Influence de la méthylation des éléments transposables sur les gènes situés à proximité

4.1 Conséquences de la perte de la méthylation de l'ADN des ET sur l'expression des gènes situés à proximité

Comme décrit dans l'introduction, il semble qu'une perte importante de la méthylation de l'ADN au niveau des séquences répétées, telle celle induite par une mutation dans le gène *ddm1*, n'ait que peu de conséquences sur la régulation des gènes (Lippman et al. 2004, Vongs et al. 1993). Cette observation est renforcée par le fait que le mutant *ddm1* ne présente pas d'altérations phénotypiques majeures (du moins dans les premières générations d'autofécondation). Cependant, ces informations sont quelque peu contradictoires avec les résultats d'une étude montrant que les ET méthylés semblent avoir un effet négatif sur l'expression des gènes à proximité desquels ils sont localisés (Hollister et al. 2011).

Afin d'analyser plus avant les conséquences de la perte de la méthylation de l'ADN au niveau des séquences répétées, j'ai participé aux travaux initiés par le précédent étudiant en thèse, Felipe Teixeira, présentés dans le manuscrit en préparation ci-après. Ma contribution principale concerne les expériences d'ARN-chip utilisant les puces à ADN CATMA réalisées afin de comparer les transcriptomes des différents mutants. J'ai également généré le triple mutant *ddm1rdr2T-sdc* et ai participé à la validation des gènes candidats potentiellement contrôlés directement par la méthylation des séquences répétées situées à proximité.

A DNA methylation-based double lock system acting over repeat sequences ensures normal development in Arabidopsis

Felipe Karam Teixeira1,2,a,*, Mathilde Etcheverry1,*, Alexis Sarazin1, Ludivine Taconnat2, Anne-Valérie Gendrel 2b, Alexandra To1,2,c, Agnès Bulski1,2,d, Jean-Pierre Renou2,d, Vincent Colot1,2†

1Institut de Biologie de l'Ecole Normale Supérieure (IBENS), Centre National de la Recherche Scientifique (CNRS) UMR8197 - Institut National de la Santé et de la Recherche Médicale (INSERM) U1024, 46 rue d'Ulm, 75230 Paris cedex, France

2Unité de Recherche en Génomique Végétale, Evry Cedex, France

Present address: aNew York University School of Medicine, Skirball Institute, New York, NY 10016, USA; bInstitut Curie, Bâtiment Biologie du Développement, 75248 PARIS Cedex 05; cINRA, Institut Jean-Pierre Bourgin, F-78026 Versailles Cedex, France; dMuséum National d'Histoire Naturelle, CNRS UMR7205, 75005 Paris, France; dINRA Centre d'Angers, Beaucouzé Cedex, France

†Corresponding author: colot@biologie.ens.fr

*These authors contributed equally to this work

Running title: DNA methylation pathways and genome control

Abstract

DNA methylation is an epigenetic mechanism that is essential for silencing repeat elements in mammals and plants. However, the extent to which repeat-associated DNA methylation impacts the expression of neighboring genes is unclear. Here we show in Arabidopsis that severe and fully penetrant phenotypic alterations and strong misregulation of hundreds of genes are only observed when compromising simultaneously the RNA-directed DNA methylation (RdDM) and DDM1-dependent maintenance of DNA methylation pathways, which both target repeat sequences specifically. However, upregulation of repeat sequences is not detectably accentuated or more widespread under these conditions that when only DDM1 function is compromised. Further analysis indicates that the two DNA methylation pathways act together to maintain the normal expression of only a very small number of genes near repeats and that these genes tend to have pleiotropic effects. These findings demonstrate that most Arabidopsis genes are not detectably sensitive to the epigenetic status of neighboring repeats and that a double-lock system ensures that sensitive genes retain proper expression.

Keywords: Epigenetics/chromatin/RNA interference/transposable elements.

Introduction

The genetic dissection of DNA methylation in Arabidopsis has uncovered a complex interplay between DNA methyltransferases (MTases), DNA demethylases, histone-modifying or remodeling enzymes, RNA interference (RNAi) components, and RNA polymerases [1]. Among the actors identified so far, the chromatin remodeler DECREASE IN DNA METHYLATION1 (DDM1; [2]) stands alone as a master regulator of DNA methylation maintenance over repeat sequences [3]. Indeed, mutations in *DDM1* lead to a loss of DNA methylation over most repeat elements present in the genome but not over genes. Furthermore, *ddm1* mutants exhibit a massive accumulation of transcripts corresponding to

repeat sequences, but do not substantially alter the expression of neighboring genes, with few exceptions [2, 4]. Consistent with these observations, early generation *ddm1* mutants exhibit only mild phenotypic alterations. However, severe but sporadic phenotypic alterations are observed in advanced generations and result from the mobilization of transposable elements (TEs) into or near genes as well as rare late onset epigenetic alteration of gene expression [5-10]. In contrast, the RNA-dependent DNA methylation (RdDM) pathway, which targets a majority of methylated repeat elements but not genes [11-14] contributes little to the overall DNA methylation level or the silencing of its targets [12, 14-17].

We have previously shown that RdDM endows its targets with the ability to regain wild type (wt) DNA methylation following *ddm1*-induced loss of DNA methylation [14]. Moreover, we demonstrated that the RdDM pathway is active in *ddm1* plants, being able to promote low-level DNA methylation over its targets. These observations indicate that DDM1-dependent DNA methylation and RdDM act concomitantly over a subset of repeat elements. Here, we show that the simultaneous disruption of these two DNA methylation pathways leads to severe, fully penetrant phenotypic alterations as well as widespread changes in gene expression. However, detailed analysis indicates that most of these defects are caused by the misexpression of a few pleiotropic genes and that most genes are insensitive to the epigenetic status of neighboring repeat elements. Taken together, our findings reveal a limited yet critical direct function of RdDM and DDM1-dependent DNA methylation in preserving normal expression of genes in Arabidopsis.

Results and Discussion

Compromising together DDM1-dependent methylation and RdDM leads to severe and fully penetrant phenotypic alterations

To understand better the genetic interplay between RdDM and DDM1-dependent DNA methylation and their impact on the expression of genes located near repeat elements, we generated mutant plants affected in these two pathways. Double mutant plants were obtained at the expected Mendelian ratio of 1/16 in the F2 progeny of a cross between *ddm1* and a mutant line for the RNA DEPENDENT RNA POLYMERASE 2 gene, which is a key component of RdDM (supplementary Table S1). In contrast to these two parental lines however, first generation *ddm1rdr2* double mutant plants exhibited severe and fully or nearly fully penetrant phenotypic alterations. These alterations were characterized in the following generation (F3) and included wrinkled leaves, thin stems, small stature, late-flowering, and partial sterility (Fig 1). Further selfing led to even more severe defects as well as to the stochastic appearance of additional phenotypes. Similar phenotypic alterations including progressive degeneracy were also observed when the *ddm1* mutation was combined with a mutation in the genes NUCLEAR RNA POLYMERASE D2 (NRPD2) or DICER-LIKE 2 (DCL2) and DCL3, which encode other key components of RdDM.(supplementary Fig S1). Thus, compromising simultaneously DDM1-dependent DNA methylation and RdDM causes a robust suite of phenotypic defects. We therefore conclude that these two pathways act redundantly to enable normal plant development.

Most genes near methylated repeats are insensitive to the simultaneous loss of DDM1-dependent methylation and RdDM

We next performed comparative transcriptome analysis of wild type, *ddm1* and *ddm1rdr2* plants, using 10-days-old seedlings and a gene-based array (CATMA) containing ~24000 gene-specific tags from Arabidopsis [18]. In agreement with previous analyses, few genes were misregulated in *rdr2* when compared to wt (Fig 2A,B; 62 upregulated and 22 downregulated; [12, 17]). Up to three times more genes were misregulated in *ddm1* and

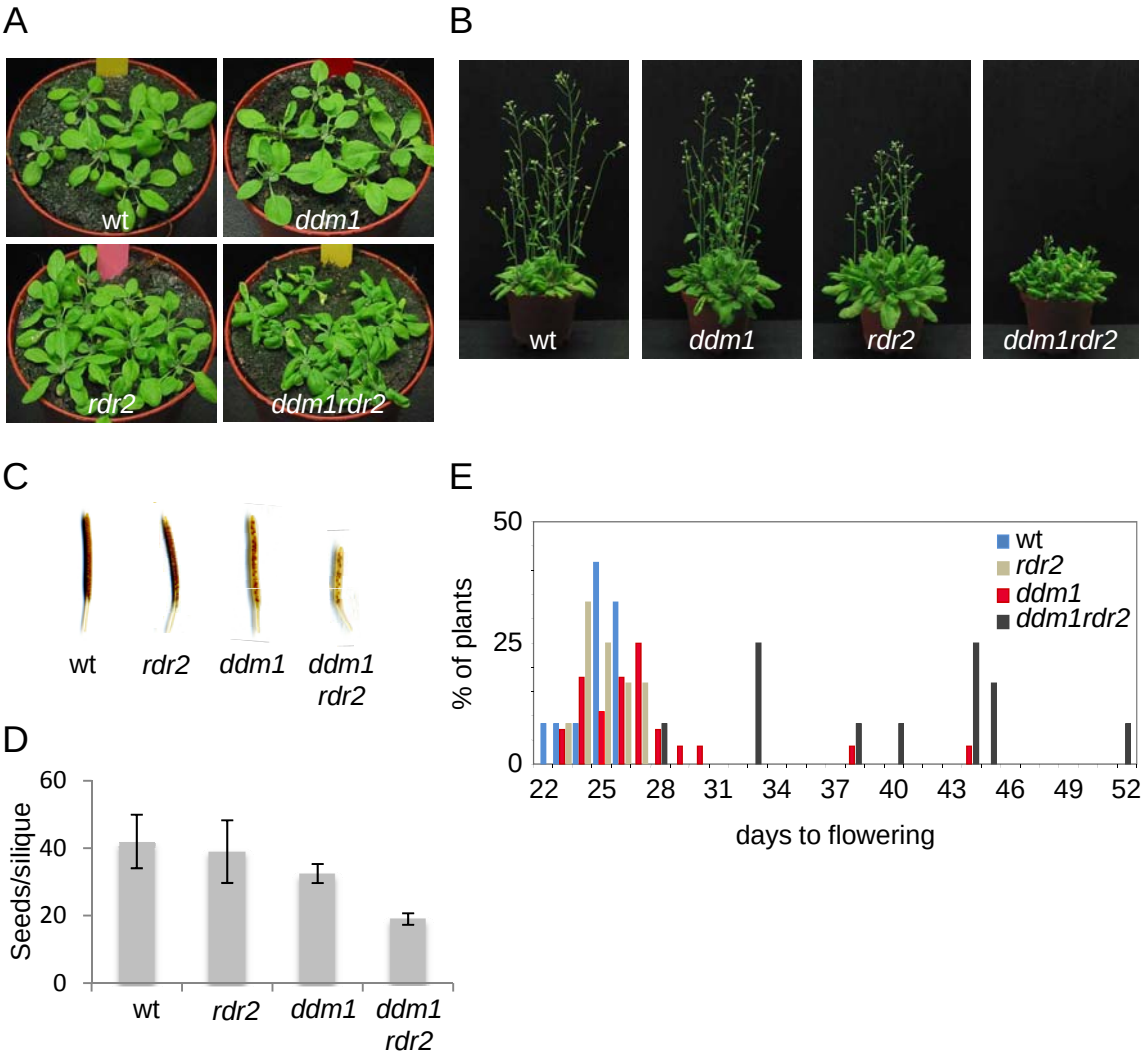


Figure 1. Combined loss of DDM1 and RNAi leads to fully penetrant phenotypic alterations. (A) wt, *ddm1*, *rdr2*, and *ddm1rdr2* plants of 23-days-old and (B) 40-days-old. (C) Siliques for each genotype, discolored using ethanol 70%. (D) Average number of seeds per silique based on 7-10 siliques. (E) Analysis of flowering time as measured by the number of days to flowering for 24 plants of each genotype.

ddm1rdr2 when these mutants were individually compared to wt, with *ddm1rdr2* exhibiting slightly more misregulated genes than *ddm1* (Fig 2A,B). Direct comparisons between single and double mutants were also carried out (supplementary Fig S2). Altogether, these analyses identified a set of ~360 genes that are misexpressed in plants compromised in one or both DNA methylation pathways. We therefore conclude that that these two pathways act redundantly to enable normal plant development.

Misregulated genes were then individually inspected for their proximity (within 1kb) to targets of DDM1-dependent methylation and RdDM, using a genome browser display of publically available single-nucleotide resolution methylome datasets [13, 19]. This revealed that less than one third (~100) of misregulated genes are close to such targets and thus likely to be directly affected by loss of DNA methylation over these sequences. As expected, this group included the three genes for which mis-expression in response to compromised DDM1-dependent methylation or RdDM has been extensively characterized, namely FLOWERING WAGENINGEN A (FWA, At4g25530; [5, 9]), SUPPRESSOR OF drm1drm2cmt3 (SDC, At2g17690; [20]), and At5g27850 [16]. Validation of the microarray results and direct examination of DNA methylation in matched samples was carried out using these three genes as well as nine randomly selected genes (supplementary Table S2, S3). With one exception (At5g22500; excluded from further analysis), RT-qPCR analysis validated the transcriptome data (supplementary Table S3). Changes in DNA methylation, as measured by McrBC-qPCR, were tightly correlation with changes in gene expression for all three previously characterized genes as well as six of the eight remaining genes that were examined (Fig 2C; supplementary Table S4). For four genes (SDC, At3g26520, FWA, and At4g30960), strongest misregulation was observed in *ddm1rdr2* and was always associated with accentuated loss of DNA methylation when compared to both ddm1 and *rdr2*. This indicates that at these genes, DDM1-dependent DNA methylation and RdDM act synergistically or additively rather than strictly redundantly in preserving normal expression. For the other direct target genes tested, expression and DNA methylation were already severely affected in either *ddm1* (At1g36240, At3g28950, and At5g26270) or *rdr2* (At1g57620 and At5g27850), with no further changes in the double mutant.

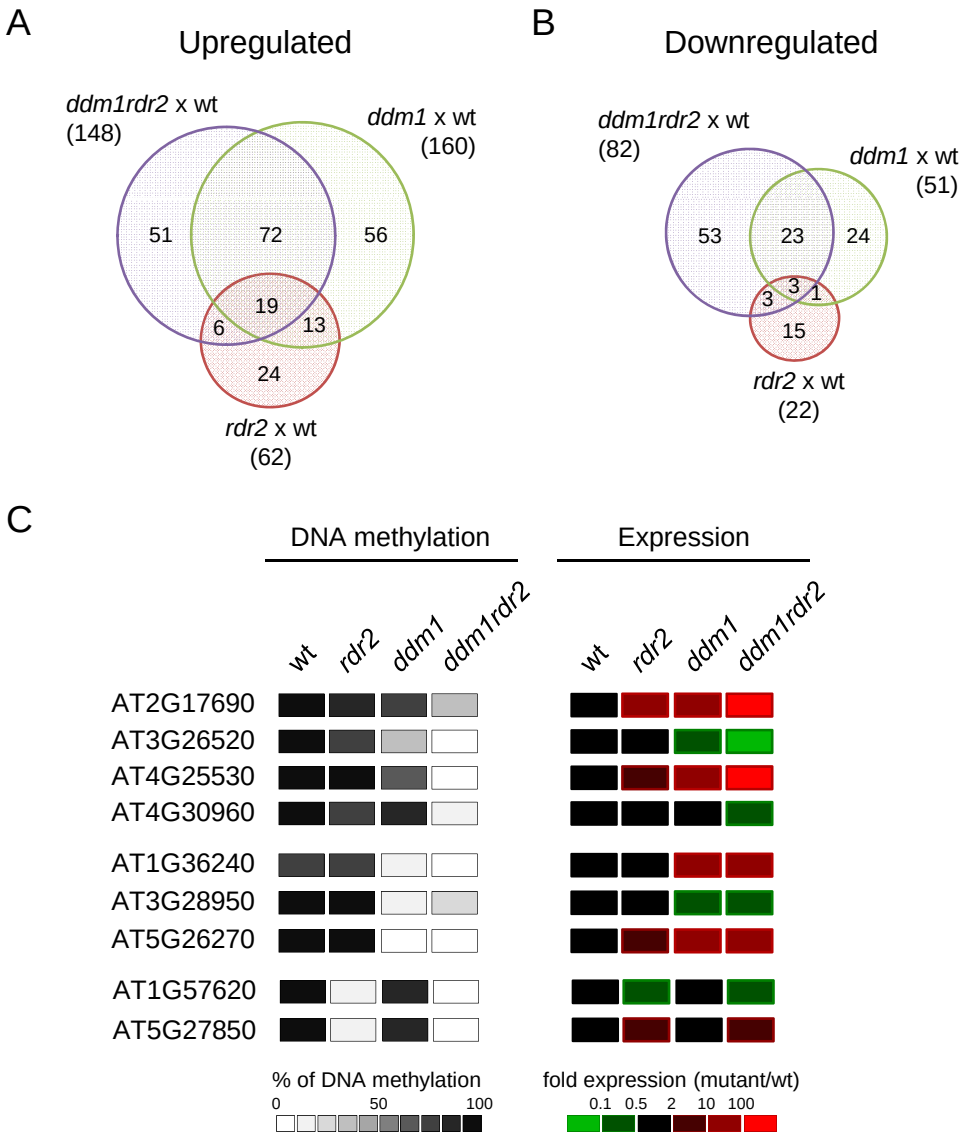


Figure 2. Genome-wide analysis allows the identification of a set of genes that are directly regulated by nearby DNA methylation. (A-B) Venn diagram representing the overlap of genes upregulated (A) and downregulated (B) in comparisons between *ddm1* and wt (green), *rdr2* and wt (red), and *ddm1rdr2* and wt (purple) as accessed by transcriptome analysis using gene-based CATMA arrays. All results are summarized in Supplemental Table S2. (C) RT-qPCR and McrBC-qPCR analyses conducted with a set of candidate genes. DNA methylation results are expressed as percentage of loss of molecules after McrBC digestion. Expression results were normalized to the average of three control genes, and are represented as fold-induction in comparison to wt levels (=1). Results are summarized in Supplemental Tables S4-S5.

Fully penetrant phenotypic alterations observed in *ddm1rdr2* are mainly caused by misregulation of the two genes FWA and SDC

It has been previously shown that ectopic overexpression of FWA and SDC is associated with late-flowering and wrinkled leaves phenotypes, respectively [5, 9, 20]. Given that loss of DNA methylation and overexpression of these two genes is most pronounced in *ddm1rdr2*, this prompted us to test the implication of FWA and SDC in the severe and fully penetrant phenotypic alterations specifically observed in *ddm1rdr2*. To this end, we crossed *ddm1rdr2* with *rdr2* and examined the DDM1/DDM1, *rdr2/rdr2* (F2) progeny (Fig 3A). These plants were recovered at the expected frequency and exhibited the late-flowering (L) and wrinkled leaf (W) phenotypes singly or in combination at frequencies compatible with Mendelian inheritance of independently segregating dominant mutations (3:1). We then characterized the DNA methylation status of the repeats associated with FWA and SDC in three (F2) DDM1/DDM1, *rdr2/rdr2* individuals with different phenotypes (Fig 3B). The wrinkled leaf, early-flowering 10H6(F2) plant was homozygous for the *sdc-dr* allele as well as for the fully methylated FWA epiallele. The 10H8(F2) plant was homozygous for the methylated SDC allele and for the *fwa-dr* allele, in agreement with its late-flowering and normal leaf phenotype. The late-flowering and wrinkled leaf 10F9(F2) plant was shown to be heterozygous at both loci, presenting average DNA methylation levels between those measured in *rdr2* and *ddm1rdr2*. These three (F2) individuals were subsequently selfed and 24 (F3) plants of each line were phenotyped (supplementary Fig S4). As expected, all 10H6 F3 progeny had wrinkled leaves and was early-flowering, while all 10H8 F3 progeny had was normal leaves and was late-flowering . As expected from the double heterozygosity of the 10F9 F2 parent, the late-flowering and wrinkled leaf phenotypes segregate independently in its F3 progeny and at the expected frequencies for two unlinked dominant mutations.

Lastly, we took advantage of a T-DNA insertion disrupting the SDC gene (*T-sdc*) to generate a triple *ddm1rdr2T-sdc* mutant (Fig 3C). Replacement of the wt SDC copy with the *T-sdc* allele sufficient to rescue the wrinkled leaf phenotype of *ddm1rdr2* plants. However, triple mutant plants still exhibited a late-flowering phenotype, consistent with the fact that this alteration is induced by ectopic expression of FWA. Altogether, our results indicate that *ddm1rdr2*-induced alterations of DNA methylation and gene expression in the imprinted FWA and SDC

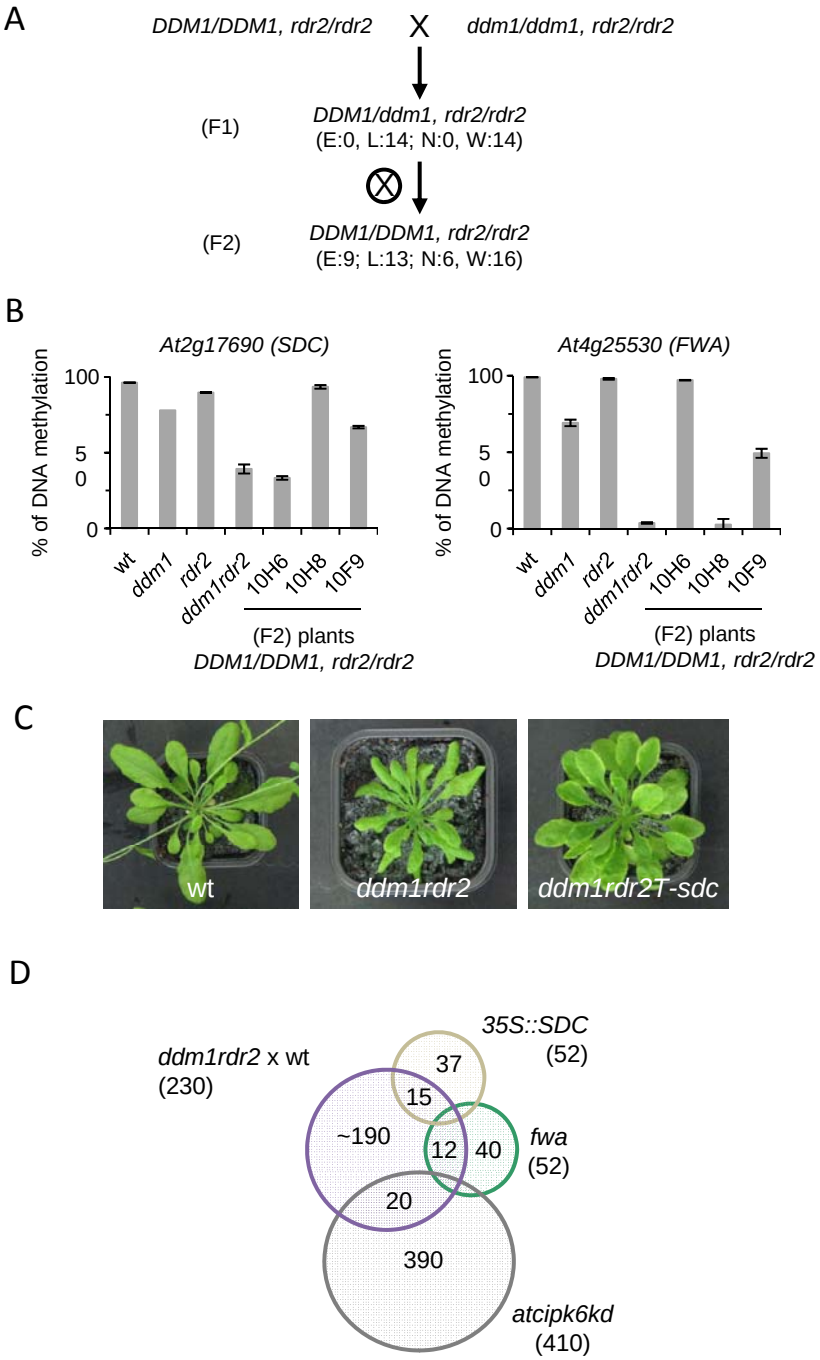


Figure 3. Phenotypic alterations and gene misregulation observed in *ddm1rdr2* directly results from misregulation of few primary targets. (A) Crossing scheme used to test the stability and segregation of late-flowering and wrinkled leaf phenotypes in a RNAi mutant background. (F1) individuals derived from *ddm1rdr2* x *rdr2* crosses were selfed and (F2) plants were genotyped. Phenotypic segregation for flowering time and leaf structure was characterized in *DDM1/DDM1*, *rdr2/rdr2* (F2) plants (E: early-flowering; L: late-flowering; N: normal leaf; W: wrinkled leaf). (B) DNA methylation over the promoter regions of *FWA* and *SDC* genes was accessed by McrBC-qPCR in wt, *ddm1*, *rdr2*, *ddm1rdr2*, and three (F2) *DDM1/DDM1*, *rdr2/rdr2* individuals: 10H6 (wrinkled leaf and early-flowering), 10H8 (normal leaf and late-flowering), and 10F9 (wrinkled leaf and late-flowering). (C) Wrinkled leaf phenotype observed in *ddm1rdr2* is rescued to wt in *ddm1rdr2T-sdc* triple mutant plants. (D) Venn diagram representing the overlap of genes misregulated in *ddm1rdr2* x *ddm1*, *35S::SDC*, *fwa*, and *atcpik6kd* (re-analyzed published data).

genes underlie much of the fully penetrant phenotypic alterations that are only observed when DDM1-dependent DNA methylation and RdDM are simultaneously compromised.

Few genes are misregulated directly as a consequence of loss of repeat-associated DNA methylation

We next tested if overexpression of FWA and SDC could also explain most of the changes in gene expression seen in *ddm1rdr2* mutant plants. To this end, we first compared the transcriptome of wt and *fwa* epimutant seedlings and identified 52 differentially expressed genes (Fig 3D). The impact of the ectopic expression of SDC was assessed using previously published data obtained with35S::SDC transgenic plants, which mimic the *sdc* epimutant [20]. Again, 52 genes showed altered expression as a result of overexpressing SDC. Additionally, we took advantage of publically available transcriptome data for another gene likely affected directly by loss of DNA methylation in *ddm1rdr2* mutant plants. This genes, called CBL-INTERACTING PROTEIN KINASE 6 (CIPK6, At4g30960) is downregulated and lose DNA methylation over its promoter region specifically in *ddm1rdr2* (Fig. 2). Disruption of the CIPK6 gene by a T-*DNA* insertion into the 3'UTR (atcipk6kd) leads to the misregulation of 410 genes [21].

When combined together, these three gene sets overlap with ~19% of the genes that are misexpressed in *ddm1rdr2* (Fig 3D).

In conclusion, our results indicate that despite most genes being insensitive to the epigenetic status of neighboring repeat sequences in Arabidopsis, DDM1-dependent DNA methylation and RdDM are critical for maintaining the normal expression of a few key genes with pleiotropic effects. However, this picture may be radically different in plant species that have repeat-rich genomes, such as maize, where phenotypic alterations and widespread gene misregulation are already observed in plants that are only defective in RdDM [22].

Methods

Plant material and growth conditions

Plants were from the A. thaliana Col-0 accession. *ddm1-2, met1-1, dcl2-1, dcl3-1, rdr2-1, nrpd1a-1,* and *nrpd1b-1* mutants have been previously described [2, 23-25]. *fwa* and 35S::SDC lines were kindly provided by Dr. Tetsuji Kakutani (NIG, Shizuoka, Japan) and Dr. Steve Jacobsen (UCLA, CA, USA), respectively [9, 20]. The SDC insertion (*T-sdc*) is Salk T-DNA_017593. Plants were grown under long-day conditions, either in vitro in liquid for 10- and 12-day-old seedlings [4], or in soil otherwise.

Transcriptome analysis

Total RNA was extracted using RNeasy Qiagen kit. Microarray analyses were conducted using the CATMA arrays [18]. Two independent biological replicates were produced for each comparison, and one technical replication with dye-swap was performed for each biological replicate. Labeling, hybridization to slides, and scanning were performed as described [27]. Normalization and statistical analysis were performed as described [28]. Published transcriptome for atcipk6kd was conducted on Affymetrix GeneChip Arabidopsis ATH1 genome array [21]. For our analysis, we only considered the differentially expressed genes that are present in the CATMA array.

Analysis of DNA methylation and transcription

DNA and RNA were extracted from seedlings using DNeasy and RNeasy Qiagen kits, respectively. Digestion with the McrBC enzyme (New England Biolabs) was performed on 500 ng of genomic DNA, as previously described [14]. Quantitative PCR (qPCR) was performed using a Roche LightCycler® 480 system machine and LightCycler® 480 SYBR Green I Master. Primers are listed in supplementary Table S5. Results were expressed as percentage of loss of molecules after McrBC digestion.

Reverse transcription (RT) was performed on 1 ug of total RNA using oligodT and Superscript II (Invitrogen). qPCR was performed as above and results were expressed as percentage of expression relative to the mean value obtained for three control genes (At2g36060; At4g29130; At5g13440; [14]). Primers are listed in supplementary Table S6.

References

1. Law, J.A. and S.E. Jacobsen, Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet, 2010. 11(3): p. 204-20.
2. Vongs, A., et al., Arabidopsis-Thaliana DNA Methylation Mutants. Science, 1993. 260(5116): p. 1926-1928.
3. Teixeira, F.K. and V. Colot, Repeat elements and the Arabidopsis DNA methylation landscape. Heredity (Edinb), 2010. 105(1): p. 14-23.
4. Lippman, Z., et al., Role of transposable elements in heterochromatin and epigenetic control. Nature, 2004. 430(6998): p. 471-6.
5. Kakutani, T., Genetic characterization of late-flowering traits induced by DNA hypomethylation mutation in Arabidopsis thaliana. Plant Journal, 1997. 12(6): p. 1447-1451.
6. Kakutani, T., et al., Developmental abnormalities and epimutations associated with DNA hypomethylation mutations. Proceedings of the National Academy of Sciences of the United States of America, 1996. 93(22): p. 12406-12411.
7. Miura, A., et al., Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. Nature, 2001. 411(6834): p. 212-4.
8. Saze, H. and T. Kakutani, Heritable epigenetic mutation of a transposon-flanked Arabidopsis gene due to lack of the chromatin-remodeling factor DDM1. Embo Journal, 2007. 26(15): p. 3641-3652.
9. Soppe, W.J.J., et al., The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. Molecular Cell, 2000. 6(4): p. 791-802.
10. Tsukahara, S., et al., Bursts of retrotransposition reproduced in Arabidopsis. Nature, 2009. 461(7262): p. 423-U125.
11. Ahmed, I., et al., Genome-wide evidence for local DNA methylation spreading from small RNA-targeted sequences in Arabidopsis. Nucleic Acids Res, 2011. 39(16): p. 6919-31.
12. Kasschau, K.D., et al., Genome-wide profiling and analysis of Arabidopsis siRNAs. Plos Biology, 2007. 5(3): p. 479-493.
13. Lister, R., et al., Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell, 2008. 133(3): p. 523-36.
14. Teixeira, F.K., et al., A Role for RNAi in the Selective Correction of DNA Methylation Defects. Science, 2009. 323(5921): p. 1600-1604.
15. Cao, X.F. and S.E. Jacobsen, Role of the Arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. Current Biology, 2002. 12(13): p. 1138-1144.
16. Huettel, B., et al., Endogenous targets of RNA-directed DNA methylation and Pol IV in Arabidopsis. Embo Journal, 2006. 25(12): p. 2828-2836.
17. Kurihara, Y., et al., Identification of the candidate genes regulated by RNA-directed DNA methylation in Arabidopsis. Biochem Biophys Res Commun, 2008. 376(3): p. 553-7.
18. Sclep, G., et al., CATMA, a comprehensive genome-scale resource for silencing and transcript profiling of Arabidopsis genes. BMC Bioinformatics, 2007. 8: p. 400.
19. Cokus, S.J., et al., Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature, 2008. 452(7184): p. 215-219.
20. Henderson, I.R. and S.E. Jacobsen, Tandem repeats upstream of the Arabidopsis endogene SDC recruit non-CG DNA methylation and initiate siRNA spreading. Genes & Development, 2008. 22(12): p. 1597-1606.
21. Tripathi, V., et al., CIPK6, a CBL-interacting protein kinase is required for development and salt tolerance in plants. Plant J, 2009. 58(5): p. 778-90.

22. Jia, Y., et al., Loss of RNA-dependent RNA polymerase 2 (RDR2) function causes widespread and unexpected changes in the expression of transposons, genes, and 24-nt small RNAs. PLoS Genet, 2009. 5(11): p. e1000737.
23. Kankel, M.W., et al., Arabidopsis MET1 Cytosine Methyltransferase Mutants. Genetics, 2003. 163(3): p. 1109-22.
24. Pontier, D., et al., Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. Genes & Development, 2005. 19(17): p. 2030-2040.
25. Xie, Z.X., et al., Genetic and functional diversification of small RNA pathways in plants. Plos Biology, 2004. 2(5): p. 642-652.
26. Turck, F., et al., Arabidopsis TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. Plos Genetics, 2007. 3(6): p. 855-866.
27. Lurin, C., et al., Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell, 2004. 16(8): p. 2089-103.
28. Gagnot, S., et al., CATdb: a public access to Arabidopsis transcriptome data from the URGV-CATMA platform. Nucleic Acids Res, 2008. 36(Database issue): p. D986-90.

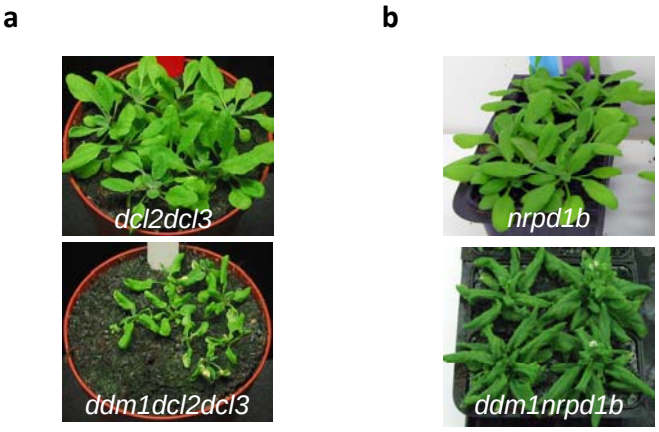
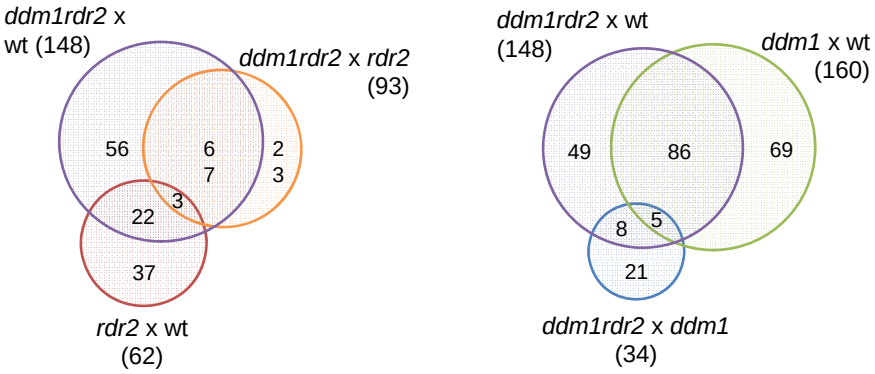


Figure S1. **Phenotypic alteration in *ddm1dcl2dcl3* and *ddm1nrpd1b* mutants. (a)** 23-days-old *dcl2dcl3* and *ddm1dcl2dcl3* plants. **(b)** 25-days-old *nrpd1b* and 40-days-old *ddm1nrpd1b* plants.

a



b

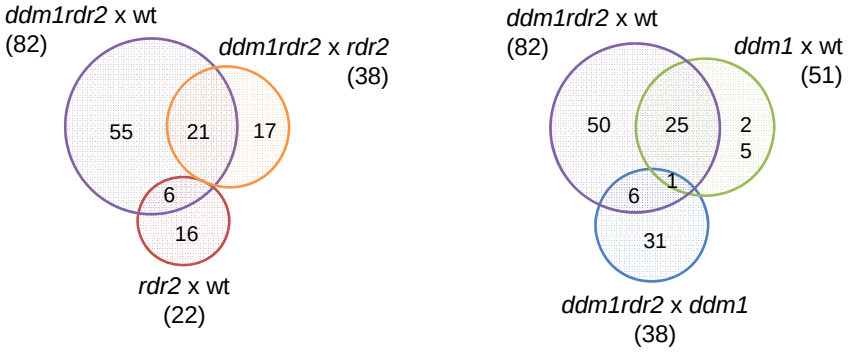
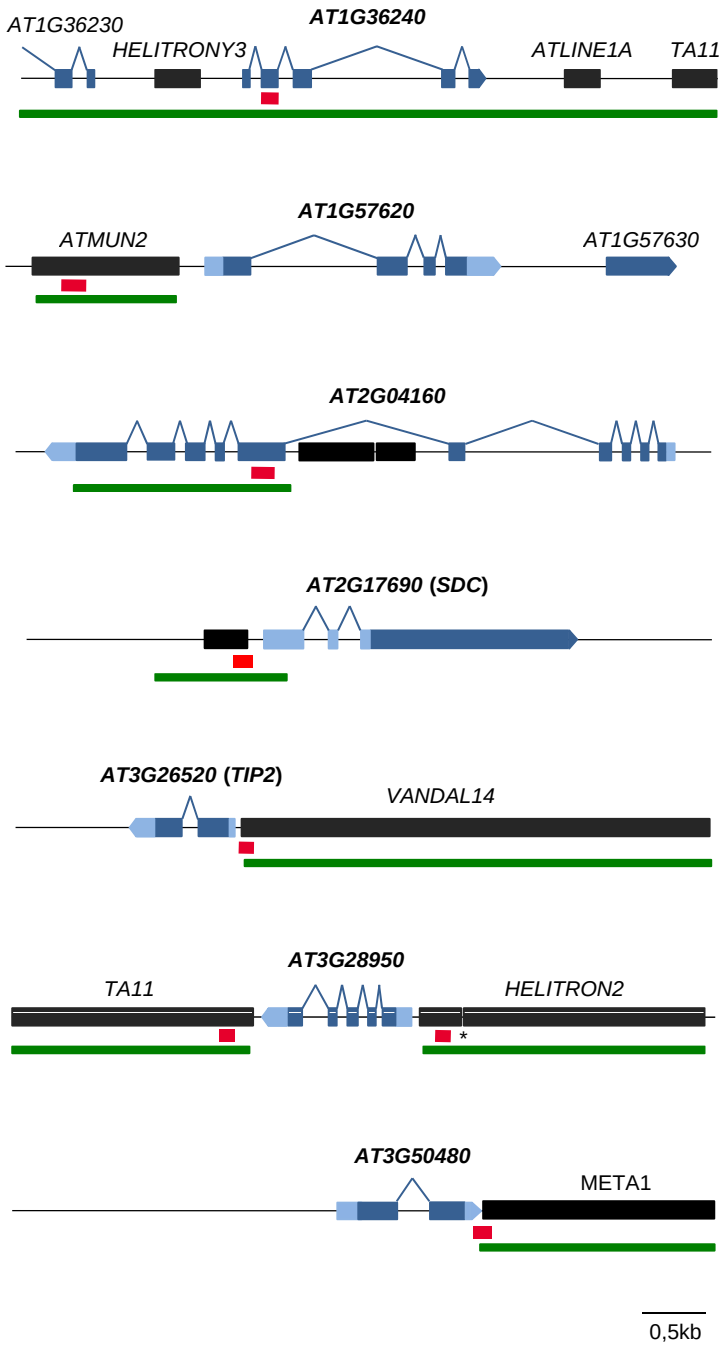


Figure S2. Additional CATMA analysis comparing single and double mutants.


AT1G36230
AT1G36240
HELITRONY3
ATLINE1A
TA11
AT1G57620
ATMUN2
AT1G57630
AT2G04160
AT2G17690 (SDC)
AT3G26520 (TIP2)
VANDAL14
AT3G28950
TA11
HELITRON2
*
AT3G50480
META1
0,5kb

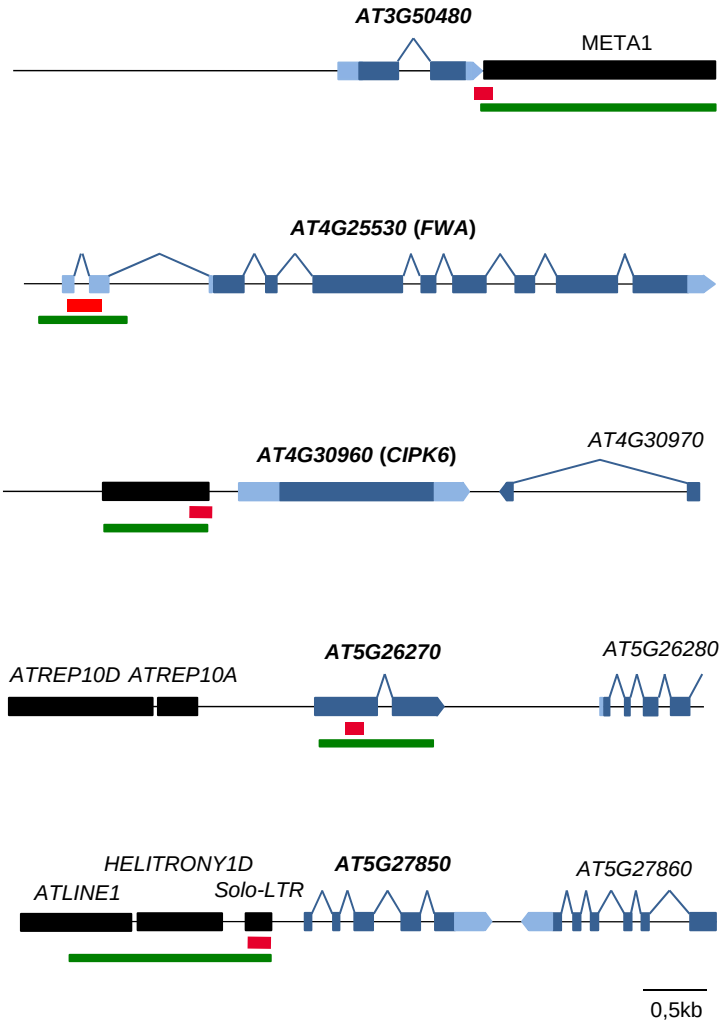


Figure S3. Genomic structure of eleven tested genes. Densely methylated sequences are indicated as green lines. Densely methylated sequences probed for DNA methylation by McrBC-qPCR are shown in red. Starts are indicate nearby probes for which changes in DNA methylation do not show correlation with changes in gene expression. Annotation: UTRs – light blue; coding sequences - dark blue; repeat elements – black.

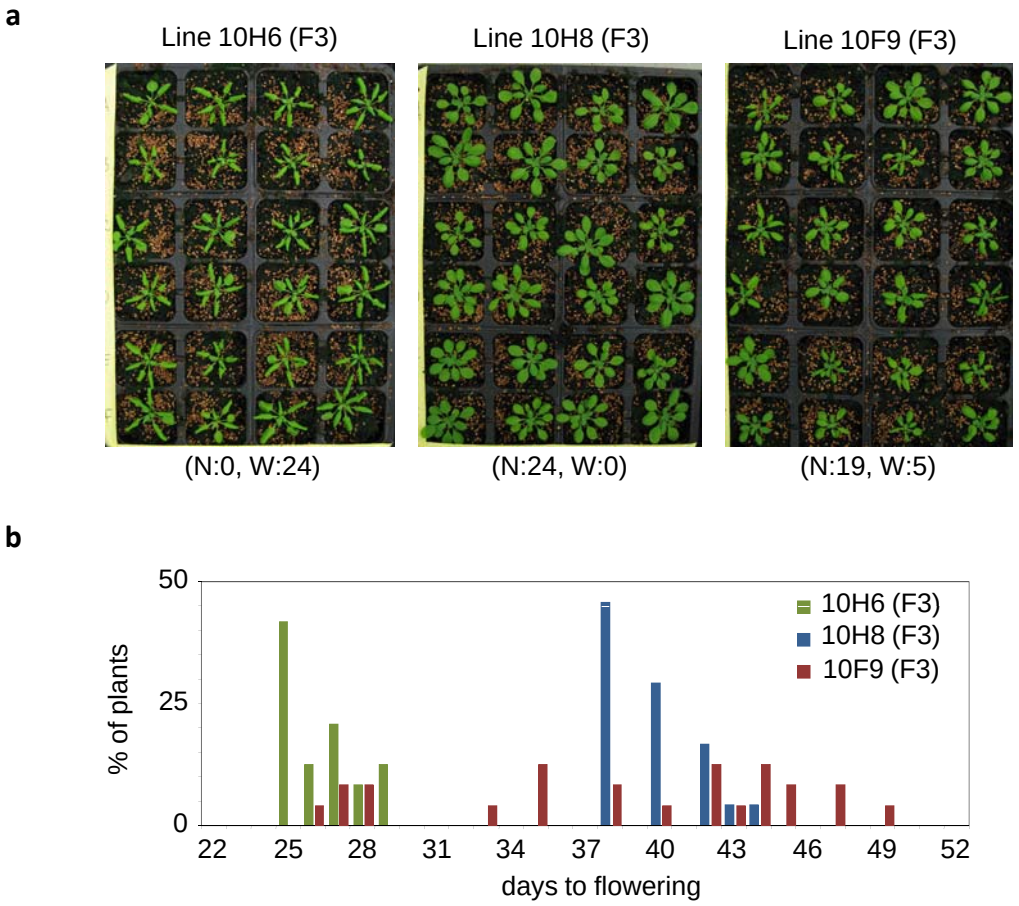


Figure S4. Segregation of leaf structure (A) and flowering time (B) phenotypes in 24 (F3) siblings originated through selfing of three (F2) *DDM1/DDM1*, *rdr2/rdr2* individuals: 10H6 (wrinkled leaf and early-flowering), 10H8 (normal leaf and late-flowering), and 10F9 (wrinkled leaf and late-flowering).

Table S1. Segregation of *DDM1* and *RDR2* alleles in F2 progeny

Genotype	**Observed**	**Expected**
DDM1/DDM1, rdr2/rdr2	9	8,75
DDM1/DDM1, RDR2/rdr2	14	17,5
DDM1/DDM1, RDR2/RDR2	12	8,75
DDM1/ddm1, rdr2/rdr2	18	17,5
DDM1/ddm1, RDR2/rdr2	36	35
DDM1/ddm1, RDR2/RDR2	19	17,5
ddm1/ddm1, rdr2/rdr2	9	8,75
ddm1/ddm1, RDR2/rdr2	18	17,5
ddm1/ddm1, RDR2/RDR2	5	8,75
Total	140	140

Table S2. List of the 12 candidate genes tested.

Gene ID	**Annotation (TAIR)**	**Position of methylated region probed**
AT1G36240	60S ribosomal protein L30	2nd exon
AT1G57620	Emp24/gp25L/p24 family protein	Upstream and Downstream
AT2G04160	Auxin-induced in root culture 3	6th exon
AT2G17690	F-box domain containing protein (SDC)	Upstream
AT3G26520	Gamma tonoplast intrinsic protein 2 (TIP2).	Upstream
AT3G28950	Avirulence induced gene (AIG) protein	Upstream and Downstream
AT3G50480	Homolog of RPW8 4 (HR4)	Downstream
AT4G25530	Homeodomain-containing transcription factor (FWA)	5' and 1st exon
AT4G30960	CBL-INTERACTING PROTEIN KINASE 6 (CIPK6)	Upstream
AT5G22500	Acyl CoA reductase/male-sterility protein (FAR1)	8th intron
AT5G26270	Unknown protein	1st exon
AT5G27850	60S ribosomal protein L18	Upstream

Table S3. Raw RT-qPCR data. Average percentage (a minimum of two measurements) of expression in relation to controls and standard deviation are presented.

Name of primer pair	% of control genes				STDEV			
	wt	*ddm1*	*rdr2*	*ddm1rdr2*	wt	*ddm1*	*rdr2*	*ddm1rdr2*
AT1G36240	1,27	50,18	1,56	50,00	0,17	0,74	0,05	0,25
AT1G57620	5,60	5,22	1,49	1,58	0,07	0,37	0,07	0,26
AT2G04160	2,83	1,51	1,43	1,14	0,14	0,09	0,03	0,03
AT2G17690	0,05	3,05	0,61	11,16	0,00	0,16	0,02	0,55
AT3G26520	981,05	241,62	716,23	40,10	69,66	6,51	24,57	2,75
AT3G28950	24,96	8,96	19,82	8,50	0,06	0,09	0,49	0,48
AT3G50480	13,68	35,23	18,69	53,42	0,47	0,52	0,18	2,09
AT4G25530	0,00	0,17	0,02	1,15	0,00	0,02	0,02	0,12
AT4G30960	3,51	2,56	2,49	1,25	0,01	0,13	0,04	0,13
AT5G22500	0,71	0,91	0,64	0,19	0,01	0,01	0,08	0,01
AT5G26270	1,02	25,53	3,27	23,90	0,10	0,75	0,11	0,47
AT5G27850	519,64	585,65	1322,71	1861,58	1,27	5,74	45,37	95,76

Table S4. Raw McrBC-qPCR data. Average percentage (a minimum of two measurements) of molecules lost through McrBC digestion and standard deviation are presented.

Name of primer pair	% of methylation				STDEV			
	wt	*ddm1*	*rdr2*	*ddm1rdr2*	wt	*ddm1*	*rdr2*	*ddm1rdr2*
AT1G36240	78,83	16,19	72,16	7,93	0,62	2,05	0,96	2,12
AT1G57620_5met	93,40	87,12	17,33	8,21	1,85	1,26	2,84	5,84
AT2G04160	99,24	7,07	99,08	20,17	0,14	10,00	0,09	0,39
AT2G17690(2)	96,27	77,93	89,74	39,25	0,02	0,00	0,05	2,98
AT3G26520_5met	98,01	39,33	75,00	2,71	0,06	0,98	0,25	3,01
AT3G28950_3met	99,54	18,49	99,19	24,21	0,09	0,40	0,04	3,55
AT3G28950_5met	98,01	85,58	62,24	41,96	0,08	0,92	0,56	0,85
AT3G50480_3met	87,24	62,37	13,54	7,63	0	0,74	0,85	3,17
AT4G25530(FWATR)	99,04	69,19	97,94	3,74	0,01	2,11	0,55	0,47
AT4G30960_5met(2)	94,88	89,88	78,79	17,34	1,38	0,25	0,00	2,03
AT5G26270	99,02	5,59	95,56	3,04	0,05	6,93	0,24	3,33
AT5G27850_5met	97,33	85,31	10,79	5,84	0,07	5,28	7,53	3,64

Table S5. List of primer pairs used for McrBC-qPCR analysis, sequence annotation and the position in relation to gene.

Name of primer pair	Chr.	Position (TAIR7)	Forward	Reverse	Annotation	Position in relation to gene
AT1G36240	1	13616208..13616307	AGTCCCATGAAGGAATCAACA	TTGGAACTGCGAAGAGATTTG	-	2nd exon
AT1G57620_5met	1	21345158..21345327	TAAGCTTGGGAAAAATGTCG	TTGGCCTGGTCAACGAAT	ATMUN2 (DNA/MuDR)	5'
AT2G04160	2	1403259..1403436	TGAGTGTGATCGCTGCCTAC	GGGGCTCCAAGAAGGATAAC	-	6th exon
AT2G17690(2)	2	7689165..7689268	CGTCAGCCCTAACACAAAACA	GCGGCGTGAGTCTCTAAGAT	Single tandem repeats	5'
AT3G26520_5met	3	9725039..9725275	GCCGAGTGTTTTCTTCTTCTTG	ACAAGTCTGCTGTCGGGAAT	VANDAL14	5'
AT3G28950_3met	3	10976761..10977039	CCCAACGCTCTATGGCTAAA	AGGATGAGCCGTTTACGATG	TA11	3'
AT3G28950_5met	3	10978833..10979036	GTCGGTATCCGTGCTACACA	CCATGCTACAAACCGAGGAA	HELITRONY2	5'
AT3G50480_3met	3	18745119..18745283	CAAAGAAATGAAGGCCCAGAT	GATGATGGGCTTAGATGATGG	META1 (LTR/Copia)	3'
AT4G25530(FWATR)	4	13038394..13038584	GCCATTGGTCCAAGTGCTAT	CGGTGCTCGTATGAATGTTG	Single tandem repeats	5' and 1st exon
AT4G30960_5met(2)	4	15066625..15066785	CCGCTGGTTTACCAGTCATAA	TAGGGCCTGACGGATTAAAA	Dispersed repeat	5'
AT5G26270	5	9206718..9206926	GCTGTGTGCAGTCCCTTCTA	GCCATGGAACGAGTAGGAGA	-	1st exon
AT5G27850_5met	5	9872531..9872826	TGCTTTCTCTTTCTCTTCTCTTTTTC	AAACCGGATAAGTATGGATGTCA	HELITRONY1D (RC/Helitron)	5'
META1_5met	4	7816672..7816958	GAGGTGATTAGCTAGCACTTTCAA	GGAAATGAGTGTGTTATTCCATTG	META1 (LTR/Copia)	-

Table S6. List of primer pairs used for RT-qPCR analysis and gene annotation

Name of primer pair	Chr.	Position (TAIR7)	Forward Primer	Reverse Primer	Annotation
AT1G36240	1	13616208..13616307	AGTCCCATGAAGGAATCAACA	TTGGAACTGCGAAGAGATTTG	60S ribosomal protein L30
AT1G57620	1	21347625..21347787	CACCTTATGGCACCGTGTTA	GCGATTCCAGTTTTCCAGTC	Emp24/gp25L/p24 family protein
AT2G04160	2	1403259..1403436	TGAGTGTGATCGCTGCCTAC	GGGGCTCCAAGAAGGATAAC	Auxin-induced in root culture 3
AT2G17690	2	7690614..7690902	ACTAACGGTTCTTGCGGAGA	CTATCGTTTTGGGAGGGTCA	F-box domain containing protein (SDC)
AT3G26520	3	9724077..9724274	TGACCCCAAGAACGGTAGTC	GCGAGTCCACCACCAATAAG	Gamma tonoplast intrinsic protein 2 (TIP2).
AT3G28950	3	10977537..10977707	CGGAGAAATTCGTGGAGACT	CAAAACCCACTCTTTACTTACTCAC	Avirulence induced gene (AIG) protein
AT3G50480	3	18744185..18744361	CCGGCTATCAGATCGTCTTC	AGCCGTCGTAAATGACTTGG	Homolog of RPW8 4 (HR4)
AT4G25530	4	13038394..13038584	GCCATTGGTCCAAGTGCTAT	CGGTGCTCGTATGAATGTTG	Homeodomain-containing transcription factor (FWA)
AT4G30960	4	15067667..15067829	CACGAAGTCATGGCGAGTAA	CACCACGGCTATGACAGAAA	CBL-INTERACTING PROTEIN KINASE 6 (CIPK6)
AT5G22500	5	7472038..7472190	AGCGGTGAACAACCTGGAT	TTGCGAGATCTCTTCATCAGAA	Acyl CoA reductase/male-sterility protein (FAR1)
AT5G26270	5	9206718..9206926	GCTGTGTGCAGTCCCTTCTA	GCCATGGAACGAGTAGGAGA	Unknown protein
AT5G27850	5	9874172..9874332	TGAAGCAGTGAAGCATTTCG	ACCAAGCAAAACGAAACCAG	60S ribosomal protein L18
Controls					
At2g36060	2	15149917..15150067	TGAAGTCGTGAGACAGCGTTG	GGGCTTCTCCATTGTTGGTC	Ubiquitin E2 variant 1C
At4g29130	4	14352280..14352408	GGCGTTTTCTGATAGCGAAAA	ATGGATCAGGCATTGGAGCT	Hexokinase (HXK1)
At5g13440	5	4308303..4308423	ACAAGCCAATTTTTGCTGAGC	ACAACAGTCCGAGTGTCATGGT	Ubiquinol-cytochrome C reductase iron-sulfur subunit

4.2 Influence des nouvelles insertions d'ET sur l'expression des gènes situés à proximité avant et après établissement du contrôle épigénétique

Les analyses présentées dans le manuscrit ci-dessus indiquent que relativement peu de gènes sont directement influencés par la méthylation des ET (et autres séquences répétées) localisés à proximité chez Arabidopsis. Quoiqu'il en soit, ces analyses ne permettent de voir que ce qui a passé le filtre de la sélection naturelle et ne nous renseigne pas sur l'influence qu'un ET tout juste inséré peut avoir sur l'expression des gènes situés à proximité. J'ai donc cherché à déterminer les conséquences que les ET nouvellement insérés pouvaient avoir sur les gènes situés à proximité dans les epiRIL avant et après acquisition du contrôle épigénétique. Je me suis focalisée sur les nouvelles insertions d'*EVADE* (*ATCOPIA93*), le seul ET pour lequel l'établissement de la méthylation de l'ADN associé à la répression transcriptionnelle a pu être clairement décrit.

Je me suis intéressée à un premier jeu de quatre gènes qui présentent des nouvelles insertions d'*EVADE* à l'état homozygote en F9 dans l'épiRIL 454 : *AT1G01010*, *AT1G68410*, *AT5G2550* et *AT5G25560* (fig. 4.1). J'ai testé l'expression de ces gènes par RT-qPCR dans des plantes individuelles de l'épiRIL 454 présentant les copies d'*EVADE* dans un état méthylé ou non. Parmi les quatre gènes étudiés, *AT5G2550* qui présente une nouvelle insertion d'*EVADE* à ~300 pb de son extrémité 3', montre une variation significative d'expression corrélée avec le changement d'état de méthylation de la copie d'*EVADE* située à proximité (fig. 4.2). Afin de déterminer si les transcrits détectés ici sont produits à partir du promoteur du gène ou s'il s'agit de transcrits « read through » initiés dans l'ET, il faudra réaliser des expériences de 5' Race. Les trois autres gènes ne sont pas affectés par l'insertion d'*ATCOPIA93* située à proximité (données non montrées). Dans Mari-Ordonez et al., 2013, les données concernant le gène *AT1G68420* sont également présentées (figure 6 de l'article). Néanmoins, étant donné qu'*EVADE* est localisé dans la séquence codante du gène (fig. 4.1), ce dernier est vraisemblablement invalidé.

Cette analyse sera étendue à plus d'insertions et poursuivie de façon systématique sur les autres epiRIL présentant des nouvelles insertions d'*ATCOPIA93*. Les domaines d'expression des candidats pourront être comparés avec celui d'*ATCOPIA93* qui est restreint à la L2.