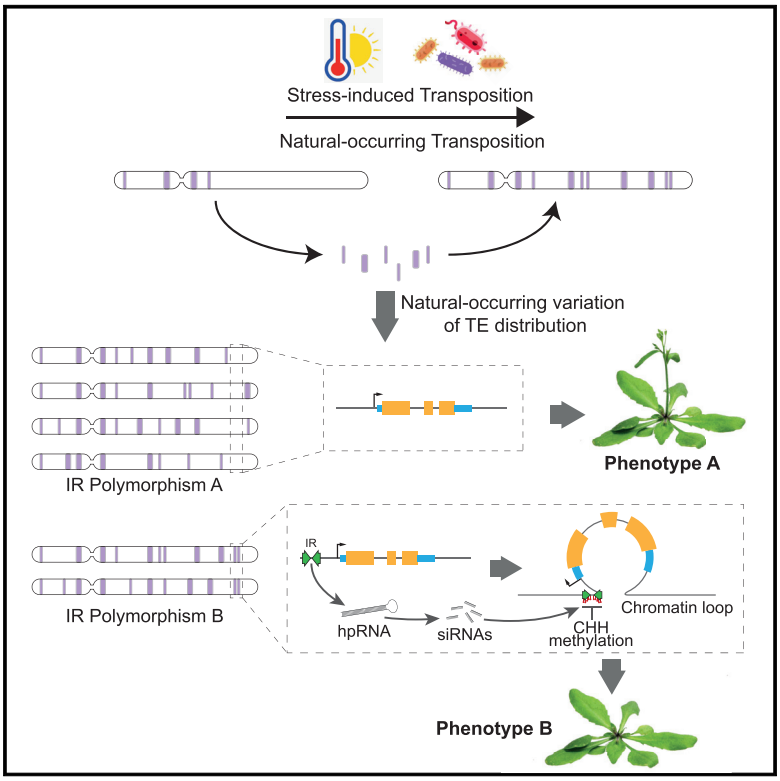
Cell Reports

Article

Polymorphic inverted repeats near coding genes impact chromatin topology and phenotypic traits in *Arabidopsis thaliana*

Graphical abstract



Highlights

- *Arabidopsis* genome contains hundreds of inverted repeats (IRs) near coding genes
- IRs produce small RNAs, triggering DNA methylation and changes in chromatin topology
- Polymorphic IRs cause natural variation in chromatin organization and gene expression
- IR editing alters short-range chromatin loops, gene expression, and linked phenotypes

Authors

Agustín L. Arce, Regina Mencia, Damian A. Cambiagno, ..., Hernán A. Burbano, Detlef Weigel, Pablo A. Manavella

Correspondence

pablomanavella@ial.santafe-conicet.gov.ar

In brief

Arce et al. show that insertion of an inverted repeat, generally a transposon, near a gene can cause changes in the local chromatin topology and host gene expression. Natural variation and gene editing validations of these elements suggest that IRs could be evolutionary agents during plant adaptation.

Arce et al., 2023, Cell Reports *42*, 112029
January 31, 2023
https://doi.org/10.1016/j.celrep.2023.112029

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Polymorphic inverted repeats near coding genes impact chromatin topology and phenotypic traits in *Arabidopsis thaliana*

Agustín L. Arce,[1,7] Regina Mencia,[1,7] Damian A. Cambiagno,[1,6] Patricia L. Lang,[2,5] Chang Liu,[4] Hernán A. Burbano,[2,3] Detlef Weigel,[2] and Pablo A. Manavella[1,8,*]

[1]Instituto de Agrobiotecnología del Litoral (CONICET-UNL), Cátedra de Biología Celular y Molecular, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, 3000 Santa Fe, Argentina
[2]Department of Molecular Biology, Max Planck Institute for Biology Tübingen, 72076 Tübingen, Germany
[3]Centre for Life's Origins and Evolution, University College London, London, UK
[4]Department of Epigenetics, Institute of Biology, University of Hohenheim, Garbenstraße 30, 70599 Stuttgart, Germany
[5]Present address: Department of Biology, Stanford University, Stanford, CA 94305, USA
[6]Present address: Unidad de Estudios Agropecuarios (UDEA), INTA-CONICET, Córdoba, Argentina
[7]These authors contributed equally
[8]Lead contact
*Correspondence: pablomanavella@ial.santafe-conicet.gov.ar
https://doi.org/10.1016/j.celrep.2023.112029

SUMMARY

Transposons are mobile elements that are commonly silenced to protect eukaryotic genome integrity. In plants, transposable element (TE)-derived inverted repeats (IRs) are commonly found near genes, where they affect host gene expression. However, the molecular mechanisms of such regulation are unclear in most cases. Expression of these IRs is associated with production of 24-nt small RNAs, methylation of the IRs, and drastic changes in local 3D chromatin organization. Notably, many of these IRs differ between *Arabidopsis thaliana* accessions, causing variation in short-range chromatin interactions and gene expression. CRISPR-Cas9-mediated disruption of two IRs leads to a switch in genome topology and gene expression with phenotypic consequences. Our data show that insertion of an IR near a gene provides an anchor point for chromatin interactions that profoundly impact the activity of neighboring loci. This turns IRs into powerful evolutionary agents that can contribute to rapid adaptation.

INTRODUCTION

Transposable elements (TEs) are widely distributed among eukaryotic genomes. In a process known as transposition, TEs move within the genome to different locations, usually copying themself as they "jump."[1] Plant genomes are particularly rich in TEs and repetitive elements, which, for example, account for 85% of the maize genome.[2] In plants, TEs are commonly silenced through DNA methylation in a process known as RNA-directed DNA methylation (RdDM), which maintains genome integrity.[3] To trigger RdDM, short RNA polymerase IV (Pol IV)-dependent TE transcripts are converted into double-stranded RNA (dsRNA) by the RNA-dependent RNA polymerase 2 (RDR2) and then to 24-nt small interfering RNAs (siRNAs) by DICER-like 3 (DCL3).[3,4] ARGONAUTE 4 (AGO4)-loaded siRNAs then direct *de novo* methylation of the TE loci by recognizing nascent Pol V transcripts there. Such methylation ultimately leads to nucleosome condensation and permanent silencing of the TE. Still, massive bursts of TE amplification have occurred in plant genomes in addition to the rarer but continued movement of individual elements.[5–7] Stress can trigger activation of TEs and fuel transposition.[8,9] TEs are thus significant contributors to genetic variation in plant genomes[8,10,11] and have been postulated as drivers of genome evolution and expansion as well as developmental plasticity and adaptation.[1,12] In addition, stress-induced mobilization of TEs generates a broad range of changes in gene expression, including the appearance of new transcript isoforms, that can be a substrate of natural selection.[13]

There are two main classes of TEs: DNA transposons and retrotransposons. The most abundant DNA transposons are miniature inverted-repeat TEs (MITEs), while the most abundant retrotransposons are long terminal repeat (LTR) retrotransposons.[1] MITEs exhibit characteristic terminal inverted repeats (TIRs) and small direct repeats (target site duplications [TSDs]) but lack transposase sequences, making them non-autonomous elements.[14,15] Many TE-derived inverted repeat (IR) insertions may not be classified as MITEs because they lack the above-mentioned components of MITEs, either because of deletions after insertion or because they were generated through a different process. MITEs are commonly situated near coding genes: for example, almost 60% of rice genes can be associated with a MITE,[7] with the MITEs often changing the expression of neighboring genes.[7,16–20] Based on this, MITEs have been proposed

to play important roles in genome evolution and gene expression.[7] One key feature of MITEs is that their transcripts can fold into hairpin-shaped dsRNAs because of the extensive sequence complementarity between IR arms. These dsRNA secondary structures are recognized and processed by DCL3 to produce 24-nt siRNAs that trigger DNA methylation without the need for Pol IV/RDR2 activity.[21–25] Thus, transcripts of these MITEs can be initiated from promoters of adjacent genes, triggering their RNA polymerase II (Pol II)-dependent DNA methylation. At the sunflower *HaWRKY6* locus, Pol II-mediated transcription of a MITE triggers methylation of its coding region and causes formation of alternative regulatory short-range chromatin loops in the locus that specifically change its expression.[25]

Three-dimensional chromatin organization has recently emerged as a critical feature in determining genome functionality, fine-tuning gene expression and developmental responses in plants.[26,27] Short-range chromatin loops reflect the interaction between relatively close regions of DNA, within a few kilobases, generally within a single locus or between adjacent loci.[28] Different from canonical regulation by cytosine methylation in the linear DNA sequence, commonly associated with gene repression, local three-dimensional chromatin organization can induce a plethora of regulatory mechanisms, including transcriptional activation/repression, transcription directionality, alternative splicing, usage of cryptic termination sites, impaired or enhanced Pol II elongation, and DNA replication and repair.[28–30]

In this study, we show that TE-derived IR elements located near genes in the *Arabidopsis thaliana* genome and mostly transcribed by Pol II can cause rearrangements of the chromatin topology, promoting formation of short-range chromatin loops. These chromatin interactions, which depend on production of IR-derived siRNAs and *de novo* DNA methylation, often translate into changes in gene expression. The presence of an IR and its associated chromatin loop near a gene does not cause a uniform regulatory effect and can either enhance or repress expression, depending on the locus and which regions within the locus are part of the loop.

Almost one-third of the identified gene-associated IRs are not conserved among a set of 216 *A. thaliana* natural accessions. Our data show that polymorphisms in IRs near genes can be coupled with a change in the chromatin topology of the region. These accession-related changes in chromatin landscapes controlled by IRs can be linked to alteration in traits commonly associated with adaptation, such as flowering. In proof-of-concept experiments, we used CRISPR-Cas9 genome editing to mimic the situation in natural accessions lacking specific IRs and demonstrated that the IRs help to shape chromatin topology, which, in turn, can control molecular and organismal phenotypes. We found that IRs downstream of *PHYC* and *CRY1* cause formation of repressive chromatin loops associated with well-defined developmental phenotypes of some natural accessions. Overall, our data demonstrate that TE-derived IRs can produce changes in chromatin topology, gene expression, and, ultimately, phenotypic changes through their capacity to trigger DNA methylation autonomously. Given the propensity of TEs to activate during stress responses and the tendency of IRs to locate near coding genes, our finding provides a scenario for these TEs to drive local adaptation and domestication by supporting rapid and sometimes drastic changes in 3D chromatin organization and gene activity after single-set mutational events.

RESULTS

TE-derived IRs located near genes produce siRNAs and trigger DNA methylation in a Pol II-dependent pathway

A TE-derived IR element located ~600 bp upstream of the *HaWRKY6* locus in the sunflower serves as an anchor point for formation of two short-range chromatin loops and thereby promotes changes in local chromatin topology.[25] The ultimate outcome is methylation of the locus because of 24-nt siRNAs produced after Pol II-dependent transcription of these IRs. In leaves, this leads to formation of a repressive loop while it promotes a second larger loop that enhances transcription in cotyledons.[25]

Because TE-derived IRs are frequently located near genes,[31] we wondered whether the *HaWRKY6* case was just one example of a more general phenomenon in plants. To evaluate whether insertion of an IR near a gene changes local chromatin topology and gene expression, we first aimed to identify all IRs neighboring protein-coding genes in the *A. thaliana* Col-0 reference genome. Using *einverted* from the European Molecular Biology Open Software Suite (EMBOSS),[32] we found a total of 885 IRs in the *A. thaliana* genome (Table S1), 634 of them near annotated protein-coding genes (222 of which were located within 500 bp upstream or downstream of a protein-coding gene, another 163 between 501 bp and 1,000 bp from a gene, and 249 between 1,001 and 3,000 bp from a gene) (Figure 1A). IRs were found within 500 bp of 260 unique genes, within 501–1,000 bp of 215 unique genes, and within 1,001–3,000 bp of 615 unique genes (Figure 1A). These IRs have a broad genome-wide distribution, with many located in the gene-rich chromosome arms and others in gene-poor/TE-rich pericentromeric regions (Figure 1B). The identified IRs represent a heterogeneous group of elements with a size ranging from 100 bp up to 1,000 bp (the length limit of our analysis; Figure S1A). Among them, we observed an overrepresentation of IRs of ~375 bp and ~1,000 bp (Figure S1A). These IRs also showed heterogeneous spacing between the repetitive elements (Figure S1B). We observed either a small spacer or near absence of any spacer (the hairpin loop) between the repeats for the smaller IRs (Figure S1B). Spacer region variability correlates with IR element length (Figure S1B). Analyzing the overlap with annotated TEs revealed that 68% of the IRs within 3,000 bp of a gene were clearly of TE origin, with most (~44%) from the MuDR superfamily and 18% from the Helitron superfamily (Figure 1C). These percentages remain invariable despite the distance from the IR to the hosting gene, although there is a moderate enrichment is LTR/Gypsy TEs between IRs not associated with genes (Figure 1C). Using plant native elongating transcripts sequencing (plaNETseq) datasets of Pol II-associated nascent transcripts,[33] we found that more than half of the identified gene-associated IRs are transcribed by Pol II (Figure 1D, dark blue and dark red). This becomes more evident for IRs closer to genes, with ~70% of those up to 500 bp away from a gene, which may indicate a stronger influence of gene-associated regulatory elements on IR transcription by Pol II. We observed that most Pol II-transcribed IRs produced siRNA

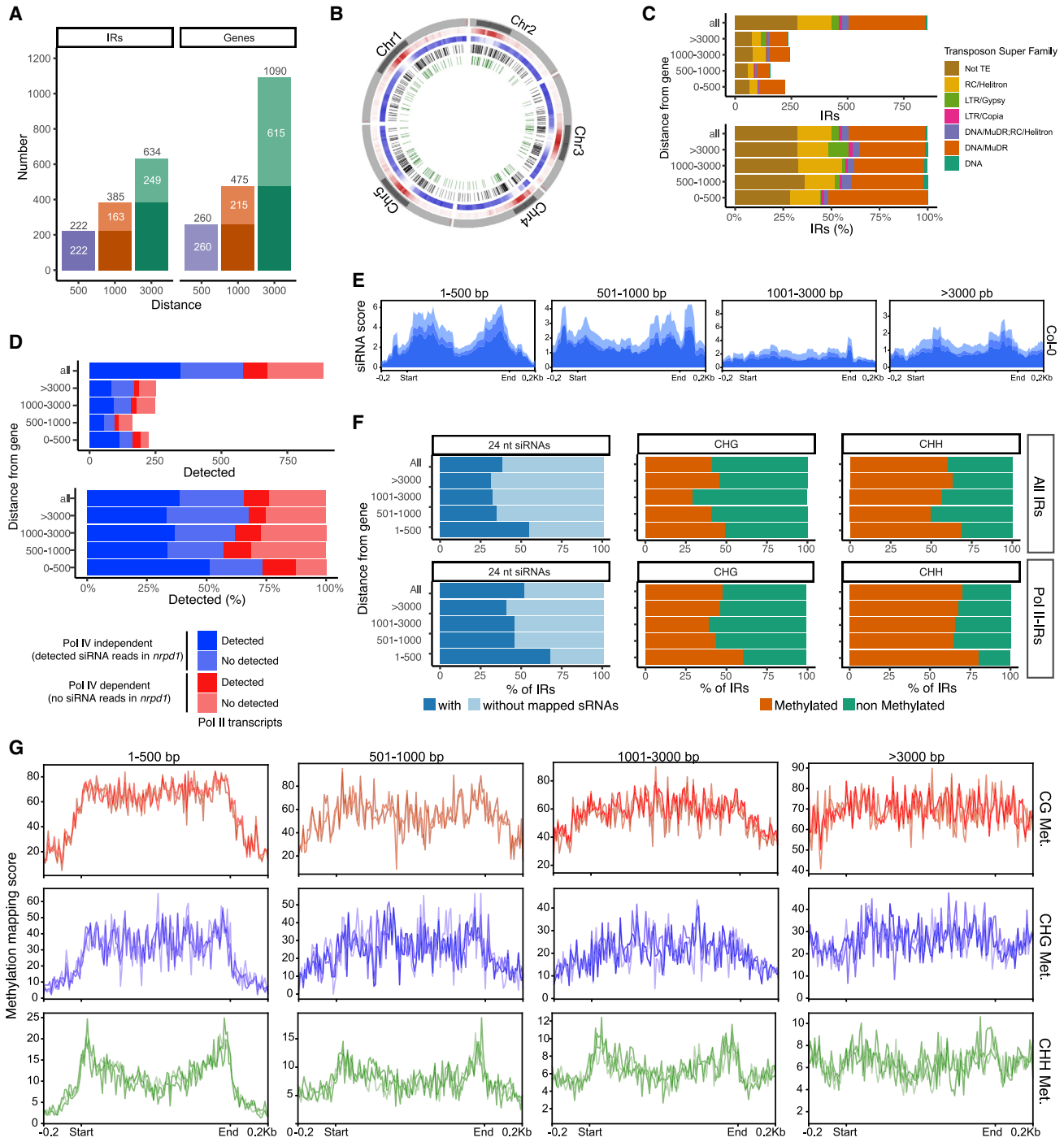


Figure 1. IRs distribution near annotated genes in the *A. thaliana* Col-0 reference genome

(A) Number of IRs detected in the Col-0 genome within 500, 1,000, and 3,000 bp of annotated genes (left) and number of associated genes within the same size window (right). Black numbers indicate the total IR and genes counts, while white numbers show the non-cumulative count.

(B) Distribution of IRs. The outermost track shows *Arabidopsis* chromosomes with the pericentromeric regions highlighted in darker gray. Shown are TE density (red) and gene density (blue) in 500-kb windows. Darker color indicates higher density. The inner two tracks show the distribution of all IRs (black) or only IRs within 3,000 bp upstream or downstream of annotated protein-coding genes (green).

(C) Classification of the identified IRs between annotated TE superfamilies. The "DNA/MuDR; RC/Helitron" category indicates IRs overlapping with TEs from both superfamilies. Shown is the total number (top) and percentage (bottom) of IR in each superfamily identified in each distance window from annotated genes.

(legend continued on next page)

even in the absence of Pol IV[34,35] (Figure 1D, dark blue bars, and Figure S1C), especially IRs close to genes. This observation confirms that these 24-nt siRNAs follow a non-canonical pathway and require Pol II transcription. Pol II dependency becomes even more evident when one also considers IRs producing siRNAs in the absence of Pol IV but that are undetected in Pol II plaNETseq analysis (Figure 1D, light blue bars). These likely represent either tissue/condition-specific Pol II transcripts or reflect the limitations of the plaNETseq tool to detect all Pol II transcripts. Supporting the idea that these IRs are transcribed from promoters of nearby genes, the fraction of IRs transcribed by Pol II increases the closer the IRs are to a gene (Figure 1D). Conversely, IRs located far from annotated genes were less likely to give rise to Pol II transcripts. Instead, their transcription likely depends only on the canonical Pol IV/RDR2 RdDM pathway (Figure 1D). We also observed a reduction in small RNA (sRNA) production in some Pol II-transcribed IRs in Pol IV mutants, although less pronounced that for Pol II-independent IRs (Figure S1C). This observation likely indicates that many Pol II transcribed IR may also follow a canonical Pol IV pathway. Whether this is a tissue/condition specific effect leading to Pol II transcription of IRs only in specific moments/tissues remains a compelling open question.

We then used sRNA and bisulfite (BS) sequencing to determine whether these IRs, especially those transcribed by Pol II and independent of the canonical Pol IV pathway, produced 24-nt siRNAs and whether production of siRNAs leads to DNA methylation. Metagene analysis of 24-nt siRNAs mapping to the IRs showed a two-peak profile, indicating that the siRNAs are produced from the IR sequences, especially for IRs closest to genes (Figure 1E). Many of the identified IRs, especially IRs located closer to genes, produced Pol II-dependent siRNAs (Figure 1F). This observation further supports the hypothesis that these siRNAs are produced from the dsRNA stem region of long Pol II transcripts folded into hairpins rather than the short and homogeneously distributed Pol IV transcripts. The BS sequencing (BS-seq) analysis revealed that most Pol II-dependent IRs, especially those closest to genes, are methylated preferentially in the CHH context (H = C, T, or A), the signature of the RdDM pathway (Figure 1F). The fraction of methylated IRs (Figure 1F) and the methylation profile obtained from the metagene analyses, with a clear profile of two peaks for CHH methylation (Figure 1G), matched the patterns observed for siRNA production (Figures 1E and 1F). In addition to methylation of IR loci, we often observed an additional peak of siRNA mapping either on the opposite border or inside many genes located near IRs (Figures S1D and S1E). These secondary mapping spots are consistent with a scenario where two adjacent methylated regions serve as anchor points for formation of short-range chromatin loops.

TE-derived IRs located near genes impact the local chromatin topology in *Arabidopsis*

To investigate whether the identified IRs, especially those producing siRNAs and DNA methylation, impact local chromatin organization, we extracted RNA and DNA from Col-0 wild-type plants; triple *dcl2, dcl3, dcl4* (*dcl234*) mutants (which are in impaired in 24-nt siRNA production);[36] and triple *drm1, drm2, cmt3* (ddc) mutants (impaired in CHH methylation)[37] and performed RNA sequencing (RNA-seq), sRNA sequencing (sRNA-seq), BS-seq, and Capture-C sequencing (CapC-seq) (statistics of the sequencing data are provided in Table S2). Because the sequencing depth required to detect short-range chromatin interaction through standard Hi-C (high-resolution chromosome conformation capture) would be enormous, we selected 290 loci containing IRs and 38 control loci (Table S3) and performed a CapC experiment to focus only on these regions and increase the chances to detect local chromatin loops. A test mapping of CapC reads confirmed the enrichment on the captured regions compared with the input Hi-C samples (Figure S2A). Collectively, the designed probes cover ~1% of the genome. After CapC, we increased the ratio of reads mapping the targeted regions 40 times on average (Figure S2B).

sRNA-seq revealed that 486 of the 634 IRs that are at most 3,000 bp from genes produced 24-nt siRNAs, with siRNA levels and mapping profiles severely reduced in *dcl234* and *ddc* mutants, as expected for this class of sRNAs (Figures 2A and 2B). The IR-derived siRNAs, whose generation requires DCL activity, appeared to be independent of IR size (Figure S2C).

CHG and CHH methylated regions strongly overlapped with Pol II-dependent IRs within 500 bp from genes (Figure 1F). Within the IR regions, we observed a drastic drop in DNA methylation, especially in the CHH context, in *dcl234* and *ddc* mutants (Figures 2C, S2D, and S2E). Such a reduction of CHH and, to some extent, CHG, but not of CG methylation in *dcl234* and *ddc*, a typical signature of RdDM mutants, paralleled the drop in siRNA production in these regions (Figure 2D). The proportion of IRs with reduced methylation in the mutants is slightly higher for those within 500 bp of genes compared with the other analyzed distance windows (Figure S2D). Changes in siRNAs and methylation were correlated, consistent with a drop of siRNAs leading to reduced methylation in RdDM-impaired mutants. Altogether, these data suggest that a large proportion of IRs located near genes and transcribed by Pol II trigger DNA methylation in *cis* through the non-canonical, Pol IV-independent RdDM pathway.

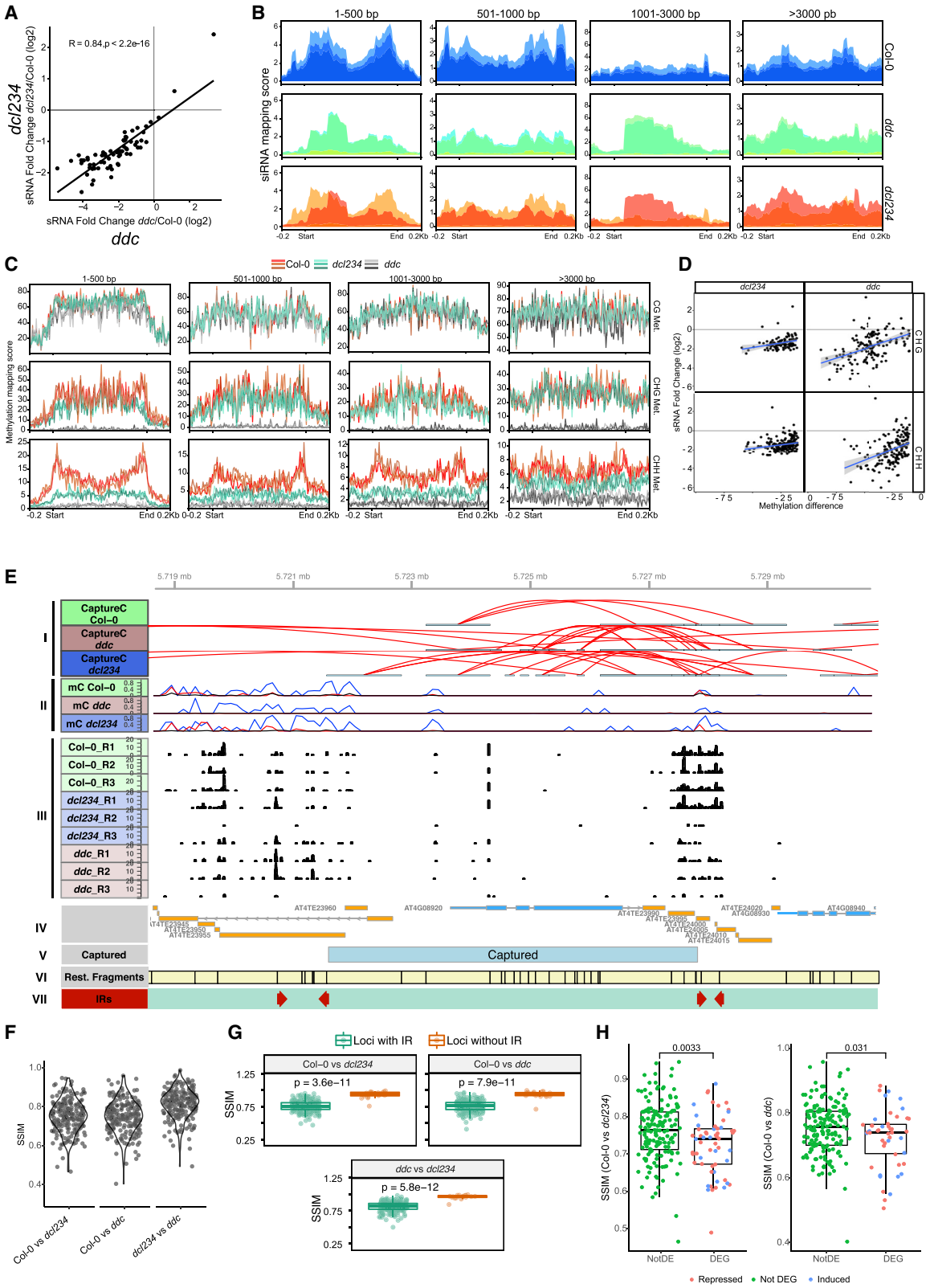
(D) Total number (top) or percentage (bottom) of IRs with mapping 24-nt siRNAs detected (blue) or not detected (red) in *nrpd1* mutants, defining Pol IV-dependent and independent IRs. For each case, IRs producing transcripts associated with Pol II, as detected by plaNETseq[33], are marked in dark colors and IRs without Pol II transcripts in light colors. All IRs are sorted depending on their range of distance to the closest protein-coding gene.

(E) Metagene analysis of 24-nt siRNAs mapping to IRs located at different ranges of distance to the closest protein-coding gene. Plots show IRs scaled from the start to the end of the IR regions plus 200 bp to each side. Three individual sRNA-seq replicates are plotted separately.

(F) Percentage of gene-adjacent IRs overlapping methylated regions, CHG (center), and CHH (right) with mean methylation levels of at least 10% and with 24-nt siRNAs mapping to the IR sequence (at least 10 reads in at least one replicate). The analysis was performed for all identified IRs (top row) or only for Pol IV-independent IRs (bottom row).

(G) Metagene analysis of CG (red), CHG (blue), and CHH (green) DNA methylation at IRs located at different ranges of distance to the closest protein-coding gene. Plots show IRs scaled from the start to the end of the IR regions plus 200 bp to each side. Three individual BS-seq replicates are plotted.

CellPress
OPEN ACCESS



(legend on next page)

To assess whether the IRs located near genes can act as regulatory elements by altering the surrounding chromatin organization, we exploited our CapC-seq data and the software CHESS (Comparison of Hi-C Experiments using Structural Similarity)to compare the structural similarity (SSIM) of the IR-hosted regions between the wild-type and mutants.[38] We also statistically determined the specific anchor points of chromatin loops in each sample using CapC-Map[39] in combination with peakC.[40] Analyzing individual loops, we found clear alterations in the chromatin topology in several randomly picked loci (Figures 2D and S3). It is important to note that we cannot detect the precise position of the anchor points of a chromatin loop using CapC-seq or chromatin conformation capture (3C) experiments. Instead, these approaches define ligated restriction fragments (Figure 2E, tracks I and VI) within which such anchor points are contained. To compare the differential methylation-related changes in loops formed near IRs, we calculated the SSIM for each captured region plus 10 kb on each end in the wild type and siRNA- or methylation-deficient mutants. The choice of a global similarity measure, the SSIM, to study short-range chromatin changes rather than comparing individual loops aimed to increase the power to detect reliable differences because random interactions increase with shorter distances and increase the methodological background noise. Moreover, other interactions, caused by dimerization of DNA-bound transcription factors or nucleosome packing, can also impact such analyses. The chromatin organization of the IR-containing loci, which should have SSIM values of $\sim$1 if similar between the genotypes, were often changed in *ddc* and *dcl234* mutants (Figure 2F), indicating that (IR-triggered) methylation has a substantial effect on the local 3D topology. Such differences are less pronounced when comparing *ddc* with *dcl234*, as could be expected from the methylation deficiency observed in both, therefore lacking the anchor regions for loop formation (Figure 2F). SSIM variation detection is not influenced by the size of the captured region (Figure S2F). We also calculated the CHESS chromatin topology variation at 38 randomly selected captured loci not containing IRs but, in most cases, showing nearby spots of DNA methylation. Different from the IR-containing loci, the 38 control loci did not change in topology between wild-type and mutant genotypes (Figures 2G and S2F). This observation indicates that the changes in local 3D topology in the *ddc* and *dcl234* mutants are not a general effect caused by the reduced DNA methylation but are rather specific to IR-containing loci. Still, other TE elements near genes could likely trigger a similar phenomenon if transcriptionally active. This is the case reported, for example, for ONSEN TEs, which, upon stress-induced mobilization and insertion, affect nearby gene expression and can lead to generation of new transcripts.[13] With this scenario, it is possible that many of these changes respond to changes in the chromatin organization triggered by the TE insertion.

Changes in chromatin topology can affect gene expression.[26] Formation of chromatin loops may repress or activate gene transcription or even trigger production of alternative transcripts.[28] Given the alterations in chromatin topology we found, we wondered whether they impact gene expression in a positive or negative manner.

Using our RNA-seq data, we found 4,305 differentially expressed genes (DEGs) in the *ddc* mutant in comparison with Col-0 and 3,636 DEGs in *dcl234*, many of them overlapping between the two mutants (Figure S4A). When we compared the SSIMs in regions with DEGs with SSIMs in regions without DEGs, we found topological differences (lower SSIM) to be significantly increased in regions that included DEGs (Figure 2H). Although there was no clear correlation between SSIMs and the changes in methylation in the IRs (Figure S4B), when we split DEGs into regions linked to differentially and not differentially methylated IRs, there were more significant topological differences in *ddc* mutants compared with the wild type for regions including DEGs and differentially methylated IRs (Figure S4C). Methylation seemed to affect the SSIM correlation with DEGs less clearly in *dcl234* (Figure S4C). Loci with altered topology

Figure 2. IR near genes produce siRNAs, trigger DNA methylation, and alter the local chromatin 3D organization

(A) Changes in siRNA levels mapping to IRs in the *ddc* and *dcl234* mutants with respect to Col-0. The axes represent log2 fold change of siRNA reads per million (RPMs) over each IR. The correlation (R) is shown in the top left corner.

(B) Metagene analysis of 24-nt siRNAs mapping IRs located at different ranges of distance to the closest protein-coding gene in Col-0 wild-type (blue), *ddc* (green), and *dcl234* (red) plants. Plots show IRs scaled from the start to the end of the IR regions plus 200 bp to each side. Three individual sRNA-seq replicates are plotted separately.

(C) Metagene analysis of CG (top), CHG (center), and CHH (bottom) methylation at IRs located at different ranges of distance to the closest protein-coding gene in Col-0 wild type (red lines), *ddc* (gray lines), and *dcl234* (green lines) plants. Plots show IRs scaled from the start to the end of the IR regions plus 200 bp to each side. Three individual BS-seq replicates are plotted for each genotype.

(D) Scatterplots showing the relation between changes in siRNA levels and DNA methylation over each gene-associated IR in both mutants, *ddc* and *dcl234*. The blue lines show the linear regression, while the gray shades show the confidence interval. Only IRs with a drop of 10% or more in methylation are shown.

(E) Region of the Col-0 genome containing the *CRY1* locus displaying the epigenetic and topological profile. (I) Chromatin interactions as detected by CapC experiments and shown by red lines, indicating interacting fragments noted in cyan. (II) Cytosine DNA methylation in CG (blue), CHG (red), and CHH (black) contexts. (III) 24-nt siRNAs mapping to the genomic regions as determined by sRNA sequencing in biological triplicates of wild-type (WT) plants and *dcl234* and *ddc* mutants. (IV) Annotated genes (cyan) and transposons (yellow) in the region. (V) Region captured by the probes designed for the CapC experiment. (VI) Potential restriction fragments after digestion. (VII) IRs identified in this region of the genome.

(F) Structural similarity (SSIM) of captured regions in *dcl234* (left) and *ddc* (center) mutants with respect to Col-0 WT or to each other (right).

(G) SSIM of captured regions in *dcl234* and *ddc* mutants with respect to Col-0 WT or each other in loci containing a gene-hosted IR (green dots) or control loci without a nearby IR (orange dots). p values were calculated with the Wilcoxon rank-sum test.

(H) SSIM of captured regions in *dcl234* (left) and *ddc* (right) mutants with respect to Col-0 grouped by regions including differentially regulated genes (DEGs) or regions without DEGs (NotDEGs). The colors of the dots indicate whether they are not differentially expressed, upregulated, or downregulated. Dots in the boxplots represent single data points, whiskers denote the minimum/maximum values (no further than 1.5× IQR from the hinge), the center represents the median, and box bounds represent the lower and upper quartiles. The p value, calculated with Student's t test, are shown on top of the boxplots.

in *ddc* and *dcl234* included up- and downregulated genes (Figure 2H). Thus, changes in chromatin topology caused by impaired RdDM machinery could be linked to opposite changes in gene expression depending on the loci. This observation, also made in rice for genes adjacent to MITEs,[7] fits with short-range chromatin loops affecting Pol II activity in different ways, depending on which part of the gene is included in the loop.[28]

In summary, our data indicate that insertion of an IR near a gene can trigger changes in the local chromatin topology that ultimately affect gene expression in a locus- and loop-dependent way.

Polymorphisms on the content of IRs near genes between *A. thaliana* accessions affect local chromatin topology

TEs are commonly silenced in plants as a means of protecting genome integrity. However, under extreme stress conditions, TE transcription can be reactivated, giving TEs the potential to jump in the genome in a process that has been proposed to help adaptation to new environments.[9,41] We wondered whether the gene-adjacent IRs that vary between natural *A. thaliana* accessions could confer adaptative potential by changing the expression of hosting loci through changes in 3D chromatin organization.

To test this hypothesis, we first used published TE polymorphism datasets[42] to detect variations in the TE content of 216 *A. thaliana* genomes.[43] Narrowing this down further to the 634 IRs located within 3,000 bp upstream or downstream of annotated genes in the Col-0 reference genome, we found 193 of these 634 IRs to be variable between the 216 accessions studied (Figure 3A). IRs without a clear TE origin are underrepresented in the collection of polymorphic IRs and were polymorphic mostly because they possess a TE insertion disrupting them in an accession different from Col-0 (Figure 3A). These 193 IRs are located within 3,000 bp of 368 annotated genes. This implies that each IR can influence two or more adjacent genes in many cases, an observation that is not surprising, given the compact nature of *A. thaliana* genomes. To investigate the effect that the variation in IR content may have on local chromatin topology and gene expression, we repeated the CapC-seq, RNA-seq, BS-seq, and sRNA-seq analyses in the Ba-1 and Hod accessions. We selected these two accessions based on the number of variable IRs present in each of them (46 and 31 IRs present in Col-0 are missing in Hod and Ba-1, respectively, with 18 missing in both accessions). With the CapC-seq experiment, we captured 290 regions comprising the genomic sequences around variable IRs, which included adjacent genes. The analysis of individual loci revealed that the presence of an IR at a locus correlated with the accumulation of 24-nt siRNAs and with distinctive short-range chromatin loops (Figures 3B and S3).

We chose five potential candidate genes (*PHV*, *PHYC*, *P5CS1*, *PHR1*, and *CRY1*) to evaluate the effect of polymorphic IRs on loop formation and gene expression. We corroborated by Sanger sequencing that each of the loci has IRs in Col-0 but not in the indicated accessions (Figures 4B, 4G, and S5). We then used 3C followed by qPCR to confirm and quantify the formation of IR-dependent chromatin loops at these loci and qRT-PCR to measure correlation with gene expression in each locus (Figures 3C, 4C, and 4I). We detected alternative chromatin loops at *PHV*, *P5CS1*, and *PHR1* (Figure 3C). In the case of *PHV*, loop 1 appeared to form independent of the associated IR, but loops 2 and 3 were missing in Ba-1, the accession without the IR, or *dcl234* mutants (Figure 3C). For *P5CS1*, loop 2 appeared independent of the presence of the IR, while formation of an intragenic loop 1 correlated with the presence of the IR or functional RdDM machinery (Figure 3C). For *PHV* and *P5CS1*, the absence of an IR appeared to be associated with enhanced gene expression (Figure 3C). In the case of *PHR1*, we found that formation of a loop 1 depends on the presence of the IR, which is missing in the Hod accessions, but loop 2 was only formed when the IR is missing, probably reflecting a hierarchy of chromatin interactions controlled by the IR (Figure 3C). Contrary to the results with *PHV* and *P5CS1*, the absence of the IR near *PHR1* caused repression of the gene. The regulatory effects caused by the presence of an IR near a gene were observed for IRs located upstream (i.e., at *PHV* and *PHR1*) and downstream of the affected gene (i.e., at *CRY1* and *P5CS1*). This particularity indicates that it is more likely that the change in chromatin topology, rather than an effect of the IR-triggered methylation per se, which could simply act through affecting binding of classical transcription factors, leads to the observed difference in gene expression.

When we analyzed the genome-wide effects of insertional IR polymorphisms over genome topology using CHESS, we also found a drastic change in chromatin folding associated with the presence/absence of an IR (Figure 3D). For this analysis, we calculated the difference between the SSIMs obtained from the comparison of Ba-1 and Hod genome topology against Col-0, as recommended by the authors of CHESS, to have a common reference to score differences between these accessions.[38] As can be observed in Figure 3D, the difference in SSIMs is greater when Ba-1 shares the IR with Col-0, having a higher SSIM, but it is absent in Hod, resulting in a lower SSIM and an increased positive difference.

These data suggest that insertion of an IR near coding genes can have an impact on local chromatin topology that affects gene expression. Aiming to explore whether this could potentially have adaptive consequences, we asked whether there is a correlation between the presence/absence of an IR near a gene and well-defined phenotypic traits recorded in the Arapheno database.[44] This analysis revealed an association between IR insertional polymorphisms near some genes with a given phenotype related to flowering, a typically adaptive trait (Figure 3E). This observation suggests that insertion of an IR near a gene can not only impact the local chromatin topology, but it may also have adaptive implications by changing the phenotypes controlled by the adjacent gene.

Insertion of an IR near genes causes strong phenotypic effects

Our data suggest that insertion of an IR near a gene can be a significant event during adaptive evolution. However, this association may not be caused by the IR itself but by adjacent polymorphisms that may be in linkage disequilibrium with the IR. In addition, even when the association is with the IR itself, it may not involve changes in genome topology. To study this

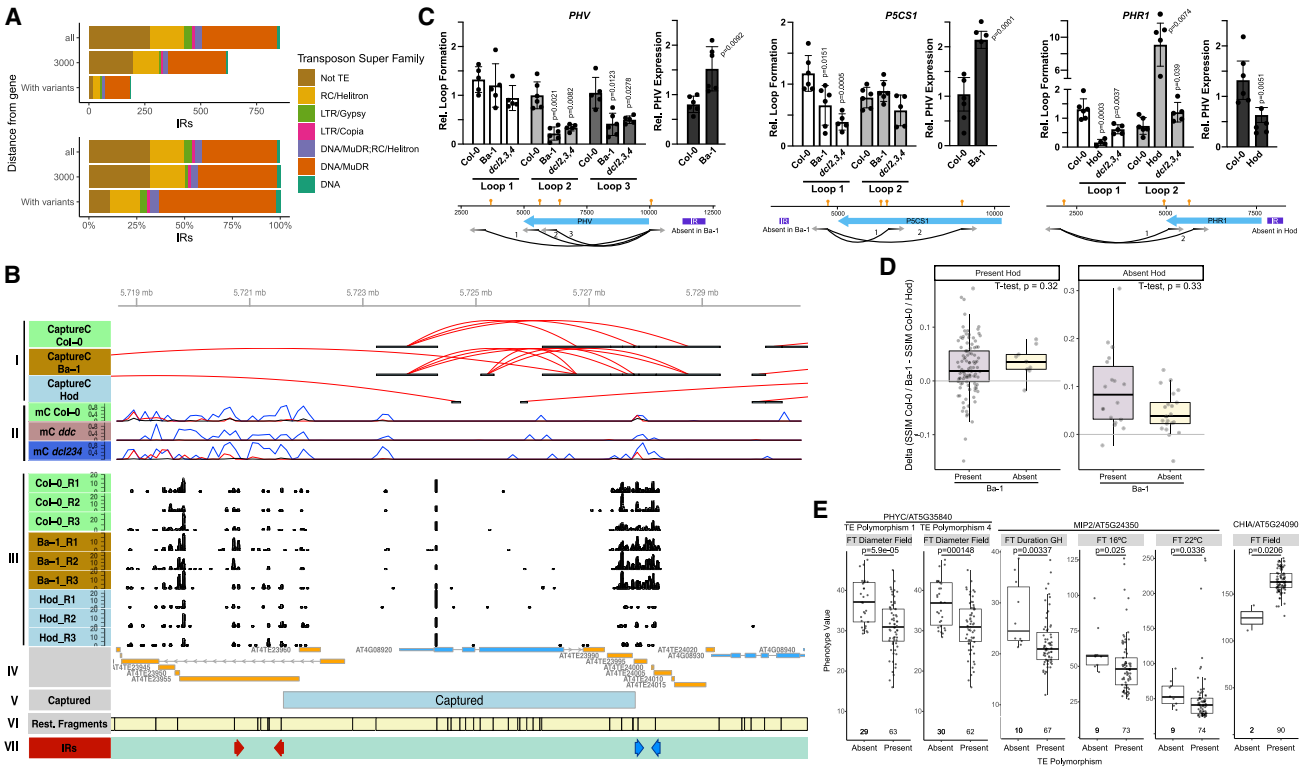


Figure 3. Insertional polymorphisms of IRs near genes cause natural variation in short-range chromatin topology

(A) Number of IRs and polymorphic IRs associated with genes (top panel) and percentage of IRs and polymorphic IRs by annotated TE superfamilies (bottom panel). Only superfamilies with at least five overlapping IRs are displayed.

(B) Region of the Col-0 genome containing the *CRY1* locus with its epigenetic and topological profile. (I) Chromatin interactions as detected by CapC experiments from Col-0 and Ba-1 plants (both contain the IR) and Hod plants (which have a deletion polymorphism in the same IR). Red lines indicate interacting cyan fragments. (II) Cytosine DNA methylation in CG (blue), CHG (red), and CHH (black) contexts. (III) 24-nt siRNAs mapping to the genomic regions as determined by sRNA sequencing in biological triplicates of WT Col-0, Ba-1, and Hod plants. (IV) Annotated genes (cyan) and transposons (yellow) in the region. (V) Region captured by the probes designed for the CapC experiment. (VI) Potential restriction fragments after digestion. (VII) IRs identified in this region of the genome. The IR in red is conserved between the three genotypes, while the IR in blue is missing in Hod.

(C) qRT-PCR with total RNA and chromatin conformation capture (3C) qPCR experiments to quantify formation of specific chromatin loops as well as expression of the affected loci in WT, Ba-1, and Hod plants and *dcl234* and *ddc* mutants. Data are presented as mean values ± SD. The p values were calculated with two-tailed unpaired Student's t test with Welch's correction. n > 4 biologically independent samples. A diagram at the bottom of each gene shows the chromatin loops amplified (shown as black lines indicating the ligated restriction fragments detected) and confirmed by Sanger sequencing for each locus. The restriction sites used for the 3C experiments are shown as orange pins. The gradient arrows show the direction from a restriction site where a chromatin loop anchor point should be located.

(D) Delta of the SSIM calculated with CHESS for the contrasts Col-0 versus Ba-1 and Col-0 versus Hod for each captured region. Each captured region is classified depending on whether the associated IR is present or absent in the three analyzed accessions.

(E) Association of flowering phenotypes (FT), including rosette diameter upon flowering, flowering duration, and flowering time at different temperatures in plants grown in the field or greenhouse (GH), with the presence or absence of an IR near selected genes. Boxplots in (D and E) show single data points as individual dots, whiskers denote the minimum/maximum values (no further than 1.5× IQR from the hinge), the center represents the median, and box bounds represent the lower and upper quartiles. The p value, calculated with the Student's t test, are shown on top of the boxplots.

possibility, we selected two loci, *PHYC* and *CRY1*, which display natural variation in the presence/absence of a close-by IR that is associated with developmental phenotypes (Figure 3E). *PHYC*- and *CRY1*-associated IRs are located downstream of the transcription termination site (TTS) (Figures 4A and 4F), making it less likely that they act directly on the promoter of the genes as general regulatory elements.

PHYC encodes a photoreceptor capable of sensing red (R) light and far-red (FR) light and is implicated in several developmental transitions, such as flowering, seed germination, and hypocotyl elongation.[45–48] In Col-0, an IR is located ~500 bp downstream of the TTS of *PHYC*, while it is missing in several natural accessions, including, for example, Ey15-2 (Figures 4A and 4B). To test whether this IR impacts *PHYC* expression and related phenotypes, we used a CRISPR-Cas9 strategy to delete a fragment of the IR, with which we can disrupt dsRNA formation without completely removing the IR sequence (Figure 4A). Three homozygous lines were obtained with a deletion of the IR fragment (Figure 4B). Supporting a role of this IR in modulating genome topology and gene expression, we detected a chromatin loop encompassing the entire *PHYC* gene in Col-0 plants that was absent in the CRISPR mutant lines and in the Ey15-2

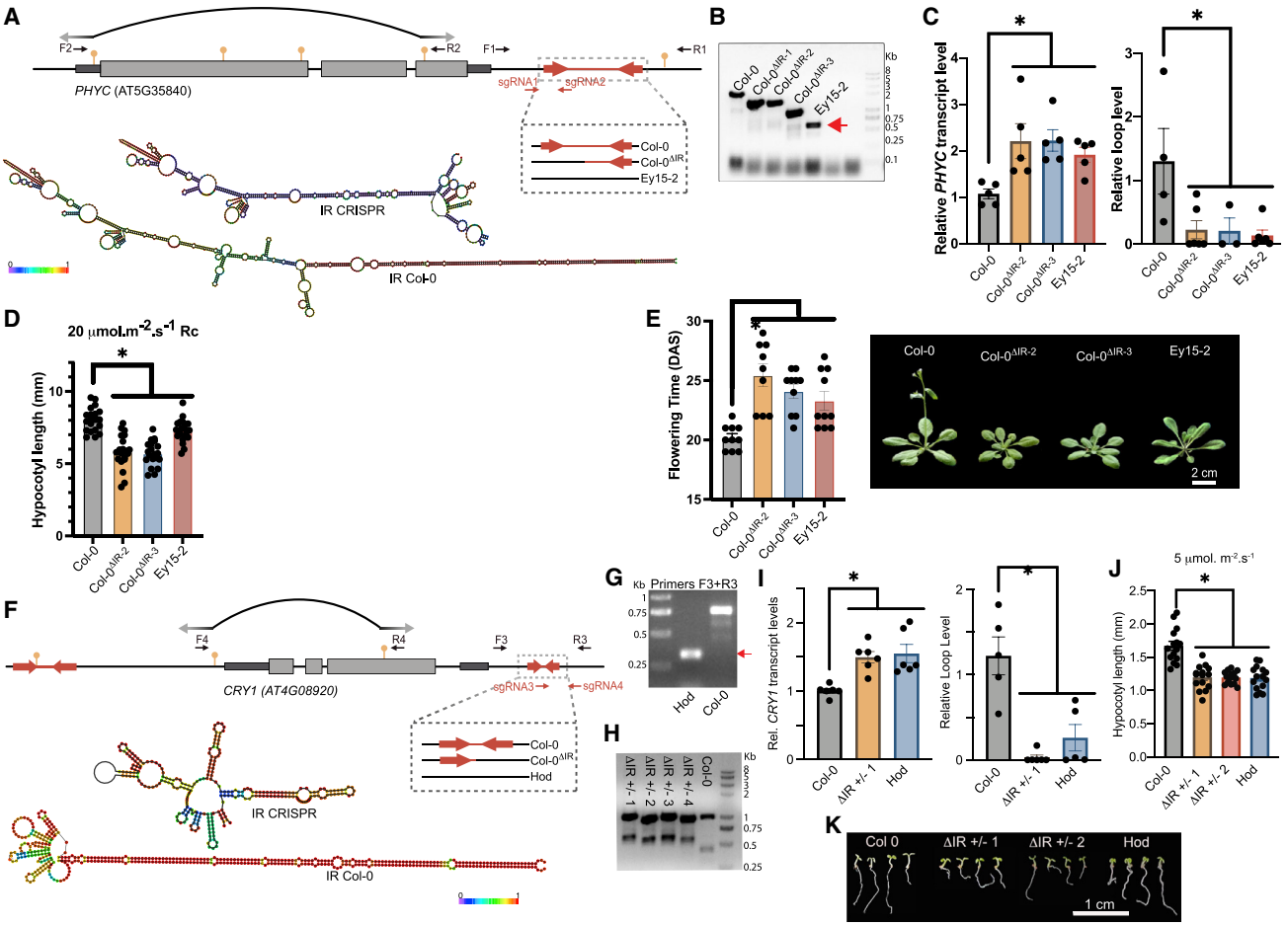


Figure 4. Insertional polymorphisms of IRs near *PHYC* and *CRY1* cause changes in locus topology, expression, and associated phenotypes

(A) *PHYC* locus diagram showing the position of the IR, restriction sites for EcoRI used for 3C experiments (yellow pins), sgRNAs used for CRISPR-Cas9 deletion (sgRNA1 and sgRNA2), and primer positions. The gradient arrows show the direction from a restriction site where a chromatin loop anchor point should be located. Predicted secondary RNA structures for the IR transcript in Col-0 plants or CRISPR-edited plants are shown using a color scale corresponding to base-pairing probability 0 (purple) to 1 (red).

(B) PCR using F1-R1 primers on genomic DNA, confirming partial deletion of the IR in the CRISPR lines and complete absence of the IR in Ey15-2 accession. A red arrowhead indicates the band sequenced to confirm the absence of the polymorphic IR in the Ey15-2 accession.

(C) *PHYC* relative expression as measured by qRT-PCR and normalized against *ACT2* (left panel). Quantification of formation of a short-range chromatin loop by 3C-qPCR analysis using EcoRI and F2-R2 primers (right panel).

(D) Hypocotyl length under a continuous R (Rc) light regimen. Seedlings were grown under Rc light (20 μmol m^{-2} s^{-1} for 4 days before measurement.

(E) Flowering time for the different genotypes expressed as days after sowing (DAS).

(F) *CRY1* locus diagram showing the position of the IRs, restriction sites for XbaI (yellow pins), sgRNAs used for CRISPR-Cas9 deletion (sgRNA3 and sgRNA4), and primer positions. Shown at the bottom are the secondary RNA structures as predicted by RNAfold for the Col-0 WT IR transcript and the CRISPR-edited outcome of the same IR. The color scale shows base-pairing probability 0 (purple) to 1 (red).

(G and H) PCR using F3-R3 primers upon genomic DNA showing successful partial deletion of the IR in CRISPR heterozygous lines and complete absence in Hod plants. A red arrowhead indicates the band sequenced to confirm the absence of the polymorphic IR in the Hod accession.

(I) *CRY1* expression analysis measured by qRT-PCR normalized against *ACT2* (left) and quantification of a chromatin loop by 3C-qPCR using XbaI and F4-R4 primers (right).

(J) Hypocotyl length in continuous blue light. Seedlings were grown under blue light (5 μmol m^{-2} s^{-1}) for 4 days before measuring.

(K) Hypocotyl phenotype of 4-day-old seedlings grown under continuous blue light (5 μmol m^{-2} s^{-1}) and quantified in (J). In all cases, data represent individual values ± SD; significant differences are indicated as *$p < 0.05$ in a one-way ANOVA.

accession, which lacks the IR (Figure 4C). The absence of the IR in Ey15-2 and in the CRISPR lines correlated with higher expression of *PHYC*, suggesting that the IR promotes formation of a repressive loop in Col-0 (Figure 4C). The CRISPR lines and Ey15-2 had altered developmental responses related to known *PHYC* functions, including delayed flowering and shortened hypocotyls under a continuous R light treatment (Figures 4D and 4E). Altogether, our data indicate that the presence of the IR next to *PHYC* has a substantial impact on gene regulation by enabling formation of a short-range chromatin loop that

represses gene expression, thereby contributing to the differential response of natural accessions to light signals.

CRY1 is a blue light receptor and participates predominately in regulation of blue light inhibition of hypocotyl elongation and anthocyanin production.[49–51] Two IRs are located near the gene, ~2,000 bp upstream of the transcription start site (TSS) and ~550 bp downstream of the TTS (Figure 4F). The second IR is variable among *A. thaliana* accessions, missing, for example, in Hod (Figure 4G). Using CRISPR-Cas9 engineering, heterozygous lines could be obtained missing a fragment of the IR located downstream of the TSS (with homozygous lines apparently not viable) (Figure 4H). Similar to *PHYC*, we detected a chromatin loop that brings together the borders of the *CRY1* gene only in plants containing the IR and that represses expression of the gene (Figure 4I). The absence of the IR in the CRISPR lines or in the natural accession Hod correlated with higher expression levels of *CRY1* and shorter hypocotyls under blue light (Figures 4I–4K), a phenotype described previously in *CRY1*-overexpressing lines.[49,52]

DISCUSSION

How organisms can adapt to a rapidly changing environment is one of the most interesting questions in evolutionary biology.[53,54] SNPs have been the main focus of many genomic studies aiming to assess the evolutionary potential of mutations, but TEs can be particularly powerful actors in rapid adaptation because single transposition events can have potentially wide-ranging consequences on gene expression and derived phenotypes.[8] On one side, the broad distribution of TEs across the genome facilitates generation of chromosomal rearrangements through ectopic recombination. Even more significantly, mobilization of TEs can disrupt, modify, or even change the expression of genes in various ways that could generate a favorable adaptive trait.[1,55] New alleles caused by TE insertions have been proposed to guarantee a consistent supply of potentially adaptable variants in response to the environment.[8] Still, many TEs inserted within gene-rich areas are quickly purged, according to population genomic surveys of TE polymorphisms.[56] This is in agreement with transposition tending to produce alleles with negative effects on fitness.

Contrary to autonomous TEs, MITEs have been found to be distributed on chromosome arms in plants, highly associated with genes, and frequently transcribed with adjacent genes.[7,57,58] In agreement with this, our study identified numerous TE-derived IRs near coding genes in *A. thaliana*. Different from rice, where more than half of the genes are associated with MITEs,[7] fewer genes are associated with IRs in *A. thaliana*. This observation is not surprising because *Arabidopsis* is an outlier regarding TE content among plants, with only 15% of its genome represented by TEs. In comparison, they account for 85% of maize and up to ~40% of rice genomes.[2,59,60]

The association between MITEs and protein-coding genes suggested that these TEs may play essential roles in genome evolution. Current evidence suggests that siRNA-triggered TE methylation tends to cause repression of neighboring genes.[61,62] However, in the case of TE-derived IRs, positive and negative effects on expression of host genes have been reported.[17–20,25] The weak correlation between methylation of MITEs and expression of adjacent genes can now be better explained in light of our findings showing that IRs located near genes affect local chromatin organization. While methylation of DNA on its own is expected to have primarily repressive effects, short-range chromatin loops could produce many different outcomes, including transcriptional repression/activation and production of alternative mRNAs.[28] In this sense, the canonical TE silencing pathway triggered by Pol IV-dependent RdDM leads to chromatin condensation and permanent repression of TE. This mechanism does not appear to be acting at IRs near genes not located in condensed regions. The localized and non-spreading methylation triggered by these IRs seems to act differently, either by recruiting specific proteins regulating gene expression[16] or, as we show here, by changing local chromatin topology. The dsRNA stem-loop structure of IR transcripts allows them to trigger production of 24-nt siRNAs when produced by Pol II transcription, different from TEs that require specific Pol IV/V transcription. Our finding that Pol II transcribes most gene-hosted IRs turns these elements into candidates for active regulatory elements in the genome and may explain why they escape the more permanent silencing mechanisms operating on other TEs. That the transcription rate of a gene has an impact on the formation/opening of a gene loop, and not the other way around, remains a possibility. We think this is unlikely because activation of genes correlates many times with chromatin loop opening but in many other instances with their formation, suggesting that transcription elongation is not sufficient to disassemble short-range chromatin loops.

Naturally occurring variation in the arrangement of inverted and tandem repeated *PAI* genes has been linked to differential patterns of *cis* and *trans* methylation and associated to differential expression of *PAI* genes in *A. thaliana* accessions.[63,64] MITEs have the potential to transpose into various locations in the genome, resulting in the presence/absence (insertional) polymorphisms between genotypes.[7,65] Such polymorphisms can be caused by insertion or excision of a MITE from a locus. However, it is unknown which of these scenarios contributes more to the genetic variation within a species. Here, we show that ~30% of the IRs located within 3,000 bp of protein-coding genes present insertional polymorphisms between 216 *A. thaliana* natural accessions. Our results also indicate that such natural variation in the gene-associated IR content can cause changes in chromatin organization that could be considered "3D polymorphisms." Our experiments using not only association between IR polymorphisms and gene expression changes but also CRISPR-Cas9-editing to demonstrate a causal relationship in several cases show that insertion of an IR near a gene can have a profound impact on chromatin organization, gene expression, and associated phenotypes. This phenomenon can potentially boost a plant's capacity to adapt to a rapidly changing environment. Because IRs can co-opt the promoters of adjacent genes to produce Pol II-derived siRNAs through a stem-loop dsRNA intermediate, they can act as autonomous regulatory elements, drastically changing the chromatin landscape of a locus upon insertion. These characteristics turn IRs into powerful elements during adaptive evolution.

Modeling has suggested faster generation of large-effect alleles because of larger transposition rates in specific populations in response to global warming.[8] Consistent with this scenario, and in a world with a rapidly changing climate, the discovery of IRs as elements shaping the 3D chromatin organization and driving genome adaptation is of great interest. Manipulation of IRs, and, in consequence, genome topology, can potentially become a powerful biotechnological tool to improve crop adaptation without the need to incorporate exogenous DNA or alter coding sequences in plants.

Limitations of the study

A. thaliana, while being the best studied plant model, has a TE-poor genome and therefore constitutes a limited model to study the effect of IRs on gene expression at a genomic level. We showed that the population of IRs is very diverse, with many being transcribed by Pol II and/or Pol IV, with different methylation patterns and sRNA levels and different structure, length, and distance from genes. Partitioning this relatively small population of elements into subclasses causes a reduction of our statistical power to detect their different effects. The power would be greater if the study were performed in a species with a higher TE content. Unfortunately, other species have fewer available resources, although our work could be used as a guide for a more directed approach.

We found clear correlations between IR methylation, sRNA production, chromatin structure, and gene expression, which are also consistent with IR variation in accessions, our gene editing proof of concept, and our previous study on the sunflower *HaWRKY6* locus. However, without identification of the mechanisms stabilizing the chromatin loops it is hard to discern the effects caused by chromatin topology and those caused by DNA modifications.

Regarding chromatin loops, quantifying short-distance interactions is challenging because specific interactions become harder to differentiate from randomly occurring interactions, which increase at these ranges. In addition, restriction enzyme distribution at each locus imposes a limitation to precisely define the anchor points of each chromatin loop. For this reason, we resorted to a global measure of chromatin topology change (the SSIM) to quantify the effects in mutants and accessions. Despite this, study of specific loci with a variety of restriction enzymes could help define potential anchor points with more precision.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Arabidopsis thaliana
- METHOD DETAILS
  - Plant material and growth conditions
  - IR detection
  - RNA-seq
  - sRNA-seq
  - Bisulfite treatment of DNA and library preparation
  - Capture-C assay
  - 3C assay and RT-PCR
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2023.112029.

ACKNOWLEDGMENTS

This work was supported by grants from Agencia Nacional de Promoción Científica y Tecnológica and Universidad Nacional del Litoral (to P.A.M.) and the Max Planck Society (to D.W.). P.A.M., A.L.A., and D.A.C. are members of CONICET; R.M. is a fellow of CONICET. We thank the Deutscher Akademischer Austauschdienst and Company of Biologists for short-term fellowships to D.A.C. and R.M., respectively. We also want to thank Danelle Seymour (UC Riverside, USA) for her valuable comments on our manuscript.

AUTHOR CONTRIBUTIONS

A.L.A. performed the majority of the analyses. D.A.C. prepared the libraries for sequencing and CapC experiments. P.L.L. and H.A.B. helped with the design of the CapC probes. P.L.L. performed the capture experiment. R.M. and D.A.C. validated the chromatin loop and created the genome-edited lines. R.M. characterized the genome-edited lines. A.L.A., D.A.C., P.A.M., and D.W. conceived the study. P.A.M. and D.W. supervised the work and secured project funding. A.L.A., R.M., D.W., and P.A.M. wrote the manuscript with help from all authors.

DECLARATION OF INTERESTS

D.W. holds equity in Computomics, which advises plant breeders. D.W. also consults for KWS SE, a plant breeder and seed producer.

Received: July 1, 2022
Revised: November 3, 2022
Accepted: January 10, 2023

REFERENCES

1. Dubin, M.J., Mittelsten Scheid, O., and Becker, C. (2018). Transposons: a blessing curse. Curr. Opin. Plant Biol. *42*, 23–29. https://doi.org/10.1016/j.pbi.2018.01.003.
2. Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. Science *326*, 1112–1115. https://doi.org/10.1126/science.1178534.
3. Matzke, M.A., Kanno, T., and Matzke, A.J.M. (2015). RNA-directed DNA methylation: the evolution of a complex epigenetic pathway in flowering plants. Annu. Rev. Plant Biol. *66*, 243–267. https://doi.org/10.1146/annurev-arplant-043014-114633.
4. Zhou, M., and Law, J.A. (2015). RNA Pol IV and V in gene silencing: rebel polymerases evolving away from Pol II's rules. Curr. Opin. Plant Biol. *27*, 154–164. https://doi.org/10.1016/j.pbi.2015.07.005.
5. Maumus, F., and Quesneville, H. (2014). Ancestral repeats have shaped epigenome and genome composition for millions of years in Arabidopsis thaliana. Nat. Commun. *5*, 4104. https://doi.org/10.1038/ncomms5104.
6. SanMiguel, P., Gaut, B.S., Tikhonov, A., Nakajima, Y., and Bennetzen, J.L. (1998). The paleontology of intergene retrotransposons of maize. Nat. Genet. *20*, 43–45. https://doi.org/10.1038/1695.

7. Lu, C., Chen, J., Zhang, Y., Hu, Q., Su, W., and Kuang, H. (2012). Miniature inverted-repeat transposable elements (MITEs) have been accumulated through amplification bursts and play important roles in gene expression and species diversity in Oryza sativa. Mol. Biol. Evol. *29*, 1005–1017. https://doi.org/10.1093/molbev/msr282.

8. Baduel, P., Leduque, B., Ignace, A., Gy, I., Gil, J., Jr., Loudet, O., Colot, V., and Quadrana, L. (2021). Genetic and environmental modulation of transposition shapes the evolutionary potential of Arabidopsis thaliana. Genome Biol. *22*, 138. https://doi.org/10.1186/s13059-021-02348-5.

9. Tittel-Elmer, M., Bucher, E., Broger, L., Mathieu, O., Paszkowski, J., and Vaillant, I. (2010). Stress-induced activation of heterochromatic transcription. PLoS Genet. *6*, e1001175. https://doi.org/10.1371/journal.pgen.1001175.

10. Collier, L.S., and Largaespada, D.A. (2007). Transposable elements and the dynamic somatic genome. Genome Biol. *8*, S5. https://doi.org/10.1186/gb-2007-8-s1-s5.

11. Deragon, J.M., Casacuberta, J.M., and Panaud, O. (2008). Plant transposable elements. Genome Dyn. *4*, 69–82. https://doi.org/10.1159/000126007.

12. Lisch, D. (2013). How important are transposons for plant evolution? Nat. Rev. Genet. *14*, 49–61. https://doi.org/10.1038/nrg3374.

13. Roquis, D., Robertson, M., Yu, L., Thieme, M., Julkowska, M., and Bucher, E. (2021). Genomic impact of stress-induced transposable element mobility in Arabidopsis. Nucleic Acids Res. *49*, 10431–10447. https://doi.org/10.1093/nar/gkab828.

14. Fattash, I., Rooke, R., Wong, A., Hui, C., Luu, T., Bhardwaj, P., and Yang, G. (2013). Miniature inverted-repeat transposable elements: discovery, distribution, and activity. Genome *56*, 475–486. https://doi.org/10.1139/gen-2012-0174.

15. Yang, G., Nagel, D.H., Feschotte, C., Hancock, C.N., and Wessler, S.R. (2009). Tuned for transposition: molecular determinants underlying the hyperactivity of a Stowaway MITE. Science *325*, 1391–1394. https://doi.org/10.1126/science.1175688.

16. Niu, C., Jiang, L., Cao, F., Liu, C., Guo, J., Zhang, Z., Yue, Q., Hou, N., Liu, Z., Li, X., et al. (2022). Methylation of a MITE insertion in the MdRFNR1-1 promoter is positively associated with its allelic expression in apple in response to drought stress. Plant Cell *34*, 3983–4006. https://doi.org/10.1093/plcell/koac220.

17. Underwood, C.J., Vijverberg, K., Rigola, D., Okamoto, S., Oplaat, C., Camp, R.H., Radoeva, T., Schauer, S.E., Fierens, J., Jansen, K., et al. (2022). A PARTHENOGENESIS allele from apomictic dandelion can induce egg cell division without fertilization in lettuce. Nat. Genet. *54*, 84–93. https://doi.org/10.1038/s41588-021-00984-y.

18. Wu, N., Yao, Y., Xiang, D., Du, H., Geng, Z., Yang, W., Li, X., Xie, T., Dong, F., and Xiong, L. (2022). A MITE variation-associated heat-inducible isoform of a heat-shock factor confers heat tolerance through regulation of JASMONATE ZIM-DOMAIN genes in rice. New Phytol. *234*, 1315–1331. https://doi.org/10.1111/nph.18068.

19. Xu, L., Yuan, K., Yuan, M., Meng, X., Chen, M., Wu, J., Li, J., and Qi, Y. (2020). Regulation of rice tillering by RNA-directed DNA methylation at miniature inverted-repeat transposable elements. Mol. Plant *13*, 851–863. https://doi.org/10.1016/j.molp.2020.02.009.

20. Zhang, H., Tao, Z., Hong, H., Chen, Z., Wu, C., Li, X., Xiao, J., and Wang, S. (2016). Transposon-derived small RNA is responsible for modified function of WRKY45 locus. Nat. Plants *2*, 16016. https://doi.org/10.1038/nplants.2016.16.

21. Ariel, F.D., and Manavella, P.A. (2021). When junk DNA turns functional: transposon-derived non-coding RNAs in plants. J. Exp. Bot. *72*, 4132–4143. https://doi.org/10.1093/jxb/erab073.

22. Cuerda-Gil, D., and Slotkin, R.K. (2016). Non-canonical RNA-directed DNA methylation. Nat. Plants *2*, 16163. https://doi.org/10.1038/nplants.2016.163.

23. Sasaki, T., Lee, T.F., Liao, W.W., Naumann, U., Liao, J.L., Eun, C., Huang, Y.Y., Fu, J.L., Chen, P.Y., Meyers, B.C., et al. (2014). Distinct and concurrent pathways of Pol II- and Pol IV-dependent siRNA biogenesis at a repetitive trans-silencer locus in Arabidopsis thaliana. Plant J. *79*, 127–138. https://doi.org/10.1111/tpj.12545.

24. Crescente, J.M., Zavallo, D., Del Vas, M., Asurmendi, S., Helguera, M., Fernandez, E., and Vanzetti, L.S. (2022). Genome-wide identification of MITE-derived microRNAs and their targets in bread wheat. BMC Genom. *23*, 154. https://doi.org/10.1186/s12864-022-08364-4.

25. Gagliardi, D., Cambiagno, D.A., Arce, A.L., Tomassi, A.H., Giacomelli, J.I., Ariel, F.D., and Manavella, P.A. (2019). Dynamic regulation of chromatin topology and transcription by inverted repeat-derived small RNAs in sunflower. Proc. Natl. Acad. Sci. USA *116*, 17578–17583. https://doi.org/10.1073/pnas.1903131116.

26. Domb, K., Wang, N., Hummel, G., and Liu, C. (2022). Spatial features and functional implications of plant 3D genome organization. Annu. Rev. Plant Biol. *73*, 173–200. https://doi.org/10.1146/annurev-arplant-102720-022810.

27. Zhang, X., and Wang, T. (2021). Plant 3D chromatin organization: important insights from chromosome conformation capture analyses of the last 10 years. Plant Cell Physiol. *62*, 1648–1661. https://doi.org/10.1093/pcp/pcab134.

28. Gagliardi, D., and Manavella, P.A. (2020). Short-range regulatory chromatin loops in plants. New Phytol. *228*, 466–471. https://doi.org/10.1111/nph.16632.

29. Grzechnik, P., Tan-Wong, S.M., and Proudfoot, N.J. (2014). Terminate and make a loop: regulation of transcriptional directionality. Trends Biochem. Sci. *39*, 319–327. https://doi.org/10.1016/j.tibs.2014.05.001.

30. Sotelo-Silveira, M., Chávez Montes, R.A., Sotelo-Silveira, J.R., Marsch-Martínez, N., and de Folter, S. (2018). Entering the next dimension: plant genomes in 3D. Trends Plant Sci. *23*, 598–612. https://doi.org/10.1016/j.tplants.2018.03.014.

31. Guo, C., Spinelli, M., Ye, C., Li, Q.Q., and Liang, C. (2017). Genome-wide comparative analysis of miniature inverted repeat transposable elements in 19 Arabidopsis thaliana ecotype accessions. Sci. Rep. *7*, 2634. https://doi.org/10.1038/s41598-017-02855-1.

32. Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. Trends Genet. *16*, 276–277. https://doi.org/10.1016/s0168-9525(00)02024-2.

33. Kindgren, P., Ivanov, M., and Marquardt, S. (2020). Native elongation transcript sequencing reveals temperature dependent dynamics of nascent RNAPII transcription in Arabidopsis. Nucleic Acids Res. *48*, 2332–2347. https://doi.org/10.1093/nar/gkz1189.

34. Ferrafiat, L., Pflieger, D., Singh, J., Thieme, M., Böhrer, M., Himber, C., Gerbaud, A., Bucher, E., Pikaard, C.S., and Blevins, T. (2019). The NRPD1 N-terminus contains a Pol IV-specific motif that is critical for genome surveillance in Arabidopsis. Nucleic Acids Res. *47*, 9037–9052. https://doi.org/10.1093/nar/gkz618.

35. Tan, L.M., Zhang, C.J., Hou, X.M., Shao, C.R., Lu, Y.J., Zhou, J.X., Li, Y.Q., Li, L., Chen, S., and He, X.J. (2018). The PEAT protein complexes are required for histone deacetylation and heterochromatin silencing. EMBO J. *37*, e98770. https://doi.org/10.15252/embj.201798770.

36. Lu, C., Kulkarni, K., Souret, F.F., MuthuValliappan, R., Tej, S.S., Poethig, R.S., Henderson, I.R., Jacobsen, S.E., Wang, W., Green, P.J., and Meyers, B.C. (2006). MicroRNAs and other small RNAs enriched in the Arabidopsis RNA-dependent RNA polymerase-2 mutant. Genome Res. *16*, 1276–1288. https://doi.org/10.1101/gr.5530106.

37. Kurihara, Y., Matsui, A., Kawashima, M., Kaminuma, E., Ishida, J., Morosawa, T., Mochizuki, Y., Kobayashi, N., Toyoda, T., Shinozaki, K., and Seki, M. (2008). Identification of the candidate genes regulated by RNA-directed DNA methylation in Arabidopsis. Biochem. Biophys. Res. Commun. *376*, 553–557. https://doi.org/10.1016/j.bbrc.2008.09.046.

38. Galan, S., Machnik, N., Kruse, K., Díaz, N., Marti-Renom, M.A., and Vaquerizas, J.M. (2020). CHESS enables quantitative comparison of chromatin contact data and automatic feature extraction. Nat. Genet. *52*, 1247–1255. https://doi.org/10.1038/s41588-020-00712-y.

39. Buckle, A., Gilbert, N., Marenduzzo, D., and Brackley, C.A. (2019). capC-MAP: software for analysis of Capture-C data. Bioinformatics *35*, 4773–4775. https://doi.org/10.1093/bioinformatics/btz480.

40. Geeven, G., Teunissen, H., de Laat, W., and de Wit, E. (2018). peakC: a flexible, non-parametric peak calling package for 4C and Capture-C data. Nucleic Acids Res. *46*, e91. https://doi.org/10.1093/nar/gky443.

41. Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I., and Paszkowski, J. (2011). An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. Nature *472*, 115–119. https://doi.org/10.1038/nature09861.

42. Stuart, T., Eichten, S.R., Cahn, J., Karpievitch, Y.V., Borevitz, J.O., and Lister, R. (2016). Population scale mapping of transposable element diversity reveals links to gene regulation and epigenomic variation. Elife *5*, e20777. https://doi.org/10.7554/eLife.20777.

43. Schmitz, R.J., He, Y., Valdés-López, O., Khan, S.M., Joshi, T., Urich, M.A., Nery, J.R., Diers, B., Xu, D., Stacey, G., and Ecker, J.R. (2013). Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population. Genome Res. *23*, 1663–1674. https://doi.org/10.1101/gr.152538.112.

44. Seren, Ü., Grimm, D., Fitz, J., Weigel, D., Nordborg, M., Borgwardt, K., and Korte, A. (2017). AraPheno: a public database for Arabidopsis thaliana phenotypes. Nucleic Acids Res. *45*, D1054–D1059. https://doi.org/10.1093/nar/gkw986.

45. Li, Q., Wu, G., Zhao, Y., Wang, B., Zhao, B., Kong, D., Wei, H., Chen, C., and Wang, H. (2020). CRISPR/Cas9-mediated knockout and overexpression studies reveal a role of maize phytochrome C in regulating flowering time and plant height. Plant Biotechnol. J. *18*, 2520–2532. https://doi.org/10.1111/pbi.13429.

46. Kippes, N., VanGessel, C., Hamilton, J., Akpinar, A., Budak, H., Dubcovsky, J., and Pearce, S. (2020). Effect of phyB and phyC loss-of-function mutations on the wheat transcriptome under short and long day photoperiods. BMC Plant Biol. *20*, 297. https://doi.org/10.1186/s12870-020-02506-0.

47. Chen, A., Li, C., Hu, W., Lau, M.Y., Lin, H., Rockwell, N.C., Martin, S.S., Jernstedt, J.A., Lagarias, J.C., and Dubcovsky, J. (2014). Phytochrome C plays a major role in the acceleration of wheat flowering under long-day photoperiod. Proc. Natl. Acad. Sci. USA *111*, 10037–10044. https://doi.org/10.1073/pnas.1409795111.

48. Nishida, H., Ishihara, D., Ishii, M., Kaneko, T., Kawahigashi, H., Akashi, Y., Saisho, D., Tanaka, K., Handa, H., Takeda, K., and Kato, K. (2013). Phytochrome C is a key factor controlling long-day flowering in barley. Plant Physiol. *163*, 804–814. https://doi.org/10.1104/pp.113.222570.

49. Liu, S., Zhang, L., Gao, L., Chen, Z., Bie, Y., Zhao, Q., Zhang, S., Hu, X., Liu, Q., Wang, X., and Wang, Q. (2022). Differential photoregulation of the nuclear and cytoplasmic CRY1 in Arabidopsis. New Phytol. *234*, 1332–1346. https://doi.org/10.1111/nph.18007.

50. Ahmad, M., Jarillo, J.A., and Cashmore, A.R. (1998). Chimeric proteins between cry1 and cry2 Arabidopsis blue light photoreceptors indicate overlapping functions and varying protein stability. Plant Cell *10*, 197–207. https://doi.org/10.1105/tpc.10.2.197.

51. Ahmad, M., Lin, C., and Cashmore, A.R. (1995). Mutations throughout an Arabidopsis blue-light photoreceptor impair blue-light-responsive anthocyanin accumulation and inhibition of hypocotyl elongation. Plant J. *8*, 653–658. https://doi.org/10.1046/j.1365-313x.1995.08050653.x.

52. He, G., Liu, J., Dong, H., and Sun, J. (2019). The blue-light receptor CRY1 interacts with BZR1 and BIN2 to modulate the phosphorylation and nuclear function of BZR1 in repressing BR signaling in Arabidopsis. Mol. Plant *12*, 689–703. https://doi.org/10.1016/j.molp.2019.02.001.

53. Hermisson, J., and Pennings, P.S. (2005). Soft sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics *169*, 2335–2352. https://doi.org/10.1534/genetics.104.036947.

54. Barrett, R.D.H., and Schluter, D. (2008). Adaptation from standing genetic variation. Trends Ecol. Evol. *23*, 38–44. https://doi.org/10.1016/j.tree.2007.09.008.

55. Friedli, M., and Trono, D. (2015). The developmental control of transposable elements and the evolution of higher species. Annu. Rev. Cell Dev. Biol. *31*, 429–451. https://doi.org/10.1146/annurev-cellbio-100814-125514.

56. Quadrana, L., Bortolini Silveira, A., Mayhew, G.F., LeBlanc, C., Martienssen, R.A., Jeddeloh, J.A., and Colot, V. (2016). The Arabidopsis thaliana mobilome and its impact at the species level. Elife *5*, e15716. https://doi.org/10.7554/eLife.15716.

57. Kuang, H., Padmanabhan, C., Li, F., Kamei, A., Bhaskar, P.B., Ouyang, S., Jiang, J., Buell, C.R., and Baker, B. (2009). Identification of miniature inverted-repeat transposable elements (MITEs) and biogenesis of their siRNAs in the Solanaceae: new functional implications for MITEs. Genome Res. *19*, 42–56. https://doi.org/10.1101/gr.078196.108.

58. Oki, N., Yano, K., Okumoto, Y., Tsukiyama, T., Teraishi, M., and Tanisaka, T. (2008). A genome-wide view of miniature inverted-repeat transposable elements (MITEs) in rice, Oryza sativa ssp. japonica. Genes Genet. Syst. *83*, 321–329. https://doi.org/10.1266/ggs.83.321.

59. Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature *408*, 796–815. https://doi.org/10.1038/35048692.

60. Li, X., Guo, K., Zhu, X., Chen, P., Li, Y., Xie, G., Wang, L., Wang, Y., Persson, S., and Peng, L. (2017). Domestication of rice has reduced the occurrence of transposable elements within gene coding regions. BMC Genom. *18*, 55. https://doi.org/10.1186/s12864-016-3454-z.

61. Hollister, J.D., Smith, L.M., Guo, Y.L., Ott, F., Weigel, D., and Gaut, B.S. (2011). Transposable elements and small RNAs contribute to gene expression divergence between Arabidopsis thaliana and Arabidopsis lyrata. Proc. Natl. Acad. Sci. USA *108*, 2322–2327. https://doi.org/10.1073/pnas.1018222108.

62. Hollister, J.D., and Gaut, B.S. (2009). Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res. *19*, 1419–1428. https://doi.org/10.1101/gr.091678.109.

63. Luff, B., Pawlowski, L., and Bender, J. (1999). An inverted repeat triggers cytosine methylation of identical sequences in Arabidopsis. Mol. Cell *3*, 505–511. https://doi.org/10.1016/s1097-2765(00)80478-5.

64. Melquist, S., Luff, B., and Bender, J. (1999). Arabidopsis PAI gene arrangements, cytosine methylation and expression. Genetics *153*, 401–413. https://doi.org/10.1093/genetics/153.1.401.

65. Lyons, M., Cardle, L., Rostoks, N., Waugh, R., and Flavell, A.J. (2008). Isolation, analysis and marker utility of novel miniature inverted repeat transposable elements from the barley genome. Mol. Genet. Genom. *280*, 275–285. https://doi.org/10.1007/s00438-008-0363-0.

66. Henderson, I.R., Zhang, X., Lu, C., Johnson, L., Meyers, B.C., Green, P.J., and Jacobsen, S.E. (2006). Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. Nat. Genet. *38*, 721–725. https://doi.org/10.1038/ng1804.

67. Wu, R., Lucke, M., Jang, Y.T., Zhu, W., Symeonidi, E., Wang, C., Fitz, J., Xi, W., Schwab, R., and Weigel, D. (2018). An efficient CRISPR vector toolbox for engineering large deletions in Arabidopsis thaliana. Plant Methods *14*, 65. https://doi.org/10.1186/s13007-018-0330-7.

68. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

69. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast

universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

70. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
71. Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics *32*, 3047–3048. https://doi.org/10.1093/bioinformatics/btw354.
72. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930. https://doi.org/10.1093/bioinformatics/btt656.
73. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.
74. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25. https://doi.org/10.1186/gb-2009-10-3-r25.
75. Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571–1572. https://doi.org/10.1093/bioinformatics/btr167.
76. Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E., Melnick, A., and Mason, C.E. (2012). methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biol. *13*, R87. https://doi.org/10.1186/gb-2012-13-10-r87.
77. Hahne, F., and Ivanek, R. (2016). Visualizing genomic data using Gviz and bioconductor. Methods Mol. Biol. *1418*, 335–351. https://doi.org/10.1007/978-1-4939-3578-9_16.
78. Harmston, N., Ing-Simmons, E., Perry, M., Barešić, A., and Lenhard, B. (2015). GenomicInteractions: an R/Bioconductor package for manipulating and investigating chromatin interaction data. BMC Genom. *16*, 963. https://doi.org/10.1186/s12864-015-2140-x.
79. Lei, Y., Lu, L., Liu, H.Y., Li, S., Xing, F., and Chen, L.L. (2014). CRISPR-P: a web tool for synthetic single-guide RNA design of CRISPR-system in plants. Mol. Plant *7*, 1494–1496. https://doi.org/10.1093/mp/ssu044.
80. Pasha, A., Subramaniam, S., Cleary, A., Chen, X., Berardini, T., Farmer, A., Town, C., and Provart, N. (2020). Araport lives: an updated framework for Arabidopsis bioinformatics. Plant Cell *32*, 2683–2686. https://doi.org/10.1105/tpc.20.00358.
81. Köster, J., and Rahmann, S. (2012). Snakemake–a scalable bioinformatics workflow engine. Bioinformatics *28*, 2520–2522. https://doi.org/10.1093/bioinformatics/bts480.
82. Kalvari, I., Nawrocki, E.P., Argasinska, J., Quinones-Olvera, N., Finn, R.D., Bateman, A., and Petrov, A.I. (2018). Non-coding RNA analysis using the rfam database. Curr. Protoc. Bioinformatics *62*, e51. https://doi.org/10.1002/cpbi.51.
83. Liu, T. (2014). Use model-based Analysis of ChIP-Seq (MACS) to analyze short reads generated by sequencing protein-DNA interactions in embryonic stem cells. Methods Mol. Biol. *1150*, 81–95. https://doi.org/10.1007/978-1-4939-0512-6_4.
84. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.
85. Liu, C. (2017). In situ Hi-C library preparation for plants to study their three-dimensional chromatin interactions on a genome-wide scale. Methods Mol. Biol. *1629*, 155–166. https://doi.org/10.1007/978-1-4939-7125-1_11.
86. Abdennur, N., and Mirny, L.A. (2020). Cooler: scalable storage for Hi-C data and other genomically labeled arrays. Bioinformatics *36*, 311–316. https://doi.org/10.1093/bioinformatics/btz540.
87. Kruse, K., Hug, C.B., and Vaquerizas, J.M. (2020). FAN-C: a feature-rich framework for the analysis and visualisation of chromosome conformation capture data. Genome Biol. *21*, 303. https://doi.org/10.1186/s13059-020-02215-9.
88. Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for computing and annotating genomic ranges. PLoS Comput. Biol. *9*, e1003118. https://doi.org/10.1371/journal.pcbi.1003118.
89. Lawrence, M., Gentleman, R., and Carey, V. (2009). rtracklayer: an R package for interfacing with genome browsers. Bioinformatics *25*, 1841–1842. https://doi.org/10.1093/bioinformatics/btp328.
90. Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). Circlize Implements and enhances circular visualization in R. Bioinformatics *30*, 2811–2812. https://doi.org/10.1093/bioinformatics/btu393.
91. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682. https://doi.org/10.1038/nmeth.2019.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
TRIzol™	Invitrogen	CAT#: 15596026
Q5® High-Fidelity DNA Polymerase	New England Biolabs	CAT#: M0491S
SuperScript II Reverse Transcriptase	Thermo Fisher	CAT#: 18064022
DNA polymerase I	New England Biolabs	CAT#: M0209S
T4 DNA polymerase	New England Biolabs	CAT#: M0203S
Klenow DNA polymerase	New England Biolabs	CAT#: M0210S
T4 polynucleotide kinase	New England Biolabs	CAT#: M0201S
DpnII	New England Biolabs	CAT#: R0543S
EcoRI	New England Biolabs	CAT#: R3101T
XbaI	New England Biolabs	CAT#: R0145T
T4 DNA ligase	Thermo Fisher	CAT#: EL0013
Proteinase K	QIAGEN	CAT#: 19131
RevertAid RT Reverse Transcription Kit	Thermo Fisher	CAT#: KI1691
Critical commercial assays		
NEBNext Poly(A) Magnetic Isolation Module	New England Biolabs	CAT#: E7490S
TruSeq sRNA Library Preparation kit	Illumina	CAT#: RS-122-2001
DNeasy plant Mini Kit	QIAGEN	CAT#: 69104
TrueSeq DNA Nano Kit	Illumina	CAT#: 20015964
Epitec Plus DNA Bisulfite Conversion Kit	QIAGEN	CAT#: 59124
Kapa Hifi Uracil + DNA polymerase	Kapa Biosystems	CAT#: 07959052001
NEBNext® Ultra™ II DNA Library Prep Kit	New England Biolabs	CAT#: E7645S
Dynabeads MyOne Streptavidin C1 beads	Invitrogen	CAT#: 65001
MyBaits system	Arbor Biosciences	N/A
RevertAid RT Reverse Transcription Kit	Thermo Fisher	CAT#: KI1691
Maxima SYBR Green qPCR Master Mix	Thermo Fisher	CAT#: K0221
Deposited data		
Small RNA sequencing reads for Arabidopsis: Col 0, Ba, Hod, *ddc* and *dcl234*.	This paper	ENA: PRJEB53956
RNA sequencing reads for Arabidopsis: Col 0, Ba, Hod, *ddc* and *dcl234*.	This paper	ENA: PRJEB53956
Bisulfite sequencing reads for Arabidopsis: Col 0, Ba, Hod, *ddc* and *dcl234*.	This paper	ENA: PRJEB53956
Capture-C sequencing reads for Arabidopsis: Col 0, Ba, Hod, *ddc* and *dcl234*.	This paper	ENA: PRJEB53956
plaNETseq datasets of RNAPII associated nascent transcripts	Kindgren et al.[33]	GEO: GSE131733
Small RNA sequencing reads for Arabidopsis *nrpd1*, Pol IV independent reads.	Tan el al.[35] Ferrafiat et al.[34]	GEO: GSE116067 ENA: PRJNA510791
Experimental models: Organisms/strains		
Arabidopsis: Col 0	Arabidopsis Biological Resource Center	CS22625
Arabidopsis: *ddc*	Kurihara et al.[37]	N/A
Arabidopsis: *dcl234*	Henderson et al.[66]	N/A
Arabidopsis: Ba-1	Arabidopsis Biological Resource Center	CS76441
Arabidopsis: Hod	Arabidopsis Biological Resource Center	CS76924

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Arabidopsis: Ey15-2	Arabidopsis Biological Resource Center	CS76399
Arabidopsis: Col $0^{\Delta IR(PHYC)}$	This paper	N/A
Arabidopsis: Col $0^{\Delta IR(CRY1)}$	This paper	N/A
Oligonucleotides		
Primers used for generating CRISPR lines are listed in Table S4.	This paper	N/A
Primers used for qPCR are listed on Table S4.	This paper	N/A
Recombinant DNA		
CRISPR/Cas9 vector toolbox	Wu et al.[67]	N/A
Plasmid *UBQ10* promoter, pcoCas9, proLacZ:LacZ sgRNA1-sgRNA2/ At2S3:mCherry (PHYC)	This paper	N/A
Plasmid *UBQ10* promoter, pcoCas9, proLacZ:LacZ sgRNA3-sgRNA4/ At2S3:mCherry (CRY1)	This paper	N/A
Software and algorithms		
einverted (EMBOSS suite version 6.6.0.0)	Rice et al.[32]	N/A
CHESS (version 0.3.7)	Galan et al.[38]	N/A
CapC-Map (version 1.1.3)	Buckle et al.[39]	N/A
peakC (version 0.2)	Geeven et al.[40]	
Trimmomatic (version 0.36)	Bolger et al.[68]	N/A
STAR (version 2.5.2b)	Dobin et al.[69]	N/A
Samtools (version 1.9)	Li et al.[70]	N/A
FastQC (version 0.11.5)	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/	N/A
MultiQC (version 1.7)	Ewels et al.[71]	N/A
featureCounts (version 1.6.2)	Liao et al.[72]	N/A
DESeq2 (version 1.36.0)	Love et al.[73]	N/A
cutadapt (version 1.9.1)	https://journal.embnet.org/index.php/embnetjournal/article/view/200	N/A
FastQC (version 0.11.4)	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/	N/A
MultiQC (version 1.7)	Ewels et al.[71]	N/A
bowtie (version 1.1.2)	Langmead et al.[74]	N/A
R statistical programming environment	R Core Team v4.1.0 https://www.r-project.org/	N/A
Bismark (version 0.22.1)	Krueger and Andrews[75]	N/A
methylKit (version 1.22.0)	Akalin et al.[76]	N/A
BRGenomics (version 1.8.0)	https://rdrr.io/bioc/BRGenomics/	N/A
Gviz (version 1.40.1)	Hahne and Ivanek[77]	N/A
GenomicInteractions (version 1.30.0)	Harmston et al.[78]	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Pablo Manavella (pablomanavella@ial.santafe-conicet.gov.ar).

Materials availability

All resources generated in this study are available from the lead contact without restriction.

Data and code availability

- All sequencing data were deposited at the European Nucleotide Archive (https://www.ebi.ac.uk/ena/) public repository with accession ENA: PRJEB53956. This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the key resources table.
- All original code generated in this study is available upon request to the lead contact.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Arabidopsis thaliana

Arabidopsis thaliana accessions Col-0, Hod, Ba1, EY15-2, *dcl234*,[36] and *ddc*[37] mutants and CRISPR mutant lines were grown on soil or petri dishes at 23°C under long-day photoperiod (16/8 h light/dark). For blue and red-light experiments seeds were sowed on petri dishes with humidified filter paper and stratified in the dark for 5 days, and then transferred to white light (80 μmol m^{-2} s^{-1}) at 23°C for 3 h and subsequently transferred to continuous red (20 μmol m^{-2} s^{-1}) or blue light (5 μmol m^{-2} s^{-1}) in LED chambers for four days.

METHOD DETAILS

Plant material and growth conditions

To obtain plants with genomic fragments deletions a CRISPR/Cas9 vector toolbox[67] was used. Specific sgRNAs, described in Table S4, were designed to obtain the IR deletions. Col-0 plants were transformed using the floral dip method, T1 plants were selected based on the presence of red fluorescence in seeds under a fluorescent dissecting microscope (Leica, Solms, Germany), non-fluorescent T2 seeds missing the transgenes were grown and genotyped by PCR to identify effective deletions. ddc and dcl234 mutants where previously described.[37,66]

For genomic DNA extraction 100 mg of fresh plant material was ground in 700 μL of extraction buffer (200 mM Tris-HCl pH8, 25 0mM EDTA, 0.5% SDS) and precipitated with isopropanol. PCR was performed using primers detailed in Table S4. RNA was extracted using TRIzol Reagent (Invitrogen) following manufacturer's recommendations.

The presence of the IR in different *A. thaliana* accessions was determined by amplification of the IR region with flanking primers with Q5 polymerase, followed by Sanger sequencing. For the deletion of the IRs neighboring *PHYC* and *CRY1*, pairs of sgRNAs targeting each locus were cloned in a CRISPR/Cas9 super module (SM) vector as described.[67] Briefly, sgRNAs targeting flanking regions of the IR were designed using the CRISPR-P Web Tool.[79] Each sgRNA was introduced into the shuffle vectors by overlap PCR with Q5 Hi-Fidelity polymerase followed by digestion of the original vector with *DpnI* (Thermo-Fisher Scientific). A destination vector harboring *UBQ10* promoter, pcoCas9, proLacZ:LacZ between both sgRNAs targeting each IR, and At2S3:mCherry for fluorescence selection in seeds was generated with the Green Gate assembly system. Destination vectors were transformed into Col-0 plants, and red fluorescence-positive seeds were isolated as hemizygous seeds. Transgene-free T2 offspring without seed fluorescence were chosen, and plants were tested by PCR and Sanger sequencing to identify IR deletion lines. All primers used are listed in Table S4.

IR detection

Inverted repeats were identified in the Col-0 *Arabidopsis thaliana* genome using *einverted* from the EMBOSS program suite.[32] The following parameters were used: maximum repeat of 1000, a gap penalty of 8, a minimum score threshold of 150, a match score of 3, and a mismatch score of −4.

RNA-seq

RNA-seq library preparation was performed as described (Cambiagno et al., 2021). An in-house scaled-down version of Illumina's TruSeq reaction was used. mRNA was purified with NEBNext Poly(A) Magnetic Isolation Module (New England Biolabs, Ipswich, MA) and heat fragmented with Elute-Prime-Fragment buffer (5× first-strand buffer, 50 ng/mL random primers). For first- and second-strand synthesis SuperScript II Reverse Transcriptase (Thermo Fisher) and DNA polymerase I (NEB) were used, respectively. T4 DNA polymerase, Klenow DNA polymerase, T4 polynucleotide kinase (NEB), and Klenow Fragment (30/50 exo-) (NEB) were used for end repair and A-tailing. Ligation of universal adapters compatible with Nextera barcodes i7 and i5 was performed with T4 DNA Ligase (NEB), and Q5 Polymerase (NEB) was used for PCR enrichment using Nextera i7 and i5 barcodes. SPRI beads were used for DNA purification in each step and size selection of the library preps. 2 × 150 bp paired-end reads were generation on the Illumina HiSeq 3000 platform.

The analysis started by quality trimming and filtering the raw reads with Trimmomatic version 0.36.[68]; They were then aligned to the *Arabidopsis thaliana* reference genome (TAIR10) using STAR version 2.5.2b,[69] which was guided by the gene and exon annotation from Araport V11 201,606,.[80] Samtools version 1.9[70]; was then used to keep only primary alignments with a minimum MAPQ of 3. Read quality before and after trimming was analyzed with FastQC (version 0.11.5; https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and, together with mapping efficiency, they were summarized with MultiQC version 1.7.[71] Read counts on each gene were then calculated with featureCounts version 1.6.2.[72] This pipeline was run with the aid of the Snakemake workflow

engine.[81] Gene counts were used to identify differentially expressed genes with DESeq2[73]; R Core Team 2022) filtering out genes with counts below 10 in all samples.

sRNA-seq

For sRNA-seq library preparation, 1 μg of total RNA was used as input for the TruSeq sRNA Library Preparation kit (Illumina) as described in the TruSeq RNA Sample Preparation v2 Guide (Illumina). A BluePippin System (Sage Science) was used for sRNA library size selections. Sequencing was performed on the Illumina HiSeq 3000 platform.

The small RNA reads generated were first cut to remove 3′adapters using cutadapt (version 1.9.1) and their quality checked using FastQC (version 0.11.4, https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and MultiQC.[71] They were then mapped with bowtie (version 1.1.2;[74]) to *A. thaliana* rRNA, tRNA, snoRNA and snRNA from RFAM (version 14.1,[82]). Unmapped reads were then mapped, also with bowtie, to the *A. thaliana* genome. Statistical analyses were performed in the R statistical programming environment (R Core Team, 2022) and graphics were produced with the ggplot package.

IRs which had 10 or more 24 nucleotide reads in its entire region were considered to have potential sRNA production. Changes in sRNA levels were calculated first calculating reads per million (RPM) mapping reads to the genome in each library, then averaging this value for all replicates in each IR, and then calculating the log2 Fold Change between the RPM in the mutant versus Col-0. For the detection of 24-nt sRNA peaks within genes or opposite to IRs (as presented in Figures S1D and S1E) we used mapped reads as input for MACS2.[83] Peaks were called with a q-value threshold of 0.05, keeping duplicates, a tag size of 24, with the nomodel parameter and an extension size of 1. Then, peak ranges were overlapped either with genes, in order to define genes including peaks, and with a 3000 bp region around the gene and opposite to the IR (opposite peak).

Bisulfite treatment of DNA and library preparation

For BS-seq, DNA extraction was performed with a DNeasy plant Mini Kit (QIAGEN). DNA was sheared to 350 bp by g by Covaris ultrasonication. Libraries were generated with an Illumina TrueSeq DNA Nano Kit. After adaptor ligation, libraries were bisulfite converted with the Epitec Plus DNA Bisulfite Conversion Kit (QIAGEN). Library enrichment was done using Kapa Hifi Uracil + DNA polymerase (Kapa Biosystems, USA). Paired-end reads (2 × 150 bp) were generated on the HiSeq 3000 platform (Illumina).

The analysis of these reads started by quality trimming and filtering them with Trimmomatic (version 0.36;[68]). Then we used the Bismark program[75] to perform the mapping of the reads to the *A. thaliana* Col-0 genome, internally done with Bowtie2,[84] the deduplication of the alignments and the extraction of the methylation results in the three contexts: CG, CHG and CHH. This output was then analyzed in R (R core team, 2022) with the methylKit package.[76] Only Cytosines with at least 4 reads were considered, and each sample was segmented with methSeg and methylation levels were calculated for those including at least 4 Cs. For Col-0, segments were collapsed for replicates using the mergeGRangesData function from the BRGenomics package (https://rdrr.io/bioc/BRGenomics/) and IRs with repeats overlapping segments with more than 10 or more percent of CHG or CHH methylation were considered methylated.

Differential methylation in the *ddc* and *dcl234* mutants was also calculated with the methylKit package. First replicates were combined with the unite function and then differential methylation was calculated with the calculateDiffMeth function, correcting for overdispersion with the MN method, using a q-value threshold of 0.1 and a differential threshold of 15%. Then IRs with repeats overlapping any of these differential segments were considered differentially methylated.

Capture-C assay

For Capture-C (Cap-C), Hi-C was performed as described.[85] Briefly, we collected 1.5 g of plant tissue, and fixed them with 1% formaldehyde. Nuclei were isolated and finally washed with NEB buffer #3. Nuclei penetration was done by resuspending the pellets in 150 μL 0.5% SDS and incubating them at 62°C for 5 min. After that, 435 μl of water and 75 μl of 10% Triton X-100 were added and incubated 37°C for 15 min. NEB buffer #3 was added to 1X, and 50 U of DpnII to digest the chromatin over night at 37°C. Incubating the digested chromatin with 10 U Klenow, dTTP, dATP, dGTP, and biotin-14-dCTP at 37°C for 2 h, cohesive ends were filled. Blunt-end ligation of chromatin was performed by adding blunt-end ligation buffer to 1X and 20 U of T4 DNA ligase at room temperature for 4 h. Nuclei were lysed with SDS buffer (50 mM Tris-HCl, 1% SDS, 10 mM EDTA, pH 8.0) and incubated with 10 μg proteinase K at 55°C for 30 min. To reverse the crosslinking, NaCl was added to reach 0.2 M and the samples were incubated at 65 °C overnight. Hi-C DNA was purified by Phenol-Chloroform-IAA method and RNAse A treated. Hi-C DNA was sheared to 500 bp with a Covaris E220 sonicator. DNA was purified and size selected (longer than 300 bp) using Ampure beads. Unligated biotin was removed in a reaction with 0.1 mM dTTP, 0.1 mM dATP and 5 U T4 DNA polymerase incubated at 20°C for 30 min. DNA was purified with Ampure beads and end-repair and adaptor ligation were performed with the NEBNext Ultra II DNA Library Prep Kit by following the manufacturer's instructions. Biotin affinity purification was then performed by using Dynabeads MyOne Streptavidin C1 beads (Invitrogen). Library amplification was done with Ultra II Q5 Master Mix with universal and selected index primers.

For the capture step, hybridization capture was performed with the MyBaits system (Arbor Biosciences) following the manufacturer's instructions. Baits of 80 nucleotides were designed on each end of the digestion fragments corresponding to the captured regions. These regions included all genes within 3000 bp of the IR and the spacer region up to the IR, excluding it. When a region was surrounded by 2 IRs, it was considered a single captured region.

Finally, Cap-C DNA was paired-end sequenced (2 × 150 bp reads) on an Illumina HiSeq 3000 instrument. The resulting reads were processed with CapC-MAP,[39] a software package specifically designed for Capture-C analysis. Initially, the Col-0 genome was segmented into DpnII restriction fragments. This produced a *bed* file with the coordinates of all restriction fragments. A second coordinate file was then prepared only with restriction fragments overlapping the captured regions (target fragments). Then, CapC-MAP was executed in the "run" mode, which performs all steps of the analysis, namely, adapter trimming, read *in silico* digestion and mapping (done with *bowtie*), deduplication, removal of invalid interactions, generation of pile-up interactions for each target fragment, normalization binning and smoothing of interactions. As additional parameters for this run, we set an exclusion zone of 500 bp, a bin size of 500 bp and a step of 250 bp. The results were then processed with the R package peakC[40] to determine statistically significant interactions. For this we ran the "combined analysis", on both replicates from each genotype, using default parameters: a window-size of 50,000 bp (region to be read), a bin-size of 250 bp, a minimum distance of 500 bp, a threshold of absolute difference of 1, and an alpha value of 0.1 for FDR. Loops were visualized with R and Gviz (Hahne and Ivanek, 2016) and GenomicInteractions.[78]

In order to evaluate consistency between replicates and the effect of sequencing depth, we performed PCA analysis with FAN-C (Figure S6). We used binned data, before and after normalization, to performed the analysis (Figure S6, upper and lower rows, respectively). In addition, we performed the same PCAs on a full re-run starting on "downsampled" fastq files (10M reads each) to independence from the deepness component of the analysis of unnormalized samples (Figure S6, right column). These PCAs show that before normalization depth is important in defining similarity in the plot, with replicates being better arranged according to genotype when starting from 10M read files (i.e., replicates of the same genotype are closer). After normalization, all genotypes are arranged in a remarkably similar way irrespective of whether the PCA was obtained using the full- or the uniform-depth read files. This indicates that the samples contain true genotypic interaction signals, with *ddc* and *dcl234*, which affect the same pathway, being very similar; and that normalization counters the effects of sequencing depth variability.

For the SSIM calculation, raw pileups produced by CapC-MAP were converted to the BG2 format, an extension of the bedGraph format, and then to the cooler format.[86] They were read with FAN-C,[87] and normalized using the VC-SQRT method on 1,000 bp bins. Finally, SSIM values were obtained using CHESS[38] on a region comprising the captured region extended by 10 kb on both extremes, and using a relative window size of 0.1.

3C assay and RT-PCR

3C was performed as described.[25] For detection of loops at *PHV* and *PHYC*, EcoRI (NEB) overnight digestion was performed; for *P5CS1*, *CRY1*, and *PHR1*, XbaI (NEB) was used. For DNA ligation, 100 U of highly concentrated T4 DNA ligase (Thermo) were used at 22°C for 5 h in a 4 mL volume. Reverse crosslinking and proteinase K treatment (QIAGEN) were performed, and phenol/chloroform method was used for DNA purification. For interaction frequency measurement, qPCR was performed using the $2-\Delta\Delta Ct$ method with *ACTIN2* as housekeeping gene.

For quantitative RT-PCR, 1 μg of total RNA was used for reverse transcription reactions using RevertAid RT Reverse Transcription Kit (Thermo Fisher Scientific). qPCR was performed using SYBR green (Thermo Scientific Maxima SYBR Green qPCR Master Mix (2x)). To calculate relative expression levels the $2^{-\Delta\Delta Ct}$ method was used with *ACTIN2* as housekeeping gene.

All primers used are listed in Table S4.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data obtained in the different analysis of the sequencing experiments was further processed and statistically analyzed in R (R Core Team, 2022) using a diversity of packages. Genomic information was handled using GenomicRanges,[88] Biostrings (https://bioconductor.org/packages/Biostrings) GenomicInteractions[78] and rtracklayer[89] packages.

Plots summarizing information were mostly performed with the ggplot and ggpubr packages. Plots of genomic regions were produced with the Gviz[77] package. Circular plots were generated with ciclize.[90]

Hypocotyl length measurements were performed using Fiji software.[91]

Data in Figures 3C and 4 are presented as individual values ±standard error of the mean (SEM). The numbers of biological replicates (n) are provided in each figure legend. Statistical significance was tested using a two-tailed, unpaired Student's t-test with Welch's correction or one-way ANOVA as specified on each figure legend.