Impact of transposable elements on methylation and gene expression across natural accessions of *Brachypodium distachyon.*

Michele Wyler[1], Christoph Stritt[1], Jean-Claude Walser[2], Célia Baroux[1] and Anne C. Roulin[1*]

[1]Institute for Plant and Microbial Biology, University of Zurich, Zollikerstrasse 107, Zurich 8008, Switzerland

[2] Genetic Diversity Centre, ETH Zürich, Universitätstrasse 16, Zurich, Switzerland

*Corresponding author: Anne C. Roulin; Institute for Plant and Microbial Biology, University of Zurich, Zollikerstrasse 107, Zurich 8008, Switzerland; +41 (0)44 63 48398; anne.roulin@botinst.uzh.ch

Abstract

Transposable elements (TEs) constitute a large fraction of plant genomes and are mostly present in a transcriptionally silent state through repressive epigenetic modifications such as DNA methylation. TE silencing is believed to influence the regulation of adjacent genes, possibly as DNA methylation spreads away from the TE. Whether this is a general principle or a context-dependent phenomenon is still under debate, pressing for studying the relationship between TEs, DNA methylation and nearby gene expression in additional plant species. Here we used the grass *Brachypodium distachyon* as a model and produced DNA methylation and transcriptome profiles for eleven natural accessions. In contrast to what is observed in *Arabidopsis thaliana*, we found that TEs have a limited impact on methylation spreading and that only few TE families are associated to a low expression of their adjacent genes. Interestingly, we found that a subset of TE insertion polymorphisms is associated with differential gene expression across accessions. Thus, although not having a global impact on gene expression, distinct TE insertions may contribute to specific gene expression patterns in *B. distachyon.*

Key words: Transposable elements, DNA methylation, gene expression, *Brachypodium distachyon*

Significance statement: Transposable elements (TEs) are a major component of plant genomes and a source of genetic and epigenetic innovations underlying adaptation to changing environmental conditions. Yet molecular evidence linking TE silencing and nearby gene expression are lacking for many plant species. We show

that in the model grass Brachypodium DNA methylation spreads over very short distances around TEs, with an influence on gene expression for a small subset of TE families.

42 **Introduction**

43 Since their discovery by Barbara McClintock in the late 1940's (McClintock 1947),
44 transposable elements (TEs) have been shown to populate a large fraction of
45 eukaryotic genomes and to be an important source of genetic and phenotypic
46 diversity. This is explained by the observation that TEs have the ability to modulate
47 nearby gene expression and can remodel genomes when activated by stress
48 (Grandbastien 1998; Grandbastien 2005; Thieme et al. 2017), thereby providing raw
49 diversity necessary for adaptation to changing environments (Bonchev and Parisod
50 2013; Rey et al. 2016). Remarkable examples of single-trait modulation by a TE
51 insertion providing a selective advantage have been reported in both plants and
52 animals (Guio et al. 2014; Van't Hof et al. 2016; Horváth et al. 2017).

53 While TEs can provide selective advantages, their mutagenic properties may
54 at the same time represent a threat for the integrity of the host genome (McClintock
55 1950; Shang et al. 2017; Thind et al. 2018). In response, genomes have evolved
56 several layers of defence to contain TE activity. In plants, transcriptional repression of
57 TEs is mediated by the RNA-directed DNA methylation (RdDM) pathway (for
58 review Sigman and Slotkin 2016; Zhang et al. 2018). Initiated by the plant-specific
59 RNA polymerase IV (*Pol IV*), TE transcription leads to the formation of 24 nt siRNA
60 targeting DNA methylation at the TE source locus. Acting as a double-edged sword,
61 DNA (and histone) methylation may spread to neighbouring regions and possibly
62 contribute the alteration of adjacent gene expression (Hollister et al. 2010; Wang et al.
63 2013).

64 Recent studies reported on the impact of TEs on DNA methylation and
65 expression of nearby genes in different plant species (Eichten et al. 2012;
66 Makarevitch et al. 2015; Eichten et al. 2016; Choi and Purugganan 2017; Wang et al.
67 2018). Yet, most of our knowledge on the regulatory mechanisms largely rely on the
68 eudicot model *Arabidopsis thaliana* (for review Sigman and Slotkin 2016; Zhang et
69 al. 2018). In *A. thaliana*, the presence of TEs tends to be associated with a reduced
70 expression of adjacent genes (Wang et al. 2013), but whether this also occurs in other
71 species remains to be demonstrated (Choi and Purugganan 2017). In this context, we
72 contributed an analysis of DNA methylation and gene expression in the model grass
73 *Brachypodium distachyon*.

Established as an important model for functional genomics, the small genome of *B. distachyon* (272 Mb; 30% of TEs) is fully sequenced and assembled into five chromosomes (The International Brachypodium Consortium 2010). Comparative genomics and cytogenetic analyses (Hasterok et al. 2019) showed that it underwent multiple and independent nested chromosome fusions which resulted in a higher gene density at the former distal regions of the inserted chromosomes. Furthermore, TEs are much more pervasive than in *A. thaliana,* and although their dynamics is well characterized in *B. distachyon* (Stritt et al. 2018; Stritt et al. 2019), little is know about their functional impact. We therefore asked to what extent TEs influence DNA methylation and gene expression in *B. distachyon,* and how these patterns differ from what is observed in *A. thaliana.* In *A. thaliana*, the impact of TEs on adjacent gene expression is extremely consistent across genetically distinct ecotypes originating from the USA (Col-0), Portugal (C24) and Ireland (Bur0; Wang et al. 2013). We thus chose Col-0 as a representative of the species. For *B. distachyon*, less information is available at such a broad geographical scale. We therefore included in our analysis the reference Bd21, which originates from Iraq, as well as ten additional accessions from Spain, France, Iraq and Turkey to further investigate potential within-species variation.

Results and Discussion

TEs largely shape genome-wide DNA methylation patterns in Brachypodium

We produced bisulfite- and RNA-sequencing data for eleven *B. distachyon* accessions. SNPs (Gordon et al. 2017; Stritt et al. 2018) and methylation profiling support their split into three distinct genetic clusters (Figure S1): Western (ABR2, Luc1, Bd30-1 and Sig2; Spain and France), Eastern (the reference accession Bd21, Bd21-3, Bd3-1, BdTR12C and Koz3; Iraq and Turkey) and accessions from a third cluster (BdTR7a, Bd1-1; Turkey) showing a delay in flowering time (Gordon et al. 2017; Stritt et al. 2018). In agreement with the genetic distances calculated with SNPs (Gordon et al. 2017; Stritt et al. 2018), the methylation-based PCA confirms that accessions from the Eastern cluster are more closely related to accessions from the Western cluster than to the two other Turkish accessions BdTR7a and Bd1-1. We also used publically available bisulfite- and RNA-sequencing data for the reference accession Col-0 of *A. thaliana* (Table S1).

107 We first estimated methylation levels (expressed as the percentage of
108 methylated cytosines), gene and TE density in 100 kb windows for the reference
109 genomes of *A. thaliana* (Col-0) and *B. distachyon* (Bd21). The circular visualisation
110 of these three genomic features reveals two distinct architectures: genes and TEs are
111 clearly compartmentalised in *A. thaliana* (Figure 1A) while largely interspersed in
112 *B. distachyon* (Figure 1B). Yet we found that DNA methylation is positively
113 correlated with TE density and negatively correlated with gene density in both species
114 (Figure S2). We then estimated the methylation level of genes and TEs, aggregating
115 the latter at the superfamily level. TE body methylation is significantly higher than
116 gene body methylation in the two species, although variability is observed among TE
117 superfamilies (Figure 1C and D, Table S2-S4 for P-values). Genome-wide DNA
118 methylation patterns are thus largely shaped by TEs in *B. distachyon,* like in other
119 flowering plants species (Zhang et al. 2006; Regulski et al. 2013; Takuno and Gaut
120 2013; Seymour et al. 2014). We also found that most TE superfamilies and genes
121 show a higher level of body methylation in *B. distachyon* than in *A. thaliana* (Figure
122 1C and D, Table S5 for P-values), confirming a low genome-wide level of body
123 methylation in *A. thaliana* compared to other species (Seymour et al. 2014, Eichten et
124 al. 2016).

125 We finally assessed how far methylation spreads around TEs. To do so, we
126 estimated methylation levels in 100 bp windows spanning 2.5 kb around each TE
127 insertion and averaged the values per TE superfamily. We also categorized TE
128 insertions depending on their location on the chromosome (arm vs. centromere) and
129 visualised the results by producing heatmaps. For the three contexts, we observe that
130 DNA methylation around TEs spreads over longer distances in gene-poor centromeric
131 regions than in distal regions of the chromosome arm in *A. thaliana* (Figure 2). In
132 *B. distachyon*, spreading mostly reaches a few 100 bp, except around centromeric
133 *Gypsy* elements where methylation tends to spread beyond 2 kb (Figure 2). Thus, in
134 contrast to our expectations, high levels of TE body methylation in *B. distachyon* are
135 generally not accompanied with a long spreading of methylation, neither in the
136 centromeres nor on the chromosome arms. This is surprising given that extended
137 methylation spreading around TE is associated to gene-poor pericentromeric regions
138 in rice (Choi and Purugganan, 2017), a species closely related to *B. distachyon*.

139 Because patterns are consistent across CpG, CHH and CHG contexts, we
140 concentrated the rest of the analysis on CpG methylation.

141

142 *Impact of TEs on nearby gene expression*

143 In *A. thaliana*, genes flanked by a TE within 2 kb show reduced expression compared
144 to genes devoid of TEs in their vicinity (Wang et al. 2013). It was proposed that this
145 may result from a combined effect of DNA methylation spreading and DNA structure
146 variation (Wang et al. 2013). For both the reference accessions Col-0 and Bd21 we
147 estimated the average level of expression for genes having a TE (regardless of the
148 type of TE) in their sequence or within windows ranging from near vicinity (0-200
149 bp) to up to 2 kb (Figure S3). While we reproduced the pattern described above with
150 publically available data in *A. thaliana*, our analysis shows that TEs do not have such
151 an effect on nearby gene expression in *B. distachyon*. To rule out potential technical
152 artefacts, we extended the analysis to the ten additional *B. distachyon* accessions we
153 sampled but did not observe any overall effect of TEs on nearby gene expression
154 either (Figure S3). This is also surprising as the distance between a given gene and its
155 closest TE is significantly smaller (Wilcoxon-test, P-value <0.0001; Figure S3) in
156 Bd21 (1,074 bp) than in Col-0 (5,457bp).

157 The effect of TEs on methylation and gene expression is known to be family-
158 dependent (Eichten et al. 2012; Choi and Purugganan 2017), which might explain
159 why no effect is observed when all TEs are pooled. We therefore repeated this
160 analysis separately for the ten most active TE families (Stritt et al. 2018) in all eleven
161 *B. distachyon* accessions to explore potential genetic background effects. We also
162 produced heatmaps as described above to assess a potential link between methylation
163 spreading around these ten TE families and nearby gene expression. Methylation
164 spreading was in a second step used to statistically cluster the ten families.
165 Eventually, as young TEs might be more efficiently targeted by silencing
166 mechanisms, we extracted the age of each copy from Stritt et al. (2019). For four of
167 the ten families, this analysis allowed to uncover clear effects on gene expression
168 following two distinct patterns (Figure 3A): a short-distance effect (< 400 bp; all P-
169 values <0.001) conveyed by two *Gypsy* families (RLG_Bdis0039 and RLG_Bdis004)
170 and one *Copia* (RLC_Bdis022) family; and a long-distance effect (1 kb; P-values
171 <0.001) for the *Gypsy* element RLG_Bdis180. However, methylation spreading is

172 unlikely to be involved as our hierarchical clustering shows that spreading varies from
173 long (RLG_Bdis004, RLG_Bdis180) to intermediate (RLG_Bdis0039) and short
174 (RLG_Bdis022) around these four families (Figure 3B). In the same manner, there is
175 no clear association neither with the age of the family (Figure 3B) nor with the
176 genetic background as all accessions show comparable gene expression patterns
177 (Figure 3A). It is nonetheless worth noting that RLG_Bdis004 and RLC_Bdis022 are
178 two of the most recently active families in the genome of *B. distachyon* (Stritt et al.
179 2019), suggesting a possible scenario where *de novo* silencing of these two elements
180 may influence nearby-gene regulation. Altogether, not more than 131 genes are
181 influenced by one of these four families, confirming that the impact of TEs on gene
182 expression is limited in *B. distachyon.*

183 We eventually investigated the impact of TE insertion polymorphisms (TIPs)
184 on gene expression. Overall, we identified 1,833 high-confidence TIPs, i.e. insertions
185 present in at least one of the ten natural accessions but absent from the reference
186 genome Bd21 (Table S6). Due to the relative small number of TIPs located nearby
187 genes, we lacked power to investigate their impact on gene expression at the family
188 level. To circumvent this difficulty, we measured the log-fold change in gene
189 expression for loci harbouring a TIP in their vicinity (<1 kb) compared to the same
190 loci without a TIP in a different accession. We found that the majority of the TIPs
191 have no effect on adjacent gene expression (Figure S4), consistent with previous
192 findings that TE polymorphisms have little impact on methylation spreading in
193 *B. distachyon* (Eichten et al. 2016). These results contrast again with what is found in
194 *A. thaliana,* where polymorphic TEs are associated with differential expression of
195 nearby-genes among natural accessions (Stuart et al. 2016). Despite the lack of global
196 effect, a subset of 32 TIPs show a significant effect on nearby gene expression at high
197 confidence (Table S7), providing candidates for further functional validation.

198

199 *Genome organization and functional impact of TEs*

200 The differences observed between the two species may result from the nature of their
201 TEs. Notably, the genome of *B. distachyon* display a much higher ratio of solo-
202 LTRs:intact copies for both conserved elements and TIPs (up to 14:1 for some
203 families; Stritt et al. 2019) compared to *A. thaliana*, where the number of intact copies

204 approximate the number of solo-LTRs (Devos et al. 2002). This may reflect a much
205 broader aspect of the TE dynamics in the two species and one could assume that
206 *B. distachyon* harbours overall a much larger proportion of non-functional TEs than
207 *A. thaliana*. As non-functional elements might be less efficiently targeted by
208 silencing, they might have in return a limited influence on their neighbouring genes.
209 This hypothesis, however, does not hold for rice. There, the ratio solo-LTR:intact
210 elements is 1.6:1 (Tian et al. 2009) and therefore does not differ much from the one
211 found in *A. thaliana.* Yet, TEs also have little impact on gene expression in rice (Choi
212 and Purugganan 2017).

213 As an alternative hypothesis, methylation spreading has been suggested to be
214 under purifying selection (Choi and Purugganan 2017) and such a mechanism may
215 largely constrain the impact of TEs on adjacent genes. Compared to *A. thaliana*, TEs
216 and genes are largely interspersed in both *B. distachyon* and rice (Garg et al. 2015)
217 and selection against the deleterious effect of TEs might be stronger in such systems.
218 We therefore suggest that the differences we observed between *A. thaliana* and
219 *B. distachyon* with regard to the functional impact of TEs are largely inherent to the
220 contrasting architecture of their genomes. In the light of the current results and recent
221 studies (Seymour et al. 2014, Choi and Purugganan 2017, Quesneville 2020), it
222 appears that genome organization varies greatly between *Brassicacea* and grasses.
223 With more than 12,000 species (Christenhusz and Byng 2016), grasses cover about
224 40% of the planet (Gibson 2009). Generalizing the patterns reported in *A. thaliana* to
225 all plants, as commonly seen in the literature, is therefore arguable. Our study is
226 calling for the analysis of a larger set of plant genomes to reflect the tremendous
227 diversity present in this kingdom.

228

229

Materials and Methods

Plant material

We used eleven accession of *B. distachyon* originating from Iraq (Bd21, Bd21-3, Bd3-1), Turkey (BdTR12c, BdTR7a, Koz3, Bd1-1) and Spain (ABR2, Sig2, Luc1, Bd30-1). Seeds were sown on soil-sand mixture in 50 ml pots. For bisulfite and for RNA sequencing, three biological replicates per accession were grown. The second and third leaves of each replicate were sampled after 24 days of 16/8 hours of day/night conditions.

Whole genome bisulfite sequencing (WGBS) and RNA-sequencing analysis

For *B. distachyon*, WGBS libraries were prepared as described in the supplementary material Method S1. For *A. thaliana,* we used publicly available data (Table S1 for references). Raw data were processed as described in Method S1 and mapped onto the reference genomes Bd21 for *B. distachyon* (Phytozome 12, https://phytozome.jgi.doe.gov) and Col-0 for *A. thaliana* (TAIR10, www.arabidopsis.org). Methylation levels were estimated following the method described in (Schultz et al. 2012; Method S1) and by averaging the percentage of methylated cytosines over the three replicates per genotype within each focal unit (e.g. 100 kb/100 bp genomic windows, TE and gene body). We drew Circos plots (Connors et al. 2009) to visualize the architecture of the two reference genomes. The PCA based on methylation data was performed using ggplot2 in R (R development core team 2013). We eventually used GLM models with logit transformation to characterize the relationship between gene or TE density and methylation levels. P-values were obtained with the R package multcomp (Hothorn et al. 2008).

For each of the eleven *B. distachyon* accessions, mRNA was extracted using the RNAeasy Plant Mini Kit from Qiagen. Libraries were prepared by the D-BSSE Basel and paired-end sequencing was performed on a Illumina HiSeq 2500. For *A. thaliana,* we again used publicly available data (Table S1). Reads were mapped to the reference genomes with Salmon quant v.0.12.0 (Patro et al. 2017). The abundance of gene expression was quantified averaging transcripts per million (TPM) per accession for each single genotype over the three replicates.

262 *Methylation spreading*

263 We estimated methylation spreading by calculating methylation levels in 100 bp
264 windows spanning 2.5 kb upstream and downstream of each TE insertion. TE
265 insertions were categorized with regard to their position in
266 centromeric/pericentromeric regions or on chromosome arms and aggregated at the
267 family- (for the ten most active TE families) or superfamily levels. Heatmaps
268 depicting methylation levels around TEs were drawn in R using ggplot2 (Wickham
269 2009). We also used the levels of methylation spreading around TEs to calculate
270 Euclidean distances between the ten most active TE families and to perform a
271 hierarchical clustering using hclust in R (R development core team 2013).
272

273 *Impact of TEs on gene expression*

274 We assessed the potential link between the presence of a TE in the vicinity of a gene
275 and expression by assigning genes to different categories depending on whether they
276 harbor a TE in their sequence or within windows ranging from near vicinity (0-200
277 bp) to up to 2 kb. Average gene TPM and standard errors were calculated per distance
278 class with all TEs pooled together as well as for the ten most active families
279 separately and plotted using the R package plotrix (Lemon 2006). For a subset of
280 candidate families, we also compared the expression level (log1p transformed) of
281 genes displaying a TE in their vicinity (0-400 bp for RLG_Bdis0039, RLG_Bdis004
282 and RLC_Bdis022; 0-1 kb for RLG_Bdis180) to the expression level of genes
283 displaying a TE further away (between 400 bp-2 kb and 1 kb-2 kb respectively) with
284 a Wilcoxon test.
285

286 *Transposable element insertion polymorphism analysis*

287 Detection of transposable element insertion polymorphisms (TIP) was performed with
288 detettore (github.com/cstritt/detettore) using sequencing data from (Gordon et al.
289 2017). Detettore was specifically developed on *B. distachyon*. It allows to output
290 different features, including the presence of target site duplication (TSD), which can
291 be used to filter out false positive. Filtering criteria were adjusted so that no TIP is
292 found when the reads of Bd21 are mapped against the reference genome. The
293 expression of genes harboring a TIP in their vicinity (<1 kb) in at least one of the

accessions was compared to the corresponding genes in the other accessions. For each single gene, difference in expression between groups was tested using a t-test with a confidence interval of 0.95. Change in gene expression was considered significant when the P-value was <0.05 and the log2 change >1.

Acknowledgment

This work was supported by the PSC-syngenta fellowship, the University Research Priority Program Evolution in Action of the University of Zurich and the Swiss National Science Foundation (PZ00P3_154724). The authors would like to thank the D-BSSE for the sequencing effort, the Genetic Diversity Center at ETH Zurich for providing access to their computing resources and Michael Thieme for his comments on the manuscript.

Data availability

All the data produced for the current study are available under the ENA project number PRJEB39651. For *A. thaliana,* references are provided in the supplementary materials. Scripts are available at https://github.com/mwylerCH.

312 **Literature**

313 Bonchev G, Parisod C. 2013. Transposable elements and microevolutionary changes
314 in natural populations. Mol. Ecol. Resour. 13:765–775.

315 Choi JY, Purugganan MD. 2017. Evolutionary Epigenomics of Retrotransposon-
316 Mediated Methylation Spreading in Rice. Mol. Evol. Biol. 35:365–382.

317 Christenhusz MJM, Bying JW. 2016. The number of known plants species in the
318 world and its annual increase. Phytotaxa 261:201-217.

319 Connors J, Krzywinski M, Schein J, Gascoyne R, Horsman D, Jones SJ, Marra MA.
320 2009. Circos : An information aesthetic for comparative genomics. Genome Res.
321 19:1639–1645.

322 Devos KM, Brown, JKM, Bennetzen JL. 2002. Genome size reduction though
323 illegitimate recombination counteracts expansion in *Arabidopsis*. Genome Res.
324 12:1075-1079.

325 Eichten S, Stuart T, Srivastava A, Borevitz J. 2016. DNA Methylation profiles of
326 diverse Brachypodium distachyon aligns with underlying genetic diversity.
327 Genome Res. 26:1520–1531.

328 Eichten SR, Ellis NA, Makarevitch I, Yeh C, Gent JI, Guo L, Mcginnis KM, Zhang
329 X, Schnable PS, Vaughn MW, et al. 2012. Spreading of Heterochromatin Is
330 Limited to Specific Families of Maize Retrotransposons. PLoS Genet 8(12):
331 e1003127.

332 Garg R, Chevala N, Shankar R, Jain M. 2015. Divergent DNA methylation patterns
333 associated with gene expression in rice cultivars with contrasting drought and
334 salinity stress response. Scientific Rep. 5:14922.

335 Gibson, D.J. (2009) Grasses and Grassland Ecology. Oxford and New York: Oxford
336 University Press.

337 Gordon SP, Contreras-moreira B, Woods DP, Marais DL Des, Burgess D, Shu S,
338 Stritt C, Roulin AC, Schackwitz W, Tyler L, et al. 2017. Extensive gene content
339 variation in the Brachypodium distachyon pan-genome correlates with
340 population structure. Nat. Commun. 8:2184.

341 Grandbastien M-A. 1998. Activation of plant retrotransposons under stress
342 conditions. Trends Plant Sci. 3:181–187.

343 Grandbastien M. 2005. Stress activation and genomic impact of Tnt1 retrotransposons
344 in Solanaceae. Cytogenet. Genome Res. 110:229–241.

345 Guio L, Barron MG, Gonzalez J. 2014. The transposable element Bari-Jheh mediates
346 oxidative stress response in Drosophila. Mol Ecol 23:2020–2030.

347 Hasterok R, Wang K, Jenkins G. 2019. Progressive refinement of the karyotyping of
348 Brachypodium genomes. New Phytol. epub ahead.

349 Hollister JD, Smith LM, Guo Y, Ott F, Weigel D, Gaut BS. 2010. Transposable
350 elements and small RNAs contribute to gene expression divergence between
351 Arabidopsis thaliana and Arabidopsis lyrata. Proc Natl Acad Sci U S A 108:
352 2322–2327.

353 Hothorn T, Bretz F, Westfall P. 2008. Simultaneous Inference in General Parametric
354 Models. Biometrical Journal, 50(3), 346–363.

Horváth V, Merenciano M, González J. 2017. Revisiting the Relationship between Transposable Elements and the Eukaryotic Stress Response. Trends in Genetics 33:832–841.

Liaw and M. Wiener. 2002. Classification and Regression by randomForest. R News 2:18-22.

Lemon J. 2006. Plotrix: a package in the red light district of R. R-News 6:8-12.

Makarevitch I, Waters AJ, West PT, Stitzer M, Hirsch CN, Ross-Ibarra J, Springer NM. 2015. Transposable elements contribute to activation of maize genes in response to abiotic stress. PLoS Genet 11:e1004915.

Mcclintock B. 1947. Cytogenetic Studies of Maize and Neurospora. Carnegie Inst Wash Yrbk 46:146–152.

McClintock B. 1950. The origin and behavior of mutable loci in maize. Proc Natl Acad Sci U S A. 36:344–355.

Negi P, Rai AN, Suprasanna P. 2016. Moving through the Stressed Genome: Emerging Regulatory Roles for Transposons in Plant Stress Response. Front Plant Sci 7:1448.

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. Nature methods 14(4)_417-418.

Quesneville H. 2020. Twenty years of transposable element analysis in the *Arabidopsis thaliana* genome. Mobile DNA. 11:28.

R development core team. 2013. R: A language and environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available from: http://www.R-project.org.

Regulski M, Lu Z, Kendall J, Donoghue MTA, Reinders J, Llaca V, Deschamps S, Smith A, Levy D, McCombie WR, et al. 2013. The maize methylome influences mRNA splice sites and reveals widespread paramutation-like switches guided by small RNA. Genome Res. 23:1651–1662.

Rey O, Danchin E, Mirouze M, Loot C, Blanchet S. 2016. Adaptation to Global Change : A Transposable Adaptation to Global Change : A Transposable Element – Epigenetics Perspective. Trends Ecol. Evol. 31(7):514-526

Schultz MD, Schmitz RJ, Ecker JR. ‘ Leveling ’ the playing field for analyses of single-base resolution DNA methylomes. Trends Genet. 28:583–585.

Seymour DK, Koenig D, Becker C, Weigel D. 2014. Evolution of DNA Methylation Patterns in the Brassicaceae is Driven by Differences in Genome Organization. PLoS Genet. 10: e1004785.

Shang Y, Yang F, Schulman AH, Zhu J, Jia Y, Wang J. 2017. Gene Deletion in Barley Mediated by LTR-retrotransposon BARE. Sci. Rep. 7: 43766.

Sigman MJ, Slotkin RK. 2016. The First Rule of Plant Transposable Element Silencing: Location, Location, Location. Plant Cell 28:304–313.

Stritt C, Gordon SP, Wicker T, Vogel JP, Roulin AC. 2018. Recent Activity in Expanding Populations and Purifying Selection Have Shaped Transposable Element Landscapes. Genome Biol. Evol.10:304–318.

398 Stritt C, Wyler M, Gimmi EL, Pippel M, Roulin AC. 2019. Diversity , dynamics and
399 effects of long terminal repeat retrotransposons in the model grass
400 Brachypodium distachyon. New Phytol. Epub ahead of print doi:
401 10.1111/nph.16308

402 Stuart T, Eichten S, Cahn J, Borevitz JO, Lister R. 2016. Population scale mapping of
403 transposable element diversity uncovers novel genetic diversity linked to gene
404 regulation and epigenomic variation. Elife 5:e20777.

405 Takuno S, Gaut BS. 2013. Gene body methylation is conserved between plant
406 orthologs and is of evolutionary consequence. Proc. Natl. Acad. Sci. U. S. A.
407 110:1797–1802.

408 The international brachypodium Consortium. 2010. Genome sequencing and analysis
409 of the model grass Brachypodium distachyon. Nature 463:763–768.

410 Thieme M, Lanciano S, Balzergue S, Daccord N, Mirouze M, Bucher E. 2017.
411 Inhibition of RNA polymerase II allows controlled mobilisation of
412 retrotransposons for plant breeding. Genome Biol. 18:134:1–10.

413 Thind AK, Wicker T, Müller T, Ackermann PM, Steuernagel B, Wulff BBH,
414 Spannagl M, Twardziok SO, Felder M, Lux T, et al. 2018. Chromosome-scale
415 comparative sequence analysis unravels molecular mechanisms of genome
416 dynamics between two wheat cultivars. Genome Biol. 19:104.

417 Tian Z, Rizzon C, Du J, Zhu L, Bennetzen JL, Jackson SA, Gaut BS, Ma J. 2009. Do
418 genetic recombination and gene density shape the pattern of DNA elimination in
419 rice long terminal repeat retrotransposons? Genome Res. 19:2221-2230.

420 Van't Hof AE, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, Hall N,
421 Darby AC, Saccheri IJ. 2016. The industrial melanism mutation in British
422 peppered moths is a transposable element. Nature 534:102–105.

423 Wang X, Weigel D, Smith LM. 2013. Transposon variants and their effects on gene
424 expression in Arabidopsis. PLoS Genet. 9:e1003255.

425 Wang Y, Liang W, Tang T. 2018. Constant conflict between Gypsy LTR
426 retrotransposons and CHH methylation within a stress-adapted mangrove
427 genome. New Phytol. 220: 922-935.

428 Wickham. H. 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag
429 New York. ISBN 978-3-319-24277-4.

430 Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, Henderson IR, Shinn
431 P, Pellegrini M, Jacobsen SE, et al. 2006. Genome-wide high-resolution
432 mapping and functional analysis of DNA methylation in arabidopsis. Cell
433 126:1189–1201.

434 Zhang H, Lang Z, Zhu J. 2018. Dynamics and function of DNA methylation in plants.
435 Nat. Rev. Mol. Cell Biol. 19:489–506.

436

437 **Figures and Tables**

438 **Figure 1: Genome organization and methylation distribution of Col-0 and Bd21**
439 Circos plot displaying methylation levels as well as gene and TE density in 100 kb
440 windows in A) *A. thaliana* and B) *B. distachyon.* Centromeric and pericentromeric
441 regions are highlighted in grey. Panel C and D display CpG methylation levels
442 (expressed as % of methylated cytosines) for the most common TE superfamilies and
443 for genes in *A. thaliana* and *B. distachyon* respectively.
444

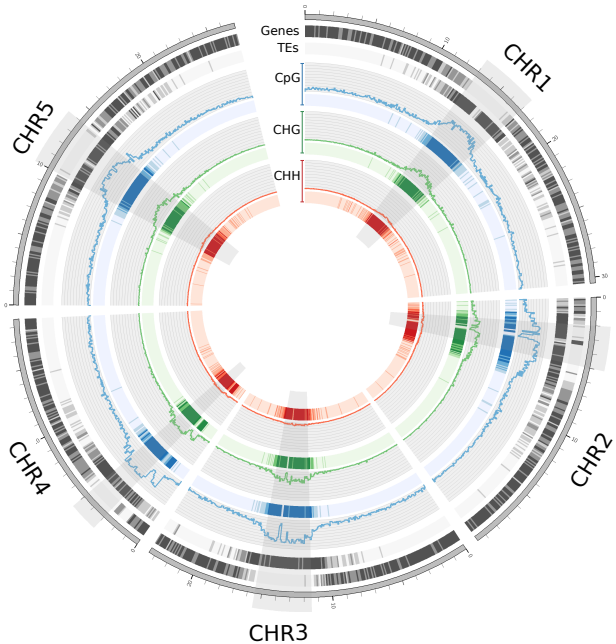
445 **Figure 2: Methylation spreading around transposable elements**
446 Heatmaps representing methylation (CpG) spreading around TEs located in the
447 pericentromere/centromere regions or on the arm of the chromosomes for *A. thaliana*
448 (N=18,521 classified TEs) and *B. distachyon* (N=62,708 classified TEs). Scales were
449 adapted for the six panels to account for the lower genome-wide methylation level in
450 *A. thaliana* and the different methylation levels for the three contexts.
451

452 **Figure 3: Impact of transposable elements on gene expression**
453 A) Average gene expression as a function of the distance to the nearest TE belonging
454 to the four LTR-retrotransposons RLG_BdisC039, RLG_BdisC004, RLG_BdisC180
455 and RLC_BdisC022 in the natural accessions of *B. distachyon*. Each coloured line
456 represents a different accession. The dashed line represents the genome-wide level of
457 expression. Boxplots display the expression of genes harbouring a TE in their vicinity
458 for two distance classes only, all accessions pooled together. Stars indicate Wilcoxon
459 test levels of significance: ***<0.001 B) Heatmap representing methylation (CpG)
460 spreading around the 10 most active LTR retrotransposon families in Bd21. Families
461 were clustered based on the level of methylation spreading (long, intermediate and
462 short) around each insertion. Names on the right of the heatmap describe the TE
463 lineage to which each family belongs. Boxplots display the age of each families
464 estimated by (Stritt et al. 2019).

Figure 1

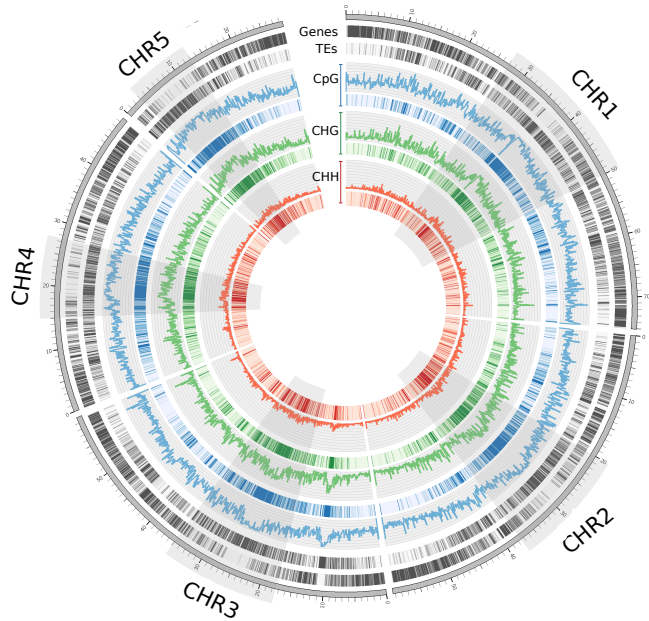
A)

A. thaliana
Col-0

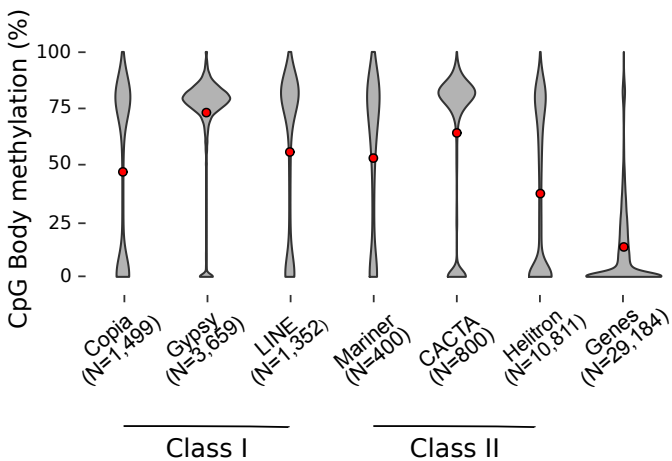


B)

B. distachyon
Bd21



C)



D)

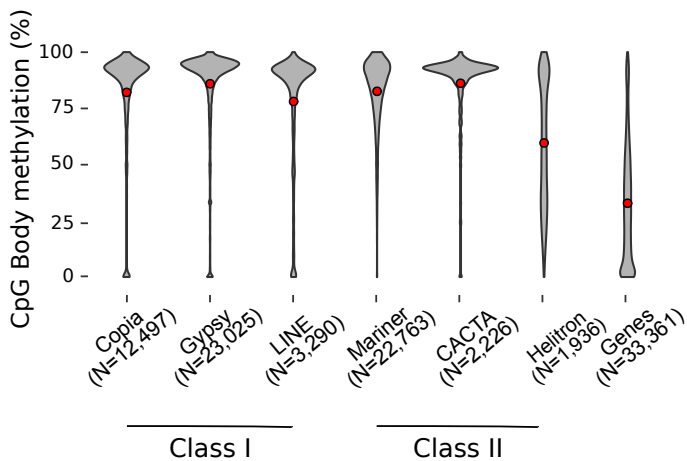


Figure 2

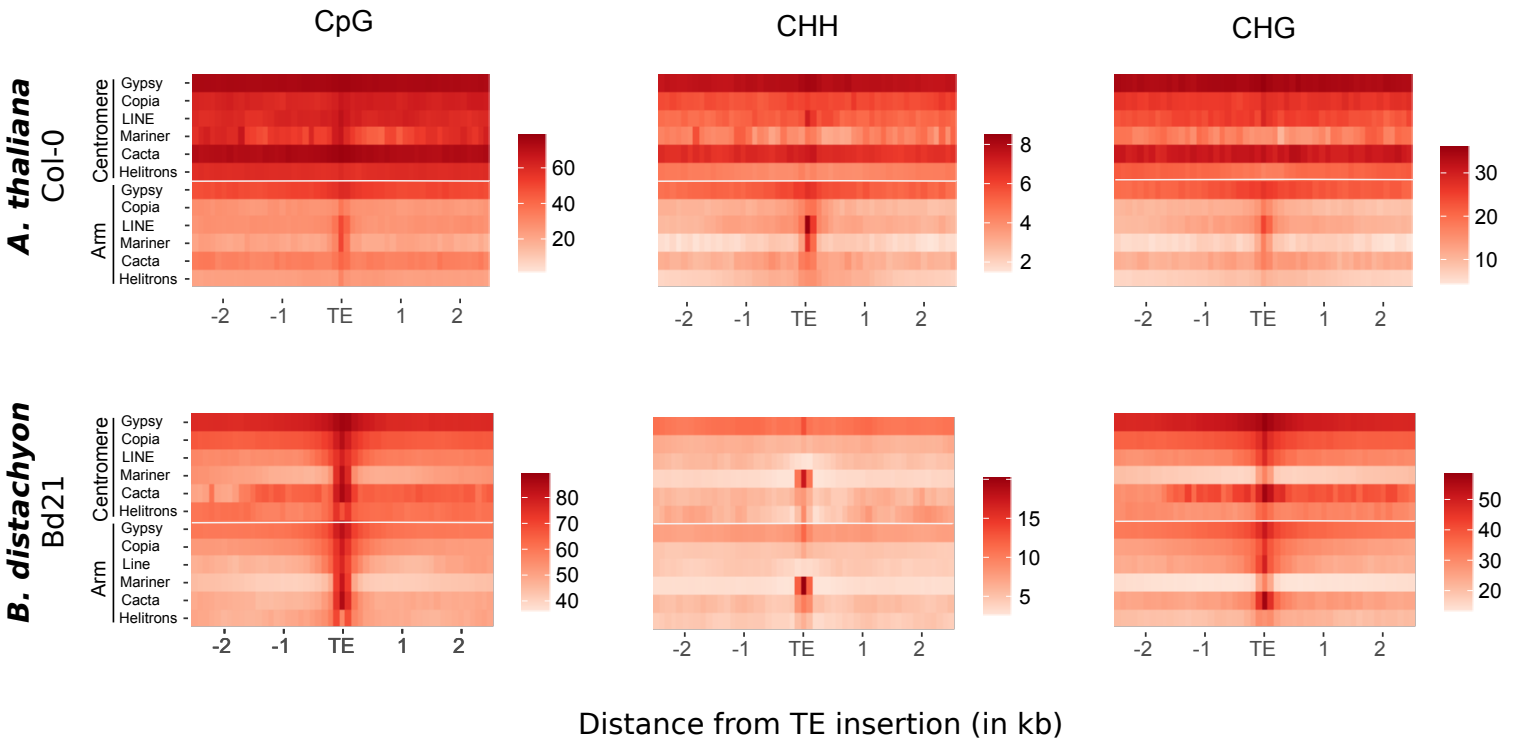
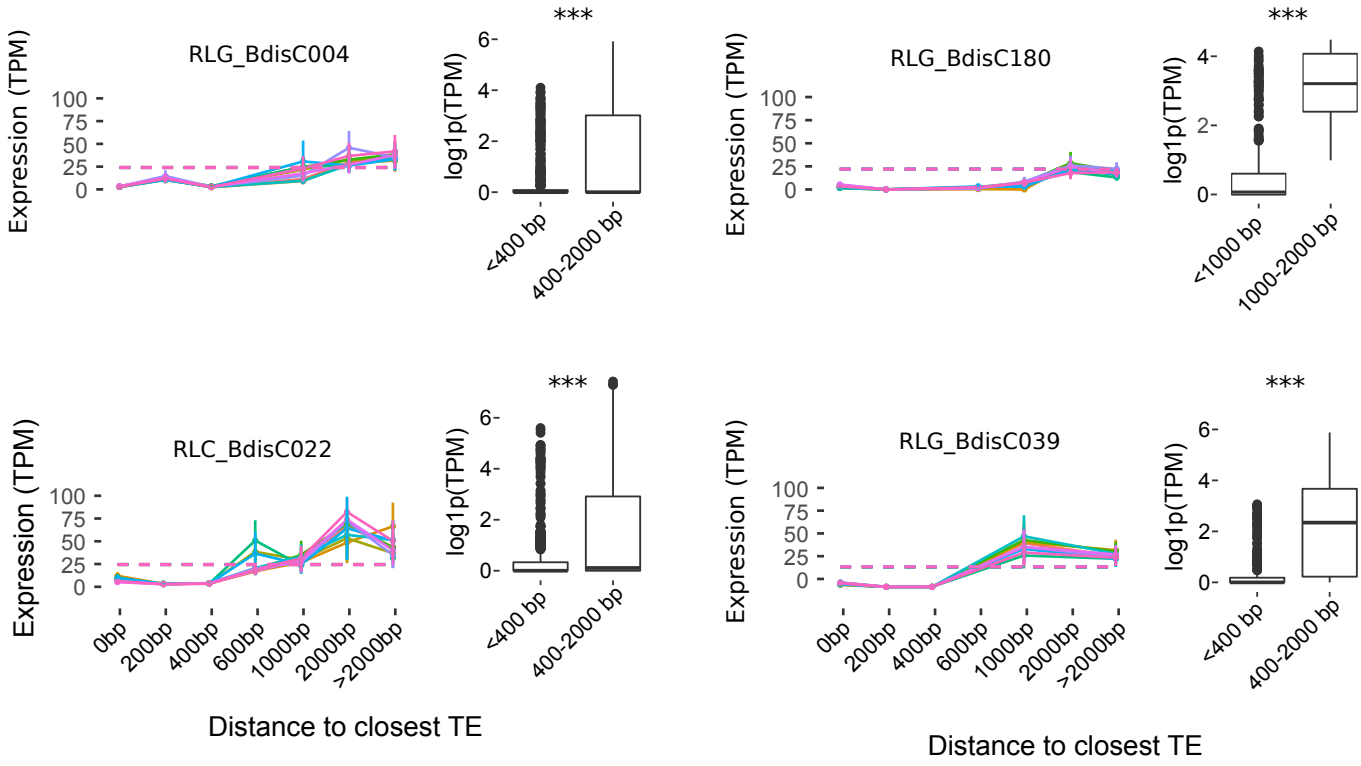
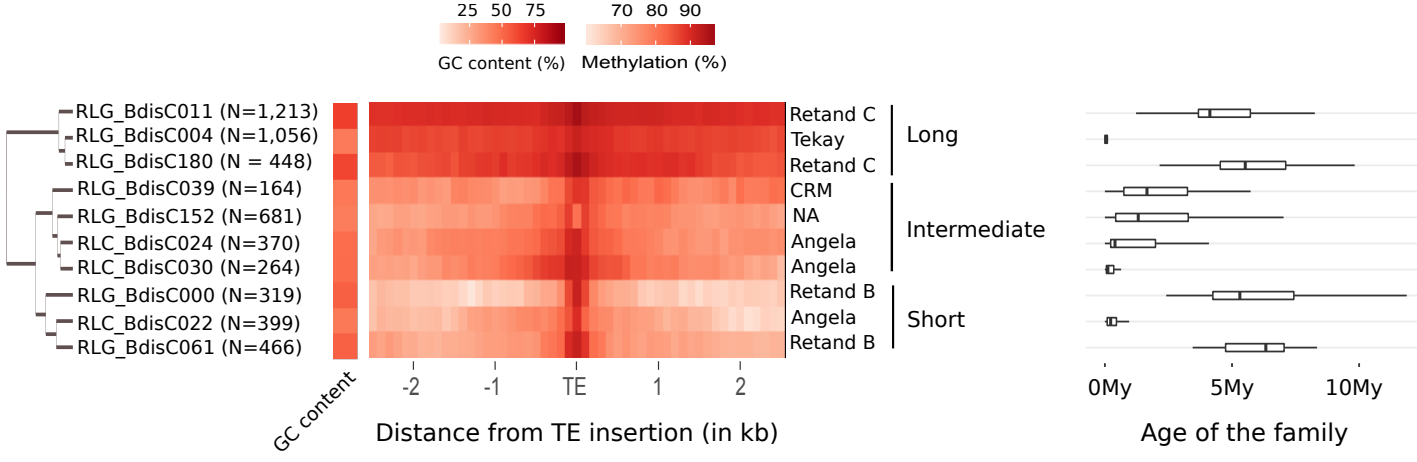

CpG
CHH
CHG
A. thaliana
Col-0
B. distachyon
Bd21
Centromere
Arm
Gypsy
Copia
LINE
Line
Mariner
Cacta
Helitrons
-2
-1
TE
1
2
Distance from TE insertion (in kb)

Figure 3

A)


RLG_BdisC004
RLG_BdisC180
RLC_BdisC022
RLG_BdisC039
Expression (TPM)
log1p(TPM)
***
<400 bp
400-2000 bp
<1000 bp
1000-2000 bp
0bp
200bp
400bp
600bp
1000bp
2000bp
>2000bp
Distance to closest TE

B)


GC content (%)
Methylation (%)
RLG_BdisC011 (N=1,213)
RLG_BdisC004 (N=1,056)
RLG_BdisC180 (N = 448)
RLG_BdisC039 (N=164)
RLG_BdisC152 (N=681)
RLC_BdisC024 (N=370)
RLC_BdisC030 (N=264)
RLG_BdisC000 (N=319)
RLC_BdisC022 (N=399)
RLG_BdisC061 (N=466)
Retand C
Tekay
Retand C
CRM
NA
Angela
Angela
Retand B
Angela
Retand B
Long
Intermediate
Short
GC content
Distance from TE insertion (in kb)
Age of the family
0My
5My
10My