S.E. (2013). Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. Cell *152*, 352-364.

Tang, K., Lang, Z., Zhang, H., and Zhu, J.K. (2016). The DNA demethylase ROS1 targets genomic regions with distinct chromatin modifications. Nat. Plants *2*, 16169.

Vanderauwera, S., De Block, M., Van de Steene, N., van de Cotte, B., Metzlaff, M., and Van Breusegem, F. (2007). Silencing of poly(ADP-ribose) polymerase in plants alters abiotic stress signal transduction. Proc.

Natl. Acad. Sci. U. S. A. *104*, 15150-15155.

Wang, G., and Pichersky, E. (2007). Nicotinamidase participates in the salvage pathway of NAD biosynthesis in Arabidopsis. Plant J. *49*, 1020-1029.

Wibowo, A., Becker, C., Marconi, G., Durr, J., Price, J., Hagmann, J., Papareddy, R., Putra, H., Kageyama, J., Becker, J., et al. (2016). Hyperosmotic stress memory in Arabidopsis is mediated by distinct epigenetically labile sites in the genome and is restricted in the male germline by DNA glycosylase activity. Elife *5*, e13546.

Wu, H., and Zhang, Y. (2014). Reversing DNA methylation: mechanisms, genomics, and biological functions. Cell *156*, 45-68.

Yamaguchi-Shinozaki, K., and Shinozaki, K. (2006). Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. Annu. Rev. Plant Biol. *57*, 781-803.

Yamamuro, C., Miki, D., Zheng, Z., Ma, J., Wang, J., Yang, Z., Dong, J., and Zhu, J.K. (2014). Overproduction of stomatal lineage cells in *Arabidopsis* mutants defective in active DNA demethylation. Nat. Commun. *5*, 4062.

Yoshida, T., Mogami, J., and Yamaguchi-Shinozaki, K. (2015). Omics approaches toward defining the comprehensive abscisic acid signaling network in plants. Plant Cell Physiol. *56*, 1043-1052.

Zhang, H., Lang, Z., and Zhu, J.K. (2018). Dynamics and function of DNA methylation in plants. Nat. Rev. Mol. Cell Biol. *19*, 489-506.

Zhang, H., and Zhu, J.K. (2012). Active DNA demethylation in plants and animals. Cold Spring Harb. Symp. Quant. Biol. *77*, 161-173.

Zhong, S., Joung, J.G., Zheng, Y., Chen, Y.R., Liu, B., Shao, Y., Xiang, J.Z., Fei, Z., and Giovannoni, J.J. (2011). High-throughput illumina

strand-specific RNA sequencing library preparation. Cold Spring Harb. Protoc. *2011*, 940-949.

Zhu, J., Kapoor, A., Sridhar, V.V., Agius, F., and Zhu, J.K. (2007). The DNA glycosylase/lyase ROS1 functions in pruning DNA methylation patterns in *Arabidopsis*. Curr. Biol. *17*, 54-59.

Zhu, J.K. (2009). Active DNA demethylation mediated by DNA glycosylases. Annu. Rev. Genet. *43*, 143-166.

Zhu, J.K. (2016). Abiotic stress signaling and responses in plants. Cell *167*, 313-324.

CHAPTER 2

Regulation of Reproductive Development by DNA Demethylases in Arabidopsis

ABSTRACT

DNA methylation is a key epigenetic mark for various cellular processes. In plants, active DNA demethylation is achieved by the DEMETER (DME) family of 5-methylcytosine glycosylases along with REPRESSOR OF SILENCING 1 (ROS1), DEMETER-LIKE 2 (DML2) and DML3. Although a genetic analysis of Arabidopsis mutants deficient in all DNA demethylases is necessary, the generation of *dme* homozygous mutants has been hindered by embryo lethality, resulting from the maternal *dme* allele-derived seed abortion. Here, I isolated the homozygous *dme* and *dme ros1 dml2 dml3* (*drdd*) mutants by *in vitro* early seed rescue to elucidate the biological functions of the entire DME family. The *drdd* mutant exhibited several developmental defects such as retarded growth, delayed floral transition, abnormal flower development and the formation of short unfertilized siliques. Transcriptome and methylome analysis revealed a significant decrease in gene expression levels accompanied by DNA hypermethylation in *drdd*, albeit to a lesser extent in *dme* and *rdd* compared to WT. These findings indicate that DNA demethylases act redundantly for the removal of DNA methylation required for proper transcriptional regulation. A total of 703 downregulated genes with increased DNA methylation at their promoters were identified, and moreover,

a subset of these genes were highly associated with cell wall expansion during pollen tube growth and morphogenesis involved in differentiation. These results suggest that the combined activity of DNA demethylases would be therefore responsible for ensuring normal fertilization and floral organ development.

INTRODUCTION

DNA methylation is a reversible but relatively stable epigenetic mark that is crucial for diverse biological processes such as cell differentiation, gene imprinting, transposon silencing and X chromosome inactivation (Huh et al., 2008; Law and Jacobsen, 2010; Smith and Meissner, 2013). Whereas DNA methylation is predominantly found in the symmetric CG context in mammals, DNA methylation in plants occurs in all sequence contexts: CG, CHG and CHH (H = A, T or C) (Law and Jacobsen, 2010). In plants, DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) catalyzes the establishment of de novo DNA methylation via a small RNA-guided pathway, referred to as RNA-directed DNA methylation (RdDM) (Matzke and Mosher, 2014). Once DNA methylation is established, different DNA methyltransferases are responsible for the maintenance of DNA methylation in each sequence context. Symmetric CG and CHG methylation are maintained by DNA METHYLTRANSFERASE 1 (MET1) and CHROMOMETHYLASE 3 (CMT3), respectively, and the maintenance of asymmetric CHH methylation depends on the activity of DRM2 as well as CMT2 (Law and Jacobsen, 2010; Zhang et al., 2018).

DNA demethylation can occur passively in a replication-dependent manner or actively through the enzymatic process. In plants, active DNA demethylation can be achieved by the plant-specific DNA glycosylases/lyases that directly excise 5-methylcytosine (5mC) and cleave the sugar-phosphate backbone creating an abasic site, followed by the base excision repair pathway (Choi et al., 2002; Gehring et al., 2006; Gong et al., 2002; Penterman et al., 2007). The Arabidopsis genome has four DNA glycosylase family members – *DEMETER* (*DME*), *REPRESSOR OF SILENCING 1* (*ROS1*), *DEMETER-LIKE 2* (*DML2*) and *DML3* (Penterman et al., 2007). *DME* is preferentially expressed in the central cell of the female gametophyte, which is required for the establishment of gene imprinting in the endosperm (Choi et al., 2002; Gehring et al., 2006). The inheritance of a maternal *dme* allele causes seed abortion, due to the failure in activating the maternal alleles of *MEDEA* (*MEA*), *FLOWERING WAGENINGEN* (*FWA*) and *FERTILIZATION INDEPENDENT SEED 2* (*FIS2*) (Gehring, 2013; Huh et al., 2008). *DME* is also expressed in the vegetative cell of the male gametophyte as well as in shoot/root apical meristems (Kim et al., 2008; Park et al., 2017), and recently, the *dme* homozygous mutant in the Landsberg *erecta* (L*er*) background was generated with the aberrant shoot and root phenotypes (Kim et al., 2021). Unlike *DME*, three DNA demethylase genes

ROS1, *DML2* and *DML3* are expressed in the vegetative tissue, and the *ros1 dml2 dml3* (*rdd*) triple mutant has no obvious developmental defects under normal growth conditions (Penterman et al., 2007). However, the *ros1* mutants exhibit abnormal stomata development due to the downregulation of *EPIDERMAL PATTERNING FACTOR 2* (*EPF2*) (Yamamuro et al., 2014). The *ros1* mutants also display hypersensitivity to abscisic acid, which is associated with the decreased expression of *NICOTINAMIDASE 3* (*NIC3*) (Kim et al., 2019b). The *ros1* and *rdd* mutants also display impaired xylem tracheary element differentiation and xylem development and increased susceptibility to bacterial and fungal pathogens (Halter et al., 2021; Le et al., 2014; Lin et al., 2020). Furthermore, DNA methylome analysis on *ros1* and *rdd* mutants revealed that ROS1, together with DML2 and DML3, prevents transcriptional silencing and the spreading of DNA methylation from transposable elements, which can also be targeted by RdDM, to nearby genes (Halter et al., 2021; Penterman et al., 2007; Tang et al., 2016).

To better understand the biological roles of DNA demethylases, the generation of a homozygous mutant defective in all four DNA demethylases is required. However, embryo lethality caused by the maternal *dme* allele acts as a barrier to generate the *dme* homozygous mutants. Recent studies attempted to produce the homozygous *dme ros1 dml2 dml3* (*drdd*) quadruple

mutant by RNA interference-mediated knockdown or the central cell-specific complementation of *DME* expression, respectively, indicating that DNA demethylases are responsible for resistance against the bacterial and fungal pathogens as well as flowering time (Schumann et al., 2019; Williams et al., 2022; Zeng et al., 2021). However, these findings still have limitations of incomplete suppression or ectopic expression of *DME* outside the central cell, respectively.

Here, I generated the homozygous *drdd* quadruple mutant via *in vitro* rescue of early seeds. The *drdd* mutant showed obvious developmental defects such as delayed growth and flowering, aberrant flower phenotypes and failure of fertilization and silique elongation. In an attempt to investigate the combined activity of all four DNA demethylases in transcriptional regulation, I performed transcriptome and methylome analysis in flowers and buds of WT, *dme*, *rdd* and *drdd*. Whereas all the mutants exhibited a significant decrease in gene expression levels, the downregulation was more prominent in the *drdd* mutant. DNA methylation levels were also considerably increased in *drdd*, and the hypermethylated regions were highly enriched in promoters and intergenic regions. These results indicate that DNA demethylases act redundantly to regulate gene expression during reproductive development. I identified a total of 703 *drdd*-downregulated genes with

hypermethylated regions at their promoter regions, some of which were highly associated with pollen tube growth and cell morphogenesis. Hence, these findings suggest that DNA demethylase-mediated DNA demethylation would contribute to normal fertilization and floral organ determination.

MATERIALS AND METHODS

Plant materials

Arabidopsis thaliana ecotype Columbia (Col-0) was used as a control, immature seeds of which were cultured *in vitro* for multiple generations. After self-fertilization of the heterozygous *dme-2* mutant in the Col-*gl* background (Choi et al., 2002), immature seeds were rescued, and then the homozygous *dme-2* mutant was generated out of > 50 rescued plants. Emasculated flowers of the triple homozygous *rdd* mutant (Penterman et al., 2007) were pollinated with the heterozygous *dme-2* mutant pollen to generate the quadruple heterozygous F1 progeny. By *in vitro* culture of early seeds over two subsequent generations, *dme* +/- *rdd* -/- mutants were obtained, and then, the quadruple homozygous *drdd* mutants were isolated out of > 150 viable plants. The wild-type (WT) and the homozygous *rdd* segregants were also isolated from F2 and F3 progenies, respectively. Primers used in genotyping are listed in Table 2-1. For RNA-seq and BS-seq analysis, paired RNA and DNA samples were collected from WT, *rdd* and *drdd* mutants whose immature seeds were rescued *in vitro* for additional 1~3 generations (WTR1, *rdd*R3 and *drdd*R1/R2) and *dme* mutants rescued for six generations, then normally

germinated in MS medium and rescued *in vitro* for two subsequent generations (*dme*R6G1R2).

In vitro culture of immature seeds

In vitro culture assay was performed as described previously (Sauer and Friml, 2008) with minor modifications. Immature seeds were dissected from early siliques with syringe needles under the microscope and transferred to *in vitro* culture medium with 5% sucrose. The culture plates were wrapped in two sheets of white paper (under dim light conditions) for 7 days and uncovered for 7 days in a growth chamber at 22°C under 16 h of fluorescent light at 30 ± 10 μmol m^{-2} s^{-1}. Seeds were transferred to half-strength MS medium, then grown in the growth chamber for two weeks and transplanted to soil. To assess the efficiency of *in vitro* culture, viable seedlings with developed green cotyledons were counted.

RNA-seq analysis

For each of the biological replicates from three or four independent lines, five flowers and ten buds were pooled, respectively. Total RNA was isolated using the QIAGEN RNeasy plant mini kit. RNA-seq libraries were constructed using the Illumina TruSeq RNA-Seq library kit. RNA-seq was

performed on the Illumina HiSeq 4000 sequencing system. 101 bp paired-end reads were filtered with Trimmomatic v.0.38 (Bolger et al., 2014) with the parameters TruSeq3-PE-2.fa:2:30:10, LEADING:15, TRAILING:15, SLIDINGWINDOW:4:18 and MINLEN:60. Filtered reads were aligned to the reference *A. thaliana* genome (TAIR 10; (Berardini et al., 2015)) using hisat2 version 2.2.1 (Kim et al., 2019a) and quantified with htseq-count version 0.6.1p1 (Anders et al., 2015). DEGs with fold change > 2 and FDR < 0.05 were identified using the edgeR package (Robinson et al., 2010).

BS-seq analysis

For each of the biological replicates from three or four independent lines, genomic DNA was extracted from flowers using the CTAB method. A total of 2 μg genomic DNA was used to generate BS-seq libraries. The library construction and 151 bp paired-end sequencing were performed with Illumina HiSeq 4000. Reads were filtered with Trimmomatic v.0.38 (Bolger et al., 2014) with the same parameters as in transcriptome analysis. Mapping and methylation calling were performed using Bismark version 0.19.1 (Krueger and Andrews, 2011). Only cytosine sites with 5× coverage were used for subsequent analysis. DMCs and DMRs were identified as described previously (Huang et al., 2013). DMRs were finally identified based on the

regions with a length ≥ 100 bp, ≥ 5 DMCs, and the mean methylation difference ≥ 0.3 for CG, ≥ 0.15 for CHG, or ≥ 0.1 for CHH.

Table 2-1. List of primers.

Name	Sequence	Purpose
dme-2_1F	CACTTGTTCCCTATGAGAGC	genotyping
dme-2_1R	CACTGATTGTGATGTTCCAC	genotyping
ros1-3_2F	CCCGCGACTCTTGATTGTTTCAGCAACTT	genotyping
ros1-3_2R	TGGAAGGGATCCGTCGTGGATTCT	genotyping
dml2-1_3F	ACCCGGAGAGTACCATTCAGACAC	genotyping
dml2-1_3R	GTGGCCAGAGGTACTTTTGAACT	genotyping
dml3-1_4F	GACGTTGCTGTAGATATGAC	genotyping
dml3-1_4R	GCCAAATCGCAAGAAGGTAAGGA	genotyping
dme-2_LB	TTGACCATCATACTCATTGCTG	genotyping
ros1-3_dml2-1_LB	CATTTTATAATAACGCTGCGGACATCTAC	genotyping
dml3-1_LB	GCATGTGAATTTCATAACCAATCTCGATACA	genotyping
DME_qRT5628f	TCGTCTCCTTGATGGTATGGA	RT-qPCR
DME_qRT5722r	GTGCCGAATTCGCTGTTT	RT-qPCR
ROS1_qRT3710f	TCTCCTGCAACAGCATCA	RT-qPCR
ROS1_qRT3847r	CATGATCCGCAAACACCT	RT-qPCR
DML2_3534f	CAAATGTTTTCAATGCAACAAGA	RT-qPCR
DML2_3616r	TTGCTGTTCTGCAAGGTATCA	RT-qPCR
DML3_2780f	TTTCGCGGAACAATTTTGA	RT-qPCR
DML3_2866r	TCCCTCATTGGTTTGGAAGT	RT-qPCR

RESULTS

Generation of the homozygous *dme* and *drdd* mutants

As seeds that inherit the maternal *dme* allele usually abort, homozygous *dme* mutants are rarely viable (Choi et al., 2002; Gehring et al., 2006). To examine the function of DME, I attempted to generate a complete loss-of-function *dme* mutant using the *in vitro* culture. Immature seeds of *dme* heterozygotes were cultured *in vitro* to overcome the seed abortion phenotype of *dme*. Although a strong mutant allele, *dme-2*, was known to cause severe embryo lethality (Choi et al., 2002), early seeds of the heterozygous *dme-2* mutants were rescued successfully (Figures 2-1A and 1B). Out of 59 rescued plants, a *dme* homozygote was obtained, then subjected to *in vitro* culture for multiple generations (Figure 2-1A).

Likewise, homozygous quadruple *drdd* mutants are embryo-lethal, which attests to the biological importance of active DNA demethylation. Homozygous *drdd* mutants have been recently generated (Schumann et al., 2019; Williams et al., 2022; Zeng et al., 2021), but these mutants still have the potential to express *DME* outside the central cell. I sought to isolate a true loss-of-function quadruple mutant to investigate the necessity of all four DNA

demethylases for plant development. The triple homozygous *rdd* mutant plants were pollinated with the heterozygous *dme* +/- pollen to produce the heterozygous quadruple F1 progeny, and early seeds of them were *in vitro*-cultured for two subsequent generations (Figures 2-1A and 1B). Out of 158 viable *dme* heterozygotes in the *rdd* background (*dme* +/- *rdd* -/-), a homozygous quadruple *drdd* mutant was isolated, together with its wild-type (WT) counterpart and homozygous *rdd* mutant (Figures 2-1B and 1C). The homozygous *drdd* mutant also exhibited a significant decrease in *DME* expression compared with WT and the *rdd* mutant, indicating that the isolated homozygous *dme* and *drdd* mutants are likely to be valuable genetic materials to examine the roles of active DNA demethylation.

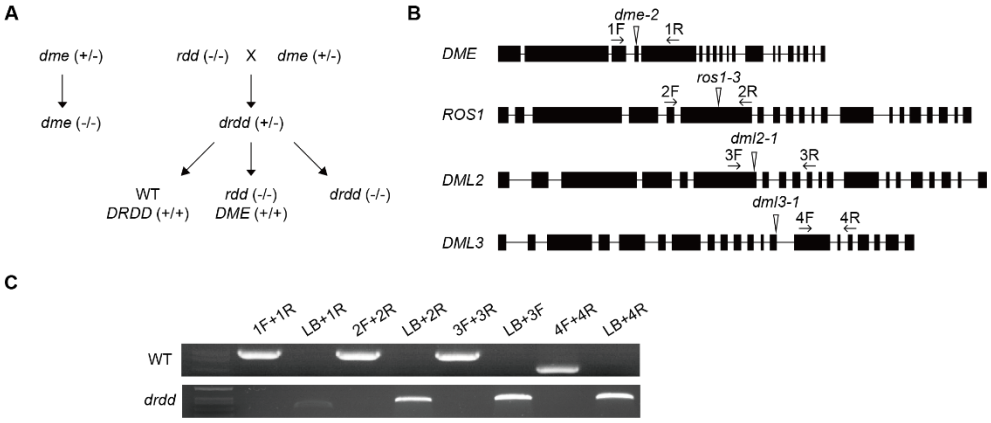

Figure 2-1. Isolation of the homozygous *dme* and *drdd* mutants.

(A) Schematic diagram of generating WT, *dme*, *rdd* and *drdd* mutants. By *in vitro* culture of immature seeds of *dme* heterozygotes, a *dme* homozygote was obtained. The *ros1-3 dml2-1 dml3-1* homozygous plants were pollinated by *dme-2* heterozygous pollen to produce heterozygous *drdd* F1 plants. After *in vitro* rescue of early seeds for subsequent generations, a *drdd* homozygote and its WT and *rdd* counterparts were isolated. (B) Schematic diagram of *DME* family gene loci. Boxes and lines indicate exons and introns, respectively. The positions of T-DNA insertions are represented by arrowheads. Arrows indicate primers used in genotyping of T-DNA alleles. (C) Genotyping of *dme*, *ros1*, *dml2* and *dml3* alleles.

The *drdd* mutant exhibits delayed growth and abnormal seed development

During reproductive development, the *drdd* quadruple mutant displayed severely retarded growth compared to WT, *dme* and *rdd* mutants (Figures 2-2A and 2-3A). To determine the effect of a complete loss of DNA demethylase activity on the transition from the vegetative to reproductive stages, I measured the time of flowering. The *drdd* mutant was late flowering under long-day conditions in comparison to WT, *dme* and *rdd* mutants (Figure 2-2). The *drdd* mutant flowered 54.2 ± 7.8 days after *in vitro* culture, whereas WT plants flowered 41.1 ± 3.6 days after *in vitro* culture (Figure 2-2B). The *dme* and *rdd* mutants also flowered later (45.0 ± 7.1 and 48.6 ± 6.8 days after *in vitro* culture, respectively) than WT, but earlier than *drdd*, indicating the functional redundancy among DNA demethylases (Figure 2-2B). Some *in vitro*-rescued plants exhibited bushy phenotypes at the vegetative stage with many deformed leaves and multiple inflorescences emerging simultaneously. The number of rosette leaves at bolting was variable but significantly higher in the *drdd* mutant compared to WT (data not shown). This observation was rather different from the previous study, where a central cell- specific *DME* complementation line in the *drdd* background showed early-flowering phenotypes (Williams et al., 2022). Furthermore, delayed growth in the *drdd*

mutant became more pronounced after flowering. In 10-week-old WT, *dme* and *rdd* mutants, the majority of mature green siliques turned yellowish, while a few young developing siliques were observed in the *drdd* mutant (Figure 2-3A). Three weeks later, most of the *drdd* siliques did not reach maturity and showed delayed senescence (Figure 2-3B). These findings demonstrate that DNA demethylases are required for the precise control of reproductive development and conceivably have an additive effect on flowering time and silique development.

I further investigated the function of active DNA demethylation in silique development. Given that WT siliques did not show any morphological defects, the effect of *in vitro* culture was presumed to be compromised. The silique length of the *dme* mutant was slightly decreased compared to WT and *rdd.* However, the *drdd* mutant produced strikingly shorter siliques (Figures 2-3A to 3D). Interestingly, most siliques of the *drdd* mutant failed to elongate, with few siliques slightly elongated and expanded (Figures 2-3B to 3E). The unfertilized and fertilized *drdd* siliques are referred to as $drdd_{UF}$ and $drdd_{F}$ siliques hereafter, respectively. The elongated siliques of the $drdd_{F}$ were still significantly shorter than WT, *dme* and *rdd* siliques (Figure 2-3F). In addition, the proportion of abnormal siliques was also higher in *rdd* in comparison to

WT and *dme*, but substantially lower than *drdd*, implying that DME activity in the *rdd* mutant may contribute to silique elongation and enlargement.

I next examined whether seed development is impaired in *drdd*. In young siliques, more than half of the seeds were aborted in *dme*, whereas the majority of *rdd* seeds were normally developed, similar to WT (Figures 2-4A, 4C and 4D). Intriguingly, only a few fertilized seeds were observed in $drdd_F$ siliques (Figures 2-4A, 4C and 4D), consistent with abnormal silique development in *drdd*. After seed maturation, all fertilized seeds of *dme* and $drdd_F$ became aborted (Figures 2-4B and 4C), in accordance with previous studies (Choi et al., 2002; Williams et al., 2022; Zeng et al., 2021). These results indicate that the *drdd* mutant showed abnormal seed and silique development.

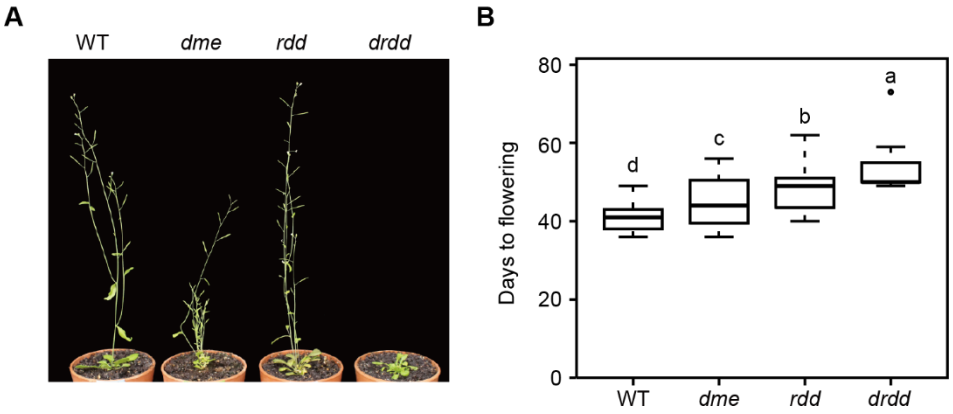

Figure 2-2. Delayed flowering in the *drdd* mutant.

(A) Phenotypes of seven-week-old plants of WT, *dme*, *rdd* and *drdd* mutants grown in long-day conditions. (B) Box plots showing the flowering time of WT, *dme*, *rdd* and *drdd* mutants under long-day conditions as measured by days of bolting. Error bars indicate ± SD (n = 13, 8, 8 and 9). Different letters indicate significant differences with one-way ANOVA followed by the post-hoc Tukey multiple comparison tests ($P < 0.05$).

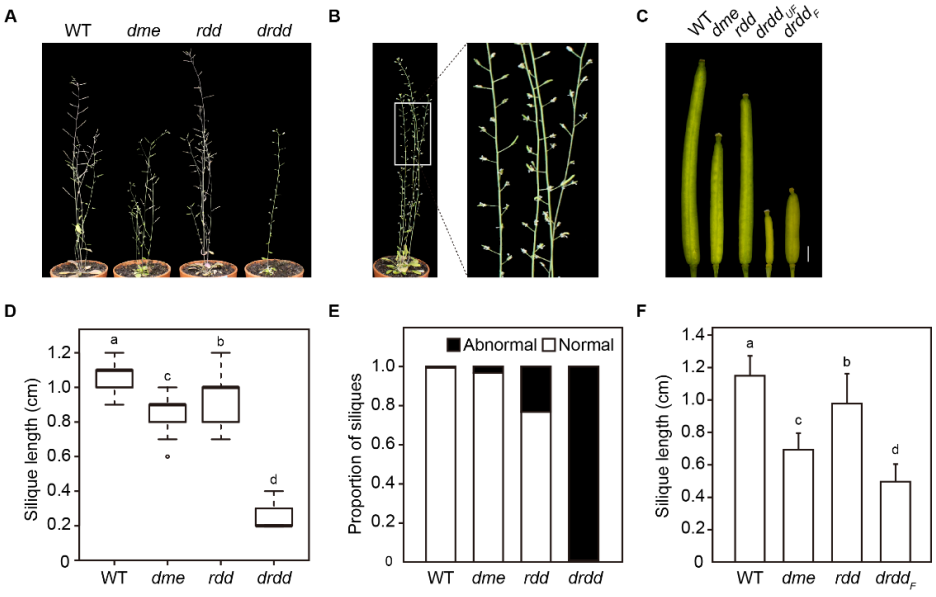

Figure 2-3. Retarded growth and abnormal silique development in the *drdd* mutant.

(A) Phenotypes of 10-week-old plants of WT, *dme*, *rdd* and *drdd* mutants. (B) Siliques of 13-week-old *drdd* mutant. (C) Representative siliques of WT, *dme*, *rdd* and *drdd* mutants. In the *drdd* mutant, the majority of siliques seemed unfertilized (left; $drdd_{UF}$) and few siliques were slightly elongated (right; $drdd_{F}$). Scale bar, 1 mm. (D) Box plots of silique lengths of WT, *dme*, *rdd* and *drdd* mutants. Four plants from each genotype were examined. Error bars indicate ± SD (n > 30). (E) Proportion of normal and abnormal siliques in WT, *dme*, *rdd* and *drdd* mutants. Five plants from each genotype were examined. Error bars indicate ± SD (n > 150). (F) Lengths of normal siliques in WT, *dme*, *rdd* and $drdd_{F}$ mutants. Four to seventeen plants from each genotype were examined. Error bars indicate ± SD (n > 15). Different letters indicate significant differences with one-way ANOVA followed by the post-hoc Tukey multiple comparison tests (P < 0.05).

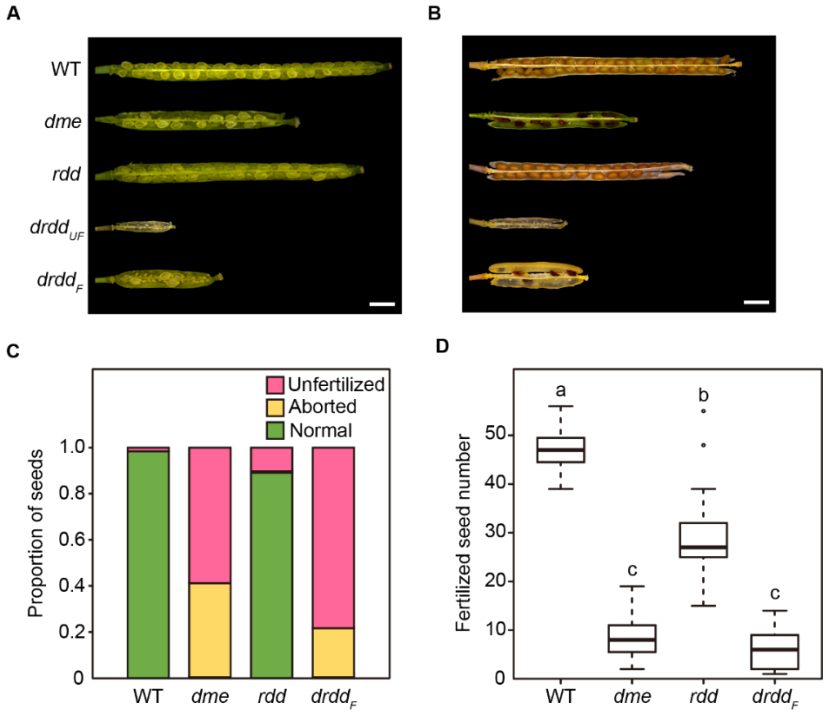

Figure 2-4. Abnormal seed development in the *drdd* mutant.

(A) Representative images of young siliques in WT, *dme*, *rdd* and *drdd* mutants. Scale bar = 1 mm. Asterisks and arrows indicate the representative of unfertilized and fertilized seeds, respectively. (B) Representative images of mature siliques in WT, *dme*, *rdd* and *drdd* mutants. Scale bar = 1 mm. Asterisks indicate the representative of aborted seeds, respectively. (C) Proportion of normal, aborted and unfertilized seeds in WT, *dme*, *rdd* and $drdd_F$ mutants. Four to six plants from each genotype were examined (n > 250). (D) Box plots of fertilized seed numbers of WT, *dme*, *rdd* and $drdd_F$ mutants. Seven to ten plants from each genotype were examined. Error bars indicate ± SD (n > 12). Different letters indicate significant differences with one-way ANOVA followed by the post-hoc Tukey multiple comparison tests (P < 0.05).

The *drdd* mutant displays aberrant floral organ development

It is assumed that an entire loss of DNA demethylase activity causes a failure of fertilization. To assess whether the defect in fertilization arises from dysregulation of floral organ development in *drdd*, I examined the flower architecture in WT and *dme*, *rdd* and *drdd* mutants. In contrast to WT flowers that consist of four sepals, four petals, six stamens and a bicarpellary pistil (Figures 2-5A and 5B), several abnormal flowers were observed in *dme* and *rdd* mutants in a stochastic manner (Table 2-2 and Figure 2-5A). Surprisingly, nearly 35 percent of the *drdd* mutants exhibited aberrant flower morphologies (Figure 2-5A).

Based on normal flower development in WT, the *in vitro* rescue effect is nearly negligible in reproductive development. In the *dme* mutant, a reduced stamen number is the most predominant phenotype (Figure 2-5D), with a lower frequency of an increase in stamen number (Figure 2-5C). I occasionally observed flowers with fused filaments (Figure 2-5E), deformed or short stamens (Figure 2-5F), petal-like stamens (Figure 2-5G) and stamen-like petals (Figure 2-5H). The inheritance of the stamenoid petal phenotype was also shown. In agreement with the previous report (Choi et al., 2002), diverse but sporadic floral abnormalities were found in *dme*. In addition, the

rdd mutant also sporadically displayed aberrant phenotypes of unfused carpels (Figure 2-5I), a stamen-carpel fusion (Figure 2-5J) and reduced or increased stamen numbers (Figures 2-5K and 5J), probably due to the accumulation of DNA methylation in multiple generations.

Remarkably, nearly half of the *drdd* mutants appeared to have combined phenotypes of abnormal flower organ development, indicating pleiotropic effects of impaired DNA demethylase activity (Table 2-2 and Figures 2-6D to 6F). I found various morphological abnormalities including homeotic conversions and alterations in the floral organ number and size in the *drdd* mutant. Nearly 50 percent of aberrant *drdd* flowers showed unfused or even split carpels (Figures 2-6A to 6C). Other carpel defects such as a stamen-fused carpel (Figure 2-6C), an extra carpel (Figure 2-6D) and a complete loss of discernible carpel (Figure 2-6E) were also observed. Approximately 30 percent of abnormal flowers in *drdd* displayed increased sepal numbers (Table 2-2 and Figure 2-6F), and nearly 26 and 23 percents of them showed an increase and decrease in stamen number, respectively (Table 2-2 and Figures 2-6D to 6G). Flowers with fused stamen filaments (Figure 2-6H), deformed or short stamens (Figures 2-6I and 6J), increased or reduced sepal numbers (Figure 2-6F) and small petals were also occasionally observed

(Table 2-2). Furthermore, the floral organs of *drdd* did not fall off shortly after anthesis, indicating delayed abscission and senescence (Figure 2-6K).

Taken together, aberrant phenotypes of floral organs were apparently more severe in *drdd* than in *dme* and *rdd*, demonstrating that all DNA demethylases act together to ensure proper floral organ development. These results suggest that the determination of the floral organ identity and boundary as well as the organ number and size might be dysregulated in the *drdd* quadruple mutant, leading to reduced fertility.

Table2-2. Phenotypic analysis of abnormal floral organ development in *dme, rdd* and *drdd* mutants.

Genotype	WT	*dme*	*rdd*	*drdd*
No. of flowers (plants) examined	25 (11)	467 (151)	94 (26)	96 (25)
No. of abnormal flowers	0	27	5	34
No. of flowers showing combined phenotypes	0	1	0	15
No. of flowers with abnormal carpels				
unfused	-	-	3	18
stamenoid	-	-	2	2
absent	-	-	-	1
No. of flowers with abnormal stamens				
increased number	-	1	1	9
decreased number	-	14	1	8
petaloid	-	4	-	-
short/deformed	-	3	-	6
fused filaments	-	1	-	1
No. of flowers with abnormal petals				
increased number	-	-	-	11
stamenoid	-	2	-	-
small	-	3	-	2
No. of flowers with abnormal sepals				
increased number	-	-	-	1
decreased number	-	-	-	2
petaloid	-	-	-	1

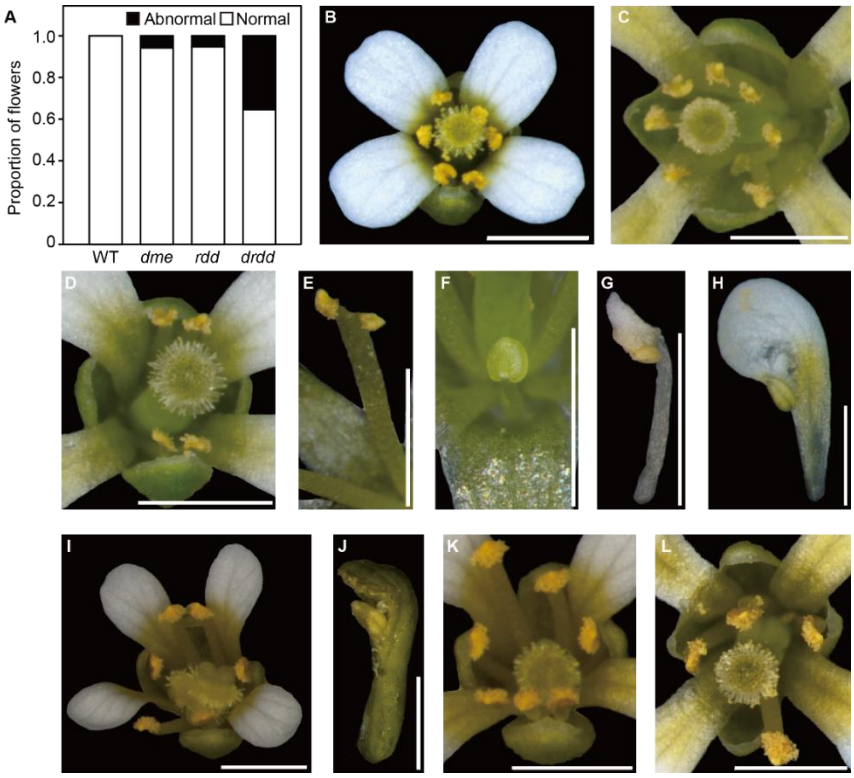

Figure 2-5. Stochastic development of floral organs in *dme* and *rdd* mutants.

(A) Proportion of normal and abnormal flowers in WT, *dme*, *rdd* and *drdd* mutants. A total of 11, 151, 26 and 25 plants were examined in WT, *dme*, *rdd* and *drdd* mutants, respectively (n = 25, 467, 94, 96). (B) Representative image of normally developed WT flower. (C-H) Abnormal flower structures in *dme* with seven stamens and a deformed stamen (C), four stamens (D), fused stamen filaments (E), a deformed stamen (F), a petal-like stamen fused with white petal tissue (G) and a stamen-like petal containing a fused pollen sac (H). (I-L) Abnormal flower morphologies in *rdd* with unfused carpels (I), a stamenoid carpel containing a fused pollen sac (J), four stamens (K) and seven stamens (L). Scale bar = 1 mm.

Figure 2-6. Combined phenotypes of floral abnormalities in the *drdd* mutant.

(A-K) Aberrant flower phenotypes with split carpels (A), opened carpels with ovules exposed (B), a stamenoid carpel containing a fused pollen sac (C), an extra carpel with eight stamens (D), an absence of discernible carpel and three stamens (E), six petals, five sepals and seven short stamens (F), eight stamens (G), fused stamen filaments (H), a deformed stamen (I), short stamens (J) and undetached floral organs after anthesis (K). Scale bar = 1 mm.

DNA demethylases redundantly regulate the expression of the target genes

To explore the role of DNA demethylases in regulating gene expression during reproductive development, RNA sequencing (RNA-seq) analysis was performed on flowers and buds of WT and *dme*, *rdd* and *drdd* mutants. I identified a total of 56, 140 and 2,516 differentially expressed genes (DEGs) in *dme*, *rdd* and *drdd* flowers, respectively, compared to WT flowers (Figure 2-7A). More than 70 percent of DEGs were downregulated in *dme*, *rdd* and *drdd* mutants (Figures 2-7A and 2-8A), consistent with the function of DNA methylases in the inhibition of transcriptional silencing (Penterman et al., 2007; Tang et al., 2016). While approximately 98 and 80 percent of downregulated genes in *dme* and *rdd* overlapped with *drdd*-downregulated genes, respectively, the majority of the *drdd*-downregulated genes appeared to be specific to the *drdd* quadruple mutant (Figure 2-8A). Among 2,028 *drdd*-downregulated genes relative to WT, 991 (48.9%) genes also showed decreased expression in *drdd* compared to *dme* and *rdd*, respectively, referred to as commonly *drdd*-downregulated genes (Figure 2-8B). This finding indicates a significant downregulation only in the *drdd* quadruple mutant, due to the combined activity of DNA demethylases in transcriptional regulation. Furthermore, 661 (32.0%) and 184 (9.1%) genes overlapped with the genes

displaying reduced expression in *drdd* compared to *rdd* and *dme* mutants, suggesting that downregulation of these genes may be DME- and RDD-dependent, respectively (Figures 2-7B and 2-8B). Together, these results demonstrate that all four DNA demethylases redundantly control the expression of a subset of genes during flower development, and DME activity might be largely responsible for the decrease in gene expression in *drdd*.

In flower buds, the number of DEGs in *drdd* (1,609 genes) was also considerably higher than that of *dme* and *rdd* (27 and 138 genes, respectively) compared to WT, more than 80 percent of which were downregulated (Figures 2-9A and 2-10A). In addition, about 35 percent of the *drdd*-downregulated genes were redundantly regulated, and the reduction of gene expression in *drdd* was dependent on the activity of DME rather than that of RDD (Figures 2-9B and 2-10B). These observations indicate that flowers and buds have highly similar transcriptome profiles.

I next conducted Gene Ontology (GO) analysis to determine whether gene expression changes caused by the *drdd* mutation are associated with biological functions. The *drdd*-downregulated genes were highly enriched in pollen tube growth and development as well as cell morphogenesis involved in differentiation (Figures 2-8C and 2-10C and Tables 2-3 and 2-4), in

agreement with the observation of abnormal flower and silique development in *drdd* (Figures 2-4 and 2-6). This is also consistent with the previous study that showed DME and ROS1 acting together to promote pollen tube progression by regulating the expression of genes with pollen tube function (Khouider et al., 2021). The *drdd*-upregulated genes were most enriched for metabolic process and cell wall biogenesis (Figure 2-8D). Moreover, to investigate whether the associated biological processes rely on the combinatorial activity of DNA demethylase, GO analysis was performed on the *drdd*-downregulated genes common to WT, *dme* and *rdd*. GO terms for pollen tube growth and development as well as cell morphogenesis involved in differentiation were also significantly enriched among the commonly downregulated genes in *drdd* (Table 2-5), suggesting that the redundant activities of DNA demethylases may contribute to the transcriptional regulation of genes involved in pollen tube development and cell differentiation.

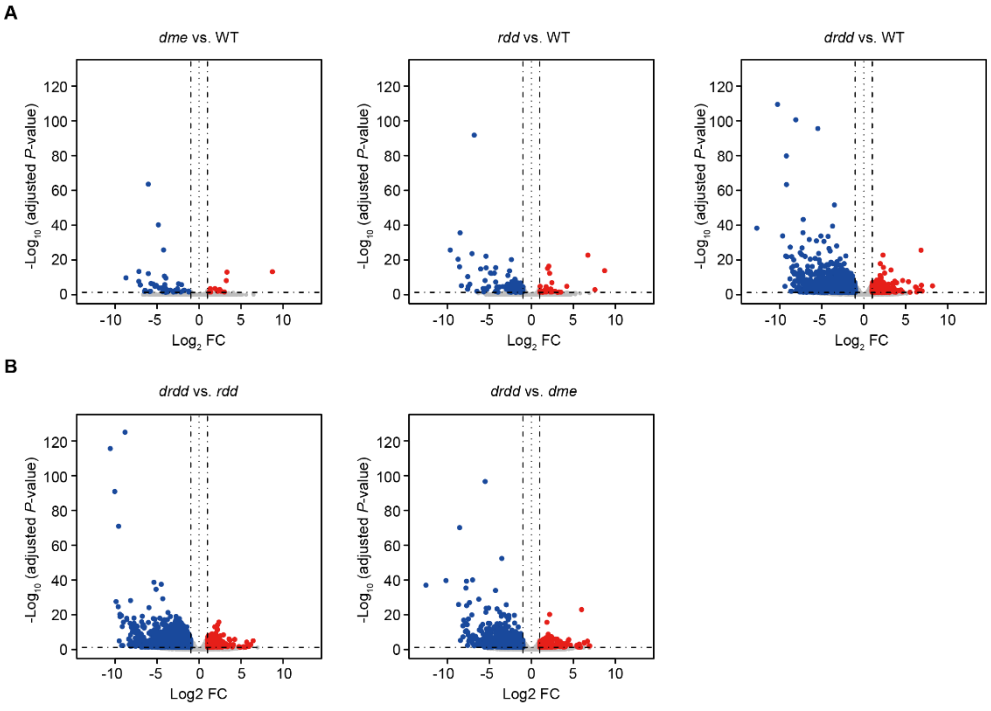

Figure 2-7. Transcriptome changes in *drdd* flowers.

(A-B) Volcano plots showing differentially expressed genes (fold change (FC) > 2, false discovery rate (FDR) < 0.05) in *dme*, *rdd* and *drdd* compared to WT (A) and in *rdd* and *dme* compared to WT (B). Blue and red dots represent downregulated and upregulated genes, respectively. Grey dots indicate the genes that are not significantly differentially expressed.

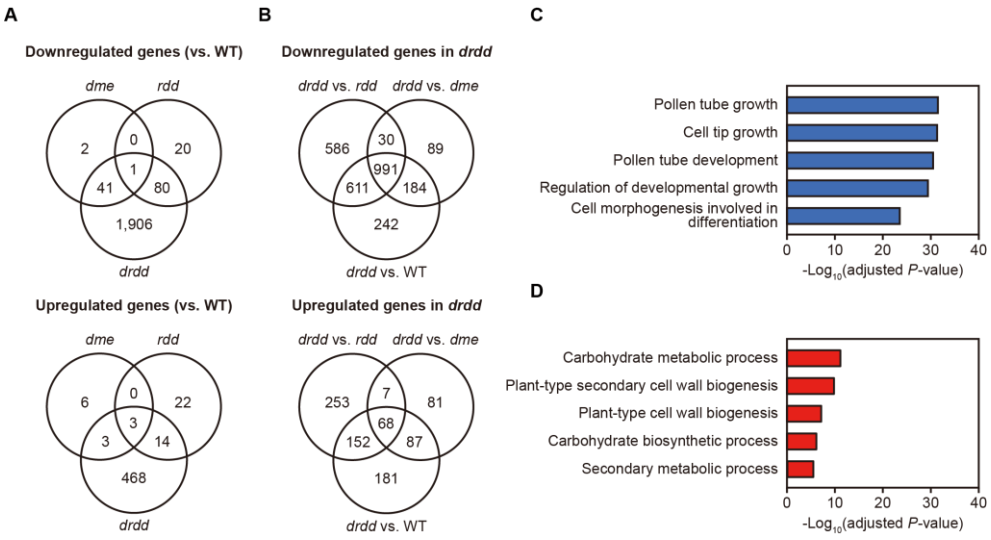

Figure 2-8. Comparative analysis of DEGs in WT, *dme*, *rdd* and *drdd* flowers.

(A-B) Venn diagrams of DEGs in *dme*, *rdd* and *drdd* mutants compared to WT (A) and DEGs in *drdd* compared to WT, *dme* and *rdd* mutants (B). (C-D) GO analysis of downregulated (C) and upregulated (D) genes in *drdd* compared to WT.

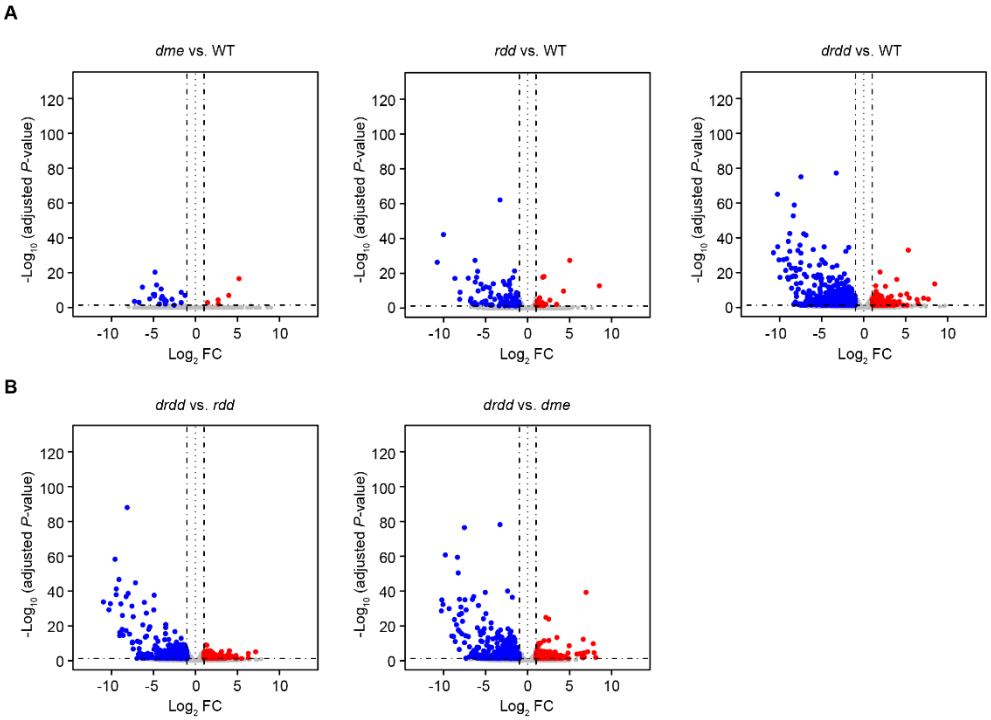

Figure 2-9. Transcriptome changes in *drdd* buds.

(A-B) Volcano plots showing differentially expressed genes (fold change (FC) > 2, false discovery rate (FDR) < 0.05) in *dme*, *rdd* and *drdd* compared to WT (A) and in *rdd* and *dme* compared to WT (B). Blue and red dots represent downregulated and upregulated genes, respectively. Grey dots indicate the genes that are not significantly differentially expressed.

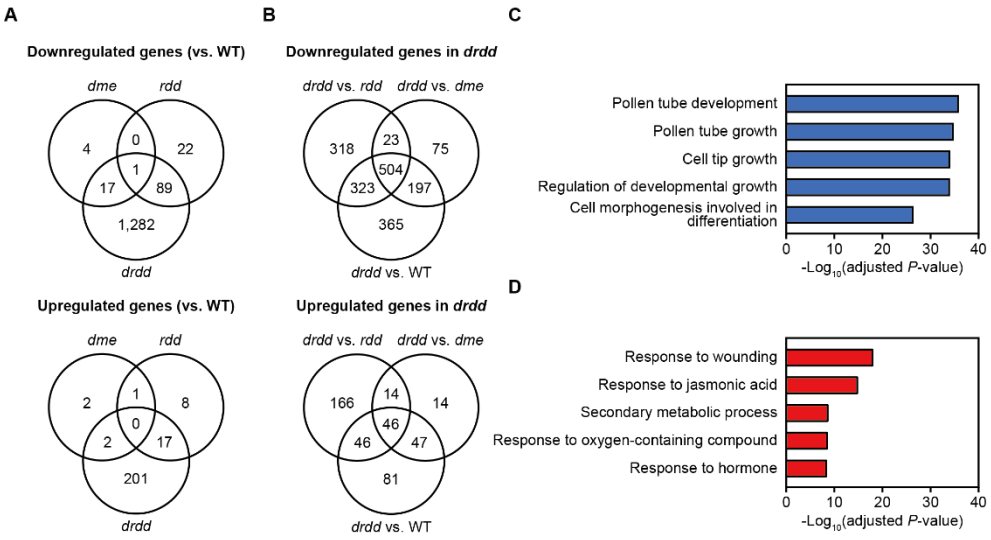

Figure 2-10. Comparative analysis of DEGs in WT, *dme*, *rdd* and *drdd* buds.

(A-B) Venn diagrams of DEGs in *dme*, *rdd* and *drdd* mutants compared to WT (A) and DEGs in *drdd* compared to WT, *dme* and *rdd* mutants (B). (C-D) GO analysis of downregulated (C) and upregulated (D) genes in *drdd* compared to WT.

Table2-3. Top 20 most enriched GO terms of downregulated genes in *drdd* flowers compared to WT.

GO ID	GO term	Count[a]	*P*-value	FDR[b]
GO:0009860	Pollen tube growth	85	7.38E-36	2.24E-32
GO:0009932	Cell tip growth	92	1.61E-35	3.25E-32
GO:0048868	Pollen tube development	96	1.91E-34	2.32E-31
GO:0048638	Regulation of developmental growth	95	2.86E-33	2.89E-30
GO:0048588	Developmental cell growth	98	1.12E-31	9.73E-29
GO:0009826	Unidimensional cell growth	115	1.73E-29	8.08E-27
GO:0048468	Cell development	109	1.60E-30	9.68E-27
GO:0000904	Cell morphogenesis involved in differentiation	92	6.43E-27	2.17E-24
GO:0060560	Developmental growth involved in morphogenesis	123	7.05E-27	2.25E-24
GO:0000902	Cell morphogenesis	119	3.30E-27	6.65E-24
GO:0030154	Cell differentiation	210	6.21E-26	1.88E-23
GO:0080092	Regulation of pollen tube growth	30	4.90E-25	7.41E-22
GO:0010769	Regulation of cell morphogenesis involved in differentiation	30	5.80E-23	7.02E-20
GO:0060284	Regulation of cell development	30	6.70E-22	6.75E-19
GO:0009617	Response to bacterium	173	8.90E-17	5.98E-14
GO:0051510	Regulation of unidimensional cell growth	32	5.20E-16	3.14E-13
GO:0042742	Defense response to bacterium	151	6.50E-16	3.57E-13
GO:0050832	Defense response to fungus	107	2.10E-14	9.77E-12
GO:0001558	Regulation of cell growth	38	2.70E-14	1.17E-11
GO:0006952	Defense response	283	2.00E-13	7.56E-11

a Number of *drdd*-downregulated genes from list

b False discovery rate (FDR) from Benjamini-Hochberg multiple tests

Table2-4. Top 20 most enriched GO terms of downregulated genes in *drdd* buds compared to WT.

GO ID	GO term	Count[a]	*P*-value	FDR[b]
GO:0009860	Pollen tube growth	85	7.38E-36	2.24E-32
GO:0009932	Cell tip growth	92	1.61E-35	3.25E-32
GO:0048868	Pollen tube development	96	1.91E-34	2.32E-31
GO:0048638	Regulation of developmental growth	95	2.86E-33	2.89E-30
GO:0048588	Developmental cell growth	98	1.12E-31	9.73E-29
GO:0009826	Unidimensional cell growth	115	1.73E-29	8.08E-27
GO:0048468	Cell development	109	1.60E-30	9.68E-27
GO:0000904	Cell morphogenesis involved in differentiation	92	6.43E-27	2.17E-24
GO:0060560	Developmental growth involved in morphogenesis	123	7.05E-27	2.25E-24
GO:0000902	Cell morphogenesis	119	3.30E-27	6.65E-24
GO:0030154	Cell differentiation	210	6.21E-26	1.88E-23
GO:0080092	Regulation of pollen tube growth	30	4.90E-25	7.41E-22
GO:0010769	Regulation of cell morphogenesis involved in differentiation	30	5.80E-23	7.02E-20
GO:0060284	Regulation of cell development	30	6.70E-22	6.75E-19
GO:0009617	Response to bacterium	173	8.90E-17	5.98E-14
GO:0051510	Regulation of unidimensional cell growth	32	5.20E-16	3.14E-13
GO:0042742	Defense response to bacterium	151	6.50E-16	3.57E-13
GO:0050832	Defense response to fungus	107	2.10E-14	9.77E-12
GO:0001558	Regulation of cell growth	38	2.70E-14	1.17E-11
GO:0006952	Defense response	283	2.00E-13	7.56E-11

a Number of *drdd*-downregulated genes from list

b False discovery rate (FDR) from Benjamini-Hochberg multiple tests

Table2-5. Top 20 most enriched GO terms of commonly downregulated genes in *drdd* flowers compared to WT.

GO ID	GO term	Count[a]	*P*-value	FDR[b]
GO:0048868	Pollen tube development	87	2.11E-40	1.28E-36
GO:0009860	Pollen tube growth	74	1.83E-38	1.59E-35
GO:0009932	Cell tip growth	79	1.05E-37	7.96E-35
GO:0048638	Regulation of developmental growth	84	1.36E-37	9.14E-35
GO:0048588	Developmental cell growth	84	2.37E-34	1.1E-31
GO:0030154	Cell differentiation	184	1.1E-35	6.69E-33
GO:0000904	Cell morphogenesis involved in differentiation	79	1.03E-29	3.9E-27
GO:0009826	Unidimensional cell growth	93	1.12E-29	4.01E-27
GO:0048468	Cell development	93	7.24E-28	2.44E-25
GO:0060560	Developmental growth involved in morphogenesis	99	2.33E-27	7.07E-25
GO:0000902	Cell morphogenesis	96	3.70E-27	1.12E-23
GO:0080092	Regulation of pollen tube growth	27	3.40E-25	6.85E-22
GO:0010769	Regulation of cell morphogenesis involved in differentiation	27	1.90E-23	2.30E-20
GO:0060284	Regulation of cell development	27	1.60E-22	1.38E-19
GO:0051510	Regulation of unidimensional cell growth	32	7.50E-21	5.67E-18
GO:0001558	Regulation of cell growth	36	6.30E-18	4.23E-15
GO:0022604	Regulation of cell morphogenesis	33	7.00E-18	4.23E-15
GO:0045595	Regulation of cell differentiation	28	6.10E-14	2.63E-11
GO:0022603	Regulation of anatomical structure morphogenesis	37	1.30E-11	4.62E-09
GO:0045490	Pectin catabolic process	22	7.30E-10	2.45E-07

a Number of *drdd*-downregulated genes from list

b False discovery rate (FDR) from Benjamini-Hochberg multiple tests

Active DNA demethylation is redundantly regulated by DNA demethylases

BS-seq was performed on flowers of WT, *dme*, *rdd* and *drdd* mutants. Differentially methylated cytosines (DMCs) and differentially methylated regions (DMRs) were identified between WT and each mutant genotype. In total, 567,025, 1,038,023 and 1,214,311 hypermethylated DMCs (hyper-DMCs) were found in *dme*, *rdd* and *drdd* compared to WT, respectively, and a larger number of hyper-DMCs were identified relative to hypomethylated DMCs (hypo-DMCs) in both *drdd* and *rdd* (Figure 2-11A). I then identified 1,228, 12,643 and 14,356 hypermethylated DMRs (hyper-DMRs) in *dme*, *rdd* and *drdd*, respectively, compared to WT (Figure 2-11B). Although DNA methylation was substantially increased in *rdd* rather than in *dme*, the *drdd* mutant exhibited a higher increase in DNA methylation, demonstrating the redundant function of DNA demethylases in removing DNA methylation (Figure 2-11B). In particular, DNA hypermethylation predominantly occurred in the CG and CHG contexts, albeit to a less extent in the CHH context, and the hyper-DMRs in all sequence contexts largely overlapped with each other (Figures 2-11B and 11C).

To assess the function of each DNA demethylase, I compared the hyper-DMRs in *dme*, *rdd* and *drdd* mutants. Approximately 40 percent of the *drdd* hyper-DMRs were specific to the *drdd* quadruple mutant, and a higher proportion of *drdd* hyper-DMRs overlapped with *rdd* hyper-DMRs than with *dme* hyper-DMRs (Figure 2-12A). I next examined DNA methylation levels of WT and mutants replicate in all sequence contexts. At the hyper-DMRs of *drdd*, the *dme* mutant exhibited less CG methylation, but similar CHG and CHH methylation compared to WT (Figure 2-12B). DNA hypermethylation also occurred in all sequence contexts of the *rdd* mutant, with an increase of DNA methylation being more prominent in *drdd* (Figure 2-12B). These results reveal that DME acts redundantly with other DNA demethylases.

I then investigated the genome-wide distribution of hyper-DMRs in each mutant. The hyper-DMRs in *dme* were localized to different genomic regions from the WT genome, with a slight increase at TEs and a decrease in introns (Figure 2-12C). Remarkably, the hyper-DMRs in *drdd* and *rdd* were highly associated with promoters and intergenic regions and less associated with gene body regions, suggesting DNA demethylases share significant portion of genomic targets in regulating gene expression (Figure 2-12C).

The *drdd* hyper-DMRs were further divided into CG-only and multi-context groups. The CG-only group shows DNA hypermethylation only in the CG context, whereas the multi-context group was hypermethylated in CG and either CHG or CHH context in the *drdd* mutant compared to WT. The CG-only hyper-DMRs in *drdd* were highly enriched at gene bodies, with no significant difference between the multi-context hyper-DMRs and the entire *drdd* hyper-DMRs (Figure 2-13A). Given that gene body methylation occurs in the CG context of the coding regions and is depleted at the transcriptional start and termination sites (Bewick and Schmitz, 2017; Zilberman, 2017), I then investigated CG methylation patterns at gene body methylated (gbM) genes and their flanking regions. In gbM genes (Bewick et al., 2016), the *drdd* mutant and WT displayed similar CG methylation levels (Figure 2-13B). In the genes overlapped with the CG-only hyper DMRs in *drdd*, referred to as DRDD gbM genes, however, DNA methylation enrichment at gene bodies was reduced in *drdd* compared to WT (Figure 2-13B). In addition, more than half of the DRDD gbM genes (448 of 995) did not overlap with the previously identified gbM genes. These findings demonstrate that DRDD gbM genes are quite different from the typical gbM genes, in accordance with the previous report (Williams et al., 2022).

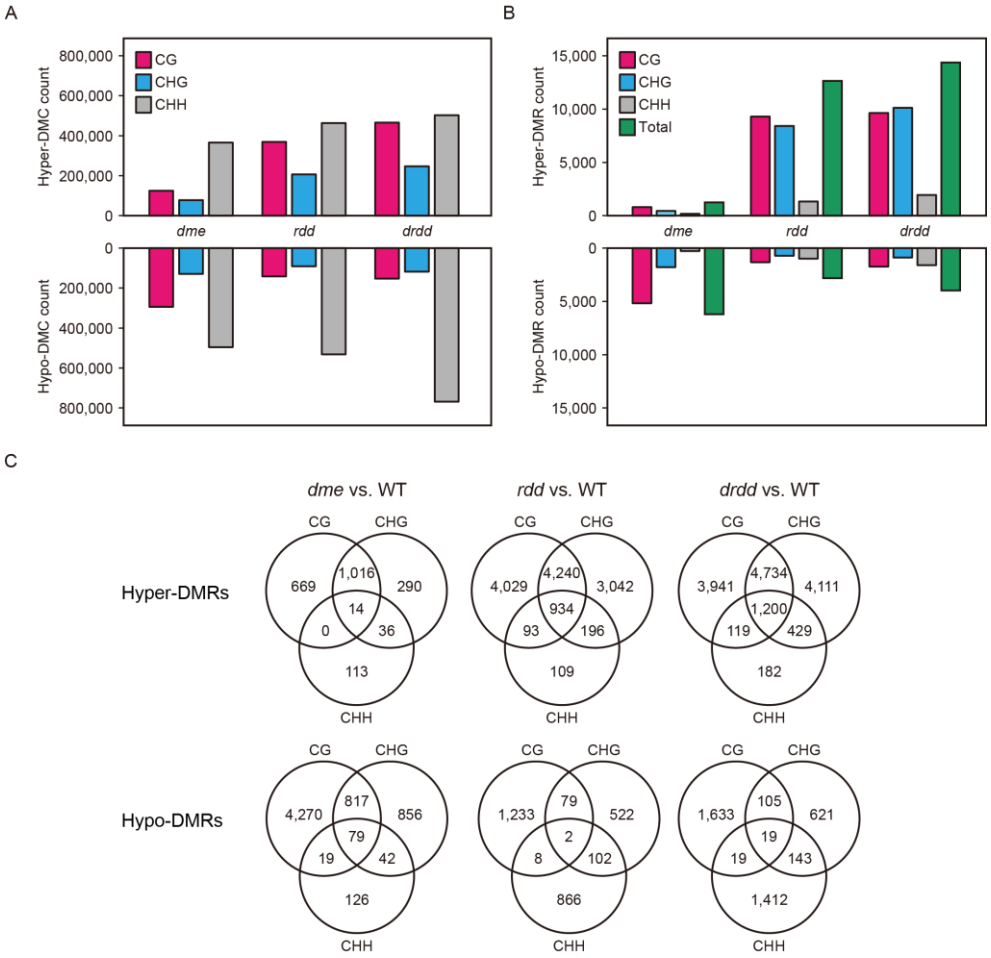

Figure 2-11. DNA methylation changes in the *drdd* mutant.

(A-B) Number of differentially methylated cytosines (DMCs) (A) and differentially methylated regions (DMRs) (B) in *dme*, *rdd* and *drdd* mutants compared to WT. (C) Venn diagrams showing the overlap of CG, CHG and CHH DMRs in *dme*, *rdd* and *drdd* mutants compared to WT.

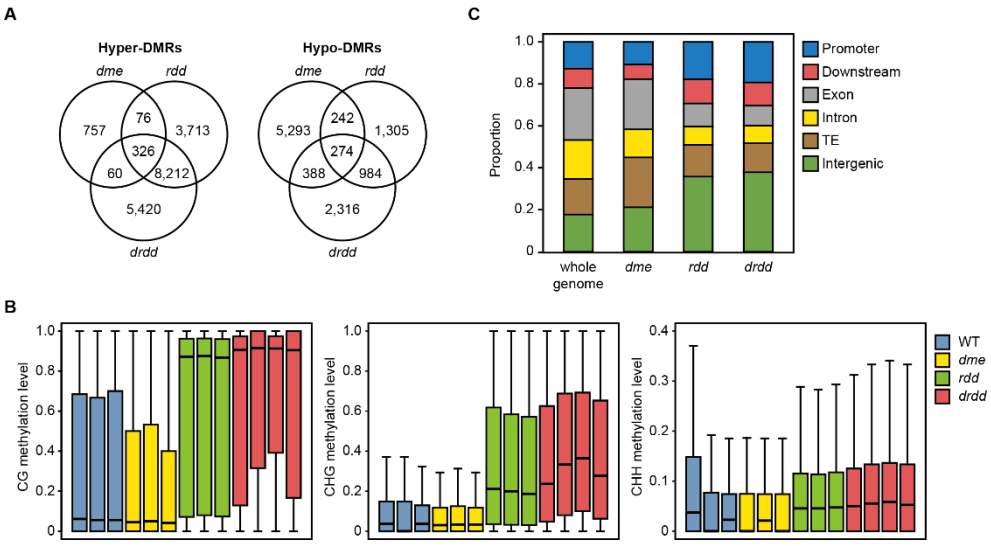

Figure 2-12. Comparative analysis of hyper-DMRs in WT, *dme*, *rdd* and *drdd* mutants.

(A) Venn diagrams indicating the overlap of DMRs among *dme*, *rdd* and *drdd* mutants. (B) Box plots of CG, CHG and CHH methylation levels of WT, *dme*, *rdd* and *drdd* mutants at the *drdd* hyper-DMRs. Replicates are shown in the same color. (C) Genomic distribution of hyper-DMRs in *dme*, *rdd* and *drdd* mutants.

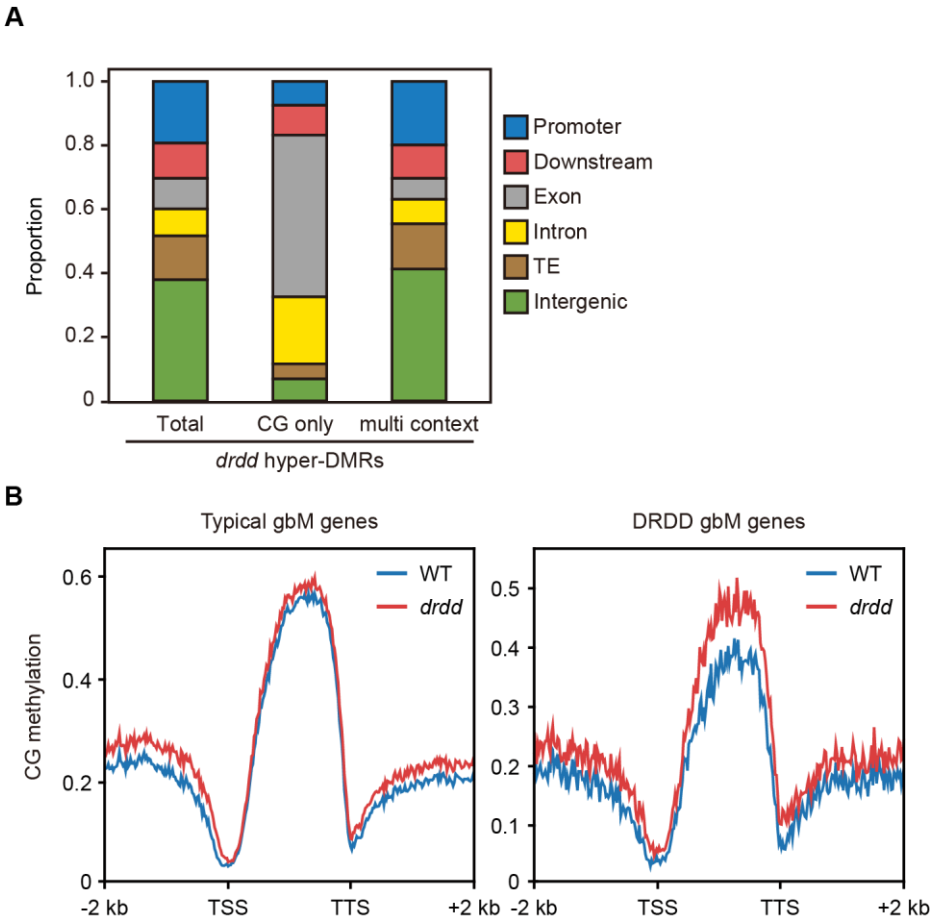

Figure 2-13. Gene body methylation (gbM) in the *drdd* mutant

(A) Genomic distribution of the *drdd* hyper-DMRs divided into CG-only and multiple context groups. (B) Metagene plots showing CG methylation levels at typical gbM genes and DRDD gbM genes in WT and *drdd*.

DNA demethylation by DNA demethylases is required for transcriptional regulation during reproductive development

To investigate whether gene expression changes induced by the *drdd* mutation were associated with DNA methylation changes, I counted the number of DEGs adjacent to the hyper-DMRs within 2 kb. The hyper-DMRs were more frequently found proximal to the downregulated genes than upregulated genes in *drdd*, and moreover, both the downregulated genes and the upregulated genes displayed a significant association with the neighboring hyper-DMRs (Figure 2-14A). I identified a total of 703 downregulated genes with hyper-DMRs in their promoter regions in *drdd*, hereafter referred to as *drdd* hyper DMR-associated genes. Hierarchical clustering analysis was performed to examine expression patterns of the hyper DMR-associated genes. More than 60 percent (444 genes) of these genes showed a significant decrease of expression only in a *drdd* quadruple mutant, approximately 25 and 10 percent (185 and 74 genes) of which displayed DME- and RDD-dependent expressions, respectively (Figure 2-14B). Notably, DNA methylation levels of hyper DMR-associated genes were substantially increased at the promoter regions, especially in the CG and CHH contexts (Figure 2-14C). These findings indicate a close relationship between DNA

hypermethylation at promoters and reduced expression in a *drdd* quadruple mutant, with functional redundancy of DNA demethylases.

I then performed GO analysis on the *drdd* hyper DMR-associated genes to clarify the biological functions under the control of active DNA demethylation. Consistent with abnormal floral development and a failure of fertilization in the *drdd* mutant (Figures 2-3 to 2-6), these hyper DMR-associated genes were significantly associated with pollen tube growth and cell differentiation (Figure 2-14D). As described previously (Zeng et al., 2021), GO terms related to biotic stress responses were also highly enriched (Table 2-6). Taken together, these findings suggest that DNA demethylase-mediated DNA demethylation may regulate the expression of genes related to pollen tube development and cell morphogenesis.

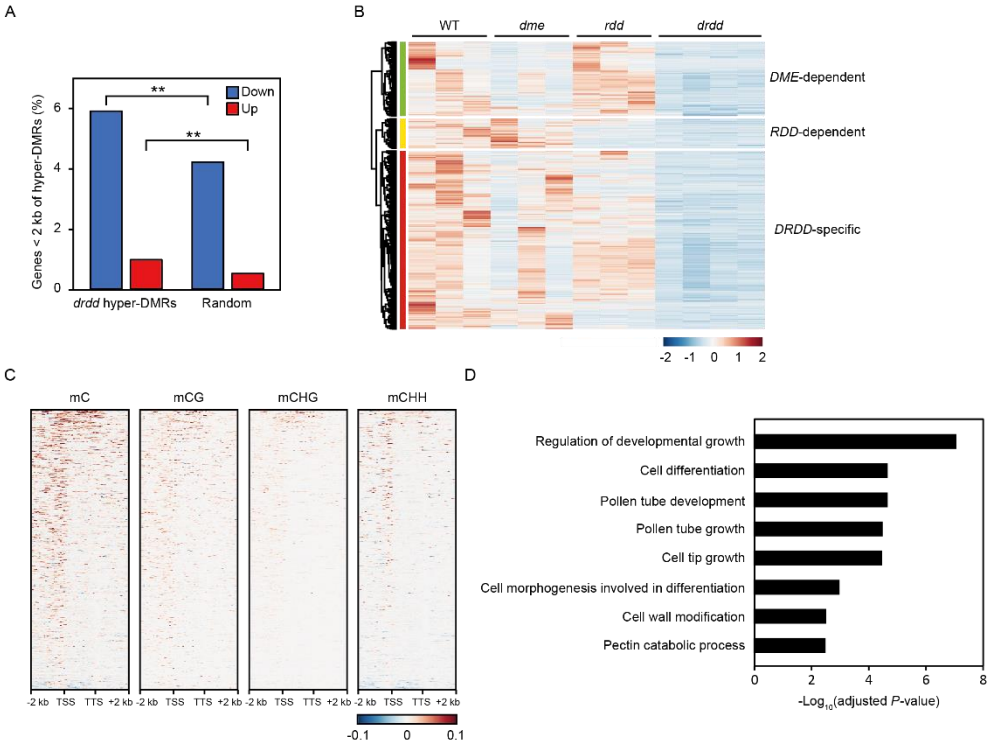

Figure 2-14. Transcriptional regulation by DNA demethylase-mediated DNA demethylation.

(A) Percentage of DEGs within 2 kb of hyper-DMRs in *drdd* compared to WT. Randomly selected windows were used as negative controls (Fisher's exact test, $^{**}P < 0.001$). (B) Hierarchical clustering of 703 hyper DMR-associated genes in *drdd*. Transcript levels in WT, *dme*, *rdd* and *drdd* replicates are shown. (C) DNA methylation levels in the 2 kb flanking regions of the *drdd* hyper DMR-associated genes. (D) GO analysis of the *drdd* hyper DMR-associated genes.

Table 2-6. Top 20 most enriched GO terms of *drdd* hyper-DMR associated genes.

GO ID	GO term	Count[a]	P-value	FDR[b]
GO:0009617	Response to bacterium	73	2.80E-11	8.49E-08
GO:0048638	Regulation of developmental growth	28	4.50E-11	9.10E-08
GO:0042742	Defense response to bacterium	64	1.20E-10	1.82E-07
GO:0006952	Defense response	116	4.20E-10	4.24E-07
GO:0098542	Defense response to other organism	101	1.00E-09	8.66E-07
GO:0043207	Response to external biotic stimulus	111	1.60E-09	9.37E-07
GO:0009751	Response to salicylic acid	35	3.70E-09	1.73E-06
GO:0014070	Response to organic cyclic compound	48	5.20E-08	2.10E-05
GO:0009620	Response to fungus	54	5.80E-08	2.20E-05
GO:0030154	Cell differentiation	64	6.80E-08	2.32E-05
GO:0048868	Pollen tube development	23	6.90E-08	2.32E-05
GO:0009860	Pollen tube growth	19	1.10E-07	3.51E-05
GO:0009932	Cell tip growth	21	1.20E-07	3.64E-05
GO:0050832	Defense response to fungus	41	3.10E-07	8.95E-05
GO:0009826	Unidimensional cell growth	29	8.90E-07	0.000225
GO:0048588	Developmental cell growth	22	3.10E-06	0.000606
GO:0031347	Regulation of defense response	40	5.00E-06	0.000948
GO:0060560	Developmental growth involved in morphogenesis	31	6.40E-06	0.001126
GO:0000904	Cell morphogenesis involved in differentiation	22	6.50E-06	0.001126
GO:0048468	Cell development	28	9.10E-06	0.001491

a Number of *drdd*-downregulated genes from list

b False discovery rate (FDR) from Benjamini-Hochberg multiple tests

DISCUSSION

In contrast to expressions of *ROS1*, *DML2* and *DML3* in vegetative tissues, the precise expression of *DME* in the central cell is responsible for endosperm development (Choi et al., 2002; Gehring et al., 2006). However, several studies have reported that DME activity is not confined to the central cell (Kim et al., 2021; Mathieu et al., 2007; Park et al., 2017; Zhang et al., 2019). To investigate the biological role of the entire DME family, central cell-specific complementation lines were generated in *dme* or *drdd* mutant backgrounds, bypassing seed abortion caused by the maternal *dme* allele (Kim et al., 2021; Williams et al., 2022; Zeng et al., 2021). In this study, I generated homozygous *dme* and *drdd* mutants by *in vitro* rescue of early seeds. A few embryos developed from the *in vitro*-cultured seeds exhibited developmental defects resulting from extra and irregular cell divisions as previously described (Sauer and Friml, 2004), and normally developed embryos also showed subtle phenotypic variations during vegetative development. Although I cannot rule out the possibility of the unexpected effect of *in vitro* rescue during vegetative growth, the *drdd* mutant displayed considerably retarded growth compared to WT and *dme* and *rdd* mutants (Figures 2-2A and 2-3A). After the transition from vegetative to reproductive

development, the effect of *in vitro* culture appeared to be compromised to some extent. WT flowers and siliques did not show any obvious phenotypes, whereas *dme* and *rdd* mutants occasionally developed abnormal flowers (Figures 2-4 and 2-5). Severe defects in floral organ determination, fertilization and silique elongation were frequently observed in the *drdd* mutant (Figures 2-4 and 2-6). These results suggest that DNA demethylases act redundantly during reproductive development.

DNA hypermethylation was observed in the *drdd* and *rdd* mutants, to a lesser extent in *dme*, compared to WT. In genic regions, DNA methylation was increased in all sequence contexts of the *drdd* mutant. In TE regions, CG and CHG methylation was increased, but CHH methylation was reduced in *drdd* (data not shown), as previously described in *rdd* (Le et al., 2014; Schumann et al., 2019). As CHH hypomethylation at TEs was often associated with repressed gene expression in *rdd*, these reports suggested that CHH methylation initiated by RdDM may function in TE silencing to repress cryptic transcription of nearby genes (Gent et al., 2013; Le et al., 2014; Schumann et al., 2019). Moreover, the *drdd* mutant exhibited a higher number of hyper-DMRs and a prominent increase in DNA methylation levels in comparison with the *rdd* mutant (Figures 2-11 and 2-12), consistent with the previous studies (Williams et al., 2022; Zeng et al., 2021). Nearly 60 percent

of the *drdd* hyper-DMRs overlapped with the *rdd* hyper-DMRs, but 40 percent of which were specific to the *drdd* quadruple mutant (Figure 2-12). These findings indicate that all four DNA demethylases would work together to remove DNA demethylation.

To examine whether the DRDD-mediated active DNA demethylation contributes to transcriptional regulation during reproductive development, I identified a total of 707 *drdd* hyper DMR-associated genes. Approximately 60 percent of these genes showed decreased expression only in *drdd*, but not in *dme* or *rdd*, with DNA hypermethylation at their promoter regions (Figure 2-14). In accordance with the failure of fertilization in the *drdd* mutant (Figures 2-3 and 2-4), GO terms related to pollen tube development, cell wall modification and pectin catabolic process were significantly enriched in the *drdd* hyper DMR-associated genes (Figure 2-14). DME and ROS1 are thought to act together for proper pollen tube growth (Khouider et al., 2021). All four candidate genes for DME/ROS1 targets in male fertility were downregulated in *drdd*, and notably, *POLLEN RECEPTOR LIKE KINASE 4* (*PRK4*) was associated with the *drdd*-hyper DMR. PRK4 is known to regulate the actin cytoskeleton at the pollen tube apex by mediating extracellular signals (Duckney et al., 2017). In addition, the *drdd* hyper-DMR associated genes involve the genes encoding expansins, pectin methylesterases and Rop-

interactive CRIB motif-containing proteins. I propose that the dysregulation of the genes involved in cell wall expansion may be responsible for abnormal pollen tube growth in the *drdd* mutant. Further experiments will be required to clarify the mechanisms underlying polarized pollen tube growth mediated by DNA demethylases.

REFERENCES

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

Berardini, T.Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., and Huala, E. (2015). The Arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. Genesis *53*, 474-485.

Bewick, A.J., Ji, L., Niederhuth, C.E., Willing, E.M., Hofmeister, B.T., Shi, X., Wang, L., Lu, Z., Rohr, N.A., Hartwig, B., et al. (2016). On the origin and evolutionary consequences of gene body DNA methylation. Proc. Natl. Acad. Sci. U. S. A. *113*, 9111-9116.

Bewick, A.J., and Schmitz, R.J. (2017). Gene body DNA methylation in plants. Curr. Opin. Plant Biol. *36*, 103-110.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114-2120.

Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J.J., Goldberg, R.B., Jacobsen, S.E., and Fischer, R.L. (2002). DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*. Cell *110*, 33-42.

Gehring, M. (2013). Genomic imprinting: insights from plants. Annu. Rev. Genet. *47*, 187-208.

Gehring, M., Huh, J.H., Hsieh, T.F., Penterman, J., Choi, Y., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (2006). DEMETER DNA glycosylase establishes *MEDEA* polycomb gene self-imprinting by

allele-specific demethylation. Cell *124*, 495-506.

Gong, Z., Morales-Ruiz, T., Ariza, R.R., Roldan-Arjona, T., David, L., and Zhu, J.K. (2002). *ROS1*, a repressor of transcriptional gene silencing in *Arabidopsis*, encodes a DNA glycosylase/lyase. Cell *111*, 803-814.

Halter, T., Wang, J., Amesefe, D., Lastrucci, E., Charvin, M., Singla Rastogi, M., and Navarro, L. (2021). The Arabidopsis active demethylase ROS1 *cis*-regulates defence genes by erasing DNA methylation at promoter-regulatory regions. Elife *10*, e62994.

Huang, C.F., Miki, D., Tang, K., Zhou, H.R., Zheng, Z., Chen, W., Ma, Z.Y., Yang, L., Zhang, H., Liu, R., et al. (2013). A pre-mRNA-splicing factor is required for RNA-directed DNA methylation in *Arabidopsis*. PLoS Genet. *9*, e1003779.

Huh, J.H., Bauer, M.J., Hsieh, T.F., and Fischer, R.L. (2008). Cellular programming of plant gene imprinting. Cell *132*, 735-744.

Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019a). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. *37*, 907-915.

Kim, J.S., Lim, J.Y., Shin, H., Kim, B.G., Yoo, S.D., Kim, W.T., and Huh, J.H. (2019b). ROS1-Dependent DNA Demethylation Is Required for ABA-Inducible *NIC3* Expression. Plant Physiol *179*, 1810-1821.

Kim, M., Ohr, H., Lee, J.W., Hyun, Y., Fischer, R.L., and Choi, Y. (2008). Temporal and spatial downregulation of *Arabidopsis* MET1 activity results in global DNA hypomethylation and developmental defects. Mol. Cells *26*, 611-615.

Kim, S., Park, J.S., Lee, J., Lee, K.K., Park, O.S., Choi, H.S., Seo, P.J., Cho, H.T., Frost, J.M., Fischer, R.L., and Choi, Y. (2021). The DME

demethylase regulates sporophyte gene expression, cell proliferation, differentiation, and meristem resurrection. Proc. Natl. Acad. Sci. U. S. A. *118*, e2026806118.

Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571-1572.

Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. *11*, 204-220.

Le, T.N., Schumann, U., Smith, N.A., Tiwari, S., Au, P.C., Zhu, Q.H., Taylor, J.M., Kazan, K., Llewellyn, D.J., Zhang, R., et al. (2014). DNA demethylases target promoter transposable elements to positively regulate stress responsive genes in *Arabidopsis*. Genome Biol. *15*, 458.

Lin, W., Sun, L., Huang, R.Z., Liang, W., Liu, X., He, H., Fukuda, H., He, X.Q., and Qian, W. (2020). Active DNA demethylation regulates tracheary element differentiation in *Arabidopsis*. Sci. Adv. *6*, eaaz2963.

Mathieu, O., Reinders, J., Caikovski, M., Smathajitt, C., and Paszkowski, J. (2007). Transgenerational stability of the *Arabidopsis* epigenome is coordinated by CG methylation. Cell *130*, 851-862.

Matzke, M.A., and Mosher, R.A. (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat. Rev. Genet. *15*, 394-408.

Park, J.S., Frost, J.M., Park, K., Ohr, H., Park, G.T., Kim, S., Eom, H., Lee, I., Brooks, J.S., Fischer, R.L., and Choi, Y. (2017). Control of DEMETER DNA demethylase gene transcription in male and female gamete companion cells in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. U. S. A. *114*, 2078-2083.

Penterman, J., Zilberman, D., Huh, J.H., Ballinger, T., Henikoff, S., and Fischer, R.L. (2007). DNA demethylation in the *Arabidopsis* genome. Proc. Natl. Acad. Sci. U. S. A. *104*, 6752-6757.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139-140.

Sauer, M., and Friml, J. (2008). In vitro culture of *Arabidopsis* embryos. Methods Mol. Biol. *427*, 71-76.

Schumann, U., Lee, J.M., Smith, N.A., Zhong, C., Zhu, J.K., Dennis, E.S., Millar, A.A., and Wang, M.B. (2019). *DEMETER* plays a role in DNA demethylation and disease response in somatic tissues of Arabidopsis. Epigenetics *14*, 1074-1087.

Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nat. Rev. Genet. *14*, 204-220.

Tang, K., Lang, Z., Zhang, H., and Zhu, J.K. (2016). The DNA demethylase ROS1 targets genomic regions with distinct chromatin modifications. Nat. Plants *2*, 16169.

Williams, B.P., Bechen, L.L., Pohlmann, D.A., and Gehring, M. (2022). Somatic DNA demethylation generates tissue-specific methylation states and impacts flowering time. Plant Cell *34*, 1189-1206.

Yamamuro, C., Miki, D., Zheng, Z., Ma, J., Wang, J., Yang, Z., Dong, J., and Zhu, J.K. (2014). Overproduction of stomatal lineage cells in *Arabidopsis* mutants defective in active DNA demethylation. Nat. Commun. *5*, 4062.

Zeng, W., Huang, H., Lin, X., Zhu, C., Kosami, K.I., Huang, C., Zhang, H., Duan, C.G., Zhu, J.K., and Miki, D. (2021). Roles of DEMETER in regulating DNA methylation in vegetative tissues and pathogen

resistance. J. Integr. Plant Biol. *63*, 691-706.

Zhang, C., Hung, Y.H., Rim, H.J., Zhang, D., Frost, J.M., Shin, H., Jang, H., Liu, F., Xiao, W., Iyer, L.M., et al. (2019). The catalytic core of DEMETER guides active DNA demethylation in *Arabidopsis*. Proc. Natl. Acad. Sci. U. S. A. *116*, 17563-17571.

Zhang, H., Lang, Z., and Zhu, J.K. (2018). Dynamics and function of DNA methylation in plants. Nat. Rev. Mol. Cell Biol. *19*, 489-506.

Zilberman, D. (2017). An evolutionary case for functional gene body methylation in plants and animals. Genome Biol. *18*, 87.

CHAPTER 3

Comparative Analysis of Genome and Epigenome Landscapes in *Brassica rapa* Subspecies

ABSTRACT

Chinese cabbage (*Brassica rapa* subsp. *pekinensis*) and turnip (*B. rapa* subsp. *rapa*) display strikingly distinct morphologies within the same species. While Chinese cabbage forms a leafy head and small roots, turnip has lobed leaves and an enlarged taproot derived from swollen hypocotyl and root tissues. However, the mechanisms that develop highly contrasting phenotypes between the two subspecies remain elusive. Considering their genomic similarity, it is presumably due to the unraveled role of epigenetic factors beyond genetic differences in subspecies-specific divergence. Here, I profiled genome-wide gene expression, chromatin accessibility, histone mark enrichment and DNA methylation in Chinese cabbage and turnip seedlings. The two subspecies showed differential expressions of approximately 8,000 genes, and more than 30 percent of accessible chromatin regions (ACRs) were subspecies-specific. The ACRs were associated with the enrichment of histone H3 lysine 27 acetylation (H3K27ac) and the depletion of DNA methylation. In addition, genes harboring multiple ACRs displayed higher expression dynamics, implying that ACRs have potential features as transcriptional enhancers. Moreover, distant ACRs (dACRs), identified as

putative distal enhancers, exhibited high sequence similarity but divergent chromatin accessibility between the two subspecies. Long terminal repeat (LTR) retrotransposons were also highly enriched in turnip-specific dACRs, suggesting the role of LTRs as an evolutionary driving force in chromatin accessibility divergence. Remarkably, motif enrichment analysis revealed that turnip-specific dACRs showed significant enrichment in transcription factor motifs related to vascular stem cell maintenance and differentiation during secondary growth, with increased H3K27ac levels. The predicted turnip-specific dACRs were validated for the transcriptional enhancer activity by reporter assay. This study provides insights into the subspecies-specific divergence of putative enhancers, thereby leading to morphotype diversification within the same species.

INTRODUCTION

Brassica species have been cultivated as economically important crops worldwide. Due to a whole genome triplication event that occurred approximately 9-15 million years ago, *Brassica* genomes have undergone diversification, followed by an extensive fractionation (Cheng et al., 2014; Wang et al., 2011). Through artificial selection during domestication, *Brassica* crops have appeared to exhibit different morphology within the same species as well as similar morphology between different species (Cheng et al., 2016). Remarkably, *Brassica rapa* L ($2n = 2x = 20$) consists of various subspecies with morphological divergence. Chinese cabbage (*B. rapa* subsp. *pekinensis*) and Pak choi (*B. rapa* subsp. *chinensis*) are the leafy vegetables, whereas turnip (*B. rapa* subsp. *rapa*) has enlarged edible roots. In addition to the oilseed field mustard (*B. rapa* subsp. *oleifera*), broccoli rabe or rapini (*B. rapa* subsp. *rapa*) have small edible buds (Gómez-Campo and Prakash, 1999; Qi et al., 2017).

The two subspecies Chinese cabbage and turnip exhibit strikingly different morphological characteristics, despite their genetic similarity that allows the formation of the hybrid between them. Whereas Chinese cabbage

forms a leafy head composed of incurved leaves with small roots, turnip shows pinnately lobed leaves, a swollen hypocotyl and a bulbous taproot, with the overall morphology similar to radish (*Raphanus sativus*) that belongs to a different genus. During the secondary growth, xylem tissues in turnip hypocotyl-tubers show thinner and less lignified cell walls compared to non-tuberizing *B. rapa*, indicating a higher meristematic activity in the vascular cambium (Liu et al., 2019). Genetic studies have attempted to identify quantitative trait loci (QTL) involved in the root development of turnip (Lou et al., 2007; Lu et al., 2008; Wu et al., 2021; Zhao et al., 2010). Moreover, comparative genomic analysis between tuber-forming morphotypes (turnip in *B. rapa* and kohlrabi in *Brassica oleracea*) and non-tuber forming morphotypes identified genomic regions under subgenome parallel selection, where a subset of genes are associated with sugar or cellulose transport and cell growth (Cheng et al., 2016). However, the mechanism to drive highly distinct morphotypes within the same *B. rapa* species cannot be fully explained by genetic factors, presumably due to a coordinative role of epigenetic factors.

Enhancers are *cis*-regulatory elements that activate expression of target genes, independent of the relative distance, location or orientation to their cognate promoter (Li et al., 2016; Long et al., 2016). A combination of

transcription factors (TFs) can bind to enhancers and recruit coactivators such as histone acetyltransferases and chromatin remodelers to increase chromatin accessibility, thereby leading to transcriptional activation of target genes (Ong and Corces, 2011; Shlyueva et al., 2014). Notably, enhancers are brought into proximity of their cognate promoters through chromatin looping, which allows long-range interactions between distal enhancers and their target genes spanning from kilobases to megabases (Schoenfelder and Fraser, 2019). In animals, active enhancers are generally depleted of nucleosomes and associated with histone H3 lysine 27 acetylation (H3K27ac), low DNA methylation levels and non-coding enhancer RNA transcripts, while inactive enhancers exhibit low chromatin accessibility and H3K27 trimethylation (H3K27me3) enrichment (Andersson et al., 2014; Kouzarides, 2007; Rada-Iglesias et al., 2011). H3K4 monomethylation (H3K4me1) is present at mammalian enhancers regardless of their activity (Rada-Iglesias et al., 2011).

Enhancers play important roles in cell identity control, development and evolutionary process by orchestrating precise spatiotemporal patterns of gene expression in a cell type- or tissue-specific manner (Long et al., 2016; Ong and Corces, 2011; Shlyueva et al., 2014). Whereas highly conserved enhancers can control fundamental biological processes including embryo development, genetic variations at enhancers can drive evolutionary changes

between species in various organisms, in many cases associated with morphological divergence (Arnold et al., 2014; Kvon et al., 2016; Prescott et al., 2015; Villar et al., 2015). In plants, phenotypic variations can arise from sequence changes in distal *cis*-regulatory elements during domestication, exemplified by inflorescence architecture in maize and diversification of flowering time in Arabidopsis accessions as well as leaf shape in the Brassicaceae family (Liu et al., 2014a; Studer et al., 2011; Vuolo et al., 2016). In addition to the sequence properties of *cis*-regulatory elements, the chromatin accessibility is indicative of active enhancers and has emerged as a key determinant of development, disease and evolution in diverse organisms (Corces et al., 2016; Corces et al., 2018; Gao et al., 2018; Zhang et al., 2021).

Due to the distance- and orientation-independent nature of enhancer activity, enhancer discovery has remained challenging, especially in plants. Although genome-wide identification of enhancer chromatin features has enabled the discovery of tens of thousands of enhancer candidates in human and other animal genomes (Andersson et al., 2014; Kvon et al., 2016), chromatin hallmarks of plant enhancers and their dynamic regulation in development and evolution remain to be clarified. By assaying chromatin accessibility, in parallel with histone mark enrichment and DNA methylation analysis, DNase-hypersensitive sites or transposase-accessible chromatin

regions have been predicted as putative enhancers in Arabidopsis, rice and maize genomes (Oka et al., 2017; Ricci et al., 2019; Zhang et al., 2012; Zhu et al., 2015). Distal *cis*-regulatory elements have also been identified based on chromatin accessibility from various angiosperm species, suggesting that they are highly conserved among species but show dynamic chromatin behaviors (Li et al., 2019; Reynoso et al., 2019).

I hypothesized that highly contrasting phenotypes between the two subspecies Chinese cabbage and turnip may arise from the coordination of epigenetic and genetic factors. Here I report the different profiles of transcriptome and epigenome in Chinese cabbage (Chiifu-401-42; CF) and Ganghwa turnip (G14; Park et al., 2019) seedlings. The majority of accessible chromatin regions (ACRs) were associated with H3K27ac enrichment and low DNA methylation levels. Furthermore, genes associated with ACRs showed a significant increase in expression dynamics, suggesting that ACRs may retain the function of transcriptional enhancers. A total of 2,972 and 2,480 distant ACRs (dACRs) were identified as enhancer candidates in CF and G14, respectively, with high sequence similarity and differential chromatin accessibility. In particular, a higher proportion of transposable elements (TEs) overlapped with G14-specific dACRs in comparison with CF-specific dACRs, among which long terminal repeat retrotransposons (LTRs)

were particularly enriched in G14-specific dACRs. Notably, G14-specific dACRs exhibited a significant enrichment of TF motifs associated with hypocotyl and root development, and a subset of dACRs were validated for transcriptional enhancer activity. These results suggest that subspecies-specific divergence in distal regulatory regions may contribute to different morphotypes during evolution.

MATERIALS AND METHODS

Plant materials

Chinese cabbage *(B. rapa* L. cv. Chiifu-401-42; CF) and a doubled haploid line of Ganghwa turnip (G14; (Park et al., 2019)) were grown on 1x Murashige and Skoog (MS) medium in a growth chamber at 22°C under 16 h of fluorescent light at 20 ± 10 μmol m^{-2} s^{-1}. For RNA-seq, whole genome bisuifite (BS)-seq, assay for transposase-accessible chromatin (ATAC)-seq, chromatin immunoprecipitation (ChIP)-seq and Hi-C analysis, 18 day-after-sowing (DAS) seedlings of CF and G14 were harvested (Figure 3-1B), as the anatomy of hypocotyl-tuber in turnips becomes distinct from non-tuber forming turnips after 16 DAS (Liu et al., 2019).

RNA-seq and analysis

For each of three biological replicates, three seedlings were pooled. Total RNA was extracted using the RNeasy Plant Mini kit (Qiagen) and treated with RNase-free DNase (Qiagen) to remove any genomic DNA contaminants. RNA-seq libraries were constructed according to the previous report (Zhong et al., 2011). RNA-seq was performed on the Illumina HiSeq 4000 sequencing system. Paired-end reads were filtered with Trimmomatic

v.0.38 (Bolger et al., 2014) with the parameters TruSeq3-PE-2.fa:2:30:10, LEADING:15, TRAILING:15, SLIDINGWINDOW:4:18 and MINLEN:60. Reads were aligned to the reference Chiifu-401-42 genome V3.0 (Wang et al., 2011) using hisat2 version 2.2.1 (Kim et al., 2019) and quantified with htseq-count version 0.6.1p1 (Anders et al., 2015). DEGs were identified using the edgeR package (Robinson et al., 2010) with fold change > 2 and FDR < 0.05.

BS-seq and analysis

A total of 5 μg of genomic DNA was used to generate BS-seq libraries using the KAPA Library kit (Roche) and EpiTect Bisulfite Kit (Qiagen). BS-seq was performed using the Illumina HiSeq 4000 sequencing system. 151 bp paired-end reads were filtered with Trimmomatic v.0.38 (Bolger et al., 2014) with the same parameters as in transcriptome analysis. Filtered reads were mapped to Chiifu-401-42 genome V3.0, and methylation was called with Bismark v0.19.1 (Krueger and Andrews, 2011). Only cytosine sites with 5× coverage were used for subsequent analysis. DMCs and DMRs were identified as described previously (Huang et al., 2013). DMRs were finally identified based on the regions with a length ≥ 100 bp, ≥ 5 DMCs, and the mean methylation difference ≥ 0.3 for CG, ≥ 0.15 for CHG, or ≥ 0.1 for CHH.

ATAC-seq and analysis

ATAC-seq was performed as described previously (Lu et al., 2017) with minor modifications. Five seedlings were chopped with a razor blade in pre-chilled lysis buffer (15 mM Tris-HCl pH7.5, 20 mM NaCl, 80 mM KCl, 0.5 mM spermine, 5 mM 2-ME, 0.2% TritonX-100). The suspension was filtered through two layers of Miracloth and a 40-μm cell strainer, then centrifuged at 2,800 g for 20 min at 4°C. The pellet was resuspended in pre-chilled lysis buffer and loaded on top of dense sucrose buffer (20 mM Tris-HCl Ph8.0, 2 mM $MgCl_2$, 2 mM EDTA, 15 mM 2-ME, 1.7 M sucrose, 0.2% TritonX-100) in 2mL tube. After centrifugation at 2,200 g for 20 min at 4°C, the nuclei were stained with 4,6-diamidino-2-phenylindole and loaded into a flow cytometer (BD FACS Aria III). A total of 50,000 to 100,000 nuclei were sorted and centrifuged at 1,500 g for 7 min at 4°C. The pellet was incubated with 2 μl of TDE1 transposase (Illumina) in TD buffer at 37°C for 30 min, then purified using a QIAquick PCR purification kit (Qiagen). The transposed DNA was amplified using NEBNext High Fidelity PCR Mix for 10-12 cycles. The optimal number of PCR cycles was determined as described previously (Bajic et al., 2017). ATAC-seq libraries were performed on the Illumina HiSeq 4000 sequencing system.

Low quality reads with > 20% of Phred score < 20 were filtered, and adapters were trimmed with cutadapt v.3.4 (Martin, 2011). Filtered reads were mapped to *B. rapa* chloroplast genome (Wang et al., 2011) with bowtie v.1.0.0 (Langmead, 2010) and reads mapped on the chloroplast genome were filtered. Remained reads were mapped to *B. rapa* V3.0 genome (Wang et al., 2011) with bowtie2 v.2.3.4.2 (Langmead and Salzberg, 2012) with default parameters. Duplicate reads were marked with picard v.2.6.0 (http://broadinstitute.github.io/picard/), and uniquely mapped reads with MAPQ ≥ 30 were used for further analysis. Enriched peaks were identified with MACS2 v.2.6.0 (Zhang et al., 2008). Sequence identity between CF and G14 and SNP density from 80 *B. rapa* accessions were calculated using GATK v.3.6.0 (McKenna et al., 2010). Matrices for heatmaps and profiles were calculated with deepTools v.3.4.3 (Ramirez et al., 2014). Motif enrichment analysis was performed using MEME-ChIP v.5.4.1 (Machanick and Bailey, 2011) and 500-bp-long dACR peak summit-centered sequences were used as input.

ChIP-seq and analysis

ChIP was performed following the previous protocols (Lee et al., 2007; Lee et al., 2014). A total of 2~3g of whole seedlings were vacuum-infiltrated

with 1% formaldehyde for crosslinking, followed by the addition of glycine for quenching the crosslinking process. Plant tissues were ground in liquid nitrogen and resuspended in nuclear extraction buffer I (0.4 M sucrose, 10 mM Tris-HCl, pH 8, 10 mM $MgCl_2$, 5 mM β-mercaptoethanol, 0.1 mM phenylmethylsulfonyl fluoride [PMSF], and 1× protease inhibitor; Roche), then filtered through Miracloth twice. After centrifugation at 2800g for 20 min at 4°C, the pellet was resuspended in nuclear extraction buffer II (0.25 M sucrose, 10 mM Tris-HCl, pH 8, 10 mM $MgCl_2$, 1% Triton X-100, 5 mM β-mercaptoethanol, 0.1 mM PMSF, and 1× protease inhibitor), then centrifuged at 12,000g for 10 min at 4°C. The pellet was resuspended in 300 μL of nuclear extraction buffer III (1.7 M sucrose, 10 mM Tris-HCl, pH 8, 0.15% Triton X-100, 2 mM $MgCl_2$, 5 mM β-mercaptoethanol, 0.1 mM PMSF, and 1× protease inhibitor) and loaded on the surface of 1.5 ml of nuclear extraction buffer III, followed by centrifugation at 16,000g for 1h at 4°C. The nuclei pellet was resuspended in lysis buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 1% SDS, and 1× protease inhibitor; Roche). Chromatin was sonicated into 0.3-0.8 kb using covaris S2x system (duty cycle: 10, intensity: 8, cycle per burst: 50, treatment time: 12 min). After preclearing with dynabeads protein G Invitrogen), 3~5 μg of anti-H3K4me1 (Abcam; xx), anti-H3K27ac (Abcam; xx), anti-H3K27me3 (Abcam; xx) or anti-H3K9me2 (Abcam; xx) antibody

was added to the chromatin solution and incubated overnight at 4°C. The precipitates were eluted from the beads, then reverse-crosslinked, followed by removal of proteins by proteinase K treatment. DNA was recovered by phenol-chloroform extraction followed by ethanol precipitation, and ChIP-seq libraries were constructed using the NEBNext Ultra II DNA Library Prep kit. ChIp-seq was performed using the Illumina HiSeq 4000 sequencing system.

Paired-end reads were filtered with Trimmomatic v.0.38 (Bolger et al., 2014) with the same parameters as in transcriptome analysis and mapped to *B. rapa* V3.0 genome with bowtie2 v.2.3.4.2 (Langmead and Salzberg, 2012) with default parameters. Peak calling was performed with MACS2 v.2.6.0 (Zhang et al., 2008) and matrices for heatmaps were calculated with deepTools v.3.4.3 (Ramirez et al., 2014).

Reporter assay

For functional validation of dACRs by protoplast transfection, the enhancer candidates were cloned into the upstream region of the minimal *CaMV 35S* promoter (−50 to −2 bp) (Zhu et al., 2015) followed by the GFP reporter construct and *nopaline synthase* terminator in the protoplast expression vector pHBT (Table 3-1). The constructs were then transformed

into protoplasts from the leaves of turnip. Protoplast transfection was performed as previously described (Yoo et al., 2007) with some modifications. The second and third leaves from four-week-old plants were cut into thin strips and immersed in 10 mL of enzyme solution (20 mM MES-KOH, pH 5.7, 0.4 M mannitol, 1.5% Cellulase R-10, 0.5% Macerozyme R-10, 10 mM $CaCl_2$, 20mM KCl and 0.1% BSA). Protoplasts were released with 20 rpm shaking for 3h at 25°C and filtered through 75 μm nylon mesh, then centrifuged at 100g for 2min at room temperature. After washing with the W5 solution (2 mM MES–KOH, pH 5.7, 154 mM NaCl, 125 mM CaCl2, 5 mM KCl), the protoplasts were kept on ice for 30 min and resuspended in MMG solution (4 mM MES–KOH, pH 5.7, 0.4 M mannitol, 15 mM $MgCl_2$). Approximately 25,000~35,000 protoplasts were transfected with 10–20 μg of the plasmid constructs containing enhancer candidates using the polyethylene glycol-calcium transfection-mediated method. After incubation in W5 solution at room temperature for 14h, GFP signals were observed using a microscope.

Table 3-1. List of primers.

Name	Sequence	Purpose
G14_peak_2478_F	TGCAACTTGAGTAAAGGCATAAGAG	Cloning
G14_peak_2478_R	AAACATCCTTACTTTAGCCCCTTAAC	Cloning
G14_peak_4506_F	CTCTCTTTCTCTTCTGTTTTCTTCAG	Cloning
G14_peak_4506_R	ACTGGAGTTTGACATCTTTATCAAAAC	Cloning
G14_peak_5781_F	AATAGCATAAATAAAATTGGGAGGC	Cloning
G14_peak_5781_R	TTCCACCAAGGGCATGTTC	Cloning
G14_peak_7379_F	AATCGGGGATTAGTTTTGTGTGTTA	Cloning
G14_peak_7379_R	TTCTTTTTTGTGTTTTTGCTTTGGG	Cloning
G14_peak_7718_F	TGGGTTATAAATAATTGGACCAAACAA	Cloning
G14_peak_7718-R	TTTATTATTCCCTTGTTAGATATTGGC	Cloning
G14_peak_7913_F	AAAAAATCACTTTTTCTTAGCCTTCC	Cloning
G14_peak_7913_R	TCATTGGAAATGCTAGACCATTGTAG	Cloning
G14_peak_8042_F	ACGAATCTGTTAGGCACGTAAAGG	Cloning
G14_peak_8042_R	GAACAAAGTGTAGAATAGATTTAGTAGC	Cloning
G14_peak_10093_F	GGCTTTCGTGATGGTTGGAGCTT	Cloning
G14_peak_10093_R	AATTGAAATGTGTGCCTTTGTCTAA	Cloning
G14_peak_11026_F	AAAATATTTAAGTCCGTATTGCGGGC	Cloning
G14_peak_11026_R	GGTCCAAACCCCATCTTTAGATC	Cloning
G14_peak_11465_F	AGACTCTGGCTTAGACTCTGGC	Cloning
G14_peak_11465_R	TTCGGTTTCGGTTCTTTGGATAC	Cloning
G14_peak_14439_F	GCAAAAATGAACTCATATAACATGAC	Cloning
G14_peak_14439_R	GAGATGACTTCCATAAAACCGC	Cloning
G14_peak_15178_F	AGTCACTGGGTTTTGTTTCGTTC	Cloning
G14_peak_15178_R	ACTCCGTTTTGACTTGTTTGAGTC	Cloning
G14_peak_15187_F	CATTAGTATTTATTAATCGTCGGAGTG	Cloning
G14_peak_15187_R	ATTTATTCCCCATCAAAATTGGAAG	Cloning
G14_peak_15472_F	GAATCGCAGCAGAACACTTG	Cloning
G14_peak_15472_R	TCTTTAACACTTTCATCTCTTTTATTAGG	Cloning
G14_peak_16990_F	CATCATTAAAACTAAGATTTAGCAGAGTG	Cloning
G14_peak_16990_R	GTATAGGTAGCAGAAGAGGCTAAC	Cloning
G14_peak_16991_F	ATGGACATATGTTTCAGTGGTGG	Cloning
G14_peak_16991_R	GGAGCATTAAACACCGTAAGTAAATG	Cloning

RESULTS

CF and G14 exhibit high genetic similarity with contrasting morphologies

Turnip (G14) displays similar morphology to radish, rather than Chinese cabbage (CF) within the same species. While CF has a compact leafy head with broad leaves and small roots, G14 has lobed leaves with long petioles and an enlarged root developed from thickened hypocotyl and root tissues with white and purple colors (Figure 3-1A). During seedling stages at 18 DAS, in contrast to white root colors in CF, the roots of G14 consist of purple and white skins (Figure 3-1B). This is consistent with the previous study between turnip and non-tuber forming Pak choi (*Brassica campestris*) (Liu et al., 2019). The anatomic analysis also revealed that the xylem cell walls in hypocotyl tissues become lignified in Pak choi, whereas the vascular cambium is still active in turnip at 16 DAS. These findings indicate the distinct hypocotyl and root development between tuber-forming and non-forming *Brassica* species at seedling stages. To assess genetic similarities, the previous study compared the sequences of coding regions in CF, G14, cabbage (*B. oleracea*) and radish (*R. sativus*) (Park et al., 2019). Turnip has a

remarkably high degree of sequence identity to Chinese cabbage (99.37 %), but to a less degree to radish (90.88%) (Figure 3-1C).

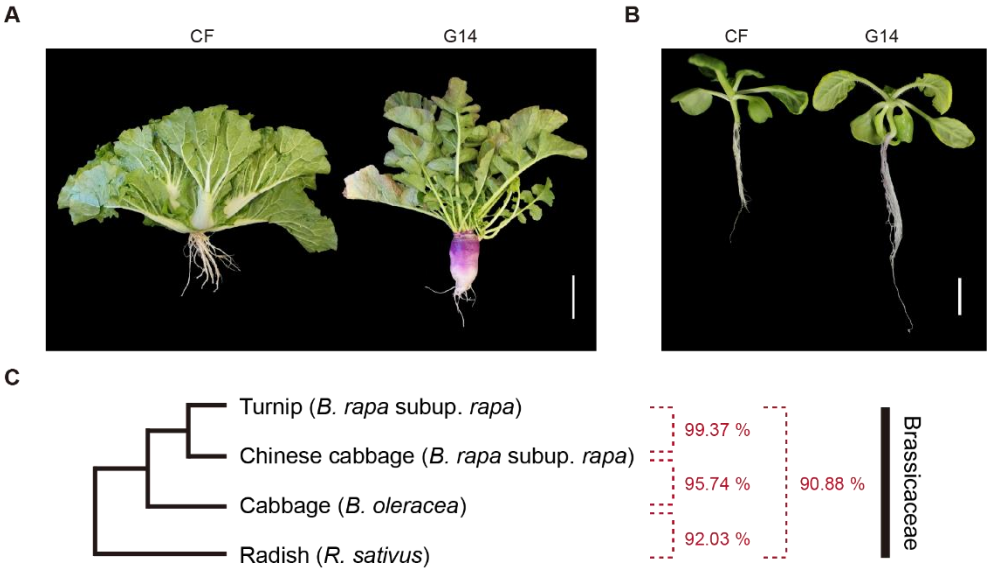

Figure 3-1. Morphological difference and genetic similarity in Chinese cabbage (CF) and turnip (G14).

(A) Phenotypes of whole plants of CF and G14 grown in the field at 40 days after planting. Scale bar = 5 cm. (B) Phenotypes of CF and G14 seedlings grown in the field at 18 day-after-sowing (DAS). Scale bar = 1 cm. (C) Sequence identity of coding sequences among Chinese cabbage (CF), turnip (G14), cabbage and radish in the Brassicaceae family.

CF and G14 seedlings have discrete profiles of transcriptome and epigenome

Based on a high sequence similarity between CF and G14, I performed RNA-sequencing (RNA-seq) to investigate whether differences in gene expression contribute to distinct phenotypes in the two subspecies. A total of 8,227 differentially expressed genes (DEGs) were identified between CF and G14 seedlings, approximately half of which were upregulated and the other half downregulated in G14 (4,038 and 4,189 genes, respectively) (Figure 3-2A). To examine the biological processes related to these DEGs, I next carried out Gene Ontology (GO) analysis. The upregulated genes in G14 were significantly enriched for oxidation-reduction process, stress response and inorganic ion transport, and the downregulated genes in G14 were most enriched for photosynthesis and metabolic process (Figure 3-2B). I then analyzed expressions of several candidate genes for tuberous organ enlargement under subgenome parallel selection in turnip and kohlrabi (*B. oleracea*) (Cheng et al., 2016). The *B. rapa* genes orthologous to *Arabidopsis EXPANSIN A7* (*EXPA7*), *EXPA5*, *ATP-BINDING CASSETTE B4* (*ABCB4*), *CELLULOSE SYNTHASE LIKE A11* (*CSLA11*), *ROOT PHOTOTROPISM 2* (*RPT2*), *ROOT HAIR DEFECTIVE 2* (*RHD2*) exhibited significantly increased expression in G14 compared to CF (Figure 3-2C). These results

suggest that differential gene expression can lead to phenotypic variations in *B. rapa* subspecies.

Given that chromatin alterations are associated with transcriptional regulation, I profiled the landscapes of chromatin accessibility, histone modifications (H3K27ac, H3K27me3, H3K4me1 and H3K9me2) and DNA methylation in CF and G14 seedlings with the assay for transposase-accessible chromatin (ATAC)-seq, chromatin immunoprecipitation (ChIP)-seq and whole genome bisulfite (BS)-seq. To investigate chromatin accessibility, a hallmark of active enhancers, a total of 21,878 and 17,773 accessible chromatin regions (ACRs) were identified in CF and G14 (Figure 3-3A). More than half of these ACRs (55 and 68 percent in CF and G14, respectively) were shared by CF and G14, but approximately 45 and 32 percent of them were specific to CF and G14, respectively (Figure 3-3B). The CF and G14 ACRs covered 3.8 and 3.1 percent of the CF reference genome (12,983,032 and 10,520,085 base pairs (bp)), respectively, and the lengths of both ACRs mainly ranged from 300 to 1,000 bp with a median of ~600 bp (Figure 3-3C). I next analyzed the genomic distribution of CF and G14 ACRs. The majority of ACRs (55.2% and 52.2% in CF and G14, respectively) were localized at promoter regions, and the rest were found in intergenic regions (~17%), downstream regions (~15%) and TEs (~7%) (Figure 3-3D). A small

portion of ACRs lied in exons and introns (Figure 3-3D). In addition, both CF and G14 ACRs displayed a lower density of single nucleotide polymorphisms (SNPs) among *B. rapa* subspecies than the surrounding regions, suggesting that ACRs may have been evolutionarily conserved for their functions (Figure 3-3E). These observations provide evidence that functionally important *cis*-regulatory elements would preferentially reside within ACRs, in accordance with prior studies (Lu et al., 2019; Maher et al., 2018).

ChIP-seq analysis in CF and G14 revealed that the active histone mark H3K27ac was predominantly located around the transcription start sites (Figure 3-4A). The repressive histone mark H3K27me3 and the transcribed gene mark H3K4me1 were distributed throughout the gene body regions (Figure 3-4A), as previously reported in other plant species (Chen et al., 2017; Li et al., 2019; Zhang et al., 2021; Zhang et al., 2009). Another histone repressive mark H3K9me2 was highly enriched in repeat regions and absent from genic regions (Figure 3-4B). I then investigated the global difference in enrichment of these histone marks between CF and G14. I identified a total of 51,602 and 53,032 H3K27ac -enriched regions, 25,742 and 29,446 H3K27me3-enriched regions, 41,955 and 41,414 H3K4me1-enriched regions and 30,887 and 36,892 H3K9me2-enriched regions in CF and G14, respectively. Among the four histone marks, H3K4me1 overlapped the most

between CF and G14 (Figure 3-4C). Approximately 20 to 25 percent of H3K27ac and H3K27me3 regions differed between CF and G14, and nearly up to 30 and 50 percent of H3K9me2 regions were specific to CF and G14, respectively, indicating differential enrichments of these histone marks in CF and G14 (Figure 3-4C).

I further examined genome-wide DNA methylation patterns in CF and G14. Among 3,672,730 differentially methylated cytosines (DMCs) and 68,334 differentially methylated regions (DMRs) between CF and G14, the numbers of hypermethylated DMCs (hyper-DMCs) and hyper-DMRs in G14 relative to CF were higher than those of hypomethylated DMCs (hypo-DMCs) and hypo-DMRs, respectively (Figures 3-5A and 5B). While the G14 hyper-DMCs were mostly found in the CHH context, more than half of the hypo-DMCs in G14 were in the CG context (Figure 3-5A). Moreover, both hyper- and hypo-DMRs in each sequence context considerably overlapped each other (Figure 3-5B). The majority of G14 hyper-DMRs were found in the CHG and CHH contexts, whereas G14 hypo-DMRs were predominantly found in the CG and CHG contexts (Figure 3-5B). Taken together, despite the partial overlap of transcriptome and epigenome landscapes between the two subspecies, CF and G14 have distinct genome-wide profiles of chromatin

accessibility, histone modification and DNA methylation, probably leading to differential gene expression.

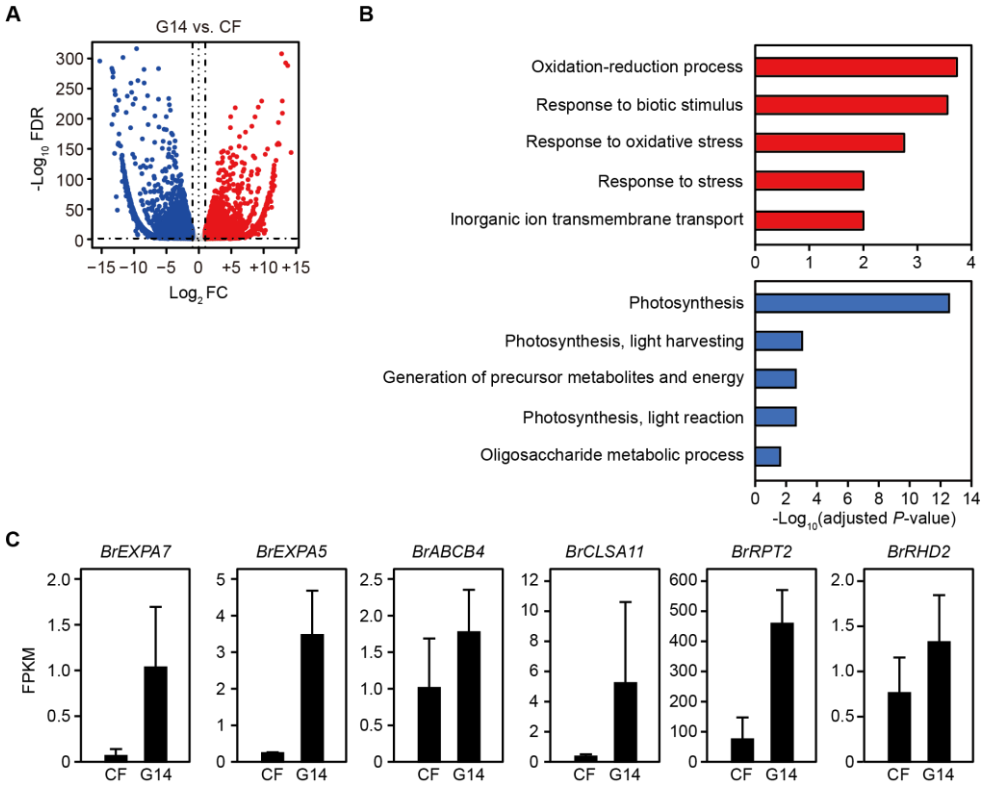

Figure 3-2. Differential gene expression between CF and G14.

(A) Volcano plot showing showing differentially expressed genes (fold change (FC) > 2, false discovery rate (FDR) < 0.05) between CF and G14. Blue and red dots represent downregulated and upregulated genes in G14, respectively. Grey dots indicate the genes that are not significantly differentially expressed. (B) The top 5 enriched GO terms for upregulated and downregulated genes in G14 compared to WT. (C) Expression of representative candidate genes for tuber organ development in turnip and kohlrabi. The orthologous genes to *EXPA7*, *EXPA5* and *ABCB4* and genes to *CSLA11*, *RPT2* and *RHD2* were located in the genomic regions under subgenome parallel selection in turnip and kohlrabi, respectively (Cheng et al., 2016).

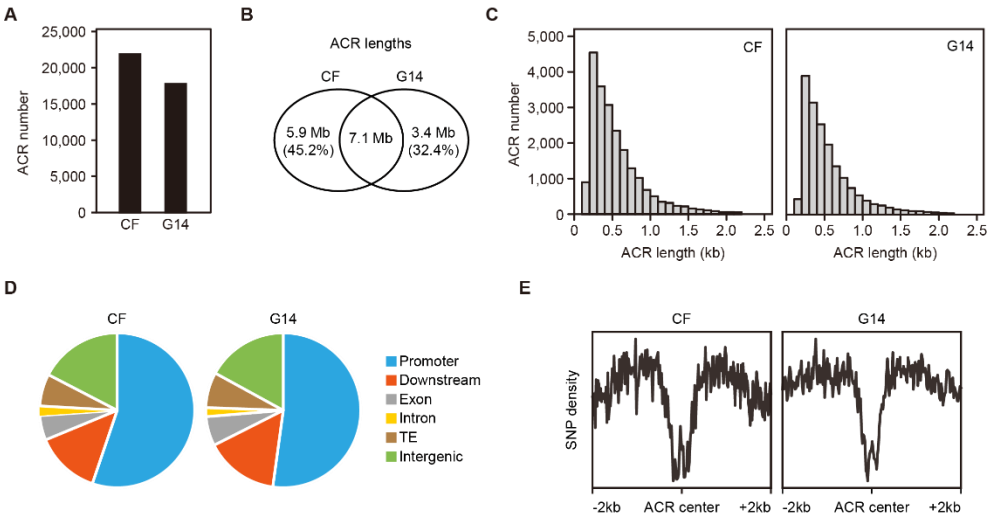

Figure 3-3. Chromatin accessibility dynamics in CF and G14.

(A) The number of ACRs in CF and G14. (B) Venn diagram showing the overlap between CF and G14 ACRs based on their lengths. (C) Lengths of CF and G14 ACRs. (D) Distribution of genomic features of ACRs in CF and G14. Promoters are defined as regions within 1 kb upstream of the transcription start sites, and downstream regions are within 500bp downstream of the transcription termination sites. Intergenic regions indicate regions more than 1 kb upstream or 500bp downstream from genic regions and do not overlap with TEs. (E) Total single nucleotide polymorphisms (SNPs) among 80 *B. rapa* accessions per 20 bp bins flanking ACR centers.

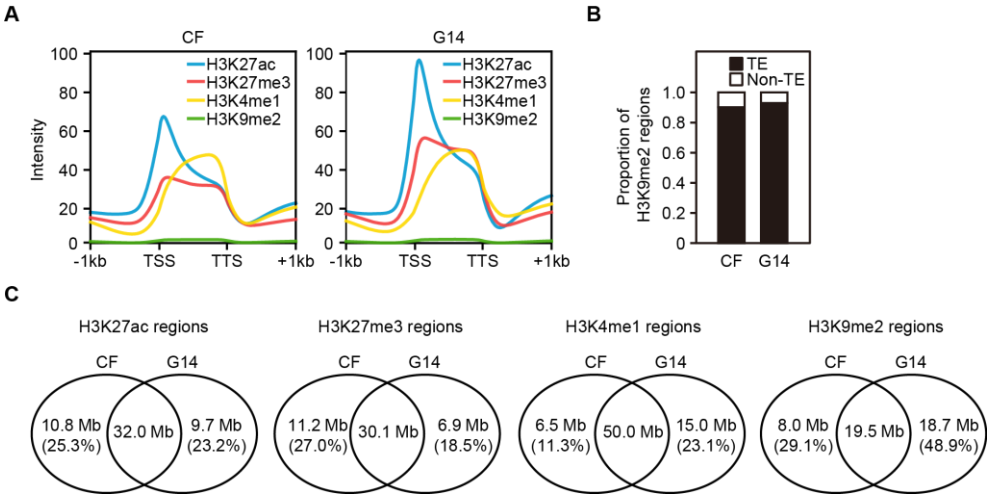

Figure 3-4. Histone modification dynamics in CF and G14.

(A) Distribution of histone modifications at genes and their 1 kb flanking regions in CF and G14. (B) Proportion of H3K9me2-enriched regions with or without TEs in CF and G14. (C) Venn diagrams representing the overlap of histone mark-enriched regions between CF and G14.

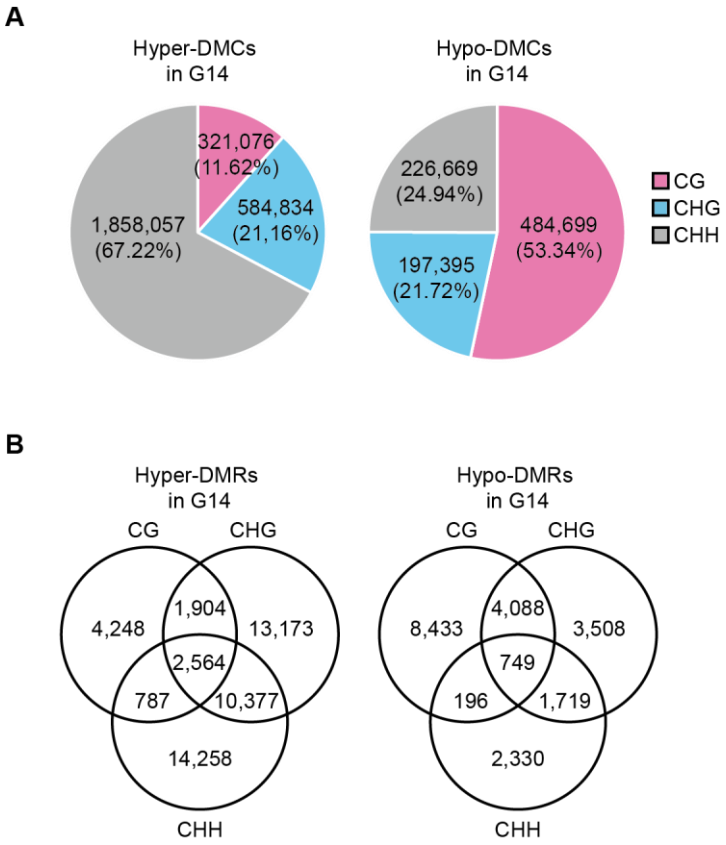

Figure 3-5. DNA methylation dynamics in CF and G14.

(A) Number of differentially methylated cytosines (DMCs) between CF and G14. (B) Venn diagrams showing the overlap of CG, CHG and CHH differentially methylated regions (DMRs) in G14 compared to CF.

CF and G14 ACRs are associated with H3K27ac enrichment and expression dynamics

In mammals, active enhancers are highly correlated with the enrichment of H3K27ac and H3K4me1 as well as the depletion of DNA methylation, while inactive enhancers exhibit the enrichment of H3K27me3 and H3K4me1 (Andersson et al., 2014; Kouzarides, 2007; Rada-Iglesias et al., 2011). In plants, however, distal enhancers are not demarcated by H3K4me1 but rather associated with H3K27ac or H3K9ac in plants (Lu et al., 2019; Oka et al., 2017; Zhu et al., 2015). To determine the chromatin features associated with chromatin accessibility in CF and G14, I investigated the distributions of histone modifications and DNA methylation at ACRs. The flanks of both CF and G14 ACRs were highly enriched with H3K27ac, with slightly decreased H3K27me3 (Figures 3-6A and 6B). H3K4me1 was enriched at the surrounding regions but markedly reduced at the centers of ACRs, where H3K9me2 was completely depleted (Figures 3-6A and 6B). DNA methylation was also remarkably reduced at ACRs in all sequence contexts (Figures 3-6A and 6B). These findings reveal that the majority of ACRs in CF and G14 are associated with H3K27ac enrichment and low DNA methylation levels.

I then sought to examine whether ACRs have the potential of enhancers to regulate transcription in CF and G14. Genes associated with ACRs displayed significantly higher expression dynamics compared to genes without ACRs (Figures 3-6C and 6D). Moreover, genes with multiple ACRs showed a significant increase in a dynamic range of expressions compared to the genes with fewer enhancers, revealing the combinatorial activities of multiple ACRs in expression repertoires. Based on the interplay between multiple enhancers in transcriptional regulation (Ong and Corces, 2011; Spitz and Furlong, 2012), these results imply that ACRs may possess functional features of enhancers in transcriptional regulation.

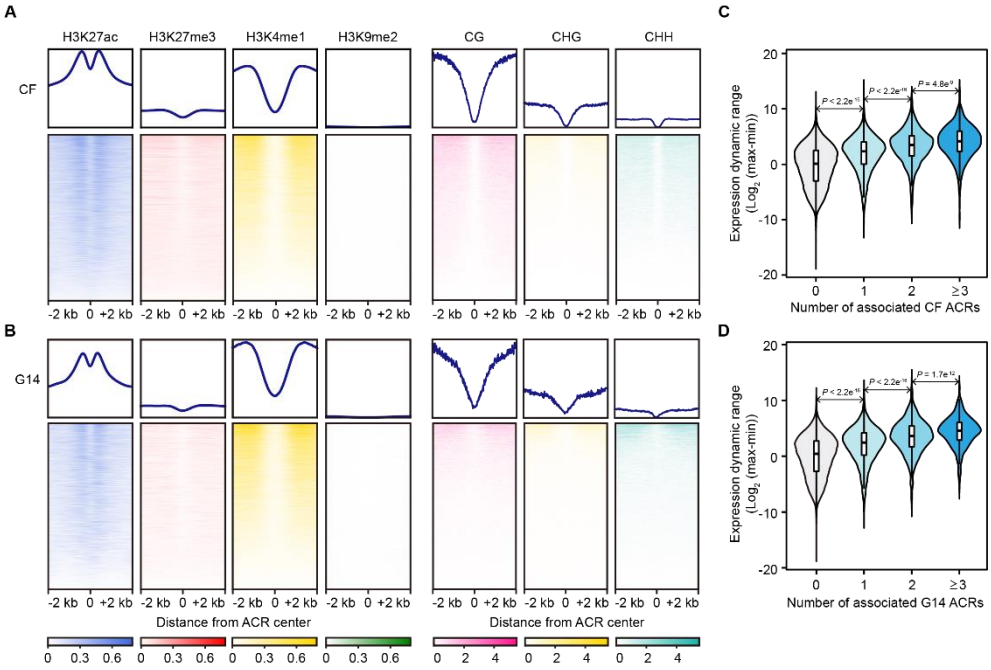

Figure 3-6. Chromatin features and gene expression associated with ACRs.

(A-B) Heatmap representation of histone modifications and DNA methylation at the ACR centers and their 2 kb flanking regions in CF (A) and G14 (B), ranked by signal intensity. Each row indicates one ATAC-seq peak. (C-D) Violin plots showing the dynamic ranges of expression for the genes with an increasing number of CF (C) and G14 (D) ACRs in their 1 kb flanking regions. Genes without any ACRs are used for comparison. *P*-value of two-tailed *t* test is given.

Putative distant enhancers are highly conserved but display differential chromatin accessibility in CF and G14

Given that enhancers can activate the expression of their target genes through long-range interactions (Schoenfelder and Fraser, 2019; Shlyueva et al., 2014), the prevalence of ACRs located distal to genes was investigated in CF and G14. The distances between ACRs and their nearest genes were distributed mostly within a range of 100 bp to 10 kb, and nearly 15 percent of the ACRs were located > 2 kb from the nearest, indicating the existence of putative enhancers that are distinguished from the promoters (Figure 3-7A). I then categorized ACRs based on the distance from their nearest genes. Approximately 40 percent of the ACRs (8,652 and 7,286 in CF and G14, respectively) were referred to as genic ACRs (gACRs) that overlapped with genes, and nearly 45 percent of the ACRs (10,254 and 8,007 in CF and G14, respectively) lied within 2 kb of genes, which are referred to as proximal ACRs (pACRs) (Figure 3-7B). Interestingly, a total of 2,972 (13.6%) and 2,480 (14.0%) dACRs were identified as putative enhancers in CF and G14, respectively (Figure 3-7B).

To determine whether the sequences of ACRs are evolutionarily conserved between CF and G14, I next examined the sequence similarity

between CF and G14 ACRs. The vast majority of ACRs showed high sequence similarity between CF and G14 each other with an average of 96.9% and 97.5%, respectively (Figure 3-7C). Notably, the sequence identity between dACRs was significantly higher than that between pACRs, indicating that dACRs are more conserved than pACRs between CF and G14 (Figure 3-7C). This is coincident with the previous reports that conserved noncoding sequences across many plant species are more strongly enriched in Arabidopsis enhancers rather than in promoters (Haudry et al., 2013; Yan et al., 2019). These observations reveal that CF and G14 retain highly conserved distal *cis*-regulatory elements.

Based on the high sequence conservation of dACRs, I further compared the chromatin accessibility between CF and G14 ACRs. All the ACRs fell into three categories. The dACRs accessible in both CF and G14 are classified as common ACRs, and the dACRs that are accessible only in CF or G14 are classified as CF-specific and G14-specific ACRs, respectively (Figure 3-7D). ACRs were identified as subspecies-specific if the genomic sequences of their subspecies counterparts were located at least 1 kb away from the nearest ACR. Notably, the proportion of subspecies-specific ACRs appeared to increase with a positive relationship with the distance from the genes, despite a great number of common ACRs (Figure 3-7E). Although dACRs exhibited higher

sequence conservation than pACRs (Figure 3-7C), a higher ratio of subspecies-specific ACRs was observed in dACRs, among which 1161 (39.1%) and 679 (27.4%) were specific to CF and G14, respectively (Figure 3-7E). Taken together, the two subspecies show divergence in chromatin accessibility in evolutionarily conserved dACRs.

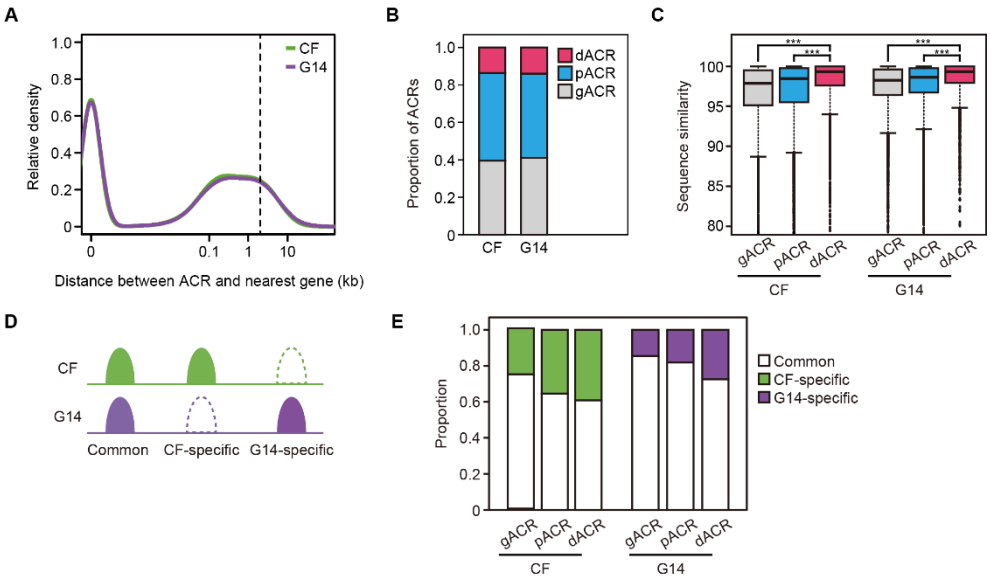

Figure 3-7. Highly conserved dACRs with divergent chromatin accessibility in CF and G14.

(A) The frequency distribution of ACR distances to the closest genes. (B) The proportion of ACRs that are classified into gACRs, pACRs and dACRs. (C) The sequence similarity of ACRs between CF and G14 (Two-tailed *t* test, $P^{***} < 2.2e^{-16}$). (D) Schematic diagram of common and specific ACRS in CF and G14. Subspecies-specific ACRs were identified if the homologous sequence of their subspecies counterpart was located more than 1 kb away from the nearest ACRs. (E) The proportion of common and subspecific-ACRs in CF and G14.

CF and G14 dACRs have conserved but distinct profiles of chromatin features

To explore chromatin features associated with CF and G14 dACRs, k-means clustering analysis was performed on dACRs with flanking histone modifications. The dACRs of CF and G14 were divided into six and four clusters, respectively. CF dACR clusters were characterized into three groups: H3K27me3, H3K27ac and depleted groups (Figure 3-8A). CF cluster 1 (*n* = 118) includes dACRs flanked by H3K27me3 and devoid of other modifications. CF cluster 2 (*n* = 320) and cluster 3 (*n* = 215) represent dACRs highly enriched with H3K27ac at their centers and slightly with H3K4me1 at either side of their flanking regions. CF cluster 4 (*n* = 2,319) includes dACRs depleted of local histone modifications. Similarly, G14 dACR clusters were classified into five groups: all modifications enriched, H3K27me3, H3K9me2, H3K27ac and depleted groups (Figure 3-8B). G14 cluster 1 (*n* = 214) represent dACRs flanked by H3K27me3. G14 cluster 2 (*n* = 96) and cluster 3 (*n* = 125) include dACRs highly enriched with H3K27ac at the center and slightly with H3K4me1 at either side of the flanking regions. G14 cluster 4 (*n* = 1,967) includes dACRs lacking flanking histone modifications. G14 cluster 5 (*n* = 14) includes dACRs showing high levels of all histone modifications at the overall flanking regions, and G14 cluster 6 (*n* = 64)

represents dACRs flanked by H3K9me2. As all enriched modifications (G14 cluster 5) and H3K9me2 only (G14 cluster 6) groups might be due to the mapping bias of G14 reads to the CF reference genome, CF and G14 dACRs seem to have similar profiles for the flanking histone modifications (Figures 3-8A and 8B). Although the patterns of histone modifications around CF and G14 dACRs are quite different from mammalian distal enhancers (Andersson et al., 2014; Kouzarides, 2007; Rada-Iglesias et al., 2011), other plant species such as rice and maize also show similar patterns of chromatin signatures at dACRs (Lu et al., 2019). Thus, the local chromatin features of dACRs are characterized into three distinct states but substantially conserved between the two subspecies.

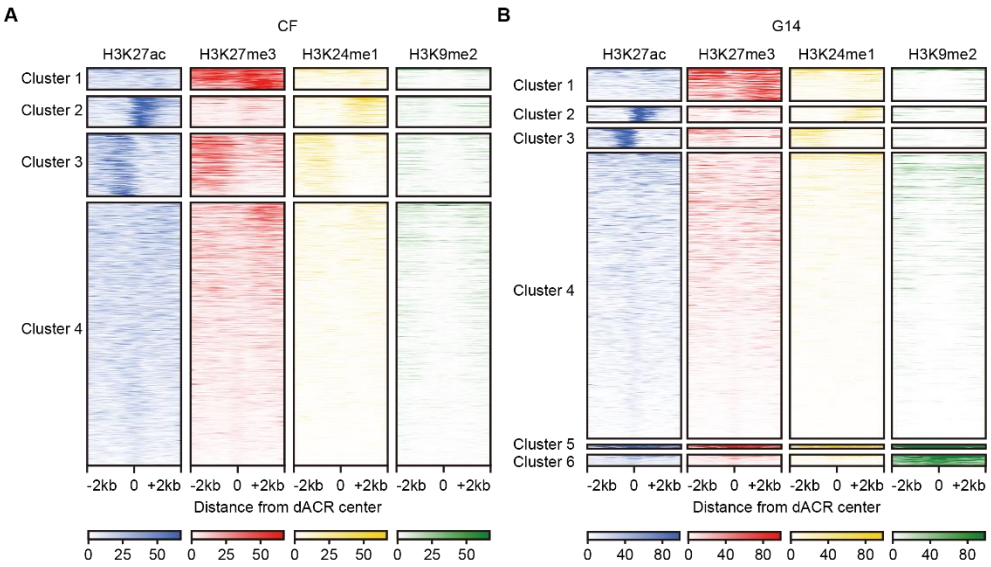

Figure 3-8. Characterization of histone modifications in CF and G14 dACRs.

(A-B) Heatmap showing the clustering analysis of CF (A) and G14 (B) dACRs based on histone modification intensities at the ATAC-peak centers and 2 kb flanking regions. Each row indicates one ATAC-seq peak.

TEs may contribute to chromatin accessibility divergence at dACRs

TEs are an important source of *cis*-regulatory elements including enhancers, which would drive the evolution of gene regulatory networks in eukaryotes (Chuong et al., 2017; Rebollo et al., 2012). To assess whether TEs shape the accessible chromatin landscape between CF and G14, the overlap between dACRs and TEs was measured. Whereas about 30~35% of CF and G14 dACRs are overlapped with TEs, the proportion of TEs was dramatically increased to nearly 50 percent in G14-specific dACRs compared to CF-specific dACRs (Figure 3-9A). Remarkably, among the TE families, long terminal repeat (LTR) retrotransposons were two-fold enriched in G14-specific dACRs (24.6%) than in CF-specific dACRs (12.2%) (Figure 3-9B), consistent with the higher abundance of LTRs in the G14 genome (Park et al., 2019). Moreover, the gypsy elements were prominently enriched in G14-specific dACR-associated LTR subfamilies (Figure 3-9C). These findings suggest that LTRs might act as an evolutionary driving force for divergent chromatin accessibility between CF and G14 dACRs.

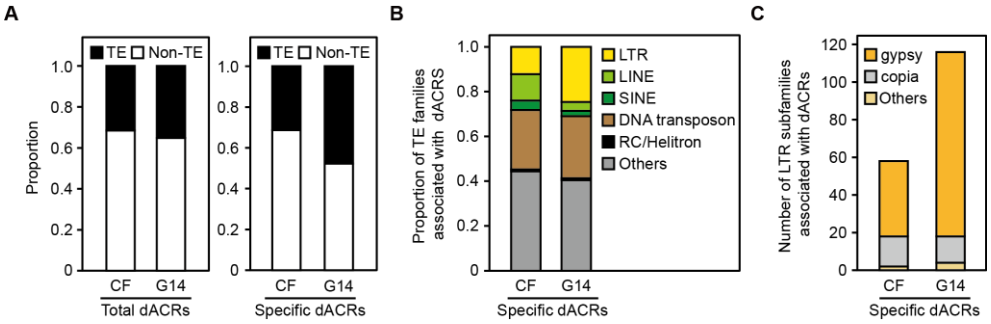

Figure 3-9. Higher LTR enrichment in G14-specific dACRs.

(A) The overlap of TEs with dACRs and subspecies-specific dACRs in CF and G14. (B) Distribution of TE families enriched in CF- and G14-specific dACRs. Long terminal repeat (LTR); Long interspersed nuclear element (LINE); Short interspersed nuclear element (SINE). (C) The number of LTR subfamilies enriched in G14-specific dACRs.

A subset of root and hypocotyl development-related TFs are significantly enriched in G14-specific dACRs

Enhancers contain binding sites for multiple transcription factors (TFs), which enables the control of gene expression in a spatiotemporal manner. I hypothesized that subspecies-specific TFs and dACRs might cooperatively regulate the gene regulatory network, thereby displaying different phenotypes within a species. MEME-ChIP analysis was performed to identify sequence motifs for TFs enriched in subspecies-specific dACRs. A total of 77 and 6 TF motifs were over-represented in CF- and G14-specific dACR centers, respectively (Tables 3-2 and 3-3 and Figure 3-10A). Interestingly, root development- and secondary growth-related TFs including BRASSINOSTEROID INSENSITIVE1-EMS-SUPPRESSOR 1 (BES1)/ BRASSINAZOLE RESISTANT1 (BZR1) HOMOLOG 3 (BEH3), NUTCRACKER (NUC) and LONESOME HIGHWAY LIKE 3 (LHL3) showed a significant enrichment only in G14-specific, but not in CF-specific dACRs (Figure 3-10A). Previous studies revealed that NUC regulates the maintenance of ground tissue identity in roots (Moreno-Risueno et al., 2015), and BEH3 and LHL3 are required for vascular stem cell maintenance and xylem differentiation, respectively, during secondary growth (Furuya et al., 2021; Ohashi-Ito et al., 2013). I next compared the motif enrichment of the

G14-specific dACR-associated TF homologs in CF-specific dACRs. While the motifs of NUC and LHL3 homologs were not significantly enriched in CF-specific dACRs, the motifs of BEH3 homologs, BZR1 and BEH4, exhibited significant enrichment for CF-specific dACRs (Figure 3-10A). Based on the previous report that BEH3 and other BES/BZR members act antagonistically to regulate vascular stem cell maintenance and differentiation (Furuya et al., 2021; Saito et al., 2018), this observation supports that CF and G14 have subspecies-specific TF regulatory networks associated with dACRs. Furthermore, the active histone mark H3K27ac was more highly enriched in G14-specific dACRs containing NUC, BEH3 and LHL3 motifs, respectively, compared to the CF counterparts (Figure 3-10B). These results suggest that G14-specific dACRs may provide binding sites for a subset of TFs associated with cell-type identity maintenance and differentiation in hypocotyl and root development.

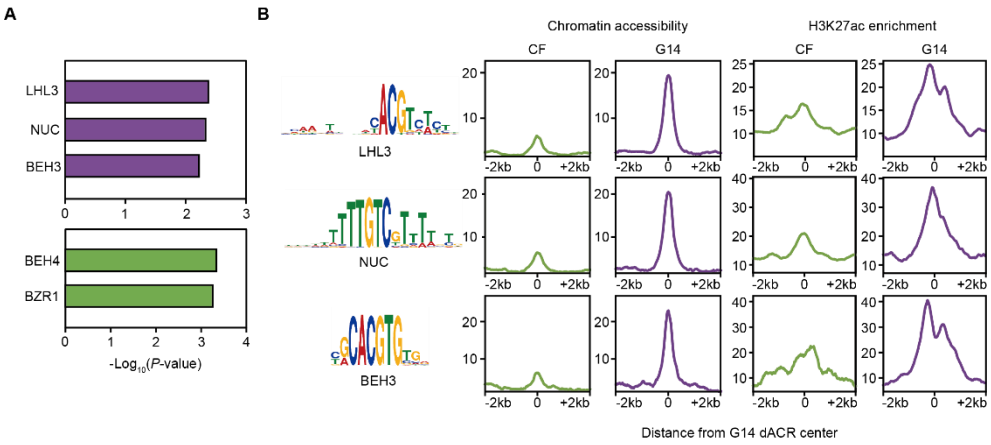

Figure 3-10. TF motif enrichment in G14-specific dACRs.

(A) Motif enrichment of representative TFs in G14-specific dACRs and their paralogous TFs in CF-specific dACRs. (B) Relative enrichment of chromatin accessibility and H3K27ac at the G14 dACRs and their corresponding regions in CF.

Table 3-2. Enrichment of TF binding motifs in CF-specific dACRs.

TF	Conserved motif	Count	*P*-value	Adjusted *P*-value
GBF3	DNWKNHSACGTGGCA	431	8.90E-10	2.20E-07
TCP9	GTGGGHCCCAC	247	3.20E-09	7.80E-07
GBF6	DWWGVTGACGTGGCA	291	1.30E-08	3.20E-06
TCP15	WDGTGGGMCCCAC	275	1.90E-08	4.70E-06
bHLH69	CGTGNKWBGCACGTG	244	2.10E-08	5.00E-06
bHLH31	VWBDHVNCACGTGBNHNWSWC	432	3.30E-08	7.80E-06
bHLH74	CGTGANDVBCACGTGMBNNCW	382	4.40E-08	1.10E-05
TCP16	YRGGTCCACMW	405	6.00E-08	1.50E-05
BIM2	CACGTGMCHHNCACG	266	7.00E-08	1.70E-05
TCP21	WGTGGGMCCCACNW	330	8.10E-08	2.00E-05
TCP9	CCATTWBHHGTGGGTCCCACM	150	1.00E-07	2.40E-05
PTF1	GTGGGGACCACWD	169	1.10E-07	2.70E-05
PTF1	WWGTGGTCCCMAH	282	2.40E-07	5.90E-05
TCP21	DGTGGGHCCCAC	199	3.50E-07	8.40E-05
BIM1	CACGTGACHHNYMY	241	3.80E-07	9.20E-05
TCP3	DTKGGGACCACHD	207	4.10E-07	1.00E-04
TCP3	DTKGGGACCACH	231	4.30E-07	1.00E-04
TCP22	WWGTGGGHCCCAC	243	4.50E-07	1.10E-04
TCP1	WGTGGGCCCCMCBTHHYMWNNHKDKHHNNY	141	5.60E-07	1.30E-04
TCP7	GTGGGRCCCAC	184	5.90E-07	1.40E-04
TCP24	DTKGGGACCACH	277	6.30E-07	1.50E-04
WRKY18	YGTTGACTTTKDH	556	8.60E-07	2.10E-04
TCP16	GTKGGGYCCAC	269	9.20E-07	2.20E-04
bZIP28	DNWKRTSACGTGGCA	378	1.10E-06	2.70E-04
TCP20	DANANRDNWNGTGGGRCCCAC	348	1.20E-06	2.80E-04
bHLH34	GTGNNNRVCACGTGBCDNHDBDH	540	1.30E-06	3.10E-04
WRKY28	DDCGTTGACTTTT	567	1.40E-06	3.40E-04
bHLH104	GVCACGTGBCDDCMNSKGSM	295	1.70E-06	4.10E-04

TF	Conserved motif	Count	*P*-value	Adjusted *P*-value
PIF7	SDKRDWGCCACGTGG	308	1.80E-06	4.40E-04
BEH4	BYRCACGTGTGNATT	169	1.90E-06	4.60E-04
AT5G59990	AATCTCRACCGTYCA	381	2.10E-06	5.10E-04
BIM2	DNHNSGBGNDNGDCACGTG	394	2.20E-06	5.40E-04
BZR1	SCRCACGTGYRHNYNHNNHB	196	2.30E-06	5.60E-04
RAP211	CYDYCDYCGCCGGCN	191	2.40E-06	5.80E-04
ANAC047	CACGT	628	2.60E-06	6.50E-04
WRKY6	DNCGTTGACTWWKDH	439	2.90E-06	7.10E-04
ERF8	NCCDCCGCCGCCGYM	182	3.00E-06	7.40E-04
bZIP16	DWWGVTSACGTGGCA	273	3.60E-06	8.60E-04
WRKY75	AAAAGTCAACGNH	439	5.10E-06	1.20E-03
TCP20	GTGGGDCCCACH	227	5.40E-06	1.30E-03
GBF3	TGCCACGTCABCWHH	255	5.70E-06	1.40E-03
bHLH74	CACGTGAY	279	6.00E-06	1.50E-03
bZIP68	TGCCACGTSABCWHH	359	6.60E-06	1.60E-03
ERF15	HDYHDYHDYMGCCGCCRY	232	6.80E-06	1.60E-03
bHLH122	WNDBCMACTTGCHH	388	6.90E-06	1.70E-03
LEP	DCCKCCGCCGYCDMHDCCKCC	123	7.30E-06	1.70E-03
ERF11	CCDCCKCCGCCGYCA	193	7.70E-06	1.90E-03
TCP22	WWGTGGGHCCCAY	257	7.80E-06	1.90E-03
At2g33710	CCDCCGCCGCCGYCR	183	8.30E-06	2.00E-03
TCP17	GTGGGGACCAC	308	9.90E-06	2.40E-03
ESE3	HGGHGGHGGCGGCGGMGGW	203	1.00E-05	2.50E-03
WRKY40	NDAAAAGTCAAMR	484	1.00E-05	2.50E-03
WRKY15	AAAAGTCAACG	475	1.10E-05	2.80E-03
WRKY21	DNCGTTGACTTTT	247	1.20E-05	2.90E-03
TCP17	GTGGTCCCCAC	250	1.30E-05	3.20E-03
RAP26	TKGCGGCGGMGGHGG	196	1.30E-05	3.20E-03
ESE3	KYGGCGGCGGMGGHG	199	1.40E-05	3.40E-03
ABF2	DWWWNTGCCACGTSWCCW	178	1.60E-05	3.80E-03

TF	Conserved motif	Count	*P*-value	Adjusted *P*-value
ATY19	YHHHAHHWHHYYCACCAACCH	138	1.60E-05	3.80E-03
WRKY29	AAAAGTCAACK	458	1.80E-05	4.50E-03
WRKY29	MAAAGTCAACKNH	580	1.90E-05	4.60E-03
WRKY43	DNCGTTGACTTTTT	441	1.90E-05	4.70E-03
WRKY17	AAAAAGTCAACGNH	319	2.00E-05	4.80E-03
ERF9	YGGCGGHKRYGGCGGCGGMGR	107	2.00E-05	4.80E-03
CRF4	CCKCCGCCGCCRCCDCMDCCD	82	2.00E-05	4.90E-03
WRKY31	DNCGTTGACTWTD	392	2.20E-05	5.40E-03
EIN3	ACCGTTR	69	2.30E-05	5.70E-03
TCP15	WKGTGGGHCCCAC	275	2.40E-05	5.80E-03
CAMTA1	AAARCGCGTGDD	257	2.40E-05	5.80E-03
WRKY18	VAARGTCAASR	558	2.60E-05	6.20E-03
ERF13	MKCMGCCGCCATWDY	189	2.60E-05	6.40E-03
WRKY27	AAAAGTCAACKNY	397	2.80E-05	6.80E-03
WRKY47	AHNHDNCGTTGACTWWDDY	591	3.10E-05	7.40E-03
WRKY26	CGTTGACTTTK	422	3.40E-05	8.30E-03
TCP14	AVADAGVDBGTGGGDCCCAC	29	3.90E-05	9.40E-03
bZIP48	GCCACGTCAGCAWH	182	4.10E-05	9.80E-03
BAM8	YCACACGTGYSAANT	205	4.30E-05	1.00E-02

Table 3-3. Enrichment of TF binding motifs in G14-specific dACRs.

TF	Conserved motif	Count	*P*-value	Adjusted *P*-value
TCP16	YRGGTCCACMW	122	8.70E-06	2.10E-03
bHLH157	HMAAWTHNDWCACGTCWCYK	362	1.80E-05	4.20E-03
TCP16	GTKGGGYCCAC	49	1.90E-05	4.50E-03
NUC	WWWWTTTTTGTCGTTTTBTD	254	1.90E-05	4.70E-03
BEH3	YGCACGTGTGR	73	2.50E-05	6.00E-03
TCP21	WGTGGGMCCCACNW	60	4.50E-05	1.10E-02

Transcriptional enhancer activities were validated in several G14-specific dACRs

To validate the enhancer activity of dACRs, a reporter assay was performed in G14 protoplasts. G14-specific dACR sequences, ranging from 400 bp to 1.2 kb, were inserted upstream of the minimal 35S promoter. An empty vector that contained the GFP reporter construct driven by the minimal 35S promoter was used as a negative control, and an enhancer element of the 35S promoter in the opposite direction (-39 to -200 bp; 35sEnR) was used as a positive control. As expected, the GFP signals were faint in the negative control, but strong in the positive control (Figures 3-11A and 11B). Among the 17 G14-specific dACRs examined, 11 (64.7%) constructs generated clear GFP signals (Table 3-4). Moreover, four of 17 dACRs, whose DNA fragments were in either forward or reverse orientation, displayed similar GFP signal patterns for both orientations (Figures 3-11C and 11D). These observations indicate that dACRs can act as putative enhancers in an orientation-independent manner.

Out of 11 validated G14-specific dACRs, five candidates have H3K27ac enriched at the center or at the flanking regions in G14 compared to CF (Figure 3-12). DNA methylation levels were also decreased only in G14 or

both in CF and G14, highly accompanied with H3K9me2 depletion (Figures 3-12A and 12B). Interestingly, these dACRs contain the LHL, NUC and BES3 motifs simultaneously, further supporting the coordinative roles of chromatin accessibility, histone modifications, DNA methylation and binding of TFs in subspecies-specific gene expression.

Table 3-4. List of G14-specific dACRs examined for reporter assay.

Construct	G14-specific dACR	Size	LHL3 motif	NUC motif	BEH3 motif	H3K 27ac	Enhancer activity
G14-specific dACR#1	G14_peak_ 2478	507	+	+	-	+	+
G14-specific dACR#2	G14_peak_ 4506	455	-	+	-	-	+
G14-specific dACR#3	G14_peak_ 5781	1137	+	+	-	+	+
G14-specific dACR#4	G14_peak_ 7379	867	+	-	-	+	-
G14-specific dACR#5	G14_peak_ 7718	728	-	+	-	-	+
G14-specific dACR#6	G14_peak_ 7913	974	+	-	-	+	+
G14-specific dACR#7	G14_peak_ 8042	826	-	+	-	+	-
G14-specific dACR#8	G14_peak_ 10093	797	+	+	+	+	-
G14-specific dACR#9	G14_peak_ 11026	552	+	-	-	-	+
G14-specific dACR#10	G14_peak_ 11465	903	+	+	+	+	+
G14-specific dACR#11	G14_peak_ 14439	492	+	-	+	+	+
G14-specific dACR#12	G14_peak_ 14888	742	+	+	-	-	+
G14-specific dACR#13	G14_peak_ 15178	518	+	+	-	+	+
G14-specific dACR#14	G14_peak_ 15187	833	+	+	-	+	-
G14-specific dACR#15	G14_peak_ 15472	584	+	+	+	+	-
G14-specific dACR#16	G14_peak_ 16990	560	+	+	-	-	-
G14-specific dACR#17	G14_peak_ 16991	981	-	-	-	-	+

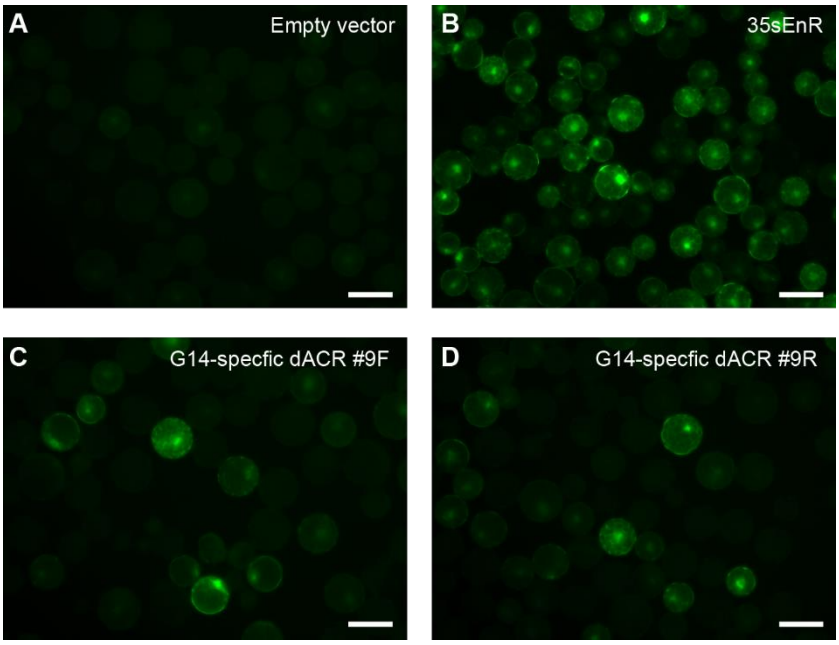

Figure 3-11. Validation of G14-specific dACRs for enhancer activity.

(A-D) GFP expression in G14 protoplasts transfected with the negative control (A), the positive control (B) and G14-specific dACR constructs in forward (C) and reverse orientations (D). Scale bar = 50 μm.

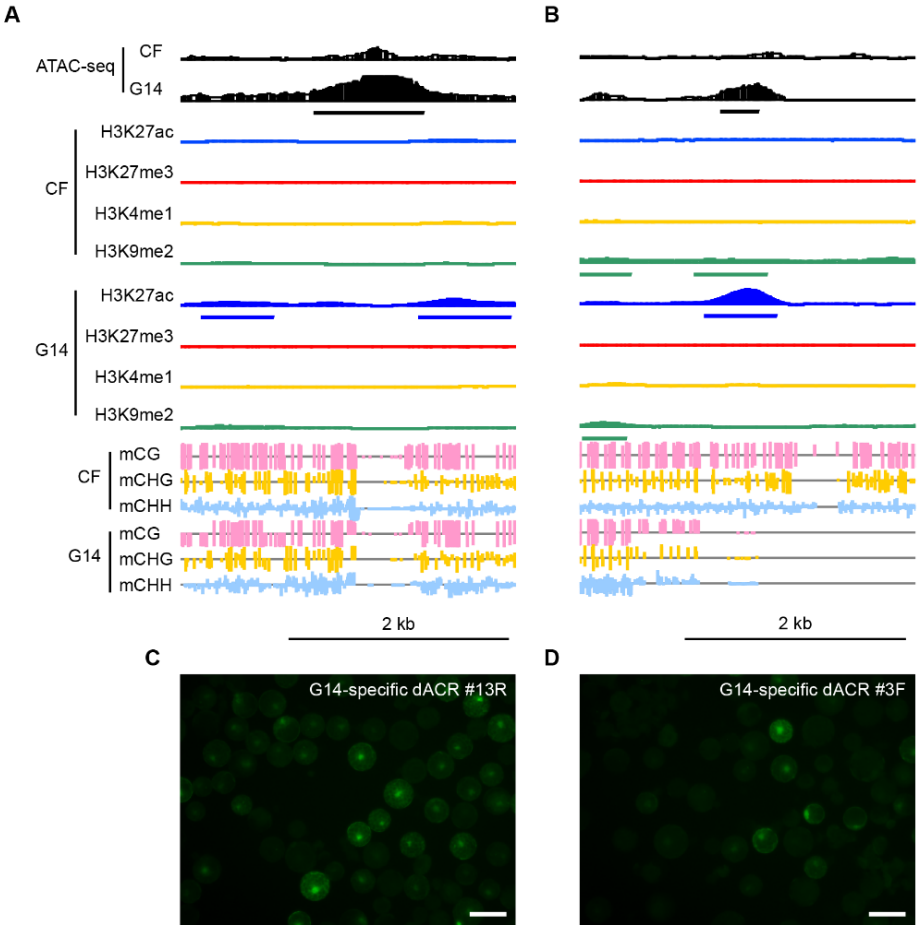

Figure 3-12. Putative enhancers associated with chromatin features and enriched TF motifs.

(A-B) Genome browser view of chromatin accessibility, histone modification enrichment, DNA methylation patterns in the G14-specific dACRs #13R (A) and #3F (B) and CF counterparts. (C-D) GFP expression in G14 protoplasts transfected with the G14-specific dACR #13R (C) and #3F (D). Scale bar = 50 μm.

G14-specific dACRs may regulate the expression of neighboring genes associated with root and hypocotyl development

Given that enhancers often contribute to the expression of nearby genes irrespective of their orientation, I examined the expression of the nearest genes located upstream and downstream of G14-specific dACRs, respectively. Out of 679 G14-specific dACRs, 171 regions (25.2%) were associated with the elevated expression of the nearest genes either upstream or downstream in G14 compared to CF. I focused on a subset of genes related to root or hypocotyl development, including the *B. rapa* genes orthologous to *Arabidopsis PIN-LIKES 5* (*PILS5*), *REVEILLE1* (*RVE1*) and *TWO OR MORE ABRES-CONTAINING GENE 2* (*TMAC2*). An auxin efflux carrier PILS5 and a MYB-like TF RVE1 regulate auxin-dependent hypocotyl and root growth (Barbez et al., 2012; Rawat et al., 2009; Sun et al., 2020), and TMAC2 plays an important role in root elongation and carbohydrate metabolism (Huang and Wu, 2007). *BrPILS5*, *BrRVE1* and *BrTMAC2* were upregulated in G14 compared to CF, with a concomitant increase in chromatin accessibility of their associated G14-specific dACRs (Figures 3-13A to 13C). The associated G14-dACRs were also validated for the enhancer function by reporter assay (Figures 3-13D to 13F). These results suggest that

the link between the G14-specific dACRs and the neighboring genes upregulated in G14 may lead to root and hypocotyl enlargement in turnip.

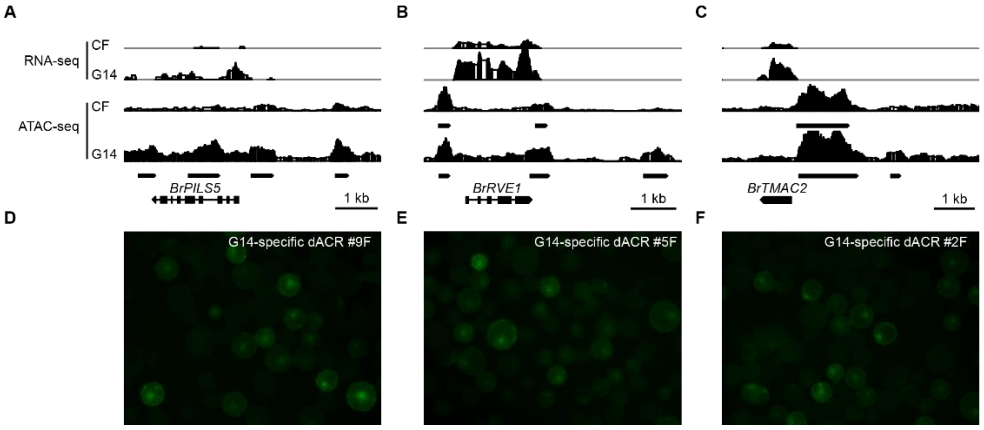

Figure 3-13. Putative enhancers associated with root and hypocotyl development-related DEGs in G14.

(A-C) Genome browser view of expression levels and chromatin accessibility around the G14-specific dACRs and CF counterparts relatively close to upregulated genes in G14. *BrRVE1* (A), *BrTMAC2* (B) and *BrPILS5* (C) are associated with hypocotyl/root elongation and carbohydrate metabolism. (D-F) GFP expression in G14 protoplasts transfected with the G14-specific dACR #5F (D), #2F (E) and #9F (F). Scale bar = 50 μm.

DISCUSSION

As a consequence of whole genome triplication followed by extensive gene loss, *Brassica* species have experienced genome and phenotype diversification (Liu et al., 2014b; Wang et al., 2011). During domestication, whereas convergent parallel selection has contributed to similar morphotypes in different *Brassica* species, the morphological divergence has emerged within the same species of *Brassica* crops (Cheng et al., 2016). Particularly, Chinese cabbage and turnip are of the same species but display extremely distinct phenotypes. Given that *cis*-regulatory elements can act as an evolutionary driving force in many organisms (Arnold et al., 2014; Prescott et al., 2015; Villar et al., 2015), I sought to elucidate the coordinative role of epigenetic and genetic factors in subspecies-specific *cis*-regulatory divergence.

Several lines of evidence indicate that ACRs contain functional *cis*-regulatory elements (Lu et al., 2019; Maher et al., 2018; Ricci et al., 2019). More than half of the CF and G14 ACRs resided in promoter regions, and ACRs showed a lower SNP density (Figure 3-3). In addition, genes associated with ACRs exhibited a significant increase in expression dynamics, and ACRs were highly enriched for TF motifs (Figures 3-6 and 3-10). Consistent

with the previous studies on plant ACRs (Lu et al., 2019; Oka et al., 2017; Zhu et al., 2015), CF and G14 ACRs were characterized by less DNA methylation and H3 lysine acetylation enrichment but not H3K4me1 (Figure 3-6). Moreover, the enhancer activities of a subset of dACRs were validated by a transient reporter assay (Figure 3-11). Based on the orientation-independent property of enhancers, in contrast to promoters, an insertion of the dACR sequence in either forward or reverse direction could upregulate the expression of the reporter gene to similar levels (Figure 3-11). These results indicate the prevalence of the ACRs containing *cis*-regulatory elements, especially enhancers. Intriguingly, dACRs displayed divergent chromatin accessibility between CF and G14 despite high sequence similarities (Figure 3-7), suggesting that dACRs, identified as putative enhancers, might contribute to subspecies diversification.

Although TEs are generally silenced and evolutionarily neutralized, if tolerated from natural selection, TEs can be co-opted as a substantial source of *cis*-regulatory elements (Chuong et al., 2017; Rebollo et al., 2012). In maize, approximately 25~30% of *cis*-regulatory elements are predicted to be derived from TEs (Oka et al., 2017; Zhao et al., 2018). Furthermore, among various angiosperms, species-specific dACRs are enriched in TEs (Lu et al., 2019. In animals, the propagation of TEs is assumed to promote species-

specific regulatory networks (Jacques et al., 2013; Sundaram et al., 2014). Notably, retrotransposons can function as novel promoters, positively regulating the expression of genes associated with agronomic traits (Studer et al., 2011; Zhang et al., 2019). Moreover, LTR retrotransposons are presumed as an evolutionary driving force in plants (Marand et al., 2021; Zhao et al., 2018). G14-specific dACRs were highly enriched in TEs, particularly in LTR retrotransposons, compared to CF-specific dACRs (Figure 3-9). These observations reveal that LTR retrotransposons may contribute to rewiring gene regulatory networks between subspecies, leading to morphotype diversification during evolution.

Multiple TFs bind to enhancers to regulate cell type-, developmental stage-, or species-specific gene expression. I found that the motifs for binding of hypocotyl and root development-related TFs were significantly enriched at G14-specific dACRs (Figure 3-10 and Table 3-3). In Arabidopsis root meristems, the cell fate regulator SHORTROOT (SHR) is expressed in the stele and moves to the daughter cell and the endodermis where it binds to SCARECROW (SCR) to generate the ground tissue identity (Cruz-Ramirez et al., 2012; Nakajima et al., 2001; Sozzani et al., 2010). SHR and SCR regulate the expression of the C2H2 TF family genes including *NUC*, *JACKDAW* (*JKD*) and *MAGPIE* (*MGP*), and in turn, NUC, JKD and MGP,

together with SCR, regulate SHR movement, thus regulating asymmetric cell divisions and cell type patterning during root development (Cui et al., 2007; Long et al., 2015; Moreno-Risueno et al., 2015). Furthermore, LHL3 and BEH3 have roles in vascular development, whose overexpression increased the number of vascular cells and cambium layers in roots and hypocotyls, respectively (Furuya et al., 2021; Ohashi-Ito et al., 2013). During secondary growth, BEH3 functions antagonistically to other BES/BZR proteins in vascular stem cell maintenance and xylem differentiation (Furuya et al., 2021; Saito et al., 2018), in accordance with the distinct motif enrichment of BEH3 in G14-specific dACRs and BZR1 and BEH4 in CF-specific dACRs, respectively (Figure 3-10 and Tables 3-2 and 3-3). In contrast to non-tuber forming pak choi (*B. rapa* subsp. *chinensis*), Asian turnip displays less lignification of xylem cell walls and active expansion of the xylem parenchyma layer, thus resulting in rapid radial growth (Liu et al., 2019). Because seedlings consist of many different cell types and thus the enriched motifs may not represent all cell types, the ChIP assay should help to further confirm the direct interaction between TFs and dACRs for turnip-specific transcriptional regulation. Therefore, this study suggests that turnip may have evolved to possess dACRs that recruit a subset of TFs associated with cell fate determination and vascular development in roots and hypocotyls, thereby

shaping subspecies-specific *cis*-regulatory networks associated with the unique morphology of turnip distinct from Chinese cabbage.

REFERENCES

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al. (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455-461.

Arnold, C.D., Gerlach, D., Spies, D., Matts, J.A., Sytnikova, Y.A., Pagani, M., Lau, N.C., and Stark, A. (2014). Quantitative genome-wide enhancer activity maps for five *Drosophila* species show functional enhancer conservation and turnover during *cis*-regulatory evolution. Nat. Genet. *46*, 685-692.

Barbez, E., Kubes, M., Rolcik, J., Beziat, C., Pencik, A., Wang, B., Rosquete, M.R., Zhu, J., Dobrev, P.I., Lee, Y., et al. (2012). A novel putative auxin carrier family regulates intracellular auxin homeostasis in plants. Nature *485*, 119-122.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114-2120.

Chen, C., Li, C., Wang, Y., Renaud, J., Tian, G., Kambhampati, S., Saatian, B., Nguyen, V., Hannoufa, A., Marsolais, F., et al. (2017). Cytosolic acetyl-CoA promotes histone acetylation predominantly at H3K27 in *Arabidopsis*. Nat. Plants *3*, 814-824.

Cheng, F., Sun, R., Hou, X., Zheng, H., Zhang, F., Zhang, Y., Liu, B., Liang, J., Zhuang, M., Liu, Y., et al. (2016). Subgenome parallel

selection is associated with morphotype diversification and convergent crop domestication in *Brassica rapa* and *Brassica oleracea*. Nat. Genet. *48*, 1218-1224.

Cheng, F., Wu, J., and Wang, X. (2014). Genome triplication drove the diversification of *Brassica* plants. Hortic. Res. *1*, 14024.

Chuong, E.B., Elde, N.C., and Feschotte, C. (2017). Regulatory activities of transposable elements: from conflicts to benefits. Nat. Rev. Genet. *18*, 71-86.

Corces, M.R., Buenrostro, J.D., Wu, B., Greenside, P.G., Chan, S.M., Koenig, J.L., Snyder, M.P., Pritchard, J.K., Kundaje, A., Greenleaf, W.J., et al. (2016). Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat. Genet. *48*, 1193-1203.

Corces, M.R., Granja, J.M., Shams, S., Louie, B.H., Seoane, J.A., Zhou, W., Silva, T.C., Groeneveld, C., Wong, C.K., Cho, S.W., et al. (2018). The chromatin accessibility landscape of primary human cancers. Science *362*, eaav1898.

Cruz-Ramirez, A., Diaz-Trivino, S., Blilou, I., Grieneisen, V.A., Sozzani, R., Zamioudis, C., Miskolczi, P., Nieuwland, J., Benjamins, R., Dhonukshe, P., et al. (2012). A bistable circuit involving SCARECROW-RETINOBLASTOMA integrates cues to inform asymmetric stem cell division. Cell *150*, 1002-1015.

Cui, H., Levesque, M.P., Vernoux, T., Jung, J.W., Paquette, A.J., Gallagher, K.L., Wang, J.Y., Blilou, I., Scheres, B., and Benfey, P.N. (2007). An evolutionarily conserved mechanism delimiting SHR movement defines a single layer of endodermis in plants. Science *316*, 421-425.

Furuya, T., Saito, M., Uchimura, H., Satake, A., Nosaki, S., Miyakawa, T., Shimadzu, S., Yamori, W., Tanokura, M., Fukuda, H., and Kondo, Y. (2021). Gene co-expression network analysis identifies BEH3 as a stabilizer of secondary vascular development in Arabidopsis. Plant Cell *33*, 2618-2636.

Gao, L., Wu, K., Liu, Z., Yao, X., Yuan, S., Tao, W., Yi, L., Yu, G., Hou, Z., Fan, D., et al. (2018). Chromatin accessibility landscape in human early embryos and its association with evolution. Cell *173*, 248-259

Gómez-Campo, C., and Prakash, S. (1999). 2 Origin and domestication. In Developments in plant genetics and breeding, (Elsevier), pp. 33-58.

Haudry, A., Platts, A.E., Vello, E., Hoen, D.R., Leclercq, M., Williamson, R.J., Forczek, E., Joly-Lopez, Z., Steffen, J.G., Hazzouri, K.M., et al. (2013). An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. Nat. Genet. *45*, 891-898.

Huang, C.F., Miki, D., Tang, K., Zhou, H.R., Zheng, Z., Chen, W., Ma, Z.Y., Yang, L., Zhang, H., Liu, R., et al. (2013). A pre-mRNA-splicing factor is required for RNA-directed DNA methylation in *Arabidopsis*. PLoS Genet. *9*, e1003779.

Huang, M.D., and Wu, W.L. (2007). Overexpression of *TMAC2*, a novel negative regulator of abscisic acid and salinity responses, has pleiotropic effects in *Arabidopsis thaliana*. Plant Mol. Biol. *63*, 557-569.

Jacques, P.E., Jeyakani, J., and Bourque, G. (2013). The majority of primate-specific regulatory sequences are derived from transposable elements. PLoS Genet *9*, e1003504.

Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. *37*, 907-915.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571-1572.

Kvon, E.Z., Kamneva, O.K., Melo, U.S., Barozzi, I., Osterwalder, M., Mannion, B.J., Tissieres, V., Pickle, C.S., Plajzer-Frick, I., Lee, E.A., et al. (2016). Progressive loss of function in a limb enhancer during snake evolution. Cell *167*, 633-642.

Kvon, E.Z., Kazmar, T., Stampfel, G., Yanez-Cuna, J.O., Pagani, M., Schernhuber, K., Dickson, B.J., and Stark, A. (2014). Genome-scale functional characterization of *Drosophila* developmental enhancers *in vivo*. Nature *512*, 91-95.

Langmead, B. (2010). Aligning short sequencing reads with Bowtie. Curr. Protoc. Bioinformatics Chapter 11, Unit 11 17.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357-359.

Li, W., Notani, D., and Rosenfeld, M.G. (2016). Enhancers as non-coding RNA transcription units: recent insights and future perspectives. Nat. Rev. Genet. *17*, 207-223.

Li, Z., Wang, M., Lin, K., Xie, Y., Guo, J., Ye, L., Zhuang, Y., Teng, W., Ran, X., Tong, Y., et al. (2019). The bread wheat epigenomic map reveals distinct chromatin architectural and evolutionary features of functional genetic elements. Genome Biol. *20*, 139.

Liu, L., Adrian, J., Pankin, A., Hu, J., Dong, X., von Korff, M., and Turck, F. (2014a). Induced and natural variation of promoter length modulates the photoperiodic response of *FLOWERING LOCUS T*. Nat. Commun.

5, 4558.

Liu, M., Bassetti, N., Petrasch, S., Zhang, N., Bucher, J., Shen, S., Zhao, J., and Bonnema, G. (2019). What makes turnips: anatomy, physiology and transcriptome during early stages of its hypocotyl-tuber development. Hortic. Res. *6*, 38.

Liu, S., Liu, Y., Yang, X., Tong, C., Edwards, D., Parkin, I.A., Zhao, M., Ma, J., Yu, J., Huang, S., et al. (2014b). The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. Nat. Commun. *5*, 3930.

Long, H.K., Prescott, S.L., and Wysocka, J. (2016). Ever-changing landscapes: transcriptional enhancers in development and evolution. Cell *167*, 1170-1187.

Long, Y., Smet, W., Cruz-Ramirez, A., Castelijns, B., de Jonge, W., Mahonen, A.P., Bouchet, B.P., Perez, G.S., Akhmanova, A., Scheres, B., et al. (2015). Arabidopsis BIRD Zinc Finger Proteins Jointly Stabilize Tissue Boundaries by Confining the Cell Fate Regulator SHORT-ROOT and Contributing to Fate Specification. Plant Cell *27*, 1185-1199.

Lou, P., Zhao, J., Kim, J.S., Shen, S., Del Carpio, D.P., Song, X., Jin, M., Vreugdenhil, D., Wang, X., Koornneef, M., and Bonnema, G. (2007). Quantitative trait loci for flowering time and morphological traits in multiple populations of *Brassica rapa*. J. Exp. Bot. *58*, 4005-4016.

Lu, G., Cao, J., Yu, X., Xiang, X., and Chen, H. (2008). Mapping QTLs for root morphological traits in *Brassica rapa* L. based on AFLP and RAPD markers. J. Appl. Genet. *49*, 23-31.

Lu, Z., Marand, A.P., Ricci, W.A., Ethridge, C.L., Zhang, X., and Schmitz, R.J. (2019). The prevalence, evolution and chromatin signatures of plant regulatory elements. Nat. Plants *5*, 1250-1259.

Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics *27*, 1696-1697.

Maher, K.A., Bajic, M., Kajala, K., Reynoso, M., Pauluzzi, G., West, D.A., Zumstein, K., Woodhouse, M., Bubb, K., Dorrity, M.W., et al. (2018). Profiling of accessible chromatin regions across multiple plant species and cell types reveals common gene regulatory principles and new control modules. Plant Cell *30*, 15-36.

Marand, A.P., Chen, Z., Gallavotti, A., and Schmitz, R.J. (2021). A *cis*-regulatory atlas in maize at single-cell resolution. Cell *184*, 3041-3055 e3021.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal *17*, 10-12.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297-1303.

Moreno-Risueno, M.A., Sozzani, R., Yardimci, G.G., Petricka, J.J., Vernoux, T., Blilou, I., Alonso, J., Winter, C.M., Ohler, U., Scheres, B., and Benfey, P.N. (2015). Transcriptional control of tissue formation throughout root development. Science *350*, 426-430.

Nakajima, K., Sena, G., Nawy, T., and Benfey, P.N. (2001). Intercellular movement of the putative transcription factor SHR in root patterning. Nature *413*, 307-311.

Ohashi-Ito, K., Matsukawa, M., and Fukuda, H. (2013). An atypical bHLH transcription factor regulates early xylem development downstream of auxin. Plant Cell Physiol. *54*, 398-405.

Oka, R., Zicola, J., Weber, B., Anderson, S.N., Hodgman, C., Gent, J.I., Wesselink, J.J., Springer, N.M., Hoefsloot, H.C.J., Turck, F., and Stam, M. (2017). Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize. Genome Biol. *18*, 137.

Ong, C.T., and Corces, V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat. Rev. Genet. *12*, 283-293.

Park, H.R., Kang, T., Yi, G., Yu, S.H., Shin, H., Kim, G.W., Park, J.E., Kim, Y.S., and Huh, J.H. (2019). Genome divergence in *Brassica rapa* subspecies revealed by whole genome analysis on a doubled-haploid line of turnip. Plant Biotechnol. Rep. *13*, 677-687.

Prescott, S.L., Srinivasan, R., Marchetto, M.C., Grishina, I., Narvaiza, I., Selleri, L., Gage, F.H., Swigut, T., and Wysocka, J. (2015). Enhancer divergence and *cis*-regulatory evolution in the human and chimp neural crest. Cell *163*, 68-83.

Qi, X., An, H., Ragsdale, A.P., Hall, T.E., Gutenkunst, R.N., Chris Pires, J., and Barker, M.S. (2017). Genomic inferences of domestication events are corroborated by written records in *Brassica rapa*. Mol. Ecol. *26*, 3373-3388.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279-283.

Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187-191.

Rawat, R., Schwartz, J., Jones, M.A., Sairanen, I., Cheng, Y., Andersson,

C.R., Zhao, Y., Ljung, K., and Harmer, S.L. (2009). REVEILLE1, a Myb-like transcription factor, integrates the circadian clock and auxin pathways. Proc. Natl. Acad. Sci. U. S. A. *106*, 16883-16888.

Rebollo, R., Romanish, M.T., and Mager, D.L. (2012). Transposable elements: an abundant and natural source of regulatory sequences for host genes. Annu. Rev. Genet. *46*, 21-42.

Reynoso, M.A., Kajala, K., Bajic, M., West, D.A., Pauluzzi, G., Yao, A.I., Hatch, K., Zumstein, K., Woodhouse, M., Rodriguez-Medina, J., et al. (2019). Evolutionary flexibility in flooding response circuitry in angiosperms. Science *365*, 1291-1295.

Ricci, W.A., Lu, Z., Ji, L., Marand, A.P., Ethridge, C.L., Murphy, N.G., Noshay, J.M., Galli, M., Mejia-Guerra, M.K., Colome-Tatche, M., et al. (2019). Widespread long-range *cis*-regulatory elements in the maize genome. Nat. Plants *5*, 1237-1249.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139-140.

Saito, M., Kondo, Y., and Fukuda, H. (2018). BES1 and BZR1 redundantly promote phloem and xylem differentiation. Plant Cell Physiol. *59*, 590-600.

Schoenfelder, S., and Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. Nat. Rev. Genet. *20*, 437-455.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nat. Rev. Genet. *15*, 272-286.

Sozzani, R., Cui, H., Moreno-Risueno, M.A., Busch, W., Van Norman, J.M., Vernoux, T., Brady, S.M., Dewitte, W., Murray, J.A., and

Benfey, P.N. (2010). Spatiotemporal regulation of cell-cycle genes by SHORTROOT links patterning and growth. Nature *466*, 128-132.

Spitz, F., and Furlong, E.E. (2012). Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. *13*, 613-626.

Studer, A., Zhao, Q., Ross-Ibarra, J., and Doebley, J. (2011). Identification of a functional transposon insertion in the maize domestication gene *tb1*. Nat. Genet *43*, 1160-1163.

Sun, L., Feraru, E., Feraru, M.I., Waidmann, S., Wang, W., Passaia, G., Wang, Z.Y., Wabnik, K., and Kleine-Vehn, J. (2020). PIN-LIKES coordinate brassinosteroid signaling with nuclear auxin input in *Arabidopsis thaliana*. Curr. Biol. *30*, 1579-1588.

Sundaram, V., Cheng, Y., Ma, Z., Li, D., Xing, X., Edge, P., Snyder, M.P., and Wang, T. (2014). Widespread contribution of transposable elements to the innovation of gene regulatory networks. Genome Res *24*, 1963-1976.

Villar, D., Berthelot, C., Aldridge, S., Rayner, T.F., Lukk, M., Pignatelli, M., Park, T.J., Deaville, R., Erichsen, J.T., Jasinska, A.J., et al. (2015). Enhancer evolution across 20 mammalian species. Cell *160*, 554-566.

Vuolo, F., Mentink, R.A., Hajheidari, M., Bailey, C.D., Filatov, D.A., and Tsiantis, M. (2016). Coupled enhancer and coding sequence evolution of a homeobox gene shaped leaf diversity. Genes Dev. *30*, 2370-2375.

Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., Bai, Y., Mun, J.H., Bancroft, I., Cheng, F., et al. (2011). The genome of the mesopolyploid crop species *Brassica rapa*. Nat. Genet. *43*, 1035-1039.

Wu, Y., Zhang, S., Zhang, H., Li, F., Li, G., Fan, C., Sun, R., and Zhang, S. (2021). QTL mapping and candidate gene identification of swollen root formation in turnip. Int. J. Mol. Sci. *22*, 653.

Yan, W., Chen, D., Schumacher, J., Durantini, D., Engelhorn, J., Chen, M., Carles, C.C., and Kaufmann, K. (2019). Dynamic control of enhancer activity drives stage-specific gene expression during flower morphogenesis. Nat. Commun. *10*, 1705.

Yoo, S.D., Cho, Y.H., and Sheen, J. (2007). *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. Nat. Protoc. *2*, 1565-1572.

Zhang, L., Hu, J., Han, X., Li, J., Gao, Y., Richards, C.M., Zhang, C., Tian, Y., Liu, G., Gul, H., et al. (2019). A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. Nat. Commun. *10*, 1494.

Zhang, Q., Guan, P., Zhao, L., Ma, M., Xie, L., Li, Y., Zheng, R., Ouyang, W., Wang, S., Li, H., et al. (2021). Asymmetric epigenome maps of subgenomes reveal imbalanced transcription and distinct evolutionary trends in *Brassica napus*. Mol. Plant *14*, 604-619.

Zhang, W., Wu, Y., Schnable, J.C., Zeng, Z., Freeling, M., Crawford, G.E., and Jiang, J. (2012). High-resolution mapping of open chromatin in the rice genome. Genome Res. *22*, 151-162.

Zhang, X., Bernatavichute, Y.V., Cokus, S., Pellegrini, M., and Jacobsen, S.E. (2009). Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. Genome Biol. *10*, R62.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zhao, H., Zhang, W., Chen, L., Wang, L., Marand, A.P., Wu, Y., and Jiang, J. (2018). Proliferation of Regulatory DNA Elements Derived

from Transposable Elements in the Maize Genome. Plant Physiol *176*, 2789-2803.

Zhao, J., Artemyeva, A., Del Carpio, D.P., Basnet, R.K., Zhang, N., Gao, J., Li, F., Bucher, J., Wang, X., Visser, R.G., and Bonnema, G. (2010). Design of a *Brassica rapa* core collection for association mapping studies. Genome *53*, 884-898.

Zhong, S., Joung, J.G., Zheng, Y., Chen, Y.R., Liu, B., Shao, Y., Xiang, J.Z., Fei, Z., and Giovannoni, J.J. (2011). High-throughput illumina strand-specific RNA sequencing library preparation. Cold Spring Harb. Protoc. *2011*, 940-949.

Zhu, B., Zhang, W., Zhang, T., Liu, B., and Jiang, J. (2015). Genome-wide prediction and validation of intergenic enhancers in Arabidopsis using open chromatin signatures. Plant Cell *27*, 2415-2426.

ABSTRACT IN KOREAN

후성유전학적 기작은 세포 분화, 내부 발달 신호 및 외부 환경 요인에 대한 반응 등 다양한 생물학적 반응을 조절하는 데 중요한 역할을 한다. DNA 메틸화, 히스톤 변형 등의 후성유전학적 변형은 DNA 염기서열의 변화 없이 염색질 구조 변화를 통해 유전자 발현을 변화시킨다. 식물은 내부 및 외부 자극을 인지하고 반응하여 후성유전학적으로 유전자 발현을 조절한다. 그리고 일부 후성유전학적 변이는 안정적으로 후대에 전달됨으로써 주요 형질에 기여하는 것으로 생각된다. 그러나 세부적인 후성유전학적 조절 기작 및 형질과 연관된 진화적 중요성은 아직 식물에서 많이 밝혀지지 않았다. DNA 메틸화 수준은 DNA 메틸화와 탈메틸화 기작에 의해 조절된다. 애기장대에서 DNA 메틸화는 DNA 탈메틸화 효소에 의해 제거되는데, DEMTER (DME), REPRESSOR OF SILENCING 1 (ROS1), DEMETER-LIKE 2 (DML2) 및 DML3 가 이를 담당한다. *DME* 는 암배우체의 중심세포에서 주로 발현되는 반면에, *ROS1*, *DML2* 및 *DML3* 는 영양조직에서 발현된다. 본 논문에서는 애기장대에서 DNA 탈메틸화 효소에 의한 환경 및 발달 반응 조절을 연구하였다. *ros1* 돌연변이체는 앱시스산 처리 시 야생형에 비해 유묘 및 뿌리 발달이 저해되었다. *ros1* 돌연변이체에서 앱시스산 유도성 유전자들의 발현이 감소하였으며, 프로모터 지역에서 DNA 메틸화 수준이 증가하였다. 따라서 ROS1 에 의한 DNA 탈메틸화는 앱시스산이 매개하는 가뭄 및 수분 스트레스 반응에 필요한 유전자들의 발현을 활성화하는 데 중요한 역할을 하는 것으로 생각된다. 더 나아가 생식 생장에 있어 DNA 탈메틸화 효소의 역할을 밝히고자 하였다. 애기장대 *dme ros1 dml2 dml3* (*drdd*) 사중 돌연변이체는 성장 지연, 개화 시기

지연, 비정상적 화기 구조 및 꼬투리 발달 등의 표현형을 보였다. 사중 돌연변이체에서는 *dme* 또는 *rdd* 돌연변이체 비해 유전자 발현 감소와 DNA 메틸화 증가가 두드러졌다. 이는 여러 DNA 탈메틸화 효소가 생식 생장에 필요한 유전자의 발현을 중복적으로 조절함을 시사한다. 또한 본 논문에서는 배추(*Brassica rapa* subsp. *pekinensis*)와 순무(*B. rapa* subsp. *rapa*)의 형태학적 다양성에 대한 연구를 수행하였다. 배추와 순무는 같은 종으로 유사한 유전체를 가지고 있지만, 상이한 표현형을 보인다. 전사체 및 후성유전체 분석을 수행하여 배추와 순무의 열린 염색질 지역은 유전자 발현 변화 범위, 히스톤 H3 의 27 번째 라이신 아세틸화 (H3 lysine 27 acetylation) 및 낮은 DNA 메틸화 수준과 연관되어 있음을 확인하였다. 유전자로부터 먼 거리에 위치하는 열린 염색질 지역은 배추와 순무 간에 DNA 염기서열이 높은 수준으로 보존되어 있었으나, 염색질 접근성과 전이인자 모티프는 상이한 양상을 보였다. 이러한 연구 결과는 아종 간의 염색질 접근성의 차이가 형태학적 다양성에 기여할 수 있음을 의미한다. 본 연구는 외부 환경 요인과 내부 발달 신호에 반응하여 일어나는 DNA 탈메틸화 효소에 의한 유전자 발현 조절 기작에 대한 이해를 높이고, 진화 과정 동안 염색질 구조 차이에서 기인하는 아종 간 형태적 차이에 대한 새로운 관점을 제시할 수 있을 것으로 기대된다.