



University of Groningen

## Computational Methods for High-Throughput Small RNA Analysis in Plants

Monteiro Morgado, Lionel

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*

Publisher's PDF, also known as Version of record

*Publication date:*

2018

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Monteiro Morgado, L. (2018). Computational Methods for High-Throughput Small RNA Analysis in Plants. [Groningen]: University of Groningen.

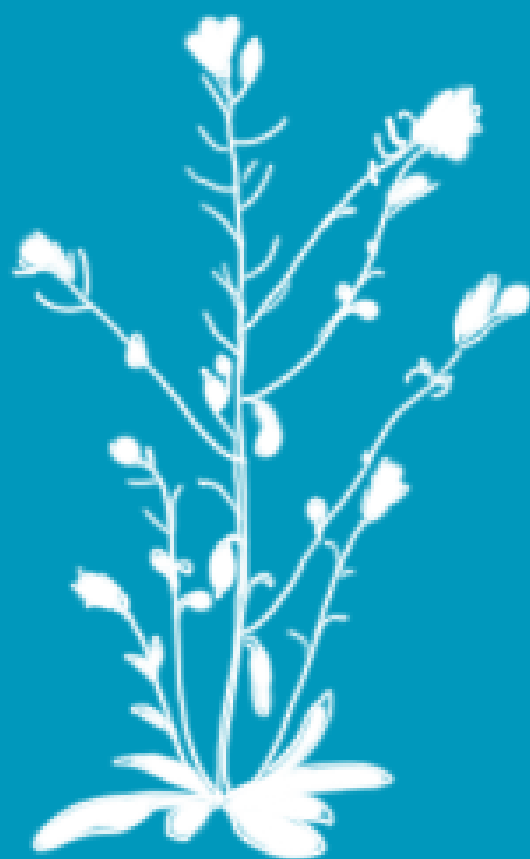
**Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

**Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



# CHAPTER 5

---

## Epigenetic mapping of the *Arabidopsis* metabolome reveals mediators of the epigenotype-phenotype map

Rik Kooke, Lionel Morgado, Frank Becker, Henriëtte van Eekelen, Qunfeng F Zhang, Ric CH de Vos, Frank Johannes, Joost JB Keurentjes

### Abstract

Identifying the sources of natural variation underlying metabolic differences between plants will enable a better understanding of plant metabolism and provide insights into the regulatory networks that govern plant growth and morphology. So far, the contribution of epigenetic variation to metabolic diversity has been largely ignored. In the present study, we utilized a panel of *Arabidopsis thaliana* epigenetic recombinant inbred lines (epiRILs) to assess the impact of epigenetic variation on the metabolic composition. Thirty epigenetic QTL (QTL<sup>epi</sup>) were detected, which partly overlap with QTL<sup>epi</sup> linked to growth and morphology. In an effort to identify causal candidate genes in the QTL<sup>epi</sup> region or their putative *trans*-targets we performed *in silico* small RNA and qPCR analyses. Differentially expressed genes were further studied by phenotypic and metabolic analyses of knockout mutants. Three genes were detected that recapitulated the detected QTL<sup>epi</sup> effects, providing strong evidence for epigenetic regulation in *cis* and in *trans*.

## 5.1 Introduction

The accumulation of secondary metabolites in specific plant tissues enables a balanced division of resources that contributes to increased fitness and competitive ability [238, 239]. Flowers form the basis of the sexual reproductive organs, and as such they are important organs for the plant to protect from herbivores and pathogens. Moreover, they serve very specialized functions, such as attracting pollinators and securing anthesis, further strengthening the need for specific chemical compounds. In this respect, it is not surprising that flowers have a much more complex metabolic profile than vegetative tissues and that defense compounds are most concentrated in the reproductive organs of plants [240–242].

Because of adaptation to various biotic and abiotic environments, extensive natural variation in phytochemical profiles exists between and within species, which can be investigated to unravel the underlying regulation of secondary metabolism [243]. The combination of genetic mapping populations with the (un)targeted analysis of large numbers of metabolites has revealed strong genetic regulation, both qualitatively and quantitatively [244–248].

Although the genetic basis of secondary metabolite variation is becoming better understood, the role of epigenetics in secondary metabolism has so far been largely overlooked. However, epigenetic variation that causes phenotypic diversity has been identified in plants and can be successfully transmitted to offspring for several generations, providing evidence for epigenetic inheritance [79, 249–254]. A number of studies have also reported a role for epigenetics in the regulation of secondary metabolism, suggesting the involvement of small RNA biosynthesis [255] and DNA methylation states [256, 257].

Epigenetic recombinant inbred lines (epiRILs) in *Arabidopsis* were especially designed to study the impact of heritable epigenetic variation on complex traits [252] and epigenetic quantitative trait loci (QTL) mapping approaches have shown that specific differentially methylated regions (DMRs) in the epiRILs can affect complex traits [251, 253, 258].

Here we analyzed the metabolic profile of 96 epiRILs using untargeted liquid chromatography–mass spectrometry (LC-MS) metabolomics of both rosette leaves and flower heads and associated the observed qualitative and quantitative variation to epigenetic variation in DNA-methylation. To explore the epigenetic mechanisms underlying the detected QTL<sup>epi</sup> effects, we tested two possible hypotheses: (1) methylation variation in promoters of genes involved in secondary metabolite regulation controls metabolic and/or

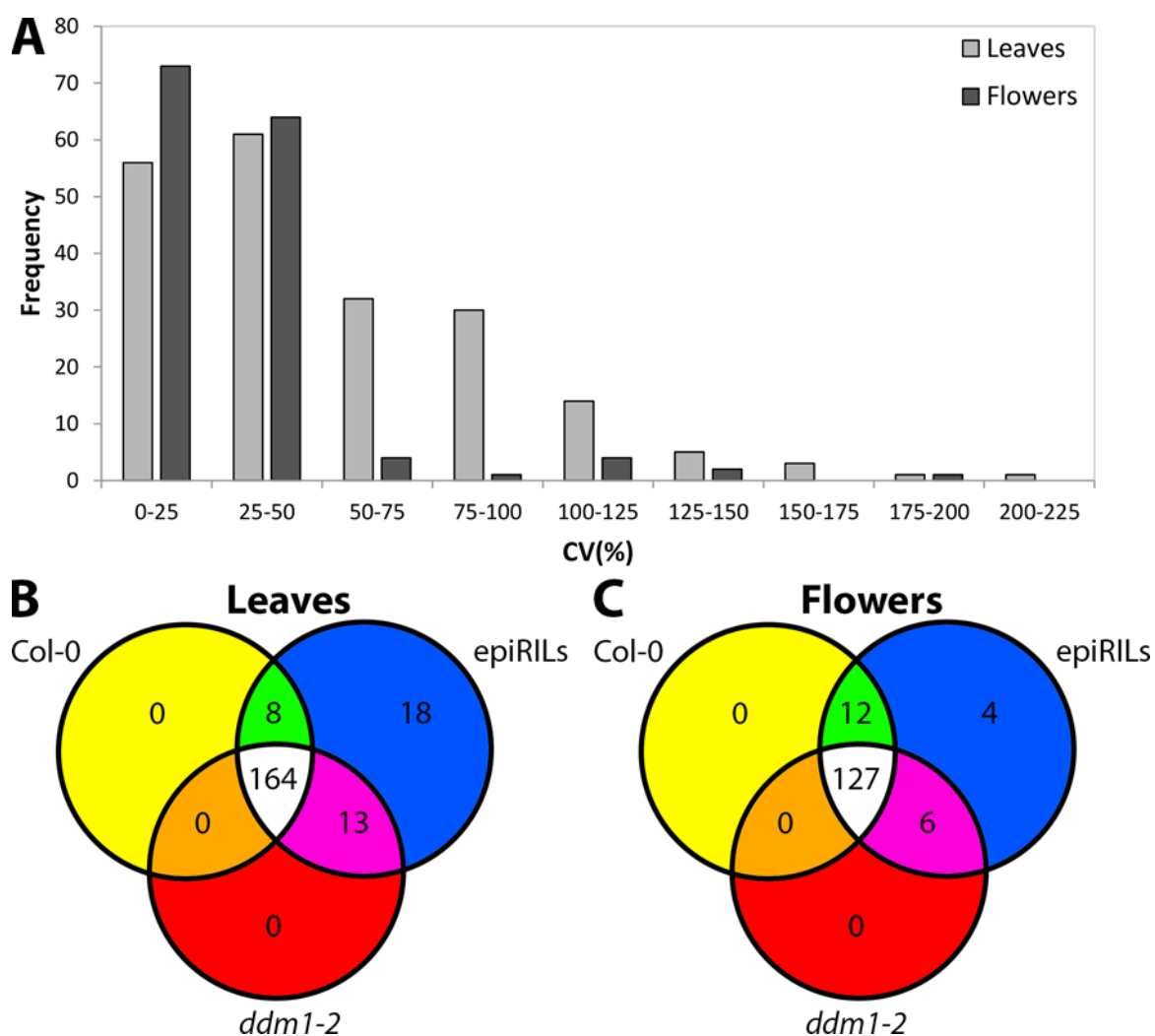
morphological variation in *cis* and (2) methylation variation in the QTL<sup>epi</sup> interval changes the production of small RNA that targets genes in *trans* leading to altered regulation of metabolic and morphological phenotypes. Supporting evidence for both hypotheses could be obtained.

## 5.2 Results

### 5.2.1 Tissue-specific epigenetic variation in plant secondary metabolism

To evaluate the effect of epigenetic variation on plant secondary metabolism, aqueous-methanol extracts of rosette leaves and flower heads from 96 epiRILs and their parents, Col-0 and *ddm1-2* (*Decreased DNA Methylation 1*) mutants, were analyzed by an essentially untargeted Liquid Chromatography Quadrupole Time-of-Flight Mass Spectrometry (LC-QTOF-MS) based metabolomics approach. This method is particularly suited for the analysis of semi-polar metabolites including glucosinolates, hydroxycinnamates, flavonoids, alkaloids, saponins and phytochemicals from various biochemical pathways present in plants [259, 260]. It is therefore not surprising that most of the annotated compounds in the used epiRIL population belong to the classes of glucosinolates and flavonoids, which are strongly enriched in Arabidopsis.

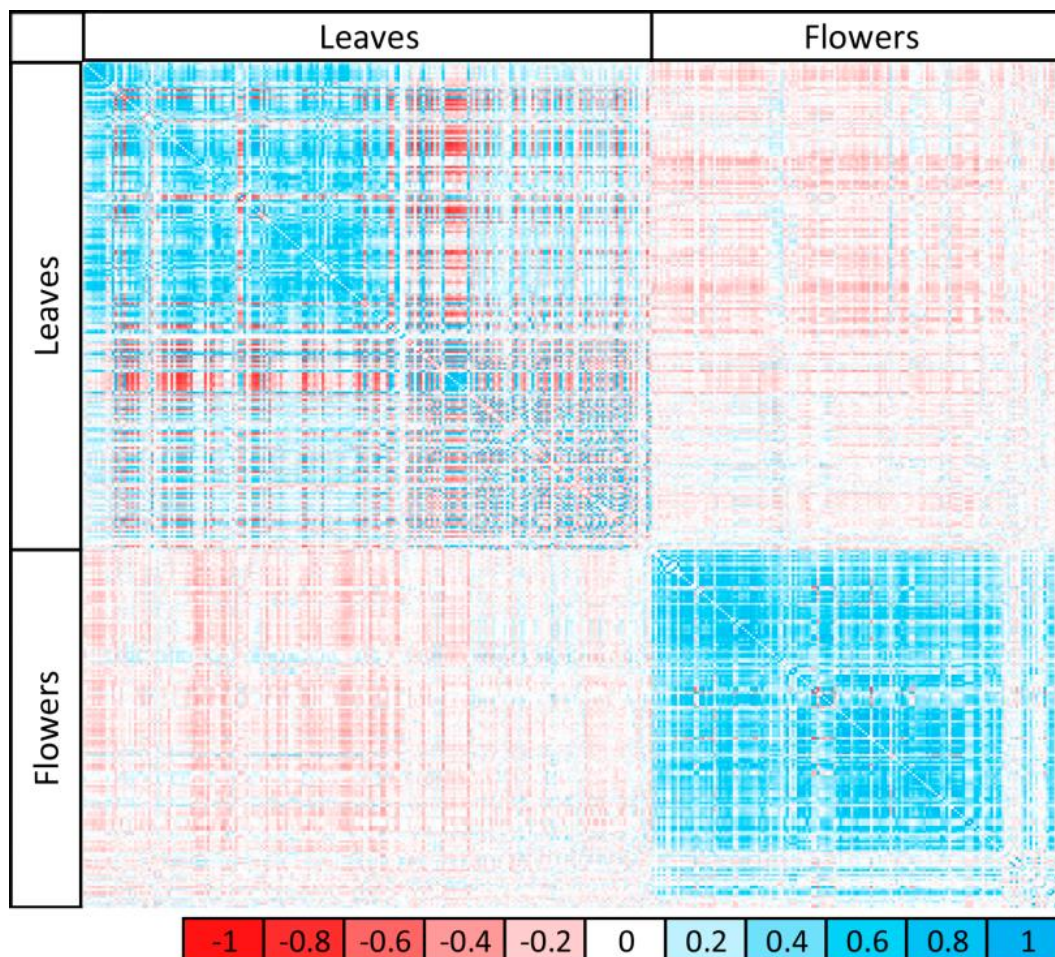
In both tissues, qualitative and quantitative variation in metabolite accumulation could be observed among the epiRILs. In the leaves, 8,955 reproducible mass signals corresponding to 203 reconstructed metabolites (mass clusters) were retrieved, using the Metalign- and MSClust-based untargeted data processing workflow [261]. The observed variation in leaf metabolites was substantial, with an average coefficient of variation (CV) of 54%, ranging from 2% to 206% (Figure 5.1A). These values are in general much higher compared to morphological variation in the same set of lines under the same conditions and should thus provide ample variation for genetic mapping [239]. Comparison of metabolic profiles showed that the vast majority of metabolites were detected in both parents and their derived epiRILs. However, in leaves, eight metabolites were only detected in Col-0, while thirteen, including quercetin-3-O-hexoside, were uniquely detected in *ddm1-2* (Figure 5.1B). Demethylation thus does not only reduce or enhance metabolite levels, but it can also cause the accumulation of additional metabolites. It should be noted that non-detection can either mean that the metabolites were not synthesized or that their concentration did not pass the



**Figure 5.1. Metabolite variation in leaves and flowers of epiRIL population.** (A) Frequency distribution of coefficient of variation (%) for all 203 leaf (light grey) and 149 flower (dark grey) metabolites detected in the Col-0 x *ddm1-2* epiRIL population using untargeted LC-QTOF-MS-based metabolomics. (B) Number of metabolites that were detected in the leaves of the parents of the population, Col-0 and *ddm1-2*, and the epiRILs (C) Number of metabolites that were detected in the flowers of the parents of the population, Col-0 and *ddm1-2*, and the epiRILs.

detection threshold limit. In the majority of cases, however, these qualitative differences were of high order of magnitude, implying substantial variation between the parents.

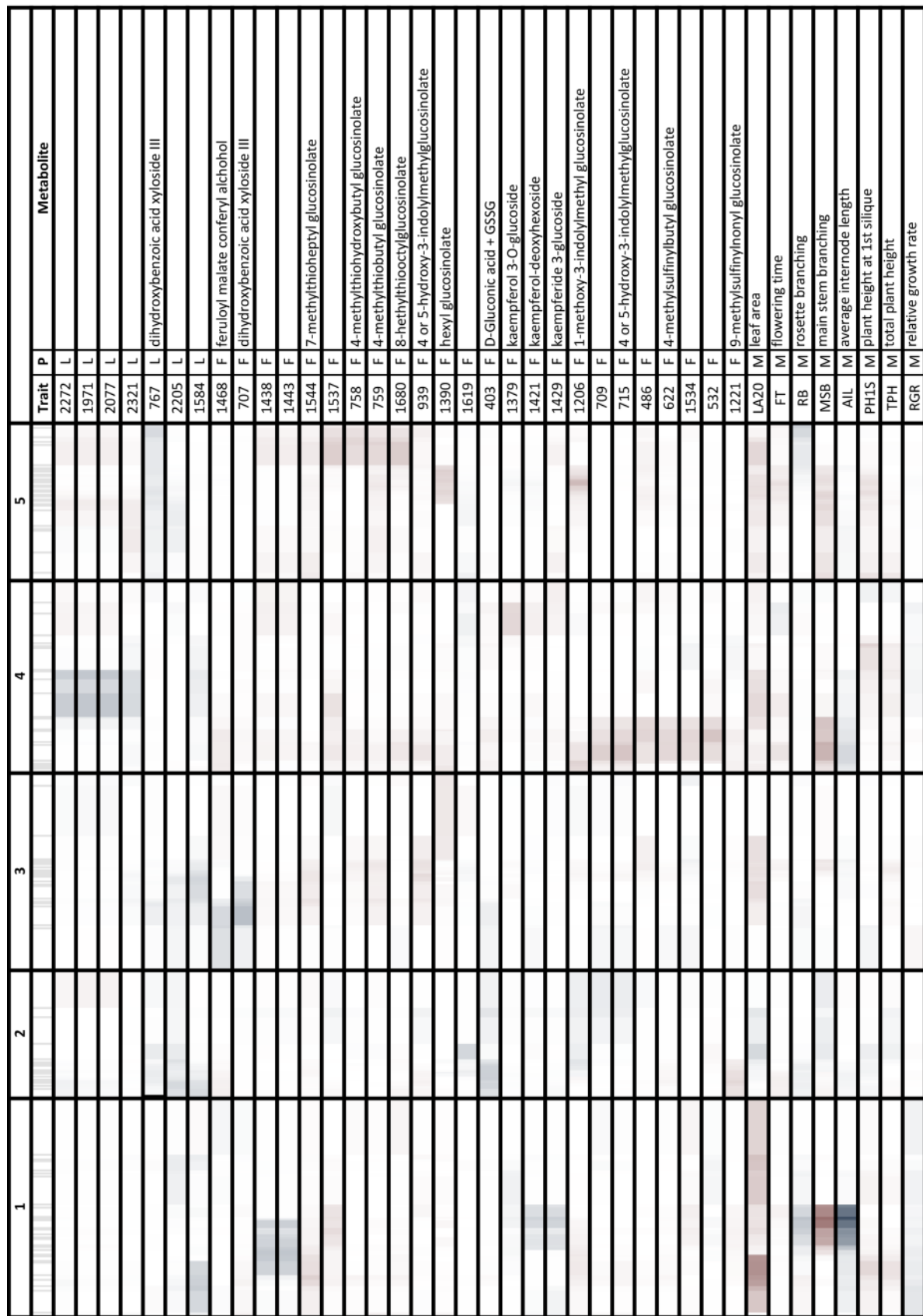
Besides qualitative and quantitative differences between the parents, eighteen metabolites were identified that were solely detected in (a part of) the epiRIL population while being absent in both parents (Figure 5.1B). These are most likely the result of epi-allelic transgressive segregation [262] that gave rise to the accumulation of novel metabolite structures, or alternatively *de novo* epigenetic or genetic variation accumulated during the development of the epiRIL population.



**Figure 5.2. Correlation matrix of detected metabolites in the epiRIL population.** Pearson correlation between metabolites within and between tissues is indicated by color intensity from -1 (red) to 1 (blue). Variation in metabolites correlates strongly within the same tissue, but correlations between different tissues is much weaker.

In the flowers, 6,738 mass signals were extracted corresponding to 149 metabolites with an average CV of 31% ranging from 8% to 191% (Figure 5.1A). Although the variation in metabolite accumulation in the flowers is somewhat lower than in the leaves, it is still high compared to morphological variation in the same epiRILs under the same conditions [253].

As was the case in leaf tissue, the majority of metabolites in flowers were detected in both parents and their derived epiRILs, whereas twelve metabolites were only detected in Col-0 flowers and six metabolites only in *dmd1-2* flowers (Figure 5.1C). Four metabolites were only detected in a portion of the epiRILs and not in the parents, suggesting that epi-allelic transgressive segregation has resulted in the accumulation of these metabolites. These findings indicate that epi-allelic variation can increase metabolic variation in a quantitative and qualitative manner in both flowers and leaves.



**Figure 5.3. QTL<sup>epi</sup> heatmap for metabolic and morphological traits.** QTL<sup>epi</sup> heatmap showing the positions of the QTL<sup>epi</sup> and the overlap with QTL<sup>epi</sup> for morphological traits divided over the 5 chromosomes. The morphological traits are described previously (Kooke et al., 2015). The thin grey lines in the second row indicate the marker positions in cM. Trait = metabolite number or morphological trait, P = phenotype group, L = leaf, F = flower, M = morphology. The legend on the right indicates the QTL LOD score between -15 (red) and 15 (blue).



Strong correlations between metabolites across all epiRILs were detected within the same tissue, but much weaker correlations occurred between metabolites in different tissues (Figure 5.2). Although the total number of correlating metabolites was quite similar in leaves and flowers (55% over 53%, respectively,  $\rho > 0.2$ ), the proportion of negative correlations between metabolites was much higher in leaves than in flowers (46% over 8%, respectively) (Figure 5.2), suggesting a stronger competition for resources in the leaves than in the flowers, possibly because of the dual role of leaves as both sink and source tissue. The high proportion of positive correlations in the flowers indicates that flowers show a much more coordinated regulation of metabolite accumulation, which might be caused by the tight developmental control and specific function of this tissue. The various glucosinolates in the flowers, for example, showed very strong positive correlations ( $P < 0.05$ ) mutually as well as with other identified compounds such as D-gluconic acid and dihydroxybenzoic acid-xyloside III. Similar positive correlations were observed among the flavonoids and between flavonoids and most other compounds, although a negative correlation of flavonoids was detected with 1-methoxy-3-indolylmethyl glucosinolate.

Although the leaf and flower tissues were not harvested from the same plant, some significant correlations ( $P < 0.05$ ) between leaf and flower metabolites (10%,  $|\rho| > 0.2$ ) could be observed, with the majority of them being negative (8.3%,  $\rho < -0.2$ ) (Figure 5.2). For example, strong negative correlations ( $-0.37 < \rho < -0.2$ ) between dihydroxybenzoic acid-xyloside III in the leaves and the majority of glucosinolates in the flowers were observed. Negative correlations were also detected between different flavonoids in leaves and flowers: quercetin-3-O-hexoside in the leaves was negatively correlated with kaempferol 3-O-glucoside, kaempferol-deoxyhexoside and kaempferide 3-glucoside in the flowers ( $-0.27 < \rho < -0.23$ ). On the other hand, the leaf quercetin-3-O-hexoside was positively correlated with 1-methoxy-3-indolylmethyl glucosinolate in the flowers. This illustrates the metabolic separation in tissue types and their functionally different roles in the life cycle of the plant demanding distinct phytochemical profiles. The wide range of quantitative variation in metabolites between the WT Col-0 and *ddm1-2* parents of the population as well as between epiRIL individuals further suggests that the methylation status is important for tissue specific metabolic control.

### 5.2.2 Site-specific differential methylation explains qualitative and quantitative metabolic variation

To gain deeper insight into the regulation of plant metabolism within the epiRIL population, QTL<sup>epi</sup> analysis was performed on all metabolites using a genetic map based on differentially methylated regions (DMRs) as physical markers [263]. In total, 34 QTL<sup>epi</sup> were identified for 30 different metabolites (Figure 5.3). Significant LOD scores of the detected QTL<sup>epi</sup> ranged from 2.49 for the unidentified compound #1537 to 4.81 for 1-methoxy-3-indolylmethyl glucosinolate, both detected in flowers and explaining 11.7% and 21.4% of the variation, respectively. The widespread quantitative and qualitative variation that was detected in the epiRILs was reflected in the detected QTL<sup>epi</sup>. For example, QTL<sup>epi</sup> were identified for two metabolites (compound #1619 in the flowers and quercetin-3-O-hexoside in the leaves) that showed qualitative variation such that the metabolite was only detected in the *ddm1-2* parent and a number of epiRILs but not in WT Col-0. Likewise, QTL<sup>epi</sup> were detected for seven metabolites for which a substantial difference (at least two-fold) between the two parents of the population could be observed. Finally, QTL<sup>epi</sup> could be detected for twenty-three metabolites that showed no clear distinction between the two parents, indicating that transgressive segregation of the epigenetic markers within the population is probably responsible for the mapped metabolic variation in these epiRILs.

Out of the thirty-four QTL<sup>epi</sup>, ten QTL<sup>epi</sup> were detected in the leaves and twenty-four in the flowers. The epigenetic variation resulted in increased or decreased metabolite content depending on the metabolite and the tissue. Sixteen of the 34 QTL<sup>epi</sup> displayed a negative effect sign, representing an increase in metabolite content between 4 and 41% in the *ddm1-2*-inherited epigenotypes. This was true for nine of the ten QTL<sup>epi</sup> detected for leaf metabolites, while this was the case for only eight of the twenty-four QTL<sup>epi</sup> detected in the flowers. All glucosinolate QTL<sup>epi</sup> displayed positive effect signs, resulting in increased glucosinolate levels in lines with the Col-0 inherited epigenotype. These analyses illustrate that the observed metabolic variation among the epiRILs can at least partly be explained by methylation variation at DMRs.

### 5.2.3 Epigenetic variation exerts pleiotropic effects on molecular and morphological traits

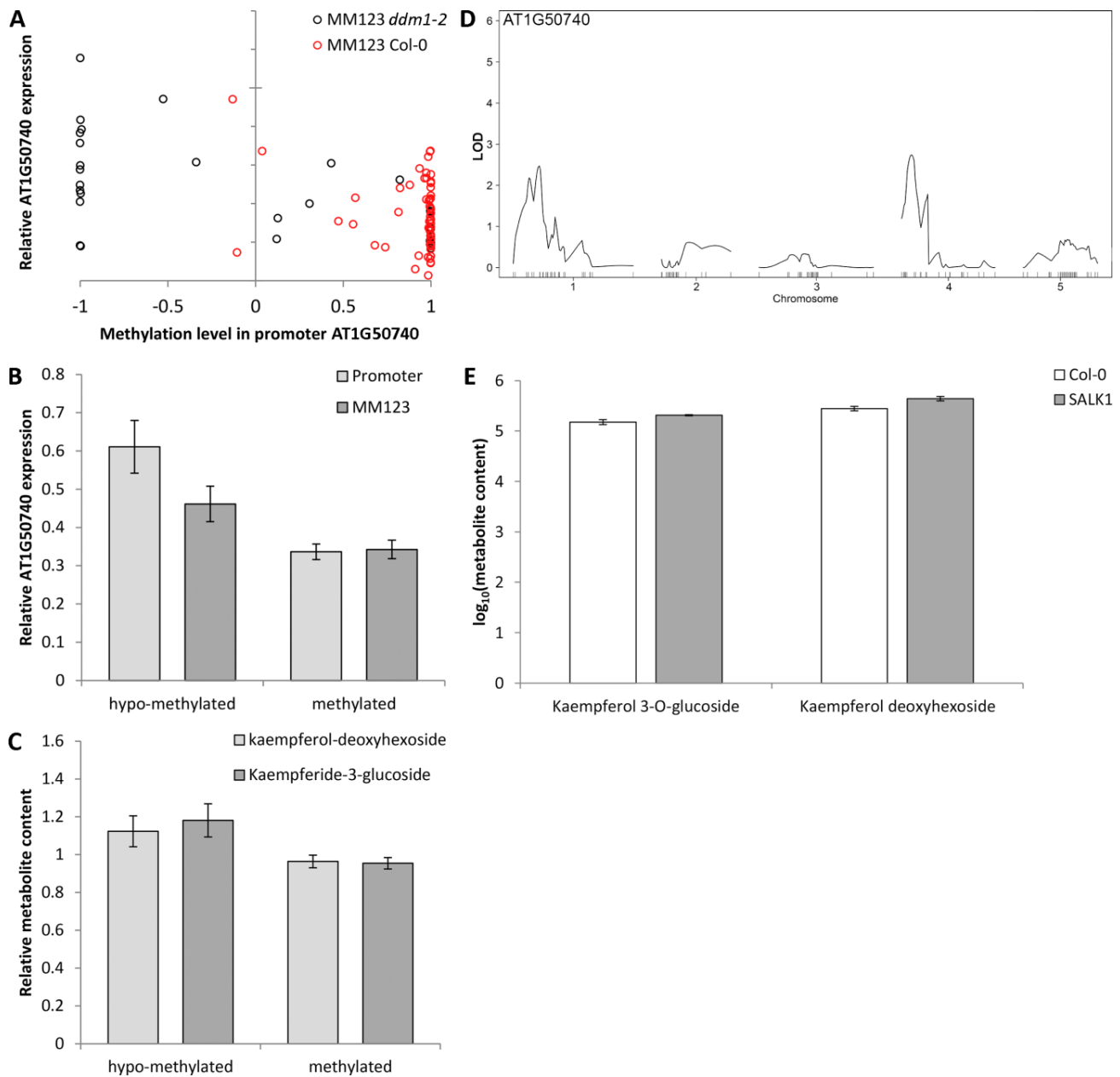
Twenty-one different QTL<sup>epi</sup> regions could be assigned, divided over the five chromosomes with many coinciding QTL<sup>epi</sup> (Figure 5.3). One QTL<sup>epi</sup> region was shared between leaf and flower metabolites, while five regions were specific for leaf metabolites and fifteen for flower metabolites. For most of the annotated compounds, QTL<sup>epi</sup> could only be detected in flowers although for quercetin-3-O-hexoside (a flavonoid) and D-gluconic acid, QTL<sup>epi</sup> could only be detected in leaves. For dihydroxybenzoic acid xyloside III, different QTL<sup>epi</sup> were identified in leaves compared to flowers, and it thus indicates differential metabolic regulation between tissues. For variation in the accumulation of flavonoids, QTL<sup>epi</sup> were detected on both chromosome 1 and 4. QTL<sup>epi</sup> explaining the variation in glucosinolate accumulation were detected at several positions in the genome and in two cases variation in different glucosinolates was found to be associated with the same genomic region. An QTL<sup>epi</sup> on chr 4 explained variation for (4- or 5-)hydroxy-3-indolylmethyl glucosinolate and 4-methylsulfinylbutyl glucosinolate, while a second QTL<sup>epi</sup> on chr 5 explained variation in 4-methylthiobutyl, 8-methylthiooctyl and 4-methylthiohydroxybutyl glucosinolates. Altogether, these QTL<sup>epi</sup> analyses suggest that epigenetics plays a significant role in regulating the tissue-specific accumulation of secondary metabolites.

Interestingly, the metabolic QTL<sup>epi</sup> identified in this study overlapped with the morphological QTL<sup>epi</sup> that were analyzed in the same experiment [253] and with morphological QTL<sup>epi</sup> detected in a previous study [251] (Figure 5.3). Twelve pleiotropic QTL<sup>epi</sup>-regions were detected, divided over the five chromosomes but with especially strong pleiotropic loci in the middle of chr 1 and 4, the start of chr 4 and the middle and lower arm of chr 5. The majority of metabolites for which a QTL<sup>epi</sup> could be detected strongly correlated with the morphological traits that mapped to the same regions. For instance, leaf area correlated significantly with both the unknown metabolite #1584 ( $r = -0.24$ ) and 7-methylthioheptyl glucosinolate ( $r = 0.22$ ) which all mapped to the same region on chr 1 (Figure 5.3). Furthermore, plant height at 1<sup>st</sup> silique correlated significantly with two as yet unknown metabolites #1443 ( $r = -0.20$ ) and #1438 ( $r = -0.21$ ) that mapped to the same region on chr 1 (Figure 5.3). For another QTL<sup>epi</sup> on chr 1, a similar observation was made as two flavonoids

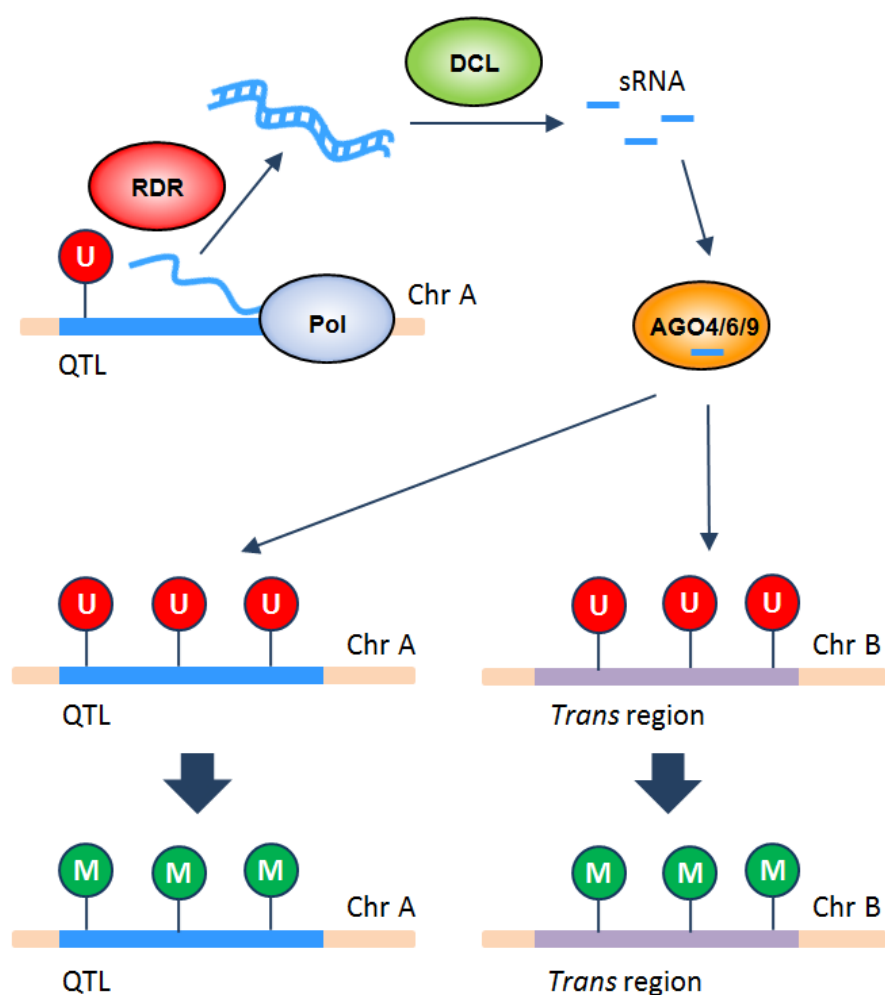
(kaempferide-3-glucoside and kaempferol deoxyhexoside) correlated significantly with both main stem branching ( $r = -0.33$  and  $-0.35$ , respectively) and average internode length ( $r = 0.22$  and  $0.20$ , respectively) (Figure 5.3). The highest correlation was detected between flowering time and kaempferol-deoxyhexoside ( $r=-0.43$ ) while both mapped to the same DMR on chr 1 (Figure 5.3). This suggests that these metabolites are strongly connected to the morphological traits and that they might be regulated by the same epigenetic mechanisms.

#### **5.2.4 Regulation of secondary metabolism in *cis* by epigenetic variation in biosynthesis genes**

To investigate epigenetically regulated candidate genes involved in secondary metabolism, we focused our attention on variation in glucosinolate and flavonoid content of the flowers, because the metabolic pathways of these metabolites are well defined and most of the QTL<sup>epi</sup> (14 out of 15) for flavonoids and glucosinolates were detected in the flowers. To narrow down the list of candidate genes, we focused on the known structural genes, regulatory genes and modifying enzymes of the glucosinolate and flavonoid pathways. Therefore, sixty-seven candidate genes were selected within the 1.5 LOD QTL<sup>epi</sup> confidence intervals, based on their involvement in glucosinolate and/or flavonoid metabolism according to the TAIR, ARACYC and KEGG databases ([www.arabidopsis.org](http://www.arabidopsis.org), [www.plantcyc.org](http://www.plantcyc.org) and [www.genome.jp/kegg](http://www.genome.jp/kegg)). To gain confidence in assigned candidate genes we next submitted each gene to a series of strict selection criteria. For all 67 genes, differentially methylated regions (DMRs) in the promoter, gene body and 1kb downstream of the candidate gene in the epiRILs were associated to their metabolic trait values. For 27 out of 67 genes, significant ( $P < 0.05$ ) associations were detected between methylation state and metabolic level. Because methylation states can be gained and lost, independent of the crossing scheme, it was investigated whether the methylation state at the DMR was in linkage with the most significant marker from the QTL<sup>epi</sup> study to determine whether the DMR of the candidate gene can explain the QTL<sup>epi</sup>. This was the case for 17 of the 27 remaining genes. From these 17 genes, we selected 9 candidates based on the relationship between gene function and metabolite pathway, position of the DMRs (promoter > gene body > downstream), presence of TEs close to DMRs and gene expression variation between



**Figure 5.4. Confirmation analyses for AT1G50740.** Methylation variation in the promoter of AT1G50740 causes variation in gene expression and metabolite content. (A) Scatterplot indicating the correlation between methylation at promoter and gene expression of AT1G50740 in all epiRILs. Red circles indicate epiRILs with wild-type allele at the DMR MM123, black circles indicate epiRILs with *ddm1-2* allele at DMR MM123. (B) Histograms indicating the association of the methylation level at the promoter of AT1G50740 (light grey) and DMR MM123 (dark grey) with the relative gene expression of AT1G50740. (C) Histograms indicating the association of the methylation level at promoter of AT1G50740 with the relative metabolite content of kaempferol-deoxyhexoside (light grey) and kaempferide-3-glucoside (dark grey). Hypo-methylated indicates a methylation level between -1 and -0.3, methylated indicates a methylation level between 0.3 and 1. (D) eQTL<sup>epi</sup> analysis for AT1G50740 in epiRILs. (E) Variation in metabolite content of kaempferol-3-O-glucoside and kaempferol deoxyhexoside in wild-type Col-0 and AT1G50740 knock-out mutant SALK1.



**Figure 5.5. Theoretical model for the regulation of DNA methylation by differential targeting of sRNA to loci in *trans*.** Changes in DNA methylation can be induced directly by differential recruitment of components of the RdDM pathway, or indirectly by post-transcriptional silencing of genes. DCL: dicer; M: methylated; Pol: RNA polymerase; RDR: RNA-dependent RNA polymerase, U: unmethylated.

Col-0 and *ddm1-2* in publicly available data [166, 229].

To determine whether the epigenetic variation was associated with variation in gene expression, Q-PCRs were performed on these nine genes in all epiRILs. Only one gene, AT1G50740, displayed a significant effect of both the DMR marker and the methylation levels around the gene on the gene expression levels ( $P < 0.05$ ) (Figure 5.4A and B). Specifically, 9 DMRs in the promoter region of AT1G50740 were significantly associated with variation in gene expression and the metabolic levels of 2 flavonoids that were associated with the QTL<sup>epi</sup> (Figure 5.4C). When the promoter of AT1G50740 is demethylated, the relative expression of the encoding gene and the metabolite content of the 2 flavonoids, kaempferol-deoxyhexoside and kaempferide-3-glucoside, is significantly increased. Although

linkage tests showed a significant effect of the DMR marker on gene expression levels ( $P < 0.05$ ), a *cis*-expressionQTL<sup>epi</sup> (eQTL<sup>epi</sup>) did not surpass the more stringent QTL<sup>epi</sup> significance threshold (Figure 5.4D). Two suggestive eQTL<sup>epi</sup> however, one close to the DMR marker on chr 1 and one on chr 4, almost passed the significance threshold, suggesting that the expression of AT1G50740 might be affected by the methylation state at these two loci (Figure 5.4D). AT1G50740 is a transmembrane protein possibly involved in defense responses and in the regulation of flavonoid biosynthetic processes [264]. To confirm the involvement of AT1G50740 in secondary metabolism a knock-out mutant was analyzed using untargeted LC-MS profiling and deep phytochemical phenotyping. Indeed, the comparison of the knock-out mutant of AT1G50740 with the Col-0 WT revealed strong effects of this gene on the levels of several flavonoids, oxylipins and 4-hydroxy-3-indolylmethyl glucosinolate (data not shown). Illustratively, the levels of kaempferol-deoxyhexoside and kaempferol-3-glucoside were doubled in the mutant compared to Col-0 wildtype (Figure 5.4E). The effects of the mutation are partly counter-intuitive, as reduced rather than increased flavonoid levels were expected upon inhibition of gene expression. Nevertheless, these findings strongly indicate that methylation in the promoter of AT1G50740 regulates gene expression and flavonoid content. Furthermore, it implies that epigenetics can play an important role in the accumulation of secondary metabolites, particularly in the specialized Arabidopsis flowers.

### 5.2.5 Regulation of secondary metabolism and plant morphology in *trans* by epigenetic variation in putative small RNA

Exploratory analyses revealed that the peak markers in the QTL<sup>epi</sup> intervals are also strongly associated with DNA methylation states at promoter regions of 324 genes in *trans*, and that these associations are not simply a signature of polygenic or epistatic selection during epiRIL inbreeding (data not shown). One molecular model that could explain these associations is that TE or repeat-associated DMRs in QTL<sup>epi</sup> intervals lead to the differential production of small RNA that affects DNA methylation maintenance at loci in *cis* but also possibly in *trans* via the canonical or non-canonical RNA directed DNA methylation (RdDM) pathways [265]. Differential targeting of sRNA to loci in *trans* could induce DNA methylation changes either

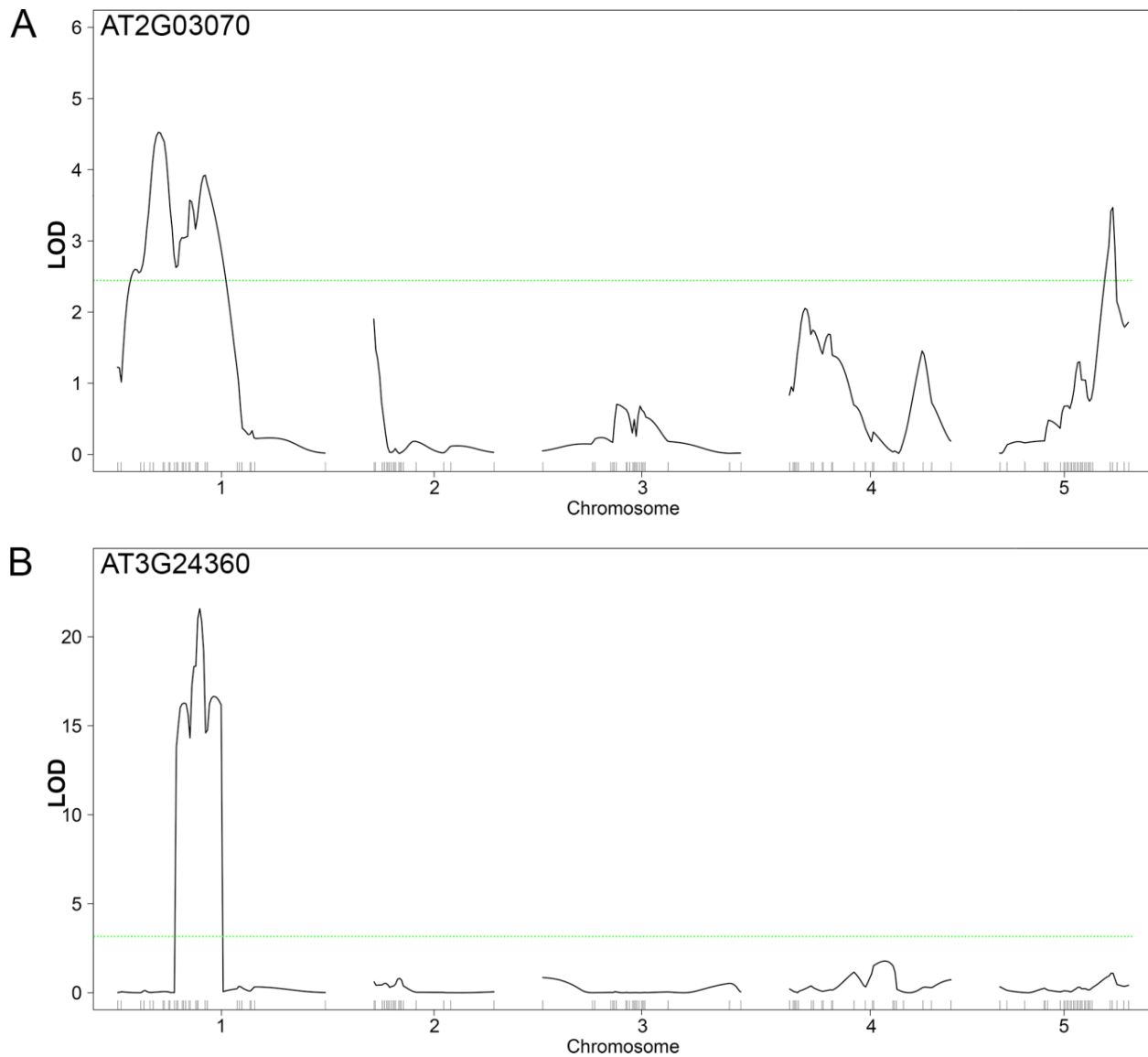
directly by altering the recruitment of components of the RdDM pathway, or indirectly by post-transcriptional silencing of genes flanking the *trans*-target loci (Figure 5.5).

Although this hypothetical mechanism is difficult to validate experimentally, a key requirement is that regions in the QTL<sup>epi</sup> intervals have sequence similarity with their putative *trans*-targets and that these target sequences match functional sRNA. To evaluate this, we searched the promoters of the 324 genes whose methylation levels correlated with those of the peak QTL<sup>epi</sup> marker for segments sharing perfect similarity with DNA regions inside their associated QTL<sup>epi</sup>. These regions were then decomposed, *in silico*, into sets of artificial sRNA (artsRNA) with a length in the range 21-24 nt to simulate candidate sRNA sequences that can map to the gene and the QTL<sup>epi</sup> interval. Selected artsRNA were then submitted to the SAILS computational framework [193] to predict whether these artsRNA have the necessary sequence properties to load into plant Argonaute (AGO) proteins 4/6/9, which are known to be involved in transcriptional silencing in plants [266]. artsRNA that had low probability for AGO-loading were discarded. The remaining artsRNAs were matched to true sRNAs from wild-type (WT) and *ddm1* sRNA libraries [83] to obtain further evidence that support these segments as real sRNA.

Thirteen of the potential artsRNA target genes were further analyzed for gene expression variation in the flower heads of the epiRIL population using Q-PCR. We used the expression levels of these genes as unique molecular phenotypes and performed a genome-wide search for expression epigenetic QTL (eQTL<sup>epi</sup>). Three of the 13 genes (AT3G24360, MED8 (AT2G03070) and AT2G16835) were significantly associated with one or multiple eQTL<sup>epi</sup>. For AT3G24360 and MED8, the detected eQTL<sup>epi</sup> mapped to regions inside the *trans*-QTL<sup>epi</sup> interval that contain artsRNA predicted to mediate transcriptional silencing (504 for AT3G24360 and 1 for MED8). Especially, the eQTL<sup>epi</sup> for AT3G24360 on chr 1 was highly significant (LOD > 20). Following this approach, we thus established a link between methylation variation in small RNAs and *trans*-genes and their level of expression (Figure 5.6A and B).

Interestingly, we found that MED8 and AT3G24360 contain TEs in their promoters and that all artsRNA originating from the QTL<sup>epi</sup> interval are complementary to TEs or their flanking sequences (< 1000 bp). In the case of AT3G24360, TEs from VANDAL families are found in all candidate regions targeted by artsRNA. From the 33 homologous sites detected, 24 (72.7%) co-localize with transposons of the VANDAL2 family and the remaining 9 (27.3%) with

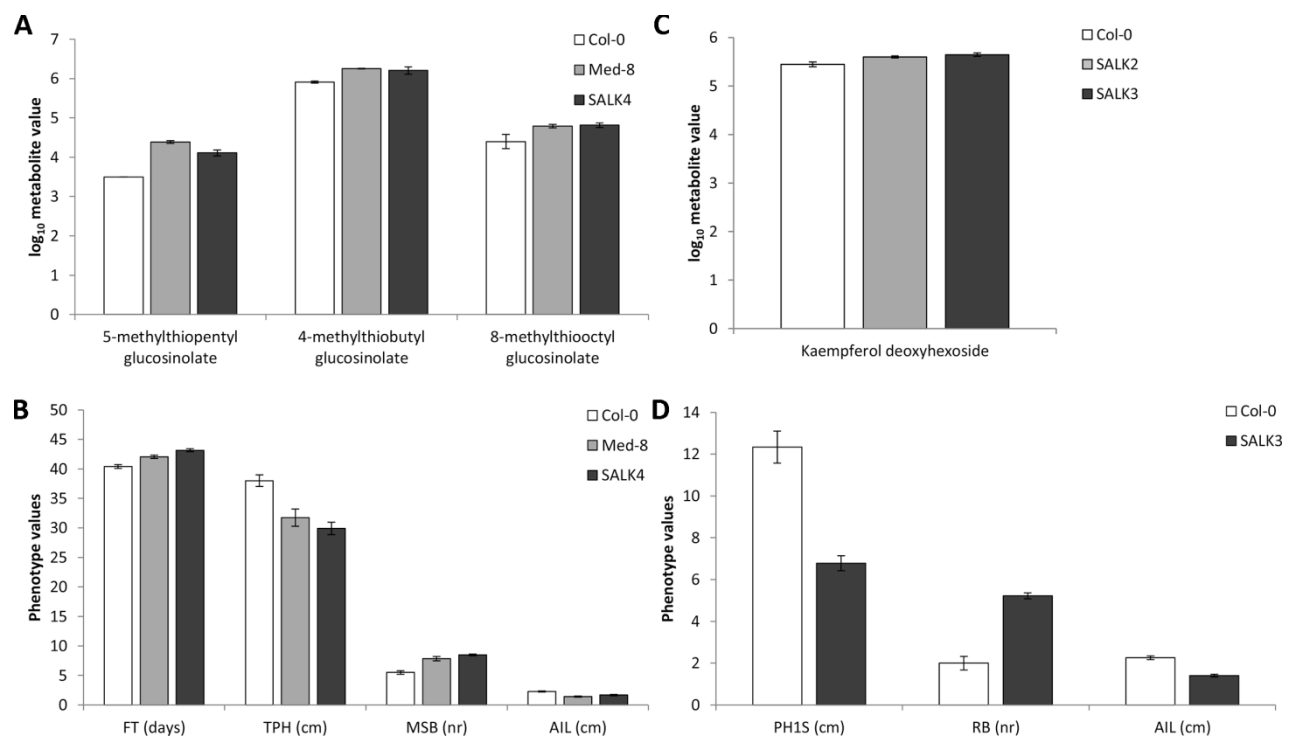




**Figure 5.6. Expression QTL analysis in epiRILs: (A) AT2G03070 and (B) AT3G24360.** RNA was extracted for 93 epiRILs, reverse transcribed to cDNA and quantified by SYBR Green qPCR. Gene expression was normalized against the reference gene TIP41 and subjected to eQTL<sup>epi</sup> analyses. The horizontal line indicates the LOD significance threshold that was calculated using 1000 random permutations with  $\alpha=0.05$  as the genome-wide type 1 error level. Marker positions are indicated on the bottom of the graph.

VANDAL2N1 members. VANDAL transposons from the MuDR superfamily in maize have been shown to modulate the expression of genes through epigenetic mechanisms [267]. To illustrate that loss of expression affects metabolic and morphological traits, two independent KO mutants for each of the two genes AT3G24360 and MED8 were analyzed by untargeted LC-MS and their metabolite profiles compared with that of the Col-0 wildtype.

Both MED8 mutants were significantly altered in the levels of 70 (KO1) and 64 (KO2) metabolites, respectively. The accumulation of the majority of these metabolites was affected in both KOs and when the metabolite levels of the two mutants were compared to Col-0 WT and the other mutants, 99 metabolites, among which various glucosinolates and flavonoids, were significantly altered ( $P < 0.05$ ) (Figure 5.7A). Although this QTL was initially detected based on differential levels of 7-methylthioheptyl glucosinolate, the levels of similar aliphatic glucosinolates were significantly increased in the flowers of both KO mutants (Figure 5.7A). Interestingly, flowering time and main stem branching were significantly increased in the mutants as well, while total plant height and average internode length were significantly reduced (Figure 5.7B). For AT3G24360, the two KO mutants significantly altered the levels of 98 (KO1) and 14 (KO2) metabolites, respectively, among which were several oxylipins, flavonoids and glucosinolates. Both KOs showed significantly increased levels for kaempferol deoxyhexoside, agreeing with the sign of effect for which the initial QTL was detected (Figure 5.7C). When the metabolite levels of both mutants were tested jointly against the Col-0 wildtype, kaempferol deoxyhexoside was the second most significantly different metabolite between mutants and wildtype. For KO1, we also observed many morphological differences compared to the wild-type. Plant height and average internode length were significantly reduced, while rosette branching was significantly increased (Figure 5.7D). The more severe effects of KO1, compared to KO2, might be due to differences in the insertion site of the T-DNA. KO1 was inserted in the promoter of AT3G24360 and covers a large part of the first exon, which strongly suggests that the gene has lost its entire function. The second KO, however, was introduced in an intron and it covers part of the very small third exon, suggesting that although heavily impaired it might still function. In addition, using the recently released methylation data from the 1001 Genomes Consortium [268], variation could be observed in the level of methylation in the promoters of the three candidate genes in natural accessions. The promoters of At1G50740 (*cis*-QTL<sup>epi</sup>) and AT2G03070 (*trans*-QTL<sup>epi</sup>) are clearly demethylated in a small subset of the natural accessions, while the promoter of AT3G24360 (*trans*-QTL<sup>epi</sup>) is demethylated in the majority of natural accessions, but methylated in a small subset. These findings indicate that variation in the level of methylation may play a role in natural settings as well.



**Figure 5.7. Metabolic and morphological trait analyses in mutants and wild-type.** (A) Metabolic values for 3 different glucosinolates (4-methylthiobutyl glucosinolate, 5-methylthiopentyl glucosinolate and 8-methylthiooctyl glucosinolate) in Col-0 wild-type and two *AT2G03070* mutants *med8* and SALK4. (B) Phenotypic trait values for flowering time (FT), total plant height (TPH), main stem branching (MSB) and average internode length (AIL) in col-0 wild-type and two *AT2G03070* mutants *med8* and SALK4. (C) Metabolic values for kaempferol deoxyhexoside in Col-0 wild-type and two *AT3G24360* mutants SALK2 and SALK3. (D) Phenotypic trait values for plant height 1<sup>st</sup> silique (PH1S), rosette branching (RB) and average internode length (AIL) in col-0 wild-type and two *AT3G24360* mutants SALK2 and SALK3.

### 5.3 Discussion

The findings presented here strongly indicate that epigenetics is at least partly involved in the regulation of plant (secondary) metabolism in both leaves and flowers of *Arabidopsis*. It must be noted, however, that the number and strength of QTL<sup>epi</sup> was considerably lower than the detected metabolic QTL in genetic studies on classical RILs [244, 269]. Although the populations that were used in those studies were substantially larger, which enhances the power to find QTL, the results strongly indicate that genetic variation is a much larger source for metabolic variation than epigenetic variation. Nonetheless, the epigenetic control of secondary metabolite content in terms of number and strength of QTL<sup>epi</sup> was much stronger in flowers than in leaves. Flowers, as reproductive organs, are important plant tissues in terms of fitness and should thus be well protected [270].

DNA methylation variation, as was also observed in natural accessions [268], is evidently the most likely reason for the observed phenotypic variation in the epiRILs. We provide evidence that methylation variation in candidate genes can either be due to methylation variation in *cis*, co-locating with the DMRs, or that methylation variation is regulated in *trans* through the activation or repression of small RNAs that target specific genes in RNA-directed DNA methylation (RdDM). Independent knock-out mutant and eQTL<sup>epi</sup> analyses further confirmed that methylation-directed loss of expression of candidate genes can cause significant variation in plant metabolism. Besides the variation in flavonoids and glucosinolates, we detected strong differences in the content of so-called Arabidopsides, such as dinor-oxophytodienoic acid monoacylglyceride (dn-OPDA-DGMG), dn-OPDA-DGMG derivatives and dn-OPDA-DGMG isomers, which contain esterified oxylipins and are precursors for the plant defense hormone jasmonic acid. The Arabidopsides may act as an important supply of oxylipins in plant defense and are strongly induced upon wounding [271, 272].

Intriguingly, the epigenetic variants underlying the QTL<sup>epi</sup> affect many phenotypic traits in parallel. The majority of QTL<sup>epi</sup> detected for morphological traits in control and stress conditions, phenotypic plasticity and secondary metabolism collapsed into 12 QTL<sup>epi</sup> regions [251, 253]. The master epigenetic regulators are most likely sRNA that became inactive through hypo-methylation in the F1 and contributed to the alteration of the methylation state at various loci in *trans*, which have maintained that state through meiosis. These sRNAs must be guiding most of the observed variation in phenotypic traits in the epiRIL population.

