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[2,4], Chang Liu[5], Hernán A. Burbano[3,4], Detlef Weigel[4], Pablo A. Manavella[1*]

[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 Biology, Stanford University, Stanford, USA.
[3]Centre for Life's Origin and Evolution, University College London, London, United Kingdom.
[4]Department of Molecular Biology, Max Planck Institute for Biology Tübingen, 72076 Tübingen, Germany.
[5]Department of Epigenetics, Institute of Biology, University of Hohenheim, Garbenstraße 30, 70599 Stuttgart, Germany
[6]Current address: Unidad de Estudios Agropecuarios (UDEA), INTA-CONICET, Córdoba, Argentina.
[7]These authors contributed equally
*To whom correspondence should be addressed. E-mail pablomanavella@ial.santafe-conicet.gov.ar

Abstract

Transposons are mobile elements that are commonly silenced to protect eukaryotic genome integrity. In plants, transposable elements (TEs) can be activated during stress conditions and subsequently insert into gene-rich regions. TE-derived inverted repeats (IRs) are commonly found near plant genes, where they affect host gene expression with potentially positive effects on adaptation. However, the molecular mechanisms by which these IRs control gene expression is unclear in most cases. Here, we identify in the *Arabidopsis thaliana* genome hundreds of IRs located near genes that are transcribed by RNA Polymerase II, resulting in the production of 24-nt small RNAs that trigger methylation of the IRs. The expression of these IRs is associated with drastic changes in the local 3D chromatin organization, which alter the expression pattern of the hosting genes. Notably, the presence and structure of many IRs differ between *A. thaliana* accessions. Capture-C sequencing experiments revealed that such variation changes short-range chromatin interactions, which translates into changes in gene expression patterns. CRISPR/Cas9-mediated disruption of two of such IRs leads to a switch in genome topology and gene expression, with phenotypic consequences. Our data demonstrate that the insertion of an IR near a gene provides an anchor point for chromatin interactions that can 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' (Dubin et al., 2018). Plant genomes are particularly rich in TEs and repetitive elements, which, for example, account for 85% of the maize genome (Schnable et al., 2009). In plants, TEs are commonly silenced through DNA methylation in a process known as RNA-directed DNA methylation (RdDM), which maintains genome integrity (Matzke et al., 2015). To trigger RdDM, short RNA Polymerase IV (RNAPIV)-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) (Matzke *et al.*, 2015; Zhou and Law, 2015). ARGONAUTE 4 (AGO4)-loaded siRNAs then direct de novo methylation of the TE loci by recognizing nascent RNAPV 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 (Lu et al., 2012; Maumus and Quesneville, 2014; SanMiguel et al., 1998). Stress can trigger the activation of TEs and fuel transposition (Baduel et al., 2021; Tittel-Elmer et al., 2010). TEs are thus significant contributors to genetic variation in plant genomes (Baduel *et al.*, 2021; Collier and Largaespada, 2007; Deragon et al.,

2008) and have been postulated as drivers of genome evolution and expansion, as well as developmental plasticity and adaptation (Dubin *et al.*, 2018; Lisch, 2013).

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 retrotransposons (LTRs) (Dubin *et al.*, 2018). MITEs exhibit characteristic terminal inverted repeats (TIRs) and small direct repeats (target site duplications, TSDs), but lack transposase sequences, making them non-autonomous elements (Fattash et al., 2013; Yang et al., 2009). Many TE-derived inverted repeated (IR) insertions may not be classified as MITE as they lack the above-mentioned components of MITEs, either due to 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 (Lu *et al.*, 2012), with the MITEs often changing the expression of neighboring genes (Lu *et al.*, 2012; Underwood et al., 2022; Wu et al., 2022; Xu et al., 2020; Zhang et al., 2016). Based on this, MITEs have been proposed to play important roles in genome evolution and gene expression (Lu *et al.*, 2012). One key feature of MITEs is that their transcripts can fold into hairpin-shaped dsRNAs due to 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 RNAPIV/RDR2 activity (Ariel and Manavella, 2021; Crescente et al., 2022; Cuerda-Gil and Slotkin, 2016; Gagliardi et al., 2019; Sasaki et al., 2014). Thus, transcripts of these MITEs can be initiated from promoters of adjacent genes, triggering their RNA Polymerase II (RNAPII)-dependent DNA methylation. At the sunflower *HaWRKY6* locus, the RNAPII-mediated transcription of a MITE triggers the methylation of its coding region and causes the formation of alternative regulatory short-range chromatin loops in the locus that specifically change its expression (Gagliardi *et al.*, 2019).

Three-dimensional chromatin organization has recently emerged as a critical feature determining genome functionality, fine-tuning gene expression and developmental responses in plants (Domb et al., 2022; Zhang and Wang, 2021). Short-range chromatin loops reflect interaction between relatively close regions of DNA, within a few kb, generally within a single locus or between adjacent loci (Gagliardi and Manavella, 2020). Different from canonical regulation by mC 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, the usage of cryptic termination sites, impaired or enhanced RNAPII elongation, DNA replication and repair (Gagliardi and Manavella, 2020; Grzechnik et al., 2014; Sotelo-Silveira et al., 2018).

112   In this study, we show that TE-derived IR elements located near genes in the
113   *Arabidopsis thaliana* genome can cause rearrangements of the chromatin
114   topology, promoting the formation of short-range chromatin loops. These
115   chromatin interactions, which depend on the production of IR-derived siRNAs and
116   *de novo* DNA methylation, often translate into changes in gene expression. The
117   presence of an IR and its associated chromatin loop near a gene does not cause
118   a uniform regulatory effect, and can either enhance or repress expression
119   depending on the locus and which regions within the locus are part of the loop.
120   Almost one third of the identified gene-associated IRs are not conserved among
121   a set of 216 *A. thaliana* natural accessions. Our data show that polymorphisms
122   in IRs near genes can be coupled with a change in the chromatin topology of the
123   region. These accession-related changes in chromatin landscapes controlled by
124   IRs can be linked to alteration in traits commonly associated with adaptation, such
125   as flowering. In proof-of-concept experiments, we used CRISPR/Cas9 genome
126   editing to mimic the situation in natural accessions lacking specific IRs, and
127   demonstrated that the IRs help to shape chromatin topology, which in turn can
128   control molecular and organismal phenotypes. We found that IRs downstream of
129   *PHYC* and *CRY1* cause the formation of repressive chromatin loops associated
130   with well-defined developmental phenotypes of some natural accessions.
131   Overall, our data demonstrate that TE-derived IRs can produce changes in
132   chromatin topology, gene expression, and ultimately phenotypic changes through
133   their capacity to trigger DNA methylation autonomously. Given the propensity of
134   TEs to activate during stress responses and the tendency of IRs to locate near
135   coding genes, our finding provides a scenario for these TEs to drive local
136   adaptation and domestication by supporting rapid and sometimes drastic
137   changes in 3D chromatin organization and gene activity after single-set
138   mutational events.
139
Results
141
TE-derived IRs located near genes impact the local chromatin topology in Arabidopsis

144   A TE-derived IR element located ~600 bp upstream of the *HaWRKY6* locus in
145   sunflower serves as an anchor point for the formation of two short-range
146   chromatin loops, and thereby promotes changes in local chromatin topology
147   (Gagliardi *et al.*, 2019). The ultimate outcome is the methylation of the locus, due
148   to 24-nt siRNAs produced after RNAPII-dependent transcription of these IRs. In
149   leaves, this leads to the formation of a repressive loop, while it promotes a
150   second, larger, loop that enhances transcription in cotyledons (Gagliardi *et al.*,
151   2019).
152   Because TE-derived IRs are frequently located near genes (Guo et al., 2017), we
153   wondered whether the *HaWRKY6* case was just one example of a more general
154   phenomenon in plants. To evaluate if the insertion of an IR near a gene changes
155   local chromatin topology and gene expression, we first aimed to identify all IRs

156 neighboring protein-coding genes in the *A. thaliana* Col-0 reference genome.
157 Using *einverted* from the EMBOSS program suite (Rice et al., 2000), we found a
158 total of 885 IRs in the *A. thaliana* genome, 634 of then near annotated protein
159 coding genes (222 of which were located within 500 bp upstream or downstream
160 of a protein coding gene, a further 163 between 500 bp and 1,000 bp from a gene,
161 and 249 between 1,000 and 3,000 bp from a gene) (Fig. 1A). IRs were found
162 within 500 bp of 260 unique genes, within 500 to 1,000 bp of 215 unique genes,
163 and within 1,000 to 3,000 bp of 615 unique genes (Fig. 1A). These IRs have a
164 broad genome-wide distribution, with many located in the gene-rich chromosome
165 arms, and others in gene-poor/TE-rich pericentromeric regions (Fig. 1B).
166 Analyzing the overlap with annotated transposable elements (TEs) revealed that
167 68% of the IRs within 3,000 bp of a gene were clearly of TE origin, with most
168 (~44%) from the MuDR superfamily and 18% from the Helitron superfamily (Fig.
169 1C). These percentages remain invariable despite of the distance from the IR to
170 the hosting gene although a moderate enrichment is LTR/Gypsy TEs between
171 IRs not associated with genes (Fig. 1C). Using the plaNETseq dataset of RNAPII
172 associated nascent transcripts (Kindgren et al., 2020), we found that more than
173 half of the identified gene-associated IRs are transcribed by RNAPII (Fig. 1D).
174 Supporting the idea that these IRs are transcribed from promoters of nearby
175 genes, the fraction of IRs transcribed by RNAPII increases the closer the IRs are
176 to a gene (Fig. 1D). Conversely, IRs located far from annotated genes were less
177 likely to give rise to RNAPII transcripts. Instead, their transcription likely depends
178 on the canonical RNAPIV/RDR2 RdDM pathway (Fig. 1D).

179 With small RNA (sRNA)- and bisulfite (BS)-sequencing, and paralleling what we
180 observed in the case of the *HaWRKY6* adjacent IR, we found that most of the
181 identified IRs near genes produce RNAPII transcripts (Fig. 1D), giving rise to 24-
182 nt siRNAs (486/634 IRs 3000bp away from genes), and are associated with CHG
183 and CHH methylation (Fig. 1E). We often observed an additional peak of
184 methylation, which could represent a second anchor point for chromatin loop
185 formation, either on the opposite border or inside of many genes located near IRs
186 (Fig. S1), once more providing a scenario similar to that of the *HaWRKY6* locus
187 where two close-by methylated regions served as anchor points for regulatory
188 short-range chromatin loops.

189 To investigate whether the identified IRs, especially those producing siRNAs and
190 DNA methylation, impact local chromatin organization, we extracted RNA and
191 DNA from Col-0 wild-type plants, triple *dcl2, dcl3, dcl4* (*dcl234*)mutants (which
192 are in impaired in 24-nt siRNA production) (Lu et al., 2006), and triple *drm1, drm2,*
193 *cmt3* (ddc) mutants (impaired in CHH methylation) (Kurihara et al., 2008), and
194 performed RNA-seq, sRNA-seq, BS-seq, and Capture-C. As the sequencing
195 depth required to detect short-range chromatin interaction through standard Hi-C
196 would be enormous, we selected 290 loci containing IRs and 40 control loci and
197 performed a Capture-C experiment to only focus on these regions and increase
198 the chances to detect local chromatin loops. A test mapping of Capture C reads
199 confirmed the enrichment on the captured regions compared to the input HiC

200 samples (Fig. S2A). Collectively, the designed probes cover ~1% of the genome.
201 After Capture-C we increased the ratio of reads mapping the targeted regions 40
202 times in average (Fig. S2B).
203 sRNA-seq revealed that 486 out of the 634 IRs 3000 bp from genes produce 24
204 nt siRNAs (Fig. 2A), with siRNA levels severely reduced in both *dcl234* and *ddc*
205 mutants, as expected for this class of small RNAs (Fig. 2A). Both CHG (~50%)
206 and CHH (~70%) methylated regions strongly overlapped with IRs within 500 bp
207 from genes (Fig. 1E), and methylation in both contexts is reduced in both in
208 *dcl234* and *ddc* for many IRs as expected by the reduction in siRNAs in these
209 genotypes (Fig. 2B). The proportion of IRs with reduced methylation in the
210 mutants is slightly higher for those within 500 bp of genes compared to the other
211 analyzed distance windows (Fig S2C). Changes in siRNAs and methylation were
212 correlated, consistent with a drop of siRNAs leading to reduced methylation in
213 RdDM impaired mutants (Fig. 2C). Altogether, these data suggest that a large
214 proportion of IRs located near genes and transcribed by RNAPII may be able to
215 trigger DNA methylation in *cis* through the non-canonical RdDM pathway.
216 As in the case of the *HaWRKY6* locus in sunflower, our data suggest that IRs
217 located near genes can act as regulatory elements changing the epigenetic
218 landscape of the region. In order to assess the impact of IR methylation on the
219 surrounding chromatin organization, we used the software CHESS to compare
220 the structural similarity (SSIM) of the IR-host regions between wild type and
221 mutants (Galan et al., 2020). We also statistically determined the specific anchor
222 points of chromatin loops in each sample using capC-Map (Buckle et al., 2019)
223 in combination with peakC (Geeven et al., 2018). Analyzing individual loops, we
224 found clear alterations in the chromatin topology in several randomly picked loci
225 (Fig. 2D and S3). To compare the differential methylation-related changes in
226 loops formed near IRs, we calculated the SSIM for each captured region plus 10
227 kb on each end in wild type and siRNA- or methylation-deficient mutants. The
228 choice of a global similarity measure, the SSIM, to study short-range chromatin
229 changes, rather than comparing individual loops, aimed at increasing the power
230 to detect reliable differences, as random interactions increase with shorter
231 distances and increase the methodological background noise. Moreover, other
232 interactions, caused by dimerization of DNA-bound transcription factors or
233 nucleosome packing, can also impact such analyses. Chromatin organization of
234 the IR-containing loci, which should have SSIM values ~1 if similar between the
235 genotypes, were often changed in *ddc* and *dcl234* mutants (Fig. 2E), indicating
236 that (IR-triggered) methylation has a substantial effect on the local 3D topology.
237 Such differences are less pronounced when comparing *ddc* with *dcl234* as could
238 be expected from the methylation deficiency observed in both, therefore lacking
239 the anchor points for loop formation (Fig. 2E).
240 Changes in chromatin topology can affect gene expression (Domb *et al.*, 2022).
241 Given the alterations in chromatin topology that we found, we wondered whether
242 they impact gene expression. Formation of a chromatin loops may repress or

243 activate gene transcription, or even trigger production of alternative transcripts
244 (Gagliardi and Manavella, 2020).
245 We found 4,305 differentially expressed genes (DEGs) in the *ddc* mutant in com-
246 parison with Col-0, and 3,636 DEGs in *dcl234*, many of them overlapping be-
247 tween the two mutants (Fig. S4A). When we compared the SSIMs in regions with
248 DEGs with SSIMs in regions without DEGs we found topological differences
249 (lower SSIM) to be significantly increased in regions that included DEGs (Fig.
250 2F). We then split those differences into regions linked to differentially and not
251 differentially methylated IRs, which revealed greater topological differences in
252 *ddc* mutants for regions both including DEGs and differentially methylated IRs
253 (Fig. S4B). Methylation seemed to affect the SSIM correlation with DEGs less
254 clearly in *dcl234* (Fig. S4C). Loci with altered topology in *ddc* and *dcl234* included
255 both up- and downregulated genes (Fig. 2F). Thus, changes in chromatin topol-
256 ogy caused by an impaired RdDM machinery could be linked to opposite changes
257 in gene expression depending on the loci. This observation, also detected in rice
258 for genes adjacent to MITEs (Lu *et al.*, 2012), fits with short-range chromatin
259 loops affecting RNAPII activity in different ways, depending on which part of the
260 gene is included in the loop (Gagliardi and Manavella, 2020).
261 In summary, our data indicated that the insertion of an IR near a gene can trigger
262 changes in the local chromatin topology that ultimately affect gene expression in
263 a locus- and loop-dependent way.
264
265 **The absence of IRs near genes between Arabidopsis accessions causes**
266 **natural variation in the local chromatin topology**
267 TEs are commonly silenced in plants as a means of protecting genome integrity.
268 However, under extreme stress conditions TE transcription can be reactivated,
269 giving TEs the potential to jump in the genome in a process that has been
270 proposed to help adaptation to new environments (Ito et al., 2011; Tittel-Elmer *et*
271 *al.*, 2010). We wondered whether the IRs we were studying could have adaptative
272 potential to change the structure of the RNAs produced by a locus or their
273 expression, through changes in the 3D chromatin organization.
274 To test this hypothesis, we first used published TE polymorphism datasets (Stuart
275 et al., 2016) to detect variations in the TE content of 216 *A. thaliana* genomes
276 (Schmitz et al., 2013). Narrowing this down further to the 634 IRs located within
277 3,000 bp upstream or downstream of annotated genes in the Col-0 reference
278 genome, we found 193 of these 634 IRs to be variable between the 216
279 accessions studied (Fig. 3A). IRs without a clear TE origin are underrepresented
280 in the collection of polymorphic IRs (Fig. 3A). These 193 IRs are located within
281 3,000 bp of 368 annotated genes. This implies that each IR can influence two or
282 more adjacent genes in many cases, an observation that is not surprising given
283 the compact nature of *A. thaliana* genomes. To investigate the effect that the
284 variation in the IR content may have on local chromatin topology and gene
285 expression, we repeated the Capture-C, RNA-seq, BS-seq, and sRNA-seq
286 analyses in the Ba1 and Hod accessions. We selected these two accessions

287 based on the number of variable IRs present in each of them (46 and 31 IRs
288 present in Col-0 are missing in Hod and Ba-1, respectively, with 18 missing in
289 both accessions). With the Capture-C experiment, we captured 290 regions
290 comprising the genomic sequences around variable IRs which included adjacent
291 genes. The analysis of individual loci revealed that the presence of an IR at a
292 locus correlated with the accumulation of 24 nt siRNAs and with distinctive short-
293 range chromatin loops (Fig. 3B and S3).
294 We chose five potential candidate genes (*PHV*, *PHYC*, *P5CS1*, *PHR1, CRY1*) to
295 evaluate the effect of polymorphic IRs on loop formation and gene expression.
296 We corroborated by Sanger sequencing that each of the loci has IRs in Col-0, but
297 not in the indicated accessions (Fig. 4B, 4G, andS5). We then used Chromatin
298 Conformation Capture (3C) followed by qPCR to confirm and quantify the
299 formation of IR-dependent chromatin loops at these loci, and RT-qPCR to
300 measure correlation with gene expression in each locus (Fig. 3C, 4C, and 4I). We
301 detected alternative chromatin loops at *PHV*, *P5CS1*, and *PHR1* (Fig. 3C). In the
302 case of *PHV*, loop 1 appeared to form independently of the associated IR, but
303 loops 2 and 3 were missing in Ba-1, the accession without the IR, or in *dcl234*
304 mutants (Fig. 3C). For *P5CS1*, loop 2 appeared independently of the presence of
305 the IR while the formation of an intragenic loop 1 correlated with the presence of
306 the IR or a functional RdDM machinery (Fig. 3C). Both for *PHV* and *P5CS1* the
307 absence of a loop-triggering IR appeared to be associated with enhanced gene
308 expression (Fig. 3C). In the case of *PHR1*, we found that the formation of a loop
309 1 depends on the presence of the IR, which is missing in the Hod accessions, but
310 loop 2 was only formed when the IR is missing, probably reflecting a hierarchy of
311 chromatin interactions controlled by the IR (Fig. 3C). Contrary to the results with
312 *PHV* and *P5CS1*, the absence of the IR near *PHR1* caused repression of the
313 gene.
314 When we analyzed the genome-wide effects of insertional IR polymorphisms over
315 genome topology using CHESS, we also found a drastic change in chromatin
316 folding associated with the presence/absence of an IR (Fig. 3D). For this analysis
317 we calculated the difference between the SSIMs obtained from the comparison
318 of Ba-1 and Hod genome topology against Col-0, as recommended by the
319 authors of CHESS, to have a common reference to score differences between
320 these accessions (Galan *et al.*, 2020). As it can be observed in Figure 3D, the
321 difference in SSIMs are greater when Ba-1 shares the IR with Col-0, having a
322 higher SSIM, but it is absent in Hod resulting in a lower SSIM and an increased
323 positive difference.
324 These data suggest that the insertion of an IR near coding genes can have an
325 impact on local chromatin topology that affects gene expression. Aiming to
326 explore whether this could potentially have adaptive consequences, we asked
327 whether there is a correlation between the presence/absence of an IR near a
328 gene and well-defined phenotypic traits recorded in the Arapheno database
329 (Seren et al., 2017). This analysis revealed an association between IR insertional
330 polymorphisms near some genes with a given phenotype related to flowering, a

331 typically adaptative trait (Fig. 3E). This observation suggests that insertion of an
332 IR near a gene can not only impact the local chromatin topology, but that it may
333 also have adaptive implications by changing the phenotypes controlled by the
334 adjacent gene.
335
336 **The insertion of an IR near genes causes strong phenotypic effects**
337 Our data suggest that the insertion of an IR near a gene can be a significant event
338 during adaptative evolution. However, this association may not be caused by the
339 IR itself, but by adjacent polymorphisms that may be in linkage disequilibrium with
340 the IR. In addition, even if the association is with the IR itself, it may not involve
341 changes in the genome topology. To study this possibility we selected two loci,
342 *PHYC* and *CRY1,* which display natural variation in the presence/absence of a
343 close-by IR that is associated with developmental phenotypes (Fig. 3E). Both
344 *PHYC* and *CRY1* associated IRs are located downstream of the transcription
345 termination site (TTS) (Fig. 4A and 4F), thus making it less likely that they act
346 directly on the promoter of the genes as general regulatory elements.
347 *PHYC* encodes a photoreceptor capable of sensing red light (R) and far-red light
348 (FR) and is implicated in several developmental transitions, such as flowering,
349 seed germination and hypocotyl elongation (Chen et al., 2014; Kippes et al.,
350 2020; Li et al., 2020; Nishida et al., 2013). In Col-0 an IR is located ~500 bp
351 downstream of the TTS of *PHYC*, while it is missing in several natural accessions,
352 including, for example, Ey15-2 (Fig. 4A and 4B). To test whether this IR impacts
353 *PHYC* expression and related phenotypes, we used a CRISPR/Cas9 strategy to
354 delete a fragment of the IR, with which we can disrupt dsRNA formation without
355 completely removing the IR sequence (Fig. 4A). Three homozygous lines were
356 obtained with a deletion of the IR fragment (Fig. 4B). Supporting a role for this IR
357 in modulating genome topology and gene expression, we detected a chromatin
358 loop encompassing the entire *PHYC* gene in Col-0 plants that was absent in the
359 CRISPR mutant lines and in the Ey15-2 accession, which lacks the IR (Fig. 4C).
360 The absence of the IR, both in Ey15-2 and in the CRISPR lines, correlated with
361 higher expression of *PHYC*, indicating that the IR promotes the formation of a
362 repressive loop in Col-0 (Fig. 4C). Both, the CRISPR lines and Ey15-2, had
363 altered developmental responses related to known *PHYC* functions, including
364 delayed flowering and shortened hypocotyls under a continuous red-light
365 treatment (Fig. 4D and 4E). Altogether, our data indicate that the presence of the
366 IR next to *PHYC* has a substantial impact on gene regulation, by enabling the
367 formation of a short-range chromatin loop that represses gene expression,
368 thereby contributing to the differential response of natural accessions to light
369 signals.

370 CRY1 is a blue light receptor and participates predominately in the regulation of
371 blue-light inhibition of hypocotyl elongation and anthocyanin production (Ahmad
372 et al., 1998; Ahmad et al., 1995; Liu et al., 2022). Two IRs are located near the
373 gene, ~2,000 bp upstream of the transcription start site (TSS) and ~550 bp

374 downstream of the TTS (Fig. 4F). The second IR is variable among *A. thaliana*
375 accession, missing, for example, in Hod (Fig. 4G). Using CRISPR/Cas9
376 engineering, heterozygous lines could be obtained missing a fragment of the IR
377 located downstream of the TSS (with homozygous lines apparently not being
378 viable) (Fig. 4H). Similar to *PHYC*, we detected a chromatin loop that brings
379 together the borders of the *CRY1* gene only in those plants containing the IR and
380 that represses the expression of the gene (Fig. 4I). The absence of the IR, both
381 in the CRISPR lines or in the natural accession Hod, correlated with higher
382 expression levels of *CRY1* and shorter hypocotyls under blue light (Fig. 4I and
383 4J), a phenotype previously described in *CRY1* overexpressing lines (He et al.,
384 2019; Liu *et al.*, 2022).
385

Discussion

387 How organisms can adapt to a rapidly changing environment is one of the most
388 interesting questions in evolutionary biology (Barrett and Schluter, 2008;
389 Hermisson and Pennings, 2005). SNPs have been the main focus of many
390 genomic studies aiming to assess the evolutionary potential of mutations, but TEs
391 can be particularly powerful actors in rapid adaptation, as single transposition
392 events can have potentially wide-ranging consequences on gene expression and
393 derived phenotypes (Baduel *et al.*, 2021). On one side, the broad distribution of
394 TEs across the genome facilitates the generation of chromosomal
395 rearrangements through ectopic recombination. Even more significantly, the
396 mobilization of TEs can disrupt, modify or even change the expression of genes
397 in various ways that could generate a favorable adaptative trait (Dubin *et al.*,
398 2018; Friedli and Trono, 2015). New alleles caused by TE insertions have been
399 proposed to guarantee a consistent supply of potentially adaptable variants in
400 response to the environment (Baduel *et al.*, 2021). Still, many TEs inserted within
401 gene-rich areas are quickly purged, according to population genomic surveys of
402 TE polymorphisms (Quadrana et al., 2016). This is in agreement with
403 transposition tending to produce alleles with negative effects.
404 Contrary to autonomous TEs, MITEs were found to be distributed on
405 chromosome arms in plants, highly associated with genes, and frequently
406 transcribed with adjacent genes (Kuang et al., 2009; Lu *et al.*, 2012; Oki et al.,
407 2008). In agreement with this, our study identified numerous TE-derived IRs near
408 coding genes in *A. thaliana*. Different from rice, where more than half of the genes
409 are associated with MITEs (Lu *et al.*, 2012), fewer genes are associated with IRs
410 in *A. thaliana*. This observation is not surprising as Arabidopsis is an outlier
411 regarding TE content among plants, with only 15% of its genome represented by
412 TEs. In comparison, they account for 85% of maize and up to ~40% of rice
413 genomes (Arabidopsis Genome, 2000; Li et al., 2017; Schnable *et al.*, 2009).
414 The association between MITEs and protein coding genes suggested that these
415 TEs may play essential roles in genome evolution. Current evidence suggests
416 that siRNA-triggered TE methylation tends to cause the repression of neighboring
417 genes (Hollister and Gaut, 2009; Hollister et al., 2011). However, in the case of

MITEs, both positive and negative effects on the expression of host genes have been reported (Gagliardi *et al.*, 2019; Underwood *et al.*, 2022; Wu *et al.*, 2022; Xu *et al.*, 2020; Zhang *et al.*, 2016). The weak correlation between methylation of MITEs and the expression of adjacent genes can now be better explained in the light of our findings showing that IRs located near genes affect the 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 (Gagliardi and Manavella, 2020).

MITEs have the potential to transpose into various locations in the genome resulting in the presence/absence (insertional) polymorphisms between genotypes (Lu *et al.*, 2012; Lyons et al., 2008). Such polymorphism can be caused by the 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 3,000bp of protein coding genes present insertional polymorphism between 216 *A. thaliana* natural accessions. Our results also indicate that such natural variation of 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 the 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 coopt the promoters of adjacent genes to produce RNAPII-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 adaptative evolution.

Modelling has suggested faster generation of large-effect alleles due to larger transposition rates in specific populations in response to global warming (Baduel *et al.*, 2021). Consistent with this scenario, and in a world of a rapidly changing climate, the discovery of IRs as elements shaping the 3D chromatin organization and driving genome adaptation is of great interest. The 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.

Material and methods

Plant material and growth conditions

Arabidopsis thaliana accessions Col 0, Hod, Ba1, EY15-2, *dcl234* (Lu *et al.*, 2006), and *ddc* (Kurihara *et al.*, 2008) mutants and CRISPR mutant lines were grown at 23ºC in long day (16/8 hours light/dark). Blue and red-light experiments were carried out as follows: 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 23ºC for 3 hours and subsequently transferred to red (20 µmol m^{-2} s^{-1}) or blue light (5 µmol m^{-2} s^{-1}) in LED chambers. Hypocotyl measurements were taken at four days using ImageJ (http://rsb.info.nih.gov/ij).

To obtain plants with genomic fragments deletions a CRISPR/Cas9 vector toolbox (Wu et al., 2018) was used. Specific sgRNAs, described in Supplemental Table S1, were designed to obtain the IRs 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.

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 S1. RNA was extracted using TRIzol^{TM} Reagent (Invitrogen) following manufacturer's recommendations.

The presence of the IR in different *A. thaliana* accessions was determine 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 (Wu *et al.*, 2018). Briefly, sgRNAs targeting flanking regions of the IR were designed using the CRISPR-P Web Tool (Lei et al., 2014). 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 Scientific). A destination vector harboring *UBQ10* promoter, pcoCas9, proLacZ:LacZ between both sgRNAs targeting each IR, and At2S3:mCherry for fluorescence selection in seeds were 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 S1

IR detection

Inverted repeats were identified in the Col-0 *Arabidopsis thaliana* genome using einverted, from the EMBOSS program suite (Rice *et al.*, 2000). The paremeters used were: maximum repeat of 1000, a gap penality 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

12

506 buffer (5x first-strand buffer, 50 ng/ml random primers). For first- and second-
507 strand synthesis SuperScript II Reverse Transcriptase (Thermo Fisher) and DNA
508 polymerase I (NEB) were used, respectively. T4 DNA polymerase, Klenow DNA
509 polymerase, T4 polynucleotide kinase (NEB), and Klenow Fragment (30/50 exo-)
510 (NEB) were used for end repair and A-tailing. Ligation of universal adapters
511 compatible with Nextera barcodes i7 and i5 was performed with T4 DNA Ligase
512 (NEB), and Q5 Polymerase (NEB) was used for PCR enrichment using Nextera
513 i7 and i5 barcodes. SPRI beads were used for DNA purification in each step and
514 size selection of the library preps. 2 x 150 bp paired-end reads were generation
515 on the Illumina HiSeq 3000 platform.

516 The analysis started by quality trimming and filtering the raw reads with
517 Trimmomatic version 0.36; (Bolger et al., 2014. They were then aligned to the
518 Arabidopsis thaliana genome (TAIR10) using STAR version 2.5.2b, {Dobin, 2013
519 #73), which was guided by the gene and exon annotation from Araport V11
520 201606, (Pasha et al., 2020). Samtools version 1.; (Li et al., 2009) was then used
521 to keep only primary alignments with a minimum MAPQ of 3. Read quality before
522 and after trimming was analyzed with FastQC (version 0.11.5;
523 https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and, together with
524 mapping efficiency, they were summarized with MultiQC version 1.7 (Ewels et
525 al., 2016). Read counts on each gene were then calculated with featureCounts
526 version 1.6.2 (Liao et al., 2014). This pipeline was run with the aid of the
527 Snakemake workflow engine (Koster and Rahmann, 2012). Gene counts were
528 used to identify differentially expressed genes with DESeq2 (Love et al., 2014);
529 R Core Team 2022) filtering out genes with counts below 10 in all samples.

530
531 **sRNA-seq**
532 For sRNA-seq library preparation, 1 µg of total RNA was used as input for
533 the TruSeq sRNA Library Preparation kit (Illumina) as described in the TruSeq
534 RNA Sample Preparation v2 Guide (Illumina). BluePippin System (Sage Science)
535 was used for sRNA library size selections. Sequencing was performed on the
536 Illumina HiSeq 3000 platform.

537 The small RNA reads generated were first cut to remove 3´adapters using
538 cutadapt (version 1.9.1) and their quality checked using FastQC (version 0.11.4,
539 https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and MultiQC
540 (Ewels *et al.*, 2016). They were then mapped with bowtie (version 1.1.2;
541 (Langmead et al., 2009)) to *A. thaliana* rRNA, tRNA, snoRNA and snRNA from
542 RFAM (version 14.1, (Kalvari et al., 2018). Unmapped reads were then mapped,
543 also with bowtie, to the *A. thaliana* genome. Statistical analyses were performed
544 in the R statistical programming environment (R Core Team, 2022) and graphics
545 were produced with the ggplot package.

546 IRs which had 10 or more 24 nucleotide reads in its entire region were
547 considered to have potential sRNA production. Changes in sRNA levels were
548 calculated first calculating reads per million (RPM) mapping reads to the genome

549 in each library, then averaging this value for all replicates in each IR, and then
550 calculating the log2 Fold Change between the RPM in the mutant versus Col-0.
551
552 **Bisulfite treatment of DNA and library preparation**
553 For BS-seq, DNA extraction was performed with DNeasy plant Mini Kit
554 (QUIAGEN). DNA was sheared to 350 bp by g by Covaris ultrasonication.
555 Libraries were generated with Illumina TrueSeq DNA Nano Kit. After adaptor
556 ligation, libraries were bisulfite converted with the Epitec Plus DNA Bisulfite
557 Conversion Kit (QUIAGEN). Library enrichment was done using Kapa Hifi Uracil+
558 DNA polymerase (Kapa Biosystems, USA). Paired-end reads (2 x 150 bp) were
559 generation on the HiSeq 3000 platform (Illumina).
560 The analysis of these reads started by quality trimming and filtering them
561 with Trimmomatic (version 0.36; (Bolger *et al.*, 2014)). Then we used the Bismark
562 program (Krueger and Andrews, 2011) to perform the mapping of the reads to
563 the *A. thaliana* Col-0 genome, internally done with Bowtie2 (Langmead and
564 Salzberg, 2012), the deduplication of the alignments and the extraction of the
565 methylation results in the three contexts: CG, CHG and CHH. This output was
566 then analyzed in R (R core team, 2022) with the methylKit package (Akalin et al.,
567 2012). Only Citosines with at least 4 reads were considered, and each sample
568 was segmented with methSeg and methylation levels were calculated for those
569 including at least 4 Cs. For Col-0, segments were collapsed for replicates using
570 the mergeGRangesData function from the BRGenomics package
571 (https://rdrr.io/bioc/BRGenomics/) and IRs with repeats overlapping segments
572 with more than 10 or more percent of CHG or CHH methylation were considered
573 methylated.
574 Differential methylation in the *ddc* and *dcl234* mutants was also calculated
575 with the mehtylKit package. First replicates were combined with the unite function
576 and then differential methylation calculated with the calculateDiffMeth function,
577 correcting for overdispertion with the MN method, using a q-value threshold of
578 0.1 and a differential threshold of 15 %. Then IRs with repeats overlapping any
579 of these differential segments was considered differentially methylated.
580
581 **Capture-C assay**
582 For Capture-C, Hi-C was performed as described (Liu, 2017). Briefly, we
583 collected 1.5 g of plant tissue, and fixed them with 1% formaldehyde. Nuclei were
584 isolated and finally washed with NEB buffer #3. Nuclei penetration was done by
585 resuspending the pellet in 150 µl 0.5% SDS and incubating them at 62 ºC for 5
586 min. After that, 435 ul of water and 75 ul of 10% Triton X-100 were added and
587 incubated 37 ºC for 15 min. NEB buffer #3 was added to 1X, and 50 U of DpnII
588 to digest the chromatin over night at 37°C. Incubating the digested chromatin with
589 10 U Klenow, dTTP, dATP, dGTP, and biotin-14-dCTP at 37°C for 2 h, cohesive
590 ends were filled. Blunt-end ligation of chromatin was performed by adding blunt-
591 end ligation buffer to 1X and 20 U of T4 DNA ligase at room temperature for 4 h.
592 Nuclei were lysed with SDS buffer (50 mM Tris-HCl, 1% SDS, 10 mM EDTA,

593 pH 8.0) and incubated with 10 μg proteinase K at 55°C for 30 min. To reverse the
594 crosslinking, NaCl was added to reach 0.2 M and the samples were incubated at
595 65 °C overnight. Hi-C DNA was purified by Phenol-Chloroform-IAA method and
596 RNAse A treated. Hi-C DNA was sheared to 500 bp with a Covaris E220
597 sonicator. DNA was purified and size selected (longer than 300 bp) using Ampure
598 beads. Unligated biotin was removed in a reaction with 0.1 mM dTTP, 0.1 mM
599 dATP and 5 U T4 DNA polymerase incubated at 20°C for 30 min. DNA was
600 purified with Ampure beads and end-repair and adaptor ligation were performed
601 with the NEBNext® Ultra™ II DNA Library Prep Kit by following the
602 manufacturer's instructions. Biotin affinity purification was then performed by
603 using Dynabeads MyOne Streptavidin C1 beads (Invitrogen). Library
604 amplification was done with Ultra II Q5 Master Mix with universal and selected
605 index primers.

606   For the Capture step, hybridization capture was performed with the
607 MyBaits system (Arbor Biosciences) following the manufacturer's instructions.
608 Baits of 80 nucleotides were designed on each end of the digestion fragments
609 corresponding to the captured regions. These regions included all genes within
610 3000 bp of the IR and the spacer region up to the IR, excluding it. When a region
611 was surrounded by 2 IRs, it was considered a single captured region.

612   Finally, Capture-C DNA was pared-end sequenced (2 × 150 bp reads) on
613 an Illumina HiSeq 3000. The resulting reads were processed with capC-MAP
614 (Buckle *et al.*, 2019), which performs the *in silico* genome digestion, read
615 alignment, the pile-up of interactions, and can generate normalized, binned and
616 smoothed profiles of interaction for each target. For this the Col-0 genome was
617 used, an exclusion zone of 500 bp, a bin size of 500 bp and a step of 250 bp. The
618 results were then processed with the R package peakC (Geeven *et al.*, 2018) to
619 determine statistically significant interactions. These loops were visualized with
620 the aid of the R package and Gviz (Hahne and Ivanek, 2016).

621   For the SSIM calculation, raw pileups were first normalized with FAN-C
622 (Kruse et al., 2020), using the VC-SQRT method on 1000 bp bins. The SSIM
623 value was obtained on a region comprising the captured region extended by 10
624 kb on both extremes, and using a relative window size of 0.1.

625

626 **Data processing, plotting and statistical analysis**

627   Data obtained in the different analysis of the sequencing experiments was
628 further processed and statistically analyzed in R (R Core Team, 2022) using a
629 diversity of packages. Genomic information was handled using GenomicRanges
630 (Lawrence et al., 2013), Biostrings
631 (https://bioconductor.org/packages/Biostrings) GenomicInteractions (Harmston
632 et al., 2015) and rtracklayer (Lawrence et al., 2009) packages.

633   Plots summarizing information were mostly performed with the ggplot and
634 ggpubr packages. Plots of genomic regions were produced with the Gviz (Hahne
635 and Ivanek, 2016) package. Circular plots were generated with ciclize (Gu et al.,
636 2014).

Data availability

All sequencing data were deposited at the European Nucleotide Archive (https://www.ebi.ac.uk/ena/) public repository with accession PRJEB53956.

3C assay and RT-PCR

3C assay was performed as described (Gagliardi *et al.*, 2019). 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 (Invitrogen) were performed, and phenol/chloroform method was used for DNA purification. For interaction frequency measurement, qPCR was performed using *ACTIN2* as housekeeping gene. All primers used are listed in Table S1.

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)). Three biological replicates were used to calculate the standard error of the mean. Standard error of the mean (SEM) was calculated using propagation of error of the $2^{-\Delta\Delta Ct}$ values and expressed in figures as two times the SEM. Statistical significance was tested using a two-tailed, unpaired Student's t-test. All primers used are listed in Table S1.

Author contribution

A.L.A. performed the majority of the analyses. D.A.C. prepared the libraries for sequencing and Capture-C experiments. P.L.L. and H.A.B. helped with the design of the Capture-C probes. P.L.L. performed the Capture experiment. R.M. and D.A.C validated the chromatin loops formations and created the CRISPR/CAS9 mutant lines. R.M. characterized the CRISPR/Cas9 mutant lines and performed validation experiments. A.L.A., D.A.C., P.A.M., and D.W. conceived this 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.

Footnotes

680
681   The authors declare no competing interest.

682
683   **References**

684

685   Ahmad, M., Jarillo, J.A., and Cashmore, A.R. (1998). Chimeric proteins between
686   cry1 and cry2 Arabidopsis blue light photoreceptors indicate overlapping
687   functions and varying protein stability. Plant Cell *10*, 197-207.
688   10.1105/tpc.10.2.197.
689   Ahmad, M., Lin, C., and Cashmore, A.R. (1995). Mutations throughout an
690   Arabidopsis blue-light photoreceptor impair blue-light-responsive anthocyanin
691   accumulation and inhibition of hypocotyl elongation. Plant J *8*, 653-658.
692   10.1046/j.1365-313x.1995.08050653.x.
693   Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E.,
694   Melnick, A., and Mason, C.E. (2012). methylKit: a comprehensive R package for
695   the analysis of genome-wide DNA methylation profiles. Genome Biol *13*, R87.
696   10.1186/gb-2012-13-10-r87.
697   Arabidopsis Genome, I. (2000). Analysis of the genome sequence of the
698   flowering plant Arabidopsis thaliana. Nature *408*, 796-815. 10.1038/35048692.
699   Ariel, F.D., and Manavella, P.A. (2021). When junk DNA turns functional:
700   transposon-derived non-coding RNAs in plants. J Exp Bot *72*, 4132-4143.
701   10.1093/jxb/erab073.
702   Baduel, P., Leduque, B., Ignace, A., Gy, I., Gil, J., Jr., Loudet, O., Colot, V., and
703   Quadrana, L. (2021). Genetic and environmental modulation of transposition
704   shapes the evolutionary potential of Arabidopsis thaliana. Genome Biol *22*, 138.
705   10.1186/s13059-021-02348-5.
706   Barrett, R.D., and Schluter, D. (2008). Adaptation from standing genetic variation.
707   Trends Ecol Evol *23*, 38-44. 10.1016/j.tree.2007.09.008.
708   Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer
709   for Illumina sequence data. Bioinformatics *30*, 2114-2120.
710   10.1093/bioinformatics/btu170.
711   Buckle, A., Gilbert, N., Marenduzzo, D., and Brackley, C.A. (2019). capC-MAP:
712   software for analysis of Capture-C data. Bioinformatics *35*, 4773-4775.
713   10.1093/bioinformatics/btz480.
714   Cambiagno, D.A., Giudicatti, A.J., Arce, A.L., Gagliardi, D., Li, L., Yuan, W.,
715   Lundberg, D.S., Weigel, D., and Manavella, P.A. (2021). HASTY modulates
716   miRNA biogenesis by linking pri-miRNA transcription and processing. Mol Plant
717   *14*, 426-439. 10.1016/j.molp.2020.12.019.
718   Chen, A., Li, C., Hu, W., Lau, M.Y., Lin, H., Rockwell, N.C., Martin, S.S.,
719   Jernstedt, J.A., Lagarias, J.C., and Dubcovsky, J. (2014). Phytochrome C plays
720   a major role in the acceleration of wheat flowering under long-day photoperiod.
721   Proc Natl Acad Sci U S A *111*, 10037-10044. 10.1073/pnas.1409795111.
722   Collier, L.S., and Largaespada, D.A. (2007). Transposable elements and the
723   dynamic somatic genome. Genome Biol *8 Suppl 1*, S5. 10.1186/gb-2007-8-s1-
724   s5.
725   Crescente, J.M., Zavallo, D., Del Vas, M., Asurmendi, S., Helguera, M.,
726   Fernandez, E., and Vanzetti, L.S. (2022). Genome-wide identification of MITE-
727   derived microRNAs and their targets in bread wheat. BMC Genomics *23*, 154.
728   10.1186/s12864-022-08364-4.

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

Deragon, J.M., Casacuberta, J.M., and Panaud, O. (2008). Plant transposable elements. Genome Dyn *4*, 69-82. 10.1159/000126007.

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. 10.1146/annurev-arplant-102720-022810.

Dubin, M.J., Mittelsten Scheid, O., and Becker, C. (2018). Transposons: a blessing curse. Curr Opin Plant Biol *42*, 23-29. 10.1016/j.pbi.2018.01.003.

Ewels, P., Magnusson, M., Lundin, S., and Kaller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics *32*, 3047-3048. 10.1093/bioinformatics/btw354.

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. 10.1139/gen-2012-0174.

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. 10.1146/annurev-cellbio-100814-125514.

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 U S A *116*, 17578-17583. 10.1073/pnas.1903131116.

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

Galan, S., Machnik, N., Kruse, K., Diaz, 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. 10.1038/s41588-020-00712-y.

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. 10.1093/nar/gky443.

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. 10.1016/j.tibs.2014.05.001.

Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). circlize Implements and enhances circular visualization in R. Bioinformatics *30*, 2811-2812. 10.1093/bioinformatics/btu393.

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. 10.1038/s41598-017-02855-1.

Hahne, F., and Ivanek, R. (2016). Visualizing Genomic Data Using Gviz and Bioconductor. Methods Mol Biol *1418*, 335-351. 10.1007/978-1-4939-3578-9_16.

Harmston, N., Ing-Simmons, E., Perry, M., Baresic, A., and Lenhard, B. (2015). GenomicInteractions: An R/Bioconductor package for manipulating and investigating chromatin interaction data. BMC Genomics *16*, 963. 10.1186/s12864-015-2140-x.

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

778 Function of BZR1 in Repressing BR Signaling in Arabidopsis. Mol Plant *12*, 689-
779 703. 10.1016/j.molp.2019.02.001.
780 Hermisson, J., and Pennings, P.S. (2005). Soft sweeps: molecular population
781 genetics of adaptation from standing genetic variation. Genetics *169*, 2335-2352.
782 10.1534/genetics.104.036947.
783 Hollister, J.D., and Gaut, B.S. (2009). Epigenetic silencing of transposable
784 elements: a trade-off between reduced transposition and deleterious effects on
785 neighboring gene expression. Genome Res *19*, 1419-1428.
786 10.1101/gr.091678.109.
787 Hollister, J.D., Smith, L.M., Guo, Y.L., Ott, F., Weigel, D., and Gaut, B.S. (2011).
788 Transposable elements and small RNAs contribute to gene expression
789 divergence between Arabidopsis thaliana and Arabidopsis lyrata. Proc Natl Acad
790 Sci U S A *108*, 2322-2327. 10.1073/pnas.1018222108.
791 Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I., and Paszkowski, J.
792 (2011). An siRNA pathway prevents transgenerational retrotransposition in plants
793 subjected to stress. Nature *472*, 115-119. 10.1038/nature09861.
794 Kalvari, I., Nawrocki, E.P., Argasinska, J., Quinones-Olvera, N., Finn, R.D.,
795 Bateman, A., and Petrov, A.I. (2018). Non-Coding RNA Analysis Using the Rfam
796 Database. Curr Protoc Bioinformatics *62*, e51. 10.1002/cpbi.51.
797 Kindgren, P., Ivanov, M., and Marquardt, S. (2020). Native elongation transcript
798 sequencing reveals temperature dependent dynamics of nascent RNAPII
799 transcription in Arabidopsis. Nucleic Acids Res *48*, 2332-2347.
800 10.1093/nar/gkz1189.
801 Kippes, N., VanGessel, C., Hamilton, J., Akpinar, A., Budak, H., Dubcovsky, J.,
802 and Pearce, S. (2020). Effect of phyB and phyC loss-of-function mutations on the
803 wheat transcriptome under short and long day photoperiods. BMC Plant Biol *20*,
804 297. 10.1186/s12870-020-02506-0.
805 Koster, J., and Rahmann, S. (2012). Snakemake--a scalable bioinformatics
806 workflow engine. Bioinformatics *28*, 2520-2522. 10.1093/bioinformatics/bts480.
807 Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and
808 methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571-1572.
809 10.1093/bioinformatics/btr167.
810 Kruse, K., Hug, C.B., and Vaquerizas, J.M. (2020). FAN-C: a feature-rich
811 framework for the analysis and visualisation of chromosome conformation
812 capture data. Genome Biol *21*, 303. 10.1186/s13059-020-02215-9.
813 Kuang, H., Padmanabhan, C., Li, F., Kamei, A., Bhaskar, P.B., Ouyang, S., Jiang,
814 J., Buell, C.R., and Baker, B. (2009). Identification of miniature inverted-repeat
815 transposable elements (MITEs) and biogenesis of their siRNAs in the
816 Solanaceae: new functional implications for MITEs. Genome Res *19*, 42-56.
817 10.1101/gr.078196.108.
818 Kurihara, Y., Matsui, A., Kawashima, M., Kaminuma, E., Ishida, J., Morosawa,
819 T., Mochizuki, Y., Kobayashi, N., Toyoda, T., Shinozaki, K., and Seki, M. (2008).
820 Identification of the candidate genes regulated by RNA-directed DNA methylation
821 in Arabidopsis. Biochem Biophys Res Commun *376*, 553-557.
822 10.1016/j.bbrc.2008.09.046.
823 Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with
824 Bowtie 2. Nat Methods *9*, 357-359. 10.1038/nmeth.1923.
825 Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and
826 memory-efficient alignment of short DNA sequences to the human genome.
827 Genome Biol *10*, R25. 10.1186/gb-2009-10-3-r25.

828 Lawrence, M., Gentleman, R., and Carey, V. (2009). rtracklayer: an R package
829 for interfacing with genome browsers. Bioinformatics *25*, 1841-1842.
830 10.1093/bioinformatics/btp328.
831 Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R.,
832 Morgan, M.T., and Carey, V.J. (2013). Software for computing and annotating
833 genomic ranges. PLoS Comput Biol *9*, e1003118. 10.1371/journal.pcbi.1003118.
834 Lei, Y., Lu, L., Liu, H.Y., Li, S., Xing, F., and Chen, L.L. (2014). CRISPR-P: a web
835 tool for synthetic single-guide RNA design of CRISPR-system in plants. Mol Plant
836 *7*, 1494-1496. 10.1093/mp/ssu044.
837 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G.,
838 Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The
839 Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.
840 10.1093/bioinformatics/btp352.
841 Li, Q., Wu, G., Zhao, Y., Wang, B., Zhao, B., Kong, D., Wei, H., Chen, C., and
842 Wang, H. (2020). CRISPR/Cas9-mediated knockout and overexpression studies
843 reveal a role of maize phytochrome C in regulating flowering time and plant
844 height. Plant Biotechnol J *18*, 2520-2532. 10.1111/pbi.13429.
845 Li, X., Guo, K., Zhu, X., Chen, P., Li, Y., Xie, G., Wang, L., Wang, Y., Persson,
846 S., and Peng, L. (2017). Domestication of rice has reduced the occurrence of
847 transposable elements within gene coding regions. BMC Genomics *18*, 55.
848 10.1186/s12864-016-3454-z.
849 Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general
850 purpose program for assigning sequence reads to genomic features.
851 Bioinformatics *30*, 923-930. 10.1093/bioinformatics/btt656.
852 Lisch, D. (2013). How important are transposons for plant evolution? Nat Rev
853 Genet *14*, 49-61. 10.1038/nrg3374.
854 Liu, C. (2017). In Situ Hi-C Library Preparation for Plants to Study Their Three-
855 Dimensional Chromatin Interactions on a Genome-Wide Scale. Methods Mol Biol
856 *1629*, 155-166. 10.1007/978-1-4939-7125-1_11.
857 Liu, S., Zhang, L., Gao, L., Chen, Z., Bie, Y., Zhao, Q., Zhang, S., Hu, X., Liu, Q.,
858 Wang, X., and Wang, Q. (2022). Differential photoregulation of the nuclear and
859 cytoplasmic CRY1 in Arabidopsis. New Phytol *234*, 1332-1346.
860 10.1111/nph.18007.
861 Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold
862 change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.
863 10.1186/s13059-014-0550-8.
864 Lu, C., Chen, J., Zhang, Y., Hu, Q., Su, W., and Kuang, H. (2012). Miniature
865 inverted-repeat transposable elements (MITEs) have been accumulated through
866 amplification bursts and play important roles in gene expression and species
867 diversity in Oryza sativa. Mol Biol Evol *29*, 1005-1017. 10.1093/molbev/msr282.
868 Lu, C., Kulkarni, K., Souret, F.F., MuthuValliappan, R., Tej, S.S., Poethig, R.S.,
869 Henderson, I.R., Jacobsen, S.E., Wang, W., Green, P.J., and Meyers, B.C.
870 (2006). MicroRNAs and other small RNAs enriched in the Arabidopsis RNA-
871 dependent RNA polymerase-2 mutant. Genome Res *16*, 1276-1288.
872 10.1101/gr.5530106.
873 Lyons, M., Cardle, L., Rostoks, N., Waugh, R., and Flavell, A.J. (2008). Isolation,
874 analysis and marker utility of novel miniature inverted repeat transposable
875 elements from the barley genome. Mol Genet Genomics *280*, 275-285.
876 10.1007/s00438-008-0363-0.

877 Matzke, M.A., Kanno, T., and Matzke, A.J. (2015). RNA-Directed DNA
878 Methylation: The Evolution of a Complex Epigenetic Pathway in Flowering Plants.
879 Annu Rev Plant Biol *66*, 243-267. 10.1146/annurev-arplant-043014-114633.
880 Maumus, F., and Quesneville, H. (2014). Ancestral repeats have shaped
881 epigenome and genome composition for millions of years in Arabidopsis thaliana.
882 Nat Commun *5*, 4104. 10.1038/ncomms5104.
883 Nishida, H., Ishihara, D., Ishii, M., Kaneko, T., Kawahigashi, H., Akashi, Y.,
884 Saisho, D., Tanaka, K., Handa, H., Takeda, K., and Kato, K. (2013). Phytochrome
885 C is a key factor controlling long-day flowering in barley. Plant Physiol *163*, 804-
886 814. 10.1104/pp.113.222570.
887 Oki, N., Yano, K., Okumoto, Y., Tsukiyama, T., Teraishi, M., and Tanisaka, T.
888 (2008). A genome-wide view of miniature inverted-repeat transposable elements
889 (MITEs) in rice, Oryza sativa ssp. japonica. Genes Genet Syst *83*, 321-329.
890 10.1266/ggs.83.321.
891 Pasha, A., Subramaniam, S., Cleary, A., Chen, X., Berardini, T., Farmer, A.,
892 Town, C., and Provart, N. (2020). Araport Lives: An Updated Framework for
893 Arabidopsis Bioinformatics. Plant Cell *32*, 2683-2686. 10.1105/tpc.20.00358.
894 Quadrana, L., Bortolini Silveira, A., Mayhew, G.F., LeBlanc, C., Martienssen,
895 R.A., Jeddeloh, J.A., and Colot, V. (2016). The Arabidopsis thaliana mobilome
896 and its impact at the species level. Elife *5*. 10.7554/eLife.15716.
897 Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular
898 Biology Open Software Suite. Trends Genet *16*, 276-277. 10.1016/s0168-
899 9525(00)02024-2.
900 SanMiguel, P., Gaut, B.S., Tikhonov, A., Nakajima, Y., and Bennetzen, J.L.
901 (1998). The paleontology of intergene retrotransposons of maize. Nat Genet *20*,
902 43-45. 10.1038/1695.
903 Sasaki, T., Lee, T.F., Liao, W.W., Naumann, U., Liao, J.L., Eun, C., Huang, Y.Y.,
904 Fu, J.L., Chen, P.Y., Meyers, B.C., et al. (2014). Distinct and concurrent pathways
905 of Pol II- and Pol IV-dependent siRNA biogenesis at a repetitive trans-silencer
906 locus in Arabidopsis thaliana. Plant J *79*, 127-138. 10.1111/tpj.12545.
907 Schmitz, R.J., He, Y., Valdes-Lopez, O., Khan, S.M., Joshi, T., Urich, M.A., Nery,
908 J.R., Diers, B., Xu, D., Stacey, G., and Ecker, J.R. (2013). Epigenome-wide
909 inheritance of cytosine methylation variants in a recombinant inbred population.
910 Genome Res *23*, 1663-1674. 10.1101/gr.152538.112.
911 Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang,
912 C., Zhang, J., Fulton, L., Graves, T.A., et al. (2009). The B73 maize genome:
913 complexity, diversity, and dynamics. Science *326*, 1112-1115.
914 10.1126/science.1178534.
915 Seren, U., Grimm, D., Fitz, J., Weigel, D., Nordborg, M., Borgwardt, K., and Korte,
916 A. (2017). AraPheno: a public database for Arabidopsis thaliana phenotypes.
917 Nucleic Acids Res *45*, D1054-D1059. 10.1093/nar/gkw986.
918 Sotelo-Silveira, M., Chavez Montes, R.A., Sotelo-Silveira, J.R., Marsch-Martinez,
919 N., and de Folter, S. (2018). Entering the Next Dimension: Plant Genomes in 3D.
920 Trends Plant Sci *23*, 598-612. 10.1016/j.tplants.2018.03.014.
921 Stuart, T., Eichten, S.R., Cahn, J., Karpievitch, Y.V., Borevitz, J.O., and Lister, R.
922 (2016). Population scale mapping of transposable element diversity reveals links
923 to gene regulation and epigenomic variation. Elife *5*. 10.7554/eLife.20777.
924 Tittel-Elmer, M., Bucher, E., Broger, L., Mathieu, O., Paszkowski, J., and Vaillant,
925 I. (2010). Stress-induced activation of heterochromatic transcription. PLoS Genet
926 *6*, e1001175. 10.1371/journal.pgen.1001175.

927   Underwood, C.J., Vijverberg, K., Rigola, D., Okamoto, S., Oplaat, C., Camp, R.,
928   Radoeva, T., Schauer, S.E., Fierens, J., Jansen, K., et al. (2022). A
929   PARTHENOGENESIS allele from apomictic dandelion can induce egg cell
930   division without fertilization in lettuce. Nat Genet *54*, 84-93. 10.1038/s41588-021-
931   00984-y.
932   Wu, N., Yao, Y., Xiang, D., Du, H., Geng, Z., Yang, W., Li, X., Xie, T., Dong, F.,
933   and Xiong, L. (2022). A MITE variation-associated heat-inducible isoform of a
934   heat-shock factor confers heat tolerance through regulation of JASMONATE ZIM-
935   DOMAIN genes in rice. New Phytol *234*, 1315-1331. 10.1111/nph.18068.
936   Wu, R., Lucke, M., Jang, Y.T., Zhu, W., Symeonidi, E., Wang, C., Fitz, J., Xi, W.,
937   Schwab, R., and Weigel, D. (2018). An efficient CRISPR vector toolbox for
938   engineering large deletions in Arabidopsis thaliana. Plant Methods *14*, 65.
939   10.1186/s13007-018-0330-7.
940   Xu, L., Yuan, K., Yuan, M., Meng, X., Chen, M., Wu, J., Li, J., and Qi, Y. (2020).
941   Regulation of Rice Tillering by RNA-Directed DNA Methylation at Miniature
942   Inverted-Repeat Transposable Elements. Mol Plant *13*, 851-863.
943   10.1016/j.molp.2020.02.009.
944   Yang, G., Nagel, D.H., Feschotte, C., Hancock, C.N., and Wessler, S.R. (2009).
945   Tuned for transposition: molecular determinants underlying the hyperactivity of a
946   Stowaway MITE. Science *325*, 1391-1394. 10.1126/science.1175688.
947   Zhang, H., Tao, Z., Hong, H., Chen, Z., Wu, C., Li, X., Xiao, J., and Wang, S.
948   (2016). Transposon-derived small RNA is responsible for modified function of
949   WRKY45 locus. Nat Plants *2*, 16016. 10.1038/nplants.2016.16.
950   Zhang, X., and Wang, T. (2021). Plant 3D Chromatin Organization: Important
951   Insights from Chromosome Conformation Capture Analyses of the Last 10 Years.
952   Plant Cell Physiol *62*, 1648-1661. 10.1093/pcp/pcab134.
953   Zhou, M., and Law, J.A. (2015). RNA Pol IV and V in gene silencing: Rebel
954   polymerases evolving away from Pol II's rules. Curr Opin Plant Biol *27*, 154-164.
955   10.1016/j.pbi.2015.07.005.

956

957

958

959

960

961

962

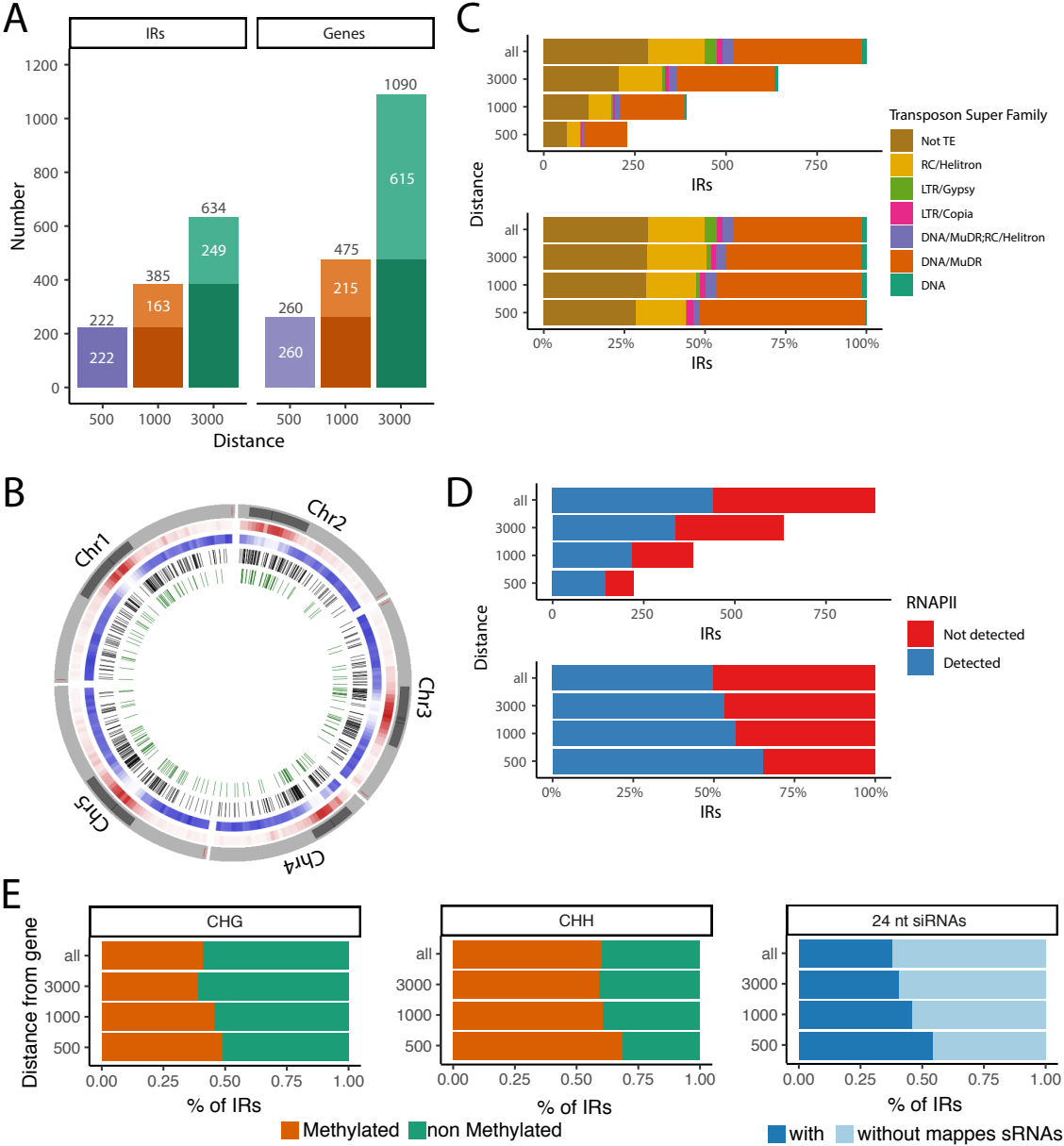963

964

965

966

967

Figure legends

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 from annotated genes (left), and the number of associated genes within the same size window from the identified IRs (right). White numbers indicate the number of IR and genes exclusive of the 0-500, 501-1000, and 1001-3000 bp windows while black numbers show the cumulative count.

B. Distribution of IRs. Outermost track, the five Arabidopsis chromosomes with the pericentromeric regions highlighted in darker gray. Then, TE density in 500 kb windows shown in red, and gene density in the same bins, in blue. In both red and blue tracks, darker color indicates higher density. The inner two tracks show

981 the distribution of all IRs (black bars) or only IRs within 3,000 bp upstream or

982 downstream of annotated protein coding genes (green bars).

983 **C.** Classification of the identified IRs between annotated TE super families, only

984 showing super families with at least five overlapping IRs. The

985 "DNA/MuDR;RC/Helitron" category indicates IRs overlapping with TEs from both

986 superfamilies. The upper panel shows the total number of IR in each superfamily

987 while the bottom panel displays the percentage of each superfamily identified in

988 each distance window from annotated genes.

989 **D.** Total number (up) or percentage (down) of IR nascent transcripts associated

990 with RNAPII as detected by plaNet-Seq (Kindgren *et al.*, 2020). IRs producing

991 transcripts associated with RNAPII are marked in blue and IRs without RNAPII

992 transcripts in red.

993 **E.** Percentage of gene-adjacent IRs with cytosine methylation in the CHG (left)

994 and CHH (center) contexts detected in at last 10% of the mapping reads and with

995 24-nt siRNAs mapping to the IR sequence (at least 10 reads in at least one
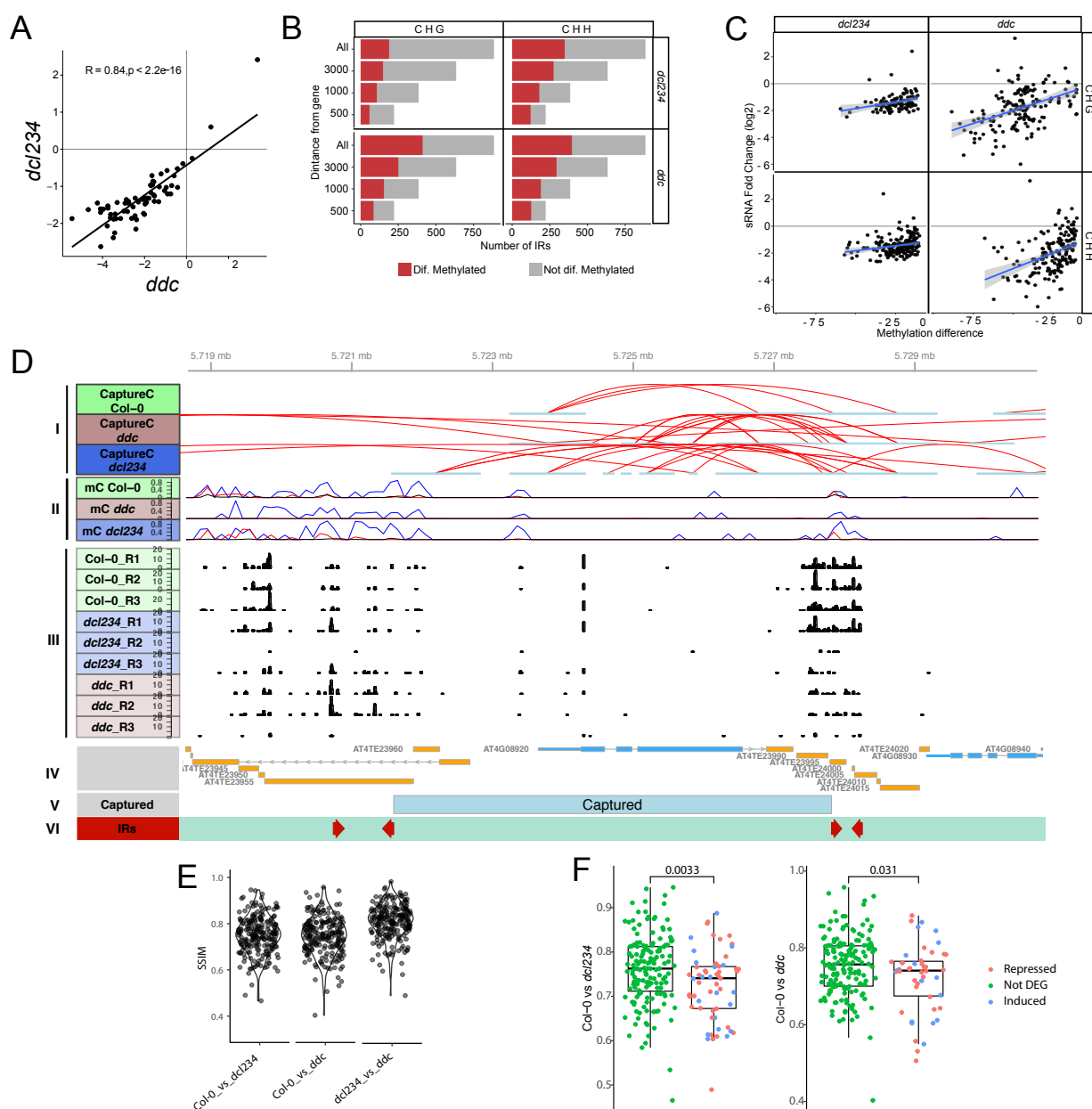
996 replicate).

997

998

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 (RPM) over each IR. The correlation (R) is shown on the upper left corner.

B. Number of gene-associated IRs showing differential DNA methylation in the CHG and CHH contexts in the *ddc* and *dcl234* mutants with respect to Col-0.

C. Scatter plots showing the relation between changes in siRNA levels and DNA methylation over each gene-associated IRs in both mutants, *ddc* and *dcl234*. The blue lines show the linear regression while the gray shades show the confidence interval.

D. Region of the Col-0 genome containing the *CRY1* locus displaying the epigenetic and topological profile. (I) Chromatin interactions as detected by Capture-C experiments and shown by red lines indicating interacting fragments.

1014   (II) Cytosine DNA methylation in CG (blue), CHG (red) and CHH (black) contexts.
1015   (III) 24 nt siRNAs mapping to the genomic regions as determined by sRNA
1016   sequencing in biological triplicates of WT plants and *dcl234* and *ddc* mutants. (IV)
1017   Annotated genes (cyan) and transposons (yellow) in the region. (V) Region
1018   captured by the probes designed for the CaptureC experiment. (IV) IRs identified
1019   in this region of the genome.

1020   **E.** Structural similarity (SSIM) of captured regions in *dcl234* and *ddc* mutants with
1021   respect to Col-0 or to each other.

1022   **F.** SSIM of captured regions in *dcl234* and *ddc* mutants with respect to Col-0
1023   grouped by regions including differentially regulated genes (DEGs) or regions
1024   without DEGs (NotDEGs). The colors of the dots represent whether they are not
1025   differentially expressed, upregulated, or downregulated.

1026

1027
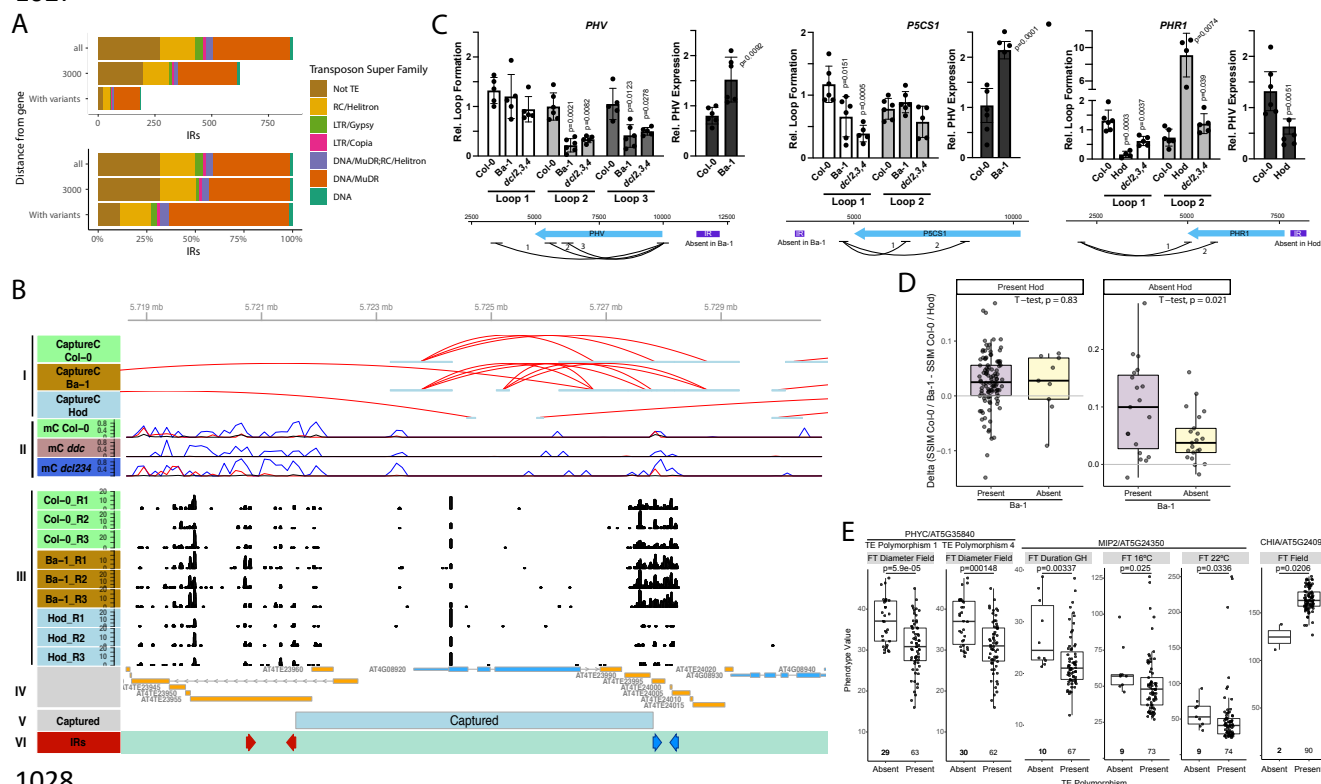
1028

1029   **Figure 3. Insertional polymorphisms of IRs near genes cause natural**
1030   **variation in short-range chromatin topology**.

1031   **A.** Number of IRs and polymorphic IRs associated with genes (upper panel).
1032   Percentage of IRs and polymorphic IRs by annotated TE superfamilies (lower
1033   panel). Only superfamilies with at least five overlapping IRs are displayed. The
1034   "DNA/MuDR;RC/Helitron" category indicates IRs overlapping with TEs from both
1035   superfamilies.

1036   **B.** Region of the Col-0 genome containing the *CRY1* locus with its epigenetic and
1037   topological profile. (I) Chromatin interactions as detected by Capture-C
1038   experiments from Col-0 and Ba-1 plants (both contain the IR) and Hod plants
1039   (which has a deletion polymorphism in the same IR). Red lines indicate

1040   interacting fragments. (II) Cytosine DNA methylation in CG (blue), CHG (red) and
1041   CHH (black) contexts. (III) 24 nt siRNAs mapping to the genomic regions as
1042   determined by sRNA sequencing in biological triplicates of wild-type Col-0, Ba-1
1043   and Hod plants. (IV) Annotated genes (cyan) and transposons (yellow) in the
1044   region. (V) Region captured by the probes designed for the Capture-C
1045   experiment. (IV) IRs identified in this region of the genome. The IR is red is
1046   conserved between the three genotypes while the IR in blue is missing in Hod.
1047   **C.** RT-qPCR with total RNA, and Chromatin Conformation Capture (3C) qPCT
1048   experiments to quantify the formation of specific chromatin loops as well as the
1049   expression of the affected loci in WT, Ba-1, Hod plants and *dcl234* and *ddc*
1050   mutants. Data are presented as mean values +/- SD. p-values were calculated
1051   with two-tailed unpaired Student's t-Test with Welch's correction. n>4 biologically
1052   independent samples. A diagram on the bottom of each gene shows the
1053   chromatin loops amplified and confirmed by Sanger sequencing for each locus.
1054   **D.** Delta of the structural similarity (SSIM) calculated with CHESS for the
1055   contrasts Col-0 vs Ba-1 and Col-0 vs Hod for each captured region. Each
1056   captured region is classified depending on whether the associated IR is present
1057   or absent in the three analyzed accessions.
1058   **E.** Association of flowering phenotypes (FT), including rosette diameter upon
1059   flowering, flowering duration, and flowering time at different temperatures either
1060   in plants grown in the field or green house (GH), with the presence or absence of
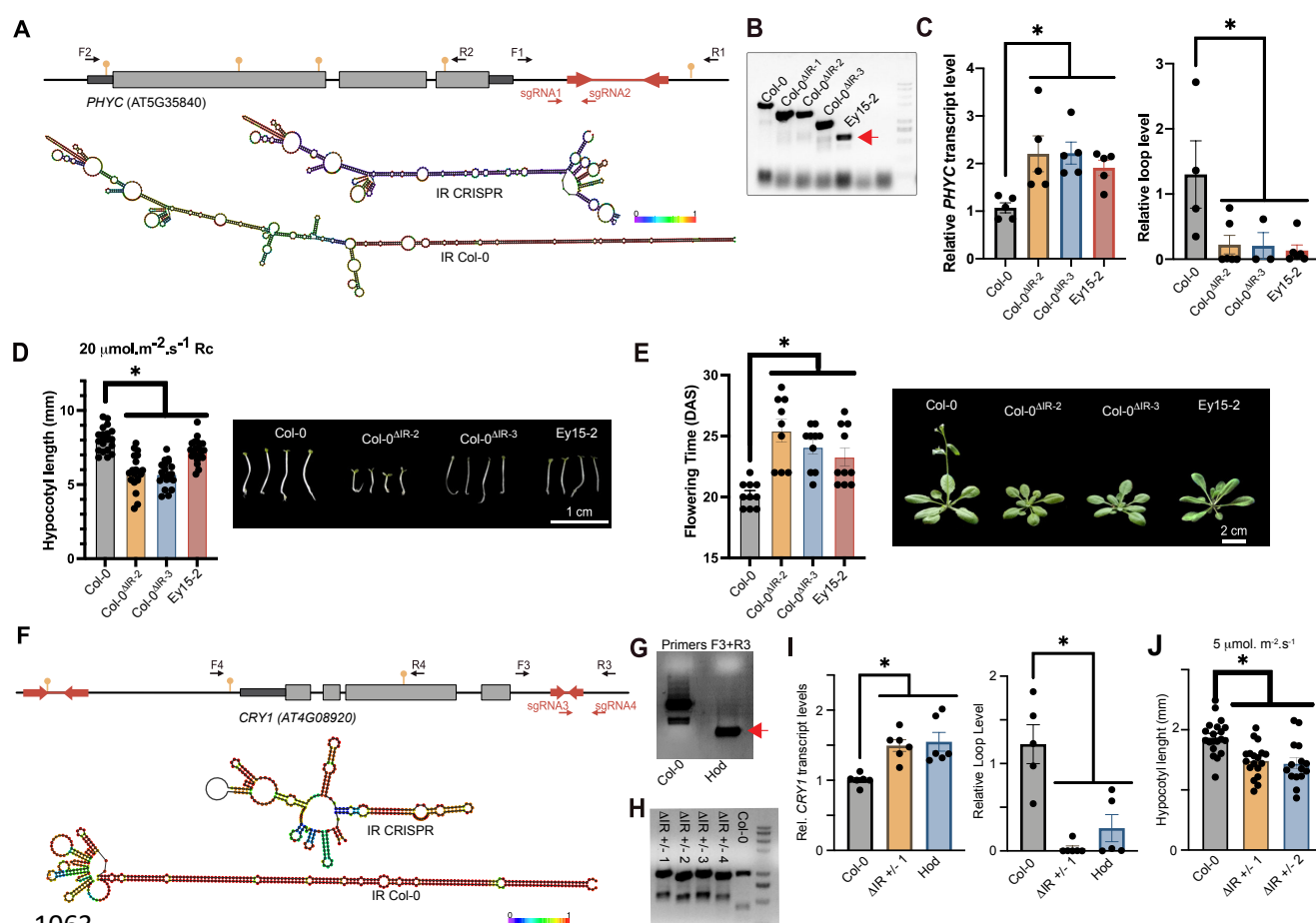1061   an IR near selected genes.
1062

Figure 4. Insertional polymorphisms of IRs near *PHYC* and *CRY1* cause changes in the loci 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), the sgRNAs used for CRISPR/Cas9 deletion (sgRNA1 and sgRNA2), and primer positions. Predicted secondary RNA structures for the IR transcript in Col-0 plants or in the 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. 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 RT-qPCR and normalized against *ACT2* (left panel). Quantification of the formation of a short-range chromatin loop by 3C-qPCR analysis using EcoRI and F2-R2 primers (right panel). Data represents individual values + SEM of five independent biological replicates.

D. Hypocotyl length under Continuous Red light (Rc) regimen. Seedlings were grown under Rc light 20 µmol m^{-2} s^{-1} for four days before measurement.

E. Flowering time for the different genotypes expressed as days after sowing (DAS). Data represents individual values ± SD, significant differences are indicated as *, for $P<0.05$ in a one-way ANOVA.

F. *CRY1* locus diagram showing the position of the IRs, restriction sites for XbaI (yellow pins), the sgRNAs used for CRISPR/Cas9 deletion (sgRNA3 and sgRNA4), and primer positions. On the bottom are the secondary RNA structures as predicted by RNAfold for the Col-0 wildtype IR transcript and the CRISPR-edited outcome of the same IR. Color scale shows base pair 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. Red arrowhead indicates the band sequenced to confirm the absence of the polymorphic IR in the Hod accession.

I. *CRY1* expression analysis measured by RT-qPCR normalized against *ACT2* on the left. Quantification of a chromatin loop by 3C-qPCR using XbaI and F4-R4 primers (right panel). Data represents average values of three independent biological replicates with SEM.

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

In all cases data represent individual values ± SD, significant differences are indicated as *, for P<0.05 in a one-way ANOVA.