The Arabidopsis "retrovirome" and its regulation by epigenetically activated small RNA

Seung Cho Lee[1,3], Evan Ernst[1,3], Benjamin Berube[2], Filipe Borges[1], Andrea Schorn[2], Jean-Sebastien Parent[1], Paul Ledon[2], Robert A. Martienssen[1, 2*]

[1] Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA

[2] Cold Spring Harbor Laboratory, 1 Bungtown Rd, Cold Spring Harbor, NY 11724, USA

[3] These authors contributed equally

[*] Correspondence: martiens@cshl.edu

Running title: The Arabidopsis retrovirome and its regulation

Abstract

In Arabidopsis, LTR-retrotransposons are activated by mutations in the chromatin remodeler DECREASE in DNA METHYLATION 1 (DDM1), giving rise to 21-22nt epigenetically activated siRNAs (easiRNAs). We purified virus-like-particles (VLPs) from *ddm1* and *ddm1rdr6* mutants in which genomic RNA is reverse transcribed into complementary DNA. Next generation short-read and long-read sequencing of VLP DNA (VLP DNA-seq) revealed a comprehensive catalog of active LTR-retrotransposons without the need for mapping transposition, and independent of genomic copy number. Linear replication intermediates of *ATCOPIA93*/*EVADE* revealed multiple central polypurine tracts (cPPT), a feature shared with HIV where cPPT promote nuclear localization. For *ATCOPIA52*, cPPT intermediates were not observed, but abundant circular DNA indicated transposon "suicide" by auto-integration within the VLP. easiRNA targeted *ATCOPIA93*/*EVADE* genomic RNA, polysome association of *GYPSY* (*ATHILA*) subgenomic RNA, and transcription via histone H3 lysine-9 dimethylation. VLP DNA-seq provides a comprehensive landscape of LTR-retrotransposons, and their control at transcriptional, post-transcriptional and reverse transcriptional levels.

Introduction

Long terminal repeat (LTR) retrotransposons are a major component of the large genomes of most animal and plant species (Huang et al., 2012; Wang et al., 2014). However, the relative activity of any individual element can only be assessed using transposition assays, or by comparing insertion sites among individuals within a population. The mouse genome, for example, contains more than one million endogenous retroviruses, of which only a handful are autonomous elements. We have developed an alternative strategy for retrotransposon discovery, using next generation sequencing of replication intermediates from viral-like particles (VLP). By sequencing intermediates from different genetic backgrounds, insights can be gained into mechanisms of genetic and epigenetic regulation.

41 VLPs have been isolated in yeast and Drosophila (Bachmann et al., 2004; Eichinger and Boeke,
42 1988; Kenna et al., 1998) as well as in plants (Bachmair et al., 2004; Jaaskelainen et al., 1999).
43 Ty1/Copia elements in plants have a single open reading frame that encodes both the GAG
44 protein, which is the capsid protein responsible for VLP formation, and the reverse transcriptase,
45 RNase H, and integrase polyprotein (POL) which are co-assembled with their genomic RNA
46 (gRNA) into VLPs (Finnegan, 2012; Pachulska-Wieczorek et al., 2016; Peterson-Burch and
47 Voytas, 2002; Sabot and Schulman, 2006). Ty3/gypsy elements also have a single GAG-POL
48 ORF, although the POL proteins are in a different order. In yeast, Ty1 uses a frameshift between
49 GAG and POL to enhance translation of GAG. In both Drosophila and plants, the Ty1/copia
50 GAG protein is translated from an abundant, alternatively spliced subgenomic RNA (Chang et al.,
51 2013; Yoshioka et al., 1990). In Arabidopsis Ty1/copia elements, the subgenomic *GAG* RNA is
52 more efficiently translated than unspliced GAG-POL transcripts, and blocking splicing leads to
53 significant reduction of GAG protein translation (Oberlin et al., 2017). After VLP formation in
54 the cytoplasm, LTR retrotransposons proliferate through tRNA-primed reverse transcription of
55 gRNA, followed by nuclear import of cDNA and integration into new loci (Schorn and
56 Martienssen, 2018) (Supplemental Fig. S1). In yeast and Arabidopsis, tRNA-iMet initiates
57 reverse transcription of the LTR from the primer binding site (PBS) to the 5' end of the R region
58 making minus-strand strong-stop DNA (Griffiths et al., 2018; Mules et al., 1998; Schorn and
59 Martienssen, 2018). RNase H degrades the template RNA upstream of the PBS, and minus-strand
60 strong-stop DNA is transferred to the 3' LTR to prime minus strand cDNA synthesis toward the
61 PBS (Supplemental Fig. S1). During the extension of minus-strand cDNA synthesis, the template
62 RNA is degraded except for an RNase H-resistant polypurine tract (PPT) near the 3' LTR
63 (Wilhelm et al., 2001). This PPT RNA fragment primes plus-strand strong-stop DNA synthesis
64 up to U5 and the PBS sequence from the translocated minus strand (Supplemental Fig. S1B).
65 Then, the plus-strand cDNA is transferred to the 5' end to prime full length double-stranded DNA.
66 Additional central PPT (cPPT) can also initiate plus-strand synthesis which is displaced by the 3'
67 end of plus-strand DNA causing DNA flaps to form during Ty1 replication (Garfinkel et al.,

68 2006). cPPT and DNA flaps have been found in the highly active lentivirus HIV-1 where they
69 play roles in nuclear import and in preventing mutagenesis by APOBEC (Hu et al., 2010;
70 VandenDriessche et al., 2002; Wurtzer et al., 2006; Zennou et al., 2000).

71

72 Inhibition of retrotransposons by small RNA has been reported in metazoans and plants, as well
73 as in fission yeast, and occurs at the transcriptional and post-transcriptional levels. In Drosophila,
74 piwi-interacting RNA (piRNA) trigger transcriptional silencing of transposons in the germline
75 (Czech et al., 2018) resembling fission yeast in this respect (Volpe et al., 2002). By contrast,
76 *Ago2* and *Dcr2* lie in the post-transcriptional pathway, and mutations result in increased somatic
77 retrotransposition (Xie et al., 2013). In mammalian embryos, 3' tRNA fragments (3'-tRF) control
78 transposition of LTR retrotransposons both post-transcriptionally and by direct inhibition of
79 reverse transcription (Schorn et al., 2017). In Arabidopsis, transcriptional activation of some LTR
80 retrotransposons by stress, or by loss of histone methylation, also requires loss of 24nt small RNA
81 and RDR2/RNA polymerase IV (Ito et al., 2011; Mirouze et al., 2009). By contrast, in *ddm1*
82 mutants and wild-type pollen, most transposons are transcriptionally activated and 24nt siRNA
83 are partly replaced by 21-22nt easiRNA (Slotkin et al., 2009). In *ddm1* mutants, easiRNA are
84 generated by RDR6 (Creasey et al., 2014; Nuthikattu et al., 2013) from the non-functional
85 *ATHILA2* and *ATHILA6* Ty3/gypsy retrotransposons but also from the functional, TY1/copia
86 element *EVADE*, and are triggered by diverse miRNA. In wild-type, retroelements generate
87 easiRNA only in pollen, where they are targeted at the PBS by miR845, and biogenesis occurs via
88 a non-canonical pathway (Borges et al., 2018).

89

90 In order to develop a comprehensive catalog of functional retrotransposons in Arabidopsis, we
91 performed VLP DNA sequencing from *ddm1* mutants, as well as genome-wide polysomal RNA
92 (translatome) and chromatin immunoprecipitation (ChIP) sequencing. VLP sequencing recovered
93 all known active retrotransposons in Arabidopsis, without the need for genome sequencing of

94 transposition events. Replication intermediates revealed profound differences between elements
95 with multiple cPPT and high integration rates, and elements with no cPPT which preferentially
96 integrated into themselves ("suicidal" auto-integration within the VLP). We examined the roles of
97 easiRNA in retrotransposon control by investigating *ddm1rdr6* double mutants (Creasey et al.,
98 2014; Lippman et al., 2004; Vongs et al., 1993). We found that some retrotransposons are
99 regulated post-transcriptionally by RNA interference, while others are regulated at the
100 transcriptional level by histone H3 lysine-9 methylation guided by small RNA. We conclude that
101 easiRNA inhibits retrotransposition at multiple levels in the replication cycle and identify features
102 of active retrotransposons that promote activity and escape from silencing.

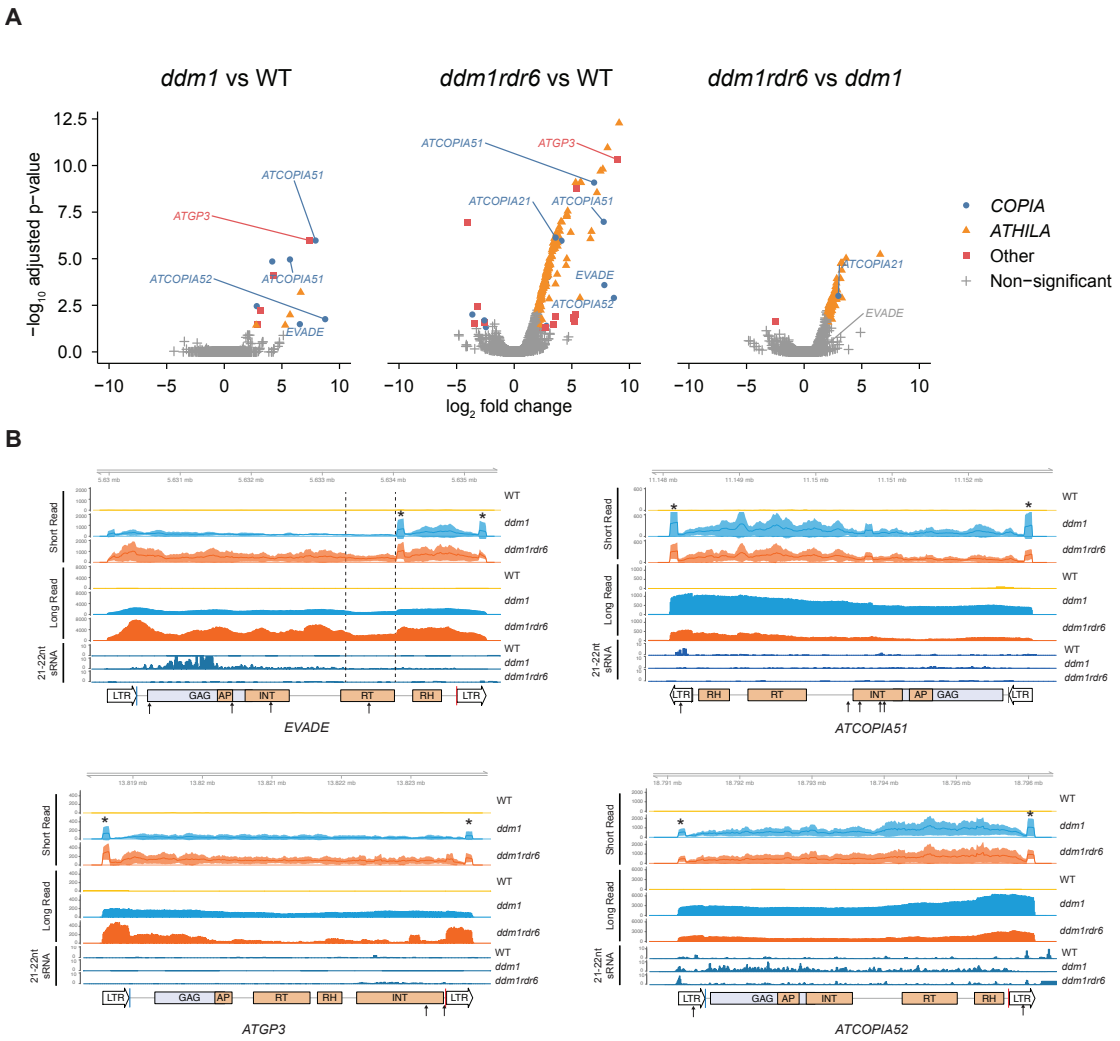
103

104 ## Results

105 ### Characterization of functional LTR retrotransposons by VLP DNA sequencing

106 Functional LTR retrotransposons form VLPs assembled from GAG proteins (Sabot and
107 Schulman, 2006) (Supplemental Fig. S1A). Reverse transcription occurs inside the VLPs, and
108 cDNA products are subsequently imported into the nucleus bound to the integrase protein. After
109 integration into new genomic loci these insertions transcribe additional gRNA. We purified VLPs
110 after treatment with DNase I (Methods), and sequenced cDNA products from wild-type, *ddm1,*
111 and *ddm1rdr6* using both short read (Illumina) and long read (Oxford Nanopore Technologies)
112 sequencing platforms (Supplemental Fig. S2). *EVADE* is one of two full length elements of the
113 *ATCOPIA93* family in *A. thaliana* Col-0, and when it is transcriptionally activated, it is the most
114 successful retroelement by far in terms of copy number increases, although transposition of
115 *ATGP3*, *ATCOPIA13*, *ATCOPIA21, ATCOPIA31*, *ATCOPIA51*, *ATCOPIA63*, *ATGP2N*, and
116 *ATRE1* have also been detected under non-stressed conditions (Ito et al., 2011; Quadrana et al.,
117 2019; Tsukahara et al., 2009). Full length VLP DNA from all of these elements was dramatically
118 enriched in *ddm1* and *ddm1rdr6* mutants consistent with active reverse transcription (Fig. 1;
119 Supplemental Figs. S2,S3; Supplemental Table S1). VLP DNA from some *ATCOPIA* families

120 were more enriched in *ddm1* than *ddm1rdr6*, likely due to transcriptional down-regulation in
121 *ddm1rdr6* (Creasey et al., 2014). VLP DNA from *ATHILA* families were enriched in *ddm1rdr6*,
122 but comprised small fragments, mostly from the LTR likely reflecting abortive retrotransposition
123 intermediates from these non-functional elements (Supplemental Fig. S3A) (Marco and Marin,
124 2008). By contrast, long read coverage of *EVADE* and other active *COPIA* elements spanned the
125 entire element, and was increased in *ddm1rdr6* (Fig. 1B; Supplemental S2B; Supplemental Table
126 S1). Furthermore, linear extrachromosomal DNA (ecDNA) was dramatically increased in
127 *ddm1rdr6* by Southern blot (Fig. 2A).



128

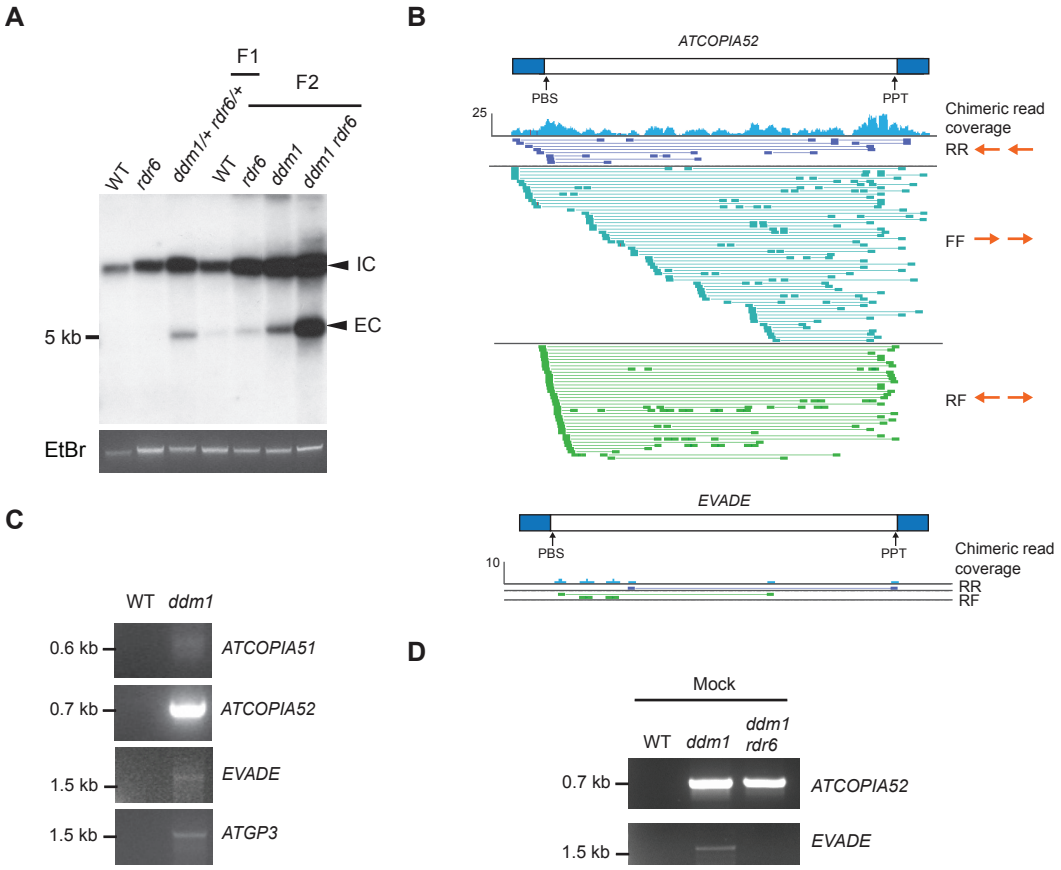
129 **Figure 1. VLP DNA-seq data of LTR retrotransposons in *ddm1* and *ddm1rdr6*.** (A) Differential
130 analysis of paired-end sequencing of VLP DNA using Illumina short read platform. The statistical
131 significance of three comparisons of wild-type (WT), *ddm1*, and *ddm1rdr6* is shown with $|\log_2$ (fold

132 change)| ≥ 2 and FDR threshold at 5%. Each point corresponds to an annotated transposable element.
133 Multiple *ATHILA* families were combined and labeled as '*ATHILA*'. (B) Coverage of short and long read
134 VLP DNA-seq at representative LTR retrotransposon loci (*EVADE*, AT5TE20395; *ATGP3*, AT1TE45315;
135 *ATCOPIA51*, AT1TE36035; *ATCOPIA52*, AT3TE76225) were plotted for *ddm1* and *ddm1rdr6*. Mean read
136 counts per million mapped reads and 95% confidence intervals of biological replicates are shown for WT
137 (yellow, n=3), *ddm1* (blue, n=2), and *ddm1rdr6* (orange, n=3) short read libraries. VLP DNA replicate
138 samples were pooled for each genotype and sequenced in aggregate by Oxford Nanopore long read
139 sequencing. In the LTR retrotransposon annotation, abbreviations for conserved protein domains within the
140 GAG-POL ORF are indicated as GAG, AP (amino peptidase), INT (integrase), RT (reverse transcriptase),
141 and RH (RNase H). Blue and red lines indicate primer binding sites (PBS) and polypurine tracts (PPT). 21-
142 22nt small RNA (sRNA) data were obtained from a previous study (Creasey et al., 2014). Target positions
143 of miRNAs are indicated as arrows (see Supplemental Table S4 for details). Central PPT (cPPT) positions
144 are indicated as dashed lines. Elevated coverage at the edges of strong-stop intermediate and flap DNA is
145 shown as asterisks above *ddm1* short read data.

146

147 cDNA can exist in both linear and circular forms, and circular forms were previously reported for
148 *EVADE* and other members of the *ATCOPIA93* family (Lanciano et al., 2017; Reinders et al.,
149 2013). We observed outward-facing paired-end read alignments from Illumina VLP-seq samples
150 mapping to *ATCOPIA51*, *ATCOPIA52*, *EVADE*, and *ATGP3*, consistent with junction-crossing
151 reads from circular templates (Fig. 2B). Outward-facing pairs appeared in the *ddm1* and
152 *ddm1rdr6* samples, but not in WT, and were present in very low numbers after read de-
153 duplication for most of the elements. Exceptionally, non-concordant read pairs were highly
154 abundant in *ATCOPIA52*. Circular ecDNA formation was confirmed by inverse PCR whose
155 products corresponded to one-LTR in size (Fig. 2C,D), and *ATCOPIA52* was by far the most
156 abundant. Double stranded one-LTR circular products are thought to be generated by integrase-
157 mediated autointegration in VLP, or as gapped intermediates in cDNA synthesis (Garfinkel et al.,
158 2006; Munir et al., 2013; Sloan and Wainberg, 2011). In contrast, two-LTR (tandem) circular
159 DNA with junction nucleotides is formed in the nucleus by non-homologous end joining and
160 enhanced when integrase is non-functional (Garfinkel et al., 2006; Sloan and Wainberg, 2011).
161 The inverse PCR products of *ATCOPIA52* were one-LTR in size, suggesting the circular DNA
162 was either a gapped double stranded circular intermediate, or else a double stranded product of

163 autointegration into same strands or opposite strands (Supplemental Fig. S1), which result in
164 deletion circles, or inversion circles, respectively (Garfinkel et al., 2006; Munir et al., 2013; Sloan
165 and Wainberg, 2011). Both inversion and deletion circles were detected in large numbers based
166 on outward facing reverse-forward and forward-forward paired end reads, respectively, indicating
167 auto-integration was the major source of these circles (Fig. 2B).



168

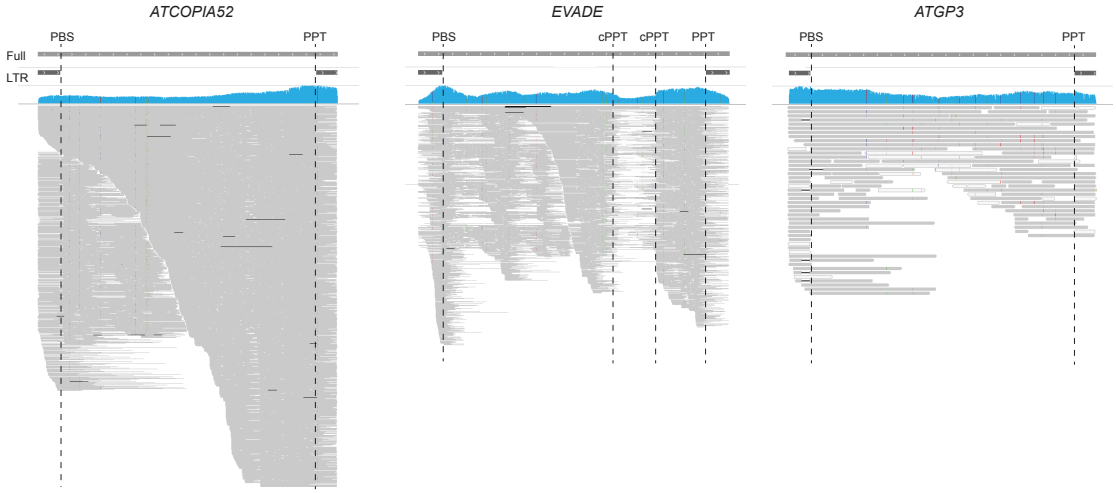
169 **Figure 2. Extrachromosomal DNA of LTR retrotransposons in *ddm1* and *ddm1rdr6*.** (A) Southern
170 blotting using an *EVADE* probe was performed with undigested genomic DNA of F1 and F2 plants from
171 the same parental lines. Integrated DNA copies (IC) and extrachromosomal DNA copies (EC) are indicated.
172 Ethidium Bromide (EtBr) staining was used for loading control. (B) Discordant short read alignments from
173 *ATCOPIA52* (AT3TE76225) and *EVADE* in *ddm1*. Read pair orientations (forward or reverse for the first
174 and second mate): RR and FF reads align in the same direction to the reference, indicating inversions, while
175 RF reads face outward, indicating circular templates. LTR regions are indicated as blue bars. (C) Inverse
176 PCR with genomic DNA to detect circular extrachromosomal DNA from *ATCOPIA51*, *ATCOPIA52*,
177 *EVADE*, and *ATGP3* in *ddm1* plants. (D) Inverse PCR with VLP DNA and reverse-forward (RF) outward
178 reading primers for *ATCOPIA52* and *EVADE*. (C-D) PCR primers are listed in Supplemental Table S6.

179 In yeast, auto-integration occurs near the central PPT (cPPT) taking advantage of a DNA flap
180 structure (Garfinkel et al., 2006). There was no strong indication of a DNA flap based on
181 polypurine sequences and read alignment in *ATCOPIA52*. We mapped individual long reads to
182 investigate the integration sites (Fig. 3; Supplemental Figs. S1C, S3B). Deletion circles are
183 predicted to have either the 5' or the 3' LTR, as well as a deleted portion of the full length cDNA,
184 up to the integration site, while inversion circles have an inverted portion separating the two LTR
185 (Garfinkel et al., 2006). Strikingly, many of the *ATCOPIA52* ONT reads fell into these categories,
186 comprising either the 5' or the 3' LTR contiguous with a truncated or inverted portion of the
187 retrotransposon (Supplemental Fig. S1C). These structural variants indicated the presence of
188 circularly permuted reads, which were presumably arbitrarily sheared during library preparation.
189 Among all the *COPIA* and *GYPSY* elements examined, only *ATCOPIA52* gave rise to large
190 numbers of these structural variants. The inversions spanned diverse regions of the element,
191 consistent with inversion circles. The deleted portions terminated at inferred autointegration sites,
192 which were distributed throughout the length of the element, consistent with the lack of a cPPT
193 flap in *ATCOPIA52*. One possibility is that nuclear import of cDNA is not efficient for
194 *ATCOPIA52,* leading to elevated autointegration inside the VLP. This could be due to mutations
195 in nuclear localization (Kenna et al., 1998), or else to reduced translation of the integrase gene
196 (see below), although read distributions were comparable for *ddm1* and *ddm1rdr6*, so easiRNA
197 likely did not play a major role.

198

199 In sharp contrast, in *EVADE* we observed discontinuous regions of read alignments flanked by
200 multiple cPPT, defined as 15-19 nt polypurine sequences (Figs. 1B,3; Supplemental Fig. S3B).
201 These regions represent active replication intermediates, generated by both minus strand and plus
202 strand strong stop DNA, as well as extension products that terminate at cPPT and DNA flaps. The
203 numbers of these intermediates, as well as their abundance, were significantly elevated in long-
204 read sequencing data from *ddm1rdr6* double mutants (Figs. 1B,3; Supplemental Fig. S3B).
205 *ATGP3* also had elevated levels of strong stop intermediates, but few if any cPPT and no circular

206 reads.



207

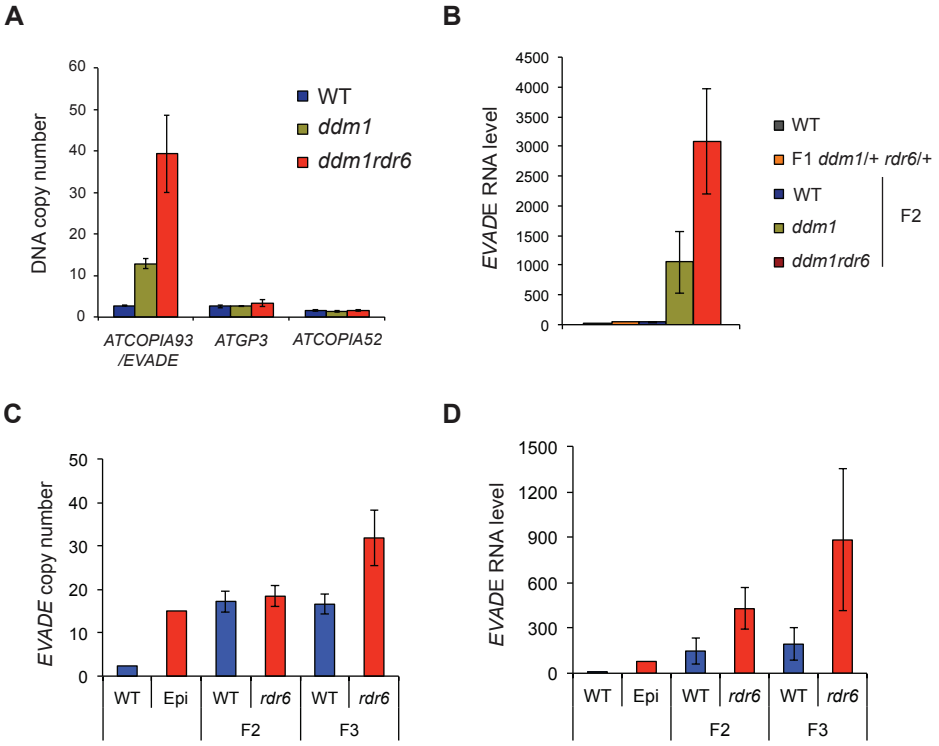
208 **Figure 3. Alignments of Oxford Nanopore long reads from *ddm1* VLP DNA.** The central polypurine
209 tract (cPPT), PBS, and PPT positions are indicated as dashed lines relative to full and LTR annotation of
210 *ATCOPIA52* (AT3TE76225), *EVADE* (AT5TE20395), and *ATGP3* (AT1TE45315). Gaps in individual
211 reads are indicated with black horizontal lines, and sequence mismatches are shown as colored dots in the
212 read alignments. Pileups of linear intermediates are observed for *EVADE*, while a continuous distribution
213 of fragment lengths is observed in *ATCOPIA52*.

214

215 **21-22nt easiRNA control retrotransposition**

216 In a previous study, DCL2/4 were shown to promote transcription of *EVADE* transgenes driven
217 by an ectopic promoter, while RDR6 had no effect, which was interpreted as evidence that
218 easiRNA might promote transposition in wild-type cells (Mari-Ordonez et al., 2013). We tested
219 whether easiRNA contribute to *EVADE* control in *ddm1* and *ddm1rdr6* mutants. Both *ddm1* and
220 *ddm1rdr6* contained higher copy numbers of *ATCOPIA93* than wild-type implying high rates of
221 *EVADE* transposition, while copy numbers of *ATGP3* and *ATCOPIA52* remained constant. Using
222 quantitative PCR, we detected an increase from 2 copies of *EVADE* in wild-type to 12 copies in
223 *ddm1* to 40 copies in *ddm1rdr6* F2 siblings (Fig. 4A). Similar increases were observed in F2 and
224 F3 progeny from backcross *rdr6* progeny that inherited active *EVADE* elements epigenetically

225 (Fig. 4C) (Mari-Ordonez et al., 2013). We detected parallel increases in gRNA levels reflecting
226 these increases in copy number (Fig. 4B,D). Consistent with gRNA levels, extrachromosomal
227 *EVADE* copies were also more abundant in *ddm1rdr6* than in *ddm1* (Fig. 2A). RNase H cleavage
228 products just upstream of the PBS, which are a hallmark of active transposition (Schorn et al.,
229 2017), were readily detected for *EVADE* in both *ddm1* and *ddm1rdr6* (Supplemental Fig. S4A,B).
230 We conclude that easiRNA actually inhibit *EVADE* retrotransposition, in *ddm1* mutants.



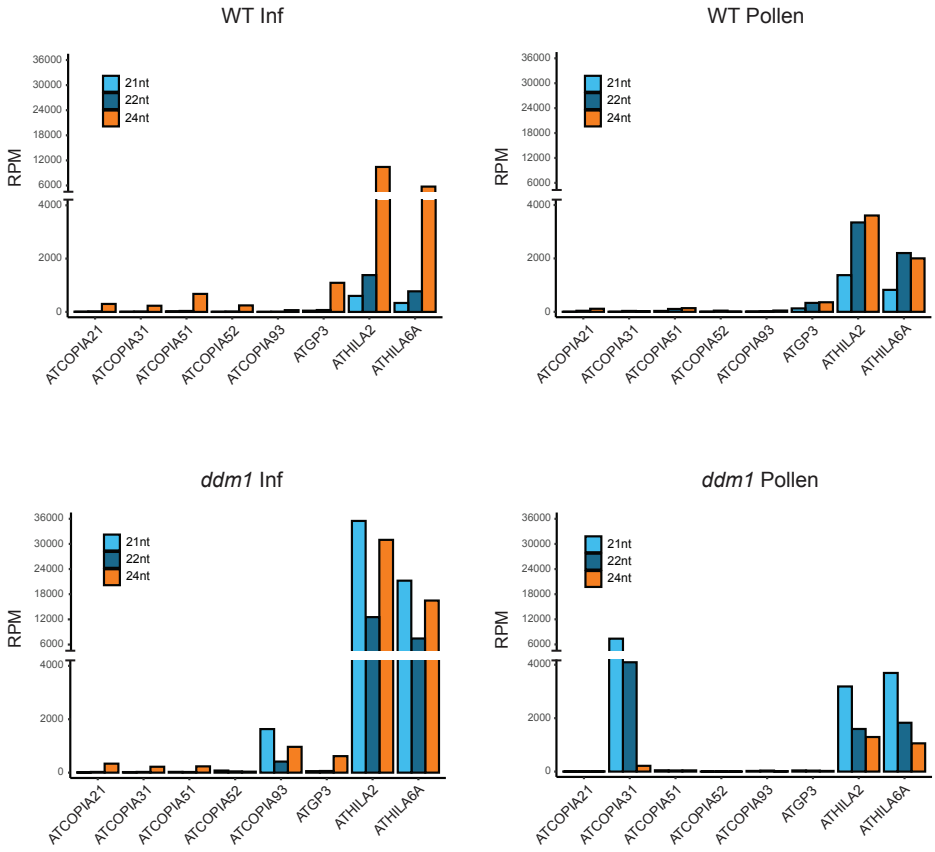
231

232 **Figure 4. DNA and RNA levels of LTR retrotransposons in *ddm1* and *rdr6* mutants.** (A) DNA copy
233 numbers of *ATCOPIA93*, *ATGP3*, and *ATCOPIA52* in *ddm1* and *ddm1rdr6* were normalized with a single
234 copy gene (AT5G13440). (B) RT-qPCR data of *EVADE* elements using POL primers. Y-axis indicates
235 relative levels of *EVADE* genomic RNA to wild-type (WT) after normalization to *ACT2*. (C-D) *EVADE*
236 DNA copy number and genomic RNA levels were analyzed in F2 and F3 progenies of F1 plants carrying
237 active *EVADE* epigenetically inherited from parental *rdr6*/+ (Epi) crossed with WT pollen. Error bars
238 indicate standard deviations (n=3).

239

240 In backcrosses to wild-type (WT) plants, *EVADE* activity is inherited epigenetically but copy

241 number increases are thought to be limited by a switch from 21nt to 24nt siRNA, accompanied by
242 re-methylation and silencing (Mari-Ordonez et al., 2013; Mirouze et al., 2009; Reinders et al.,
243 2013). Interestingly, active *EVADE* elements can be re-silenced through the female gametophyte,
244 but not through the male gametophyte (Reinders et al., 2013) where easiRNA normally
245 accumulate (Borges et al., 2018; Slotkin et al., 2009). We sequenced small RNA from wild type
246 and *ddm1* flower buds and pollen, and found that 21-22nt easiRNA from *ATCOPIA93*/*EVADE*
247 were abundant in *ddm1* inflorescence tissues, but absent from pollen (Fig. 5). In contrast,
248 *ATHILA2* and *ATHILA6A* easiRNA were present in wild type pollen (Slotkin et al., 2009), while
249 *ATCOPIA31* 21-22nt easiRNA were strongly upregulated in *ddm1* pollen. Thus the absence of
250 *EVADE* easiRNA in pollen must be due to transcriptional repression independent of DDM1, and
251 likely accounts for the lack of paternal re-silencing (Reinders et al., 2013).



252

253 **Figure 5. Small RNA profiles of representative LTR retrotransposons.** 21, 22, and 24nt small RNA
254 levels in inflorescence tissues and pollen of wild-type (WT) and *ddm1*. Reads per million (RPM) was

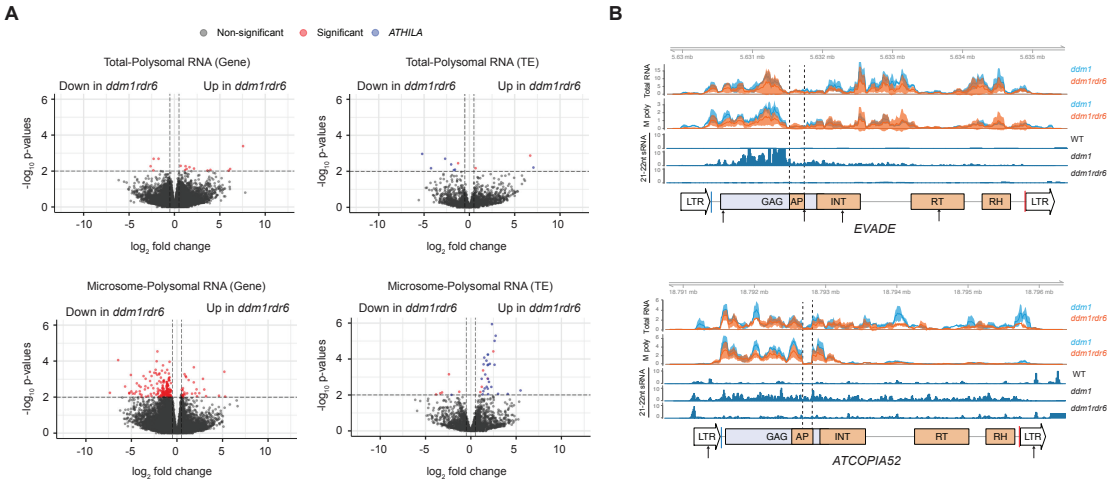
255 calculated from entire elements including LTR and coding sequences.

256

257 **Post-transcriptional suppression by easiRNA**

258 Since easiRNA in *ddm1* mutants depend on AGO1 (Nuthikattu et al., 2013), and AGO1 represses
259 translation of target mRNA (Li et al., 2013), we tested whether easiRNA can affect translation
260 efficiency of transposon transcripts. Translating ribosome affinity immunopurification (TRAP)
261 RNAseq has been utilized to estimate polysomal occupancy and translation efficiency in plants
262 (Juntawong et al., 2014). Furthermore, microsome-polysomal fractionation has revealed that
263 microRNA-dependent translational control takes place on the endoplasmic reticulum (Li et al.,
264 2013). We generated TRAP lines of *35S:FLAG-RPL18* in *ddm1* and *ddm1rdr6* mutant
265 backgrounds, and performed total RNAseq, total-polysomal RNAseq, and microsome-polysomal
266 RNAseq. The polysomal RNA occupancy (Polysomal RNA / Total RNA) was obtained for 3903
267 transposable elements defined as open reading frames from TAIR10 annotation (see Methods).
268 As for the comparison between *ddm1* and *ddm1rdr6*, we could detect the effect of the *rdr6*
269 mutation in microsome-polysomal RNAseq data for known targets of RDR6, such as *ARF4*
270 (Marin et al., 2010), and for a handful of transposons (Fig. 6A; Supplemental Fig. S5;
271 Supplemental Tables S2,S3). Among 31 up-regulated transposons in *ddm1rdr6* relative to *ddm1*,
272 26 elements belonged to *ATHILA* LTR retrotransposon families (Supplemental Table S3), which
273 are a major source of RDR6-dependent easiRNA. Although *ATHILA* elements in *A. thaliana*
274 cannot transpose, a subgenomic mRNA encoding ORF2 (the "env" gene) is spliced from the full
275 length mRNA (Havecker et al., 2004; Wright and Voytas, 2002), and was enriched on polysomes
276 (Supplemental Fig. S5; Supplemental Table S3). This subgenomic RNA is targeted extensively
277 by miRNA which trigger easiRNA production (Creasey et al., 2014). Interestingly, the other 3
278 elements were *ATENSPM3*, *LINE1_6* and *VANDAL3*, all of which have been identified as active
279 elements in *ddm1* mutants, or in population level studies of transposon variants (Stuart et al.,
280 2016). These non-LTR and DNA transposons are also targets of miRNA and generate RDR6-

281 dependent easiRNA (Creasey et al., 2014). *EVADE* easiRNA are generated from the GAG
282 subgenomic RNA (Mari-Ordonez et al., 2013), but polysomal occupancy was not increased in
283 *ddm1rdr6* (Fig. 6B). GAG subgenomic mRNA from *ATCOPIA52* was highly enriched in
284 polysomes, consistent with previous studies (Oberlin et al., 2017), whereas the relative abundance
285 of *EVADE* POL transcripts on polysomes indicates higher translation rates of integrase and
286 reverse transcriptase (Oberlin et al., 2017). Unlike for *ATHILA*, polysome association of *COPIA*
287 transcripts were unaffected by RDR6.



288

289 **Figure 6. Translatome profiles of *ddm1* and *ddm1rdr6*.** (A) Differential analysis of polysomal RNAseq
290 data between *ddm1* and *ddm1rdr6*. Polysomal RNAseq values were normalized by total RNA seq values to
291 reflect polysomal enrichment (Methods). Red dots indicate significantly regulated genes or transposable
292 elements (TE) by cut-off values of $|\log_2 (\text{fold change})| > 0.5$ and p-values < 0.01 which include *ARF4* as an
293 internal control. Significantly regulated *ATHILA* family elements are labeled with blue dots. (B) Total RNA
294 and microsome-polysomal RNA (M poly) levels are shown for *EVADE* (*ATCOPIA93*; AT5TE20395) and
295 *ATCOPIA52* (AT3TE76225). Mean read counts per million mapped reads and 95% confidence intervals of
296 three biological replicates are shown for *ddm1* (blue) and *ddm1rdr6* (orange). Conserved protein domains,
297 PBS and PPT, small RNA profiles and miRNA target sites are indicated as in Fig. 1.

298

299 easiRNA require miRNA triggers that target these transcripts (Creasey et al., 2014), and
300 *ATCOPIA52* LTRs were targeted by a single miRNA in the R region of the LTR. Consistent with
301 this miRNA acting as a trigger, easiRNA accumulated along the length of the mRNA between the

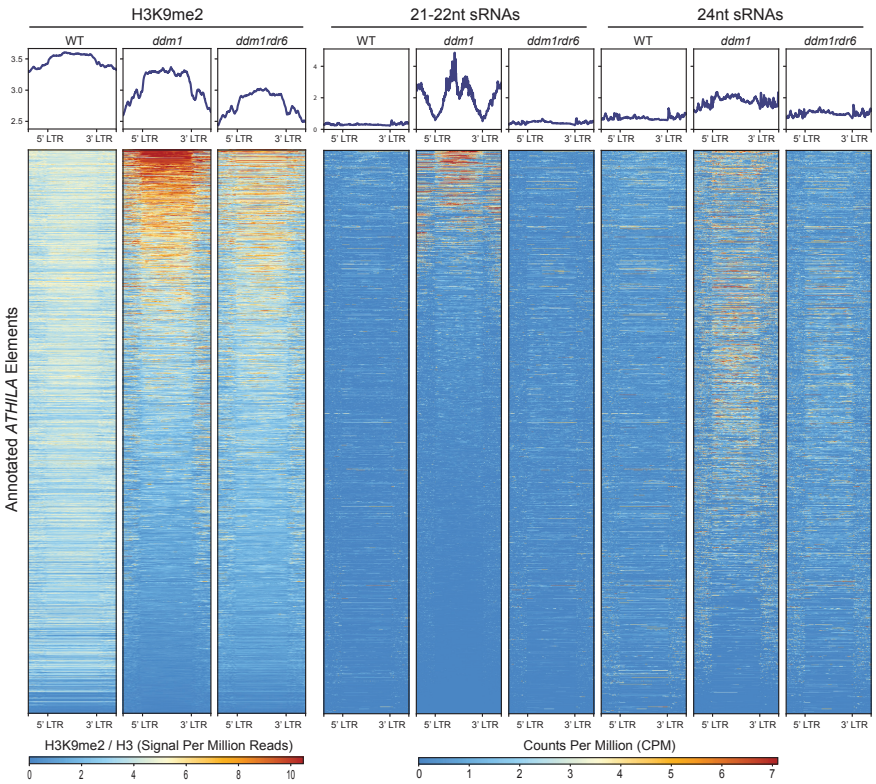
302 LTRs (Fig. 6B; Supplemental Table S4). In the case of *EVADE*, 4 miRNA were predicted to
303 target the gRNA somewhere along its length. Remarkably, miR2938 was predicted to target the
304 start codon of the GAG gene immediately 5' of the easiRNA cluster, while miR5648-5p targets
305 the 3' end of the easiRNA cluster (Supplemental Fig. S4C,D; Supplemental Table S4). *EVADE*
306 easiRNAs were down-regulated in *ddm1dcl1* as compared to *ddm1* (Supplemental Fig. S4E)
307 suggesting that miRNA were involved (Creasey et al., 2014). miR2938 and miR5648-5p
308 expression were reported in pollen and root cells (Breakfield et al., 2012; Grant-Downton et al.,
309 2009). We did not detect miRNA-mediated cleavage by PARE-seq (Creasey et al., 2014) or by
310 RACE-PCR, but secondary siRNA, such as easiRNA, do not require cleavage so long as miRNA
311 recognition recruits RdRP (Axtell et al., 2006; de Felippes et al., 2017). Consistent with induction
312 without cleavage, *EVADE* easiRNA were not phased (Arribas-Hernandez et al., 2016). miR5663
313 was detected in inflorescence tissues (Supplemental Fig. S4F), and targets the *EVADE* intron near
314 the splice acceptor site (Supplemental Fig. S4C). Interestingly, the level of unspliced RNA was
315 increased in *ddm1dcl1* mutants (Supplemental Fig. S4G), indicating that miR5663 might target
316 unspliced gRNA and promote the accumulation of spliced GAG RNA, but further experiments
317 would be required to demonstrate this requirement. Negative regulation of *P*-element splicing by
318 piRNA has been reported in Drosophila (Teixeira et al. 2017). *ATCOPIA21* and *ATCOPIA51* had
319 no strongly predicted miRNA targets, and easiRNA were barely detected in somatic tissues (Fig.
320 1B; Supplemental Fig. S3A) (Oberlin et al., 2017) accounting for lack of regulation by *RDR6*. In
321 contrast, significant levels were detected in pollen (Fig. 5) (Borges et al., 2018), where most
322 gypsy and copia class retrotransposons are targeted by miR845, a pollen specific miRNA that
323 targets the primer binding site (Borges et al., 2018).

324

325 **Transcriptional repression by easiRNA**

326 Both 21/22nt easiRNA and especially 24nt siRNA can direct RNA directed DNA methylation in
327 plants, via AGO6 and AGO4, respectively (Borges and Martienssen, 2015). However, genome

328 wide bisulphite sequencing revealed few if any differences in DNA methylation between *ddm1*
329 and *ddm1rdr6*, as *ddm1* mutants already had very low levels of DNA methylation (Creasey et al.,
330 2014). In many organisms, histone modification can also be guided by small RNA, especially
331 histone H3 lysine-9 dimethylation (Fagegaltier et al., 2009; Gu et al., 2012; Martienssen and
332 Moazed, 2015; Volpe et al., 2002). We therefore performed H3K9me2 ChIP sequencing in *ddm1*
333 and *ddm1rdr6*, and compared this to wild type. We found that *ATHILA* elements, which matched
334 by far the most abundant easiRNA, had ectopic H3K9me2 in *ddm1* mutants, which was absent in
335 *ddm1rdr6* mutants (Fig. 7; Supplemental Fig. S6). Furthermore, the levels of small RNA
336 correlated extremely well with the levels of H3K9me2 found at individual *ATHILA* elements (Fig.
337 7; Supplemental Fig. S6). Interestingly, *COPIA* elements actually gained H3K9me2 in the
338 absence of easiRNA (Supplemental Fig. S7), along with previously reported increases in 24nt
339 siRNA and reduced transcript levels (Creasey et al., 2014). We therefore conclude that in the
340 absence of DDM1, 21/22nt easiRNA and 24nt siRNA can guide H3K9me2 at *GYPSY* and *COPIA*
341 elements respectively.



342

343 **Figure 7. *ATHILA* family elements gain RDR6-dependent H3K9me2 in *ddm1*.** H3K9me2 signal at
344 transposable elements from multiple *ATHILA* families was analyzed in wild-type (WT), *ddm1*,
345 and *ddm1rdr6* genotypes and correlated with previously published small RNA data (Creasey et al., 2014).
346 RDR6-dependent gains in H3K9me2 co-localize with increased 21-22nt siRNAs in *ddm1*. Plots depict
347 transposable elements annotations scaled to 5kb, as well as 2kb upstream and downstream of each feature.
348 H3K9me2 ChIP data was normalized by H3, and small RNA data was normalized by counts per million.
349

350 ## Discussion

351 Next generation sequencing of VLP DNA detected all known functional LTR retrotransposons in
352 Arabidopsis, as well as some non-functional ones. Full length VLP DNA from *ATCOPIA* and
353 *ATGP* families (Fig. 1B; Supplemental Fig. S3A) corresponded to relatively young and low-copy
354 elements known to transpose. Ancient *ATHILA* elements did not make full length VLP DNA
355 confirming these gypsy retrotransposons are non-functional (Havecker et al., 2004; Marco and
356 Marin, 2008), but short products matching the LTR appeared to correspond to aborted strong stop
357 replication intermediates (Supplemental Fig. S3A). Interestingly, similar LTR fragments from
358 *ATHILA2* comprise an important family of dispersed centromeric satellite repeats known as 106B
359 (May et al., 2005; Thompson et al., 1996), and retrotransposition might account for their origin.
360 Thus functional and non-functional retrotransposons could be readily distinguished even though
361 non-functional *ATHILA* elements are present in copy numbers 3 to 4 orders of magnitude higher
362 than active *ATCOPIA* and *ATGP* elements. As for *ATCOPIA52*, non-productive one-LTR circular
363 DNA, corresponding to autointegration "suicide" products, markedly accumulated in the VLP at
364 levels far higher than productive retrotransposons such as *EVADE*. In contrast, two-LTR circular
365 products of *ATCOPIA52* were very rare, whereas small amounts of *EVADE* two-LTR products
366 were present as previously described (Reinders et al., 2013), presumably due to recombination of
367 non-integrated copies by host DNA repair enzymes in the nucleus. Both short read and long read
368 sequencing revealed that these auto-integration products in *ATCOPIA52* VLP led to non-
369 functional deletion and inversion circles, accounting for lack of transposition.
370

371 Our study shows that RDR6-dependent easiRNA inhibit retrotransposition at multiple levels: via
372 post-transcriptional silencing of genomic RNA, by translational suppression of subgenomic RNA,
373 and by controlling transcription via histone modification. *ATHILA* elements are no longer
374 functional, but they are the primary source of easiRNA which arise by miRNA targeting of a
375 spliced subgenomic RNA encoding the "ENV" protein (Creasey et al., 2014). These easiRNA
376 inhibit polysome association of this subgenomic RNA, and also inhibit transcript levels by
377 guiding histone H3K9me2. This transcriptional silencing occurred in the absence of DNA
378 methylation in *ddm1* mutants. In plants, RNAi dependent histone modification is thought to
379 depend entirely on RNA dependent DNA methylation, found in asymmetric CHH contexts. As
380 CHH methylation stays more or less the same in *ddm1*, while H3K9me2 is increased (Fig. 7), this
381 might indicate the existence of a novel pathway for RNA guided histone methylation, resembling
382 that found in Drosophila, *C.elegans* and fission yeast, which lack DNA methylation. Further
383 investigation will be required to establish if such a pathway exists.

384

385 In contrast to *ATHILA*, linear extrachromosomal copies of *EVADE* accumulated in *ddm1* and
386 were further enriched by mutations in *RDR6*. Like *ATHILA*, *EVADE* is targeted by 3 or 4 miRNA
387 that likely trigger easiRNA from the subgenomic GAG gene transcript, which is found associated
388 with polysomes (Oberlin et al., 2017). However, association of the GAG mRNA with polysomes
389 was unaffected in *ddm1rdr6* mutants. Instead, levels of gRNA increased 3-fold, suggesting that
390 *EVADE* easiRNA act postranscriptionally to target gRNA directly. *ATCOPIA52* easiRNA arose
391 from full-length gRNA between the two LTR. Polysomal association of full length *EVADE* GAG-
392 POL is far more abundant than *ATCOPIA52* GAG-POL, although both were unchanged in the
393 absence of *RDR6* (Fig. 6). As the INT protein is translated from this transcript, this could
394 contribute to lack of nuclear integration of *ATCOPIA52* relative to *EVADE*. Thus, while easiRNA
395 have a significant impact on *COPIA* gRNA accumulation, and so inhibit increases in copy
396 number, they have only limited impact on translation. 22nt tRNA-derived small RNA fragments
397 (3'CCA-tRFs) were recently shown to inhibit endogenous retroviruses (ERV) in mammalian cells

398 by targeting the PBS by RNA interference (Schorn et al., 2017), and it is possible that *EVADE*
399 easiRNA may have a similar function in plants.

400

401 In conclusion, next generation long-read and short-read sequencing of VLP DNA has revealed
402 features that distinguish functional and non-functional replication intermediates, and provides a
403 powerful tool for identifying active transposons from complex genomes, and for investigating
404 molecular signatures of LTR retrotransposons. One such feature is the central PPT (cPPT), which
405 is present in *EVADE* but absent in *ATCOPIA52*. cPPT are hallmarks of the most active
406 retrotransposons including Ty1 in yeast, as well as HIV and other lentiviruses, where cPPT are
407 thought to be important for nuclear import of cDNA (VandenDriessche et al., 2002; Zennou et al.,
408 2000). Our work shows that these features may play a significant role in the activity of *EVADE*,
409 the most active retrotransposon in Arabidopsis, and that their absence may account for the lack of
410 nuclear integration of *ATCOPIA52*, and high levels of "suicide" by autointegration. By
411 comparing VLP sequencing, transcriptome sequencing and translatome sequencing we have been
412 able to establish the multiple levels at which easiRNA regulate the Arabidopsis retrovirome. Our
413 methods are widely applicable to other plant and animal models and to human cells, especially
414 those with genomes that contain very large numbers of non-functional LTR retrotransposons.

415

416 **Acknowledgements**

417 We thank Vincent Colot, Leandro Quadrano, Tetsuji Kakutani and all members of the
418 Martienssen laboratory for discussions. Research in the Martienssen laboratory is supported by
419 the US National Institutes of Health (NIH) grant R01 GM067014, The National Science
420 Foundation Plant Genome Research Program, and by the Howard Hughes Medical Institute. The
421 authors acknowledge assistance from the Cold Spring Harbor Laboratory Shared Resources,
422 which are funded in part by the Cancer Center Support Grant (5PP30CA045508).

423 ## Author Contributions

424 SCL, EE and RM designed the study; SCL and EE performed the experiments; SCL, EE, BB, and
425 AS analyzed the data and its significance; SCL, EE and RM wrote the manuscript.

426

427 ## Disclosure declaration

428 The authors declare no competing interest.

429

430 ## Methods

431 **Plant materials**

432 All genotypes in this study are Col-0 background including wild-type, *dcl1-11*, *ddm1-2*, and *rdr6-*
433 *15*. Genotyping primers are listed in Supplemental Table S6. Homozygous plants of *ddm1-2* and
434 *ddm1-2 rdr6-15* were generated from heterozygous *ddm1-2* backcrossed five times with Col-0
435 (*ddm1-2* BC5), and their 2nd generation was used for VLP DNA-seq experiments. For polysomal
436 RNAseq experiments, inbred *ddm1-2* was independently crossed to *35S:FLAG-RPL18* and to
437 *rdr6-15 35S:FLAG-RPL18*. The F3 plants were used for polysomal RNA purification.

438

439 **gDNA extraction and DNA analyses**

440 Whole inflorescence stems of 4 week-old Arabidopsis plants were frozen and ground in liquid
441 nitrogen. Total gDNA was isolated using Nucleon PhytoPure kit (GE healthcare). *EVADE* DNA
442 copy number was quantified using qPCR with *EVADE* qPCR primers and single copy gene
443 primers as reference (the primers are listed in Supplemental Table S6). Southern blotting was
444 performed using *EVADE* Probe B as described (Mirouze et al., 2009).

445 **Chromatin immunoprecipitation (ChIP)**

446 ChIP was performed with two biological replicates of 10-d-old seedlings using H3K9me2
447 (Abcam; ab1220) and H3 (Abcam; ab1791) antibodies, following a previously described protocol
448 (Ingouff et al., 2017). Sequencing libraries were prepared using NEBNext Ultra II DNA Library
449 Prep Kit for Illumina (New England Biolabs) with size selection for ~200 bp insert DNA. The
450 ChIP-seq libraries were sequenced using Illumina NextSeq High Output SR 76 with 76-cycle
451 single reads. Two biological replicates were prepared and sequenced for each genotype of interest.
452 Prior to alignment, adapter trimming was performed using Trimmomatic (Bolger, et al., 2014)
453 and read quality was assessed with FastQC
454 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Reads were aligned to the TAIR10
455 reference genome using BWA-MEM (https://arxiv.org/abs/1303.3997) with default parameters.
456 Only primary alignments were retained, and optical and PCR duplicates were removed using
457 Samtools (Li et al., 2009). Peak calling was performed using MACS2 (Zhang et al., 2008) broad
458 peak calling with a q-value cutoff of 0.05 and normalization by signal per million reads. Peaks
459 that were differentially regulated across genotypes were identified using MAnorm (Shao et al.,
460 2012) and confirmed between biological replicates. Annotation of these differentially regulated
461 peaks was performed using a combination of BEDOPS (Neph et al., 2012) tools and custom
462 scripts. Deeptools (Ramirez et al., 2014) was used to visualize the data.

463

464 **RNA extraction and RT-qPCR**

465 Total RNA was isolated from the same tissues used for gDNA extraction with Direct-zol RNA
466 MiniPrep Plus (Zymo Research). DNase I was treated on column. cDNA was synthesized with
467 SuperScript VILO Master Mix (Thermo Fisher Scientific). qPCR was performed using iQ SYBR
468 Green Supermix. Primers are listed in Supplemental Table S6.

469

470 **Polysomal RNA-seq**

471 Total polysome was isolated using ribosome immunopurification as described previously
472 (Mustroph et al., 2009; Mustroph et al., 2013). Briefly, inflorescence tissues of *FLAG-RPL18*
473 lines were ground in liquid nitrogen and transferred to polysome extraction buffer (PEB). Cell
474 debris was removed by centrifugation and filtering with miracloth. The supernatant was taken and
475 transferred to pre-washed EZview anti-FLAG agarose beads (Sigma) for 2 h at 4 °C. The agarose
476 beads containing polysomes were washed once with PEB and three times with washing buffer.
477 Polysomes were eluted using 3X FLAG peptide (Sigma) and used for RNA extraction with
478 Direct-zol RNA miniprep kit (Zymo Research) including DNase I treatment. Ribosomal RNA
479 (rRNA) in the samples was depleted by Ribo-Zero Magnetic Kit (Plant Leaf) (Epicentre). Then,
480 rRNA free samples were used for RNA-seq library preparation using ScriptSeq v2 RNA-Seq
481 Library Preparation Kit (EPicentre). Microsome-polysomal RNA was obtained using a previously
482 described method with some modifications (Li et al., 2013). Briefly, 2 g frozen tissues were
483 suspended to 7 ml microsome extraction buffer (MEB). After removing cell debris by filtration
484 with micracloth and centrifugation at 10,000g for 15 min at 4°C, the supernatant was transferred
485 on the top of 1.7M/0.6M sucrose cushions and applied to ultracentrifugation using swing rotor at
486 140,000g for 1 h at 4°C. The microsome fraction of the 1.7M/0.6M layer interface was harvested
487 and diluted 10 times by MEB and centrifuged at 140,000g for 0.5 h at 4 °C to obtain microsome
488 pellet. The pellet was re-suspended with 8 ml PEB and used for ribosome immunopurification
489 and RNA-seq library preparation as described above. The PE 101 sequencing data was obtained
490 using Illumina HiSeq 2000 platform. The paired-end reads were mapped to Arabidopsis TAIR10
491 genome using Tophat and the polysome occupancy (Polysomal RNA / Total RNA) was
492 calculated using systemPipeR package (Backman and Girke, 2016) with raw count data obtained
493 by Cuffnorm.

494

495

496 **VLP DNA-seq**

497 Virus-like-particles were purified using modified method reported previously (Bachmair et al.,
498 2004). 4 g of 4 week-old whole inflorescence stems were ground with 10 ml of ice-cold VLP
499 extraction buffer and 10 ml of sea sand on ice. 10 ml of the extraction buffer and Triton X-100
500 were added and mixed. The slurry was transferred to a 50 ml tube and centrifuged for 5 min at
501 180g and 4 °C. The supernatant was carefully transferred onto 5 ml of prechilled 15% sucrose, 10
502 mM potassium phosphate buffer, pH 7.2 and ultracentrifuged for 1.5 h at 109,000g and 4 °C
503 using fixed angle rotor. The pellet was washed with the 15% sucrose buffer and resuspended with
504 4 ml particle suspension buffer to obtain VLP fractions. To remove non-VLP DNA, 0.5 ml of the
505 VLP sample was treated with 5 µl of 1 mg/ml DNase I at 37°C for 10 min. 20 µl of 0.25 M
506 EDTA, 50 µl of 10% SDS, 25 µl of 10 mg/ml proteinase were added and incubated at 65°C for 10
507 min. VLP DNA was purified by 0.5 ml equilibrated (pH 8.0) phenol:chloroform:IAA (25:24:1)
508 mixture three times and with 0.5 ml chloroform:IAA (24:1) once. The last aqueous fraction was
509 transferred into a new 1.5ml tube and used for 100% ethanol precipitation with 40 µl 3M sodium
510 acetate, pH 7.0. The DNA pellet was washed with 70% ethanol, dried, and resuspended with 100
511 µl TE buffer. 1 µl of RNase A (10 mg/ml) was added to the VLP DNA sample and incubated 10
512 min. The treated DNA sample was purified using DNA Clean & Concentrator (Zymo Research).
513 The DNA was sheared to 650 bp using Covaris S220 and subsequently used for DNA-seq library
514 preparation with NEBNext Ultra DNA Library Prep Kit (New England Biolabs). The paired-end
515 sequencing datasets with 101 nt read length (PE101) were obtained by Illumina HiSeq 2000.
516 Adapters were trimmed from raw reads with Skewer (Jiang et al., 2014) in paired-end mode and
517 read pairs with both mates longer than 25 nt were retained. Reads were aligned to the TAIR10
518 genome with STAR (Dobin et al., 2013) in two-pass mode to improve spliced alignment at
519 unannotated introns. Intact bacteria co-purified with VLP, as indicated by large numbers of reads
520 mapping to bacterial genomes (up to 95% in WT), and these were discarded. Reads mapping
521 equally well to multiple locations were randomly assigned, and chimeric/split read alignments
522 were output separately from concordant alignments. Optical and PCR duplicates were removed

523 from the alignments with picard-tools (http://broadinstitute.github.io/picard). Counts of reads
524 mapping to the TAIR10 transposon annotations were computed with featurecounts (Liao et al.,
525 2014). Pairwise differential expression at TAIR10 transposon loci was tested across three wild-
526 type, two *ddm1*, and three *ddm1rdr6* replicates using quasi-likelihood F-tests in edgeR (Robinson
527 et al., 2010), controlling FDR at 5% and a $\log_2$(fold-change) threshold of 2.

528

529 Oxford Nanopore Technologies (ONT) long-read libraries were prepared as follows: 10 ng per
530 genotype of purified VLP DNA extract was pooled from the replicate samples and initially
531 amplified following the conditions in the "1D Low-input genomic DNA with PCR" (SQK-
532 LSK108) protocol with reagents. End-repair, dA-tailing and PCR adapter ligation were performed,
533 followed by 16 cycles of PCR amplification. PCR products were purified and concentrated with
534 Ampure XP beads (Agencourt), and 300 ng of eluate per sample was carried through to library
535 preparation following the "1D Genomic DNA by Ligation" protocol with SKQ-LSK109 reagents.
536 Libraries were loaded onto r9.4 (FLO-MIN106) flow cells and sequenced on a GridION X5.
537 Basecalling was performed offline with Guppy v2.3.1 using the default r9.4.1 model. Using
538 porechop (https://github.com/rrwick/Porechop), ONT sequencing adapters were trimmed from 5'
539 ends, 3' ends, and the middle of raw reads. Reads with middle adapters were split. Remaining
540 reads longer than 100 bp were aligned to the TAIR10 reference with minimap2 (Li, 2018) for
541 coverage and read alignment plots. Structural variants were called on NGMLR (Sedlazeck et al.,
542 2018) alignments using Sniffles (Sedlazeck et al., 2018) with default parameters, except
543 minimum read support was reduced to 3.

544

545 **5' RACE PCR**

546 5' RACE PCR was performed using FirstChoice RLM-RACE Kit (Thermo Fisher Scientific)
547 without the treatments of calf intestine alkaline phosphatase and tobacco acid pyrophosphatase. A

548 gene-specific primer was used for cDNA synthesis after adaptor ligation (Supplemental Table S6).
549 1st and 2nd nested PCR was performed with the primers are listed.

550

551 **Small RNA-seq data**

552 Small RNA-seq libraries from inflorescence and pollen for comparisons of 21, 22, and 24nt small
553 RNA between wild-type and *ddm1* were prepared as previously described (Borges et al., 2018).
554 Wild-type pollen sample was previously deposited in the Gene Expression Omnibus (GEO)
555 database (GSM2829912). Briefly, small RNAs were purified by running total RNA from pollen
556 and inflorescence tissues on acrylamide gels (15% polyacrylamide, 7 M urea) with size-selection
557 of 18-to- 30-nt regions. Small RNAs were extracted from the gel bands using Trizol LS (Life
558 Technologies) and Direct-zol columns (Zymo Research). Libraries were prepared with the
559 TruSeq small RNA sample preparation kit (Illumina) and sequenced in Illumina MiSeq platform.
560 Data analysis was done as previously reported (Borges et al., 2018). 21-22nt small RNA datasets
561 from inflorescence (Creasey et al., 2014) were obtained from NCBI GEO accession GSE52951.
562 After adapter trimming with Skewer, reads were quality filtered with fastp (Chen et al., 2018) and
563 aligned to the TAIR10 genome with ShortStack (Axtell, 2013) with default parameters except "--
564 bowtie_m 1000 --ranmax 50".

565

566 **LTR Retrotransposon Annotation**

567 GenomeTools was used to structurally annotate retrotransposons across the TAIR10 genome.
568 First, LTRharvest (Ellinghaus et al., 2008) was run to detect LTR sequences with at least 85%
569 similarity separated by 1-15 kbp flanked by target site duplications and the TGCA motif. Then,
570 LTRdigest (Steinbiss et al., 2009) was run to annotate internal transposon features including the
571 PBS, PPT, and GAG and POL protein homology.

572

573 **Genome Browser Figures**

574 Genome-wide read coverage for VLP DNA, small RNA, total and polysomal RNA libraries was
575 calculated with bamCoverage from deepTools (Ramirez et al., 2014) and normalized to reads per
576 nucleotide per million mapped reads and plotted across the genome with Gviz (Hahne and Ivanek,
577 2016) or IGV (Thorvaldsdottir et al., 2013).

578

579 **Data access**

580 The datasets generated during and/or analyzed during the current study are available at NCBI
581 (GEO study: GSE128932).

582

583 **References**

584 Arribas-Hernandez L, Marchais A, Poulsen C, Haase B, Hauptmann J, Benes V, Meister G,
585 Brodersen P. 2016. The Slicer Activity of ARGONAUTE1 Is Required Specifically for the
586 Phasing, Not Production, of Trans-Acting Short Interfering RNAs in Arabidopsis. *Plant Cell* **28:**
587 1563-1580.

588

589 Axtell MJ. 2013. ShortStack: comprehensive annotation and quantification of small RNA genes.
590 *RNA* **19:** 740-751.

591

592 Axtell MJ, Jan C, Rajagopalan R, Bartel DP. 2006. A two-hit trigger for siRNA biogenesis in
593 plants. *Cell* **127:** 565-577.

594

595 Bachmair A, Garber K, Takeda S, Sugimoto K, Kakutani T, Hirochika H. 2004. Biochemical
596 analysis of long terminal repeat retrotransposons. *Methods Mol Biol* **260:** 73-82.

597

598 Bachmann AS, Corpuz G, Hareld WP, Wang G, Coller BA. 2004. A simple method for the rapid
599 purification of copia virus-like particles from Drosophila Schneider 2 cells. *J Virol Methods* **115:**
600 159-165.
601
602 Backman T, Girke T. 2016. systemPipeR: NGS workflow and report generation environment.
603 *BMC Bioinformatics* **17:** 388.
604
605 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence
606 data. *Bioinformatics* **30**: 2114-2120.
607
608 Borges F, Martienssen RA. 2015. The expanding world of small RNAs in plants. *Nat Rev Mol*
609 *Cell Biol* **16:** 727-741.
610
611 Borges F, Parent JS, van Ex F, Wolff P, Martinez G, Kohler C, Martienssen RA. 2018.
612 Transposon-derived small RNAs triggered by miR845 mediate genome dosage response in
613 Arabidopsis. *Nat Genet* **50:** 186-192.
614
615 Breakfield NW, Corcoran DL, Petricka JJ, Shen J, Sae-Seaw J, Rubio-Somoza I, Weigel D, Ohler
616 U, Benfey PN. 2012. High-resolution experimental and computational profiling of tissue-specific
617 known and novel miRNAs in Arabidopsis. *Genome Res* **22:** 163-176.
618
619 Chang W, Jaaskelainen M, Li SP, Schulman AH. 2013. BARE retrotransposons are translated and
620 replicated via distinct RNA pools. *PLoS One* **8:** e72270.
621
622 Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor.
623 *Bioinformatics* **34:** i884-i890.
624

625 Creasey KM, Zhai J, Borges F, Van Ex F, Regulski M, Meyers BC, Martienssen RA. 2014.
626 miRNAs trigger widespread epigenetically activated siRNAs from transposons in Arabidopsis.
627 *Nature* **508:** 411-415.
628
629 Czech B, Munafo M, Ciabrelli F, Eastwood EL, Fabry MH, Kneuss E, Hannon GJ. 2018. piRNA-
630 Guided Genome Defense: From Biogenesis to Silencing. *Annu Rev Genet* **52:** 131-157.
631
632 Dai X, Zhuang Z, Zhao PX. 2018. psRNATarget: a plant small RNA target analysis server (2017
633 release). *Nucleic Acids Res* **46:** W49-W54.
634
635 de Felippes FF, Marchais A, Sarazin A, Oberlin S, Voinnet O. 2017. A single miR390 targeting
636 event is sufficient for triggering TAS3-tasiRNA biogenesis in Arabidopsis. *Nucleic Acids Res* **45:**
637 5539-5554.
638
639 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras
640 TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29:** 15-21.
641
642 Eichinger DJ, Boeke JD. 1988. The DNA intermediate in yeast Ty1 element transposition
643 copurifies with virus-like particles: cell-free Ty1 transposition. *Cell* **54:** 955-966.
644
645 Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for de
646 novo detection of LTR retrotransposons. *BMC Bioinformatics* **9:** 18.
647
648 Fagegaltier D, Bouge AL, Berry B, Poisot E, Sismeiro O, Coppee JY, Theodore L, Voinnet O,
649 Antoniewski C. 2009. The endogenous siRNA pathway is involved in heterochromatin formation
650 in Drosophila. *Proc Natl Acad Sci* **106:** 21258-21263.
651

652 Finnegan DJ. 2012. Retrotransposons. *Curr Biol* **22:** R432-437.

653

654 Garfinkel DJ, Stefanisko KM, Nyswaner KM, Moore SP, Oh J, Hughes SH. 2006.
655 Retrotransposon suicide: formation of Ty1 circles and autointegration via a central DNA flap. *J*
656 *Virol* **80:** 11920-11934.

657

658 Grant-Downton R, Le Trionnaire G, Schmid R, Rodriguez-Enriquez J, Hafidh S, Mehdi S, Twell
659 D, Dickinson H. 2009. MicroRNA and tasiRNA diversity in mature pollen of *Arabidopsis*
660 *thaliana*. *BMC Genomics* **10:** 643.

661

662 Griffiths J, Catoni M, Iwasaki M, Paszkowski J. 2018. Sequence-Independent Identification of
663 Active LTR Retrotransposons in Arabidopsis. *Mol Plant* **11:** 508-511.

664

665 Gu SG, Pak J, Guang S, Maniar JM, Kennedy S, Fire A. 2012. Amplification of siRNA in
666 *Caenorhabditis elegans* generates a transgenerational sequence-targeted histone H3 lysine 9
667 methylation footprint. *Nat Genet* **44:** 157-164.

668

669 Hahne F, Ivanek R. 2016. Visualizing Genomic Data Using Gviz and Bioconductor. *Methods Mol*
670 *Biol* **1418:** 335-351.

671

672 Havecker ER, Gao X, Voytas DF. 2004. The diversity of LTR retrotransposons. *Genome Biol* **5:**
673 225.

674

675 Hu C, Saenz DT, Fadel HJ, Walker W, Peretz M, Poeschla EM. 2010. The HIV-1 central
676 polypurine tract functions as a second line of defense against APOBEC3G/F. *J Virol* **84:** 11981-
677 11993.

678

679 Huang CR, Burns KH, Boeke JD. 2012. Active transposition in genomes. *Annu Rev Genet* **46:**
680 651-675.
681
682 Ingouff M, Selles B, Michaud C, Vu TM, Berger F, Schorn AJ, Autran D, Van Durme M,
683 Nowack MK, Martienssen RA, et al. 2017. Live-cell analysis of DNA methylation during sexual
684 reproduction in Arabidopsis reveals context and sex-specific dynamics controlled by
685 noncanonical RdDM. *Genes Dev* **31:** 72-83.
686
687 Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, Paszkowski, J. 2011. An siRNA pathway
688 prevents transgenerational retrotransposition in plants subjected to stress. *Nature* **472:** 115-119.
689
690 Jaaskelainen M, Mykkanen AH, Arna T, Vicient CM, Suoniemi A, Kalendar R, Savilahti H,
691 Schulman AH. 1999. Retrotransposon BARE-1: expression of encoded proteins and formation of
692 virus-like particles in barley cells. *Plant J* **20:** 413-422.
693
694 Jiang H, Lei R, Ding SW, Zhu S. 2014. Skewer: a fast and accurate adapter trimmer for next-
695 generation sequencing paired-end reads. *BMC Bioinformatics* **15:** 182.
696
697 Juntawong P, Girke T, Bazin J, Bailey-Serres J. 2014. Translational dynamics revealed by
698 genome-wide profiling of ribosome footprints in Arabidopsis. *Proc Natl Acad Sci* **111:** E203-212.
699
700 Kenna MA, Brachmann CB, Devine SE, Boeke JD. 1998. Invading the yeast nucleus: a nuclear
701 localization signal at the C terminus of Ty1 integrase is required for transposition in vivo. *Mol*
702 *Cell Biol* **18:** 1115-1124.
703
704 Lanciano S, Carpentier MC, Llauro C, Jobet E, Robakowska-Hyzorek D, Lasserre E, Ghesquiere
705 A, Panaud O, Mirouze M. 2017. Sequencing the extrachromosomal circular mobilome reveals

706 retrotransposon activity in plants. *PLoS Genet* **13:** e1006630.

707

708 Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34:** 3094-
709 3100.

710

711 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,
712 Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools.
713 *Bioinformatics* ***25*****:** 2078-2079.

714

715 Li S, Liu L, Zhuang X, Yu Y, Liu X, Cui X, Ji L, Pan Z, Cao X, Mo B*,* et al*.* 2013. MicroRNAs
716 inhibit the translation of target mRNAs on the endoplasmic reticulum in Arabidopsis. *Cell* **153:**
717 562-574.

718

719 Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for
720 assigning sequence reads to genomic features. *Bioinformatics* **30:** 923-930.

721

722 Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, McCombie WR, Lavine K, Mittal V,
723 May B, Kasschau KD*,* et al*.* 2004. Role of transposable elements in heterochromatin and
724 epigenetic control. *Nature* **430:** 471-476.

725

726 Marco A, Marin I. 2008. How Athila retrotransposons survive in the Arabidopsis genome. *BMC*
727 *Genomics* **9:** 219.

728

729 Mari-Ordonez A, Marchais A, Etcheverry M, Martin A, Colot V, Voinnet O. 2013.
730 Reconstructing de novo silencing of an active plant retrotransposon. *Nat Genet* **45:** 1029-1039.

731

732 Marin E, Jouannet V, Herz A, Lokerse AS, Weijers D, Vaucheret H, Nussaume L, Crespi MD,

733 Maizel A. 2010. miR390, Arabidopsis TAS3 tasiRNAs, and their AUXIN RESPONSE FACTOR
734 targets define an autoregulatory network quantitatively regulating lateral root growth. *Plant Cell*
735 **22:** 1104-1117.
736
737 Martienssen R, Moazed, D. 2015. RNAi and heterochromatin assembly. *Cold Spring Harb*
738 *Perspect Biol* **7:** a019323.
739
740 May BP, Lippman ZB, Fang Y, Spector DL, Martienssen RA. 2005. Differential regulation of
741 strand-specific transcripts from Arabidopsis centromeric satellite repeats. *PLoS Genet* **1:** e79.
742
743 Mirouze M, Reinders J, Bucher E, Nishimura T, Schneeberger K, Ossowski S, Cao J, Weigel D,
744 Paszkowski J, Mathieu O. 2009. Selective epigenetic control of retrotransposition in Arabidopsis.
745 *Nature* **461:** 427-430.
746
747 Mules EH, Uzun O, Gabriel A. 1998. In vivo Ty1 reverse transcription can generate replication
748 intermediates with untidy ends. *J Virol* **72:** 6490-6503.
749
750 Munir S, Thierry S, Subra F, Deprez E, Delelis O. 2013. Quantitative analysis of the time-course
751 of viral DNA forms during the HIV-1 life cycle. *Retrovirology* **10:** 87.
752
753 Mustroph A, Juntawong P, Bailey-Serres J. 2009. Isolation of plant polysomal mRNA by
754 differential centrifugation and ribosome immunopurification methods. *Methods Mol Biol* **553:**
755 109-126.
756
757 Mustroph A, Zanetti ME, Girke T, Bailey-Serres J. 2013. Isolation and analysis of mRNAs from
758 specific cell types of plants by ribosome immunopurification. *Methods Mol Biol* **959:** 277-302.
759

760 Neph S, Kuehn MS, Reynolds AP, Haugen E, Thurman RE, Johnson AK, Rynes E, Maurano MT,
761 Vierstra J, Thomas S, et al. 2012. BEDOPS: high-performance genomic feature operations.
762 *Bioinformatics* **28:** 1919-1920.
763
764 Nuthikattu S, McCue AD, Panda K, Fultz D, DeFraia C, Thomas EN, Slotkin RK. 2013. The
765 initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22
766 nucleotide small interfering RNAs. *Plant Physiol* **162:** 116-131.
767
768 Oberlin S, Sarazin A, Chevalier C, Voinnet O, Mari-Ordonez A. 2017. A genome-wide
769 transcriptome and translatome analysis of Arabidopsis transposons identifies a unique and
770 conserved genome expression strategy for Ty1/Copia retroelements. *Genome Res* **27:** 1549-1562.
771
772 Pachulska-Wieczorek K, Le Grice SF, Purzycka KJ. 2016. Determinants of Genomic RNA
773 Encapsidation in the Saccharomyces cerevisiae Long Terminal Repeat Retrotransposons Ty1 and
774 Ty3. *Viruses* **8:** 193.
775
776 Peterson-Burch BD, Voytas DF. 2002. Genes of the Pseudoviridae (Ty1/copia retrotransposons).
777 *Mol Biol Evol* **19:** 1832-1845.
778
779 Quadrana L, Etcheverry M, Gilly A, Caillieux E, Madoui MA, Guy J, Bortolini Silveira A,
780 Engelen S, Baillet V, Wincker P, et al. 2019. Transposition favors the generation of large effect
781 mutations that may facilitate rapid adaption. *Nat Commun* **10:** 3421.
782
783 Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T. 2014. deepTools: a flexible platform for
784 exploring deep-sequencing data. *Nucleic Acids Res* **42:** W187-191.
785
786 Reinders J, Mirouze M, Nicolet J, Paszkowski J. 2013. Parent-of-origin control of

787 transgenerational retrotransposon proliferation in Arabidopsis. *EMBO Rep* **14:** 823-828.

788

789 Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential
790 expression analysis of digital gene expression data. *Bioinformatics* **26:** 139-140.

791

792 Sabot F, Schulman AH. 2006. Parasitism and the retrotransposon life cycle in plants: a
793 hitchhiker's guide to the genome. *Heredity* (Edinb) **97:** 381-388.

794

795 Schorn AJ, Gutbrod MJ, LeBlanc C, Martienssen R. 2017. LTR-Retrotransposon Control by
796 tRNA-Derived Small RNAs. *Cell* **170:** 61-71 e11.

797

798 Schorn AJ, Martienssen R. 2018. Tie-Break: Host and Retrotransposons Play tRNA. *Trends Cell*
799 *Biol* **28:** 793-806.

800

801 Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, Schatz MC.
802 2018. Accurate detection of complex structural variations using single-molecule sequencing. *Nat*
803 *Methods* **15:** 461-468.

804

805 Shao Z, Zhang Y, Yuan GC, Orkin SH, Waxman DJ. 2012. MAnorm: a robust model for
806 quantitative comparison of ChIP-Seq data sets. *Genome Biol* **13:** R16.

807

808 Sloan RD, Wainberg MA. 2011. The role of unintegrated DNA in HIV infection. *Retrovirology*
809 **8:** 52.

810

811 Slotkin RK, Vaughn M, Borges F, Tanurdzic M, Becker JD, Feijo JA, Martienssen RA. 2009.
812 Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* **136:**
813 461-472.

814

815 Steinbiss S, Willhoeft U, Gremme G, Kurtz S. 2009. Fine-grained annotation and classification of
816 de novo predicted LTR retrotransposons. *Nucleic Acids Res* **37:** 7002-7013.
817

818 Stuart T, Eichten SR, Cahn J, Karpievitch YV, Borevitz JO, Lister R. 2016. Population scale
819 mapping of transposable element diversity reveals links to gene regulation and epigenomic
820 variation. *Elife* **5:** e20777.
821

822 Teixeira FK, Okuniewska M, Malone CD, Coux R-X, Rio DC, Lehmann R. 2017. piRNA-
823 mediated regulation of transposon alternative splicing in soma and germline. *Nature* **552:** 268-
824 272.

825

826 Thompson HL, Schmidt R, Dean C. 1996. Identification and Distribution of Seven Classes of
827 Middle-Repetitive DNA in the *Arabidopsis Thaliana* Genome, *Nucleic Acids Res* **24:** 3017-3022.
828

829 Thorvaldsdottir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-
830 performance genomics data visualization and exploration. *Brief Bioinform* **14:** 178-192.
831

832 Tsukahara S, Kobayashi A, Kawabe A, Mathieu O, Miura A, Kakutani T. 2009. Bursts of
833 retrotransposition reproduced in Arabidopsis. *Nature* **461:** 423-426.
834

835 VandenDriessche T, Thorrez L, Naldini L, Follenzi A, Moons L, Berneman Z, Collen D, Chuah
836 MK. 2002. Lentiviral vectors containing the human immunodeficiency virus type-1 central
837 polypurine tract can efficiently transduce nondividing hepatocytes and antigen-presenting cells in
838 vivo. *Blood* **100:** 813-822.
839

840 Volpe TA, Kidner C, Hall IM, Teng G, Grewal SI, Martienssen RA. 2002. Regulation of

841 heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297:** 1833-1837.
842
843 Vongs A, Kakutani T, Martienssen RA, Richards EJ. 1993. *Arabidopsis thaliana* DNA
844 methylation mutants. *Science* **260:** 1926-1928.
845
846 Wang W, Haberer G, Gundlach H, Glasser C, Nussbaumer T, Luo MC, Lomsadze A,
847 Borodovsky M, Kerstetter RA, Shanklin J, et al. 2014. The *Spirodela polyrhiza* genome reveals
848 insights into its neotenous reduction fast growth and aquatic lifestyle. *Nat Commun* **5:** 3311.
849
850 Wilhelm M, Uzun O, Mules EH, Gabriel A, Wilhelm FX. 2001. Polypurine tract formation by
851 Ty1 RNase H. *J Biol Chem* **276:** 47695-47701.
852
853 Williams L, Carles CC, Osmont KS, Fletcher JC. 2005. A database analysis method identifies an
854 endogenous trans-acting short-interfering RNA that targets the Arabidopsis ARF2, ARF3, and
855 ARF4 genes. *Proc Natl Acad Sci* **102:** 9703-9708.
856
857 Wright DA, Voytas DF. 2002. Athila4 of Arabidopsis and Calypso of soybean define a lineage of
858 endogenous plant retroviruses. *Genome Res* **12:** 122-131.
859
860 Wurtzer S, Goubard A, Mammano F, Saragosti S, Lecossier D, Hance AJ, Clavel F. 2006.
861 Functional central polypurine tract provides downstream protection of the human
862 immunodeficiency virus type 1 genome from editing by APOBEC3G and APOBEC3B. *J Virol*
863 **80:** 3679-3683.
864
865 Xie W, Donohue RC, Birchler JA. 2013. Quantitatively increased somatic transposition of
866 transposable elements in Drosophila strains compromised for RNAi. *PLoS One* **8:** e72163.
867

868 Yoshioka K, Honma H, Zushi M, Kondo S, Togashi S, Miyake T, Shiba T. 1990. Virus-like
869 particle formation of Drosophila copia through autocatalytic processing. *EMBO J* **9:** 535-541.
870
871 Zennou V, Petit C, Guetard D, Nerhbass U, Montagnier L, Charneau P. 2000. HIV-1 genome
872 nuclear import is mediated by a central DNA flap. *Cell* **101:** 173-185.
873
874 Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM,
875 Brown M, Li W, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9:** R137.
876