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.*

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 that belong to three genetic clusters: Western (ABR2, Koz3, Bd30-1 and Sig2; Spain), 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). We also used publically available bisulfite- and RNA-sequencing data for the reference accession Col-0 of *A. thaliana* (Table S1).

We first estimated methylation levels (expressed as the percentage of methylated cytosines), gene and TE density in 100 kb windows for the reference genomes of *A. thaliana* (Col-0) and *B. distachyon* (Bd21). The circular visualisation of these three genomic features reveals two distinct architectures: genes and TEs are clearly compartmentalised in *A. thaliana* (Figure 1A) while largely interspersed in *B. distachyon* (Figure 1B). Yet we found that DNA methylation is positively correlated with TE density and negatively correlated with gene density in both species (Figure S1). We then estimated the methylation level of genes and TEs, aggregating the latter at the superfamily level. TE body methylation is significantly higher than gene body methylation in the two species, although variability is observed among TE superfamilies (Figure 1C and D, Table S2-S4 for P-values). Genome-wide DNA methylation patterns are thus largely shaped by TEs in *B. distachyon,* like in other

107 flowering plants species (Zhang et al. 2006; Regulski et al. 2013; Takuno and Gaut
108 2013; Seymour et al. 2014). We also found that most TE superfamilies and genes
109 show a higher level of body methylation in *B. distachyon* than in *A. thaliana* (Figure
110 1C and D, Table S5 for P-values), confirming a low genome-wide level of body
111 methylation in *A. thaliana* compared to other species (Seymour et al. 2014, Eichten et
112 al. 2016).

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

127

128 *Impact of TEs on nearby gene expression*

129 In *A. thaliana*, genes flanked by a TE within 2 kb show reduced expression compared
130 to genes devoid of TEs in their vicinity (Wang et al. 2013). It was proposed that this
131 may result from a combined effect of DNA methylation spreading and DNA structure
132 variation (Wang et al. 2013). For both the reference accessions Col-0 and Bd21 we
133 estimated the average level of expression for genes having a TE (regardless of the
134 type of TE) in their sequence or within windows ranging from near vicinity (0-200
135 bp) to up to 2 kb (Figure S2). While we reproduced the pattern described above with
136 publically available data in *A. thaliana*, our analysis shows that TEs do not have such
137 an effect on nearby gene expression in *B. distachyon*. To rule out potential technical
138 artefacts, we extended the analysis to the ten additional *B. distachyon* accessions we

139 sampled but did not observe any overall effect of TEs on nearby gene expression
140 either (Figure S2).

141 The effect of TEs on methylation and gene expression is known to be family-
142 dependent (Eichten et al. 2012; Choi and Purugganan 2017), which might explain
143 why no effect is observed when all TEs are pooled. We therefore repeated this
144 analysis separately for the ten most active TE families (Stritt et al. 2018) in all eleven
145 *B. distachyon* accessions to explore potential genetic background effects. We also
146 produced heatmaps as described above to assess a potential link between methylation
147 spreading around these ten TE families and nearby gene expression. Methylation
148 spreading was in a second step used to statistically cluster the ten families.
149 Eventually, as young TEs might be more efficiently targeted by silencing
150 mechanisms, we extracted the age of each copy from Stritt et al. (2019). For four of
151 the ten families, this analysis allowed to uncover clear effects on gene expression
152 following two distinct patterns (Figure 3A): a short-distance effect (< 400 bp; all P-
153 values <0.001) conveyed by two *Gypsy* families (RLG_Bdis0039 and RLG_Bdis004)
154 and one *Copia* (RLC_Bdis022) family; and a long-distance effect (1 kb; P-values
155 <0.001) for the *Gypsy* element RLG_Bdis180. However, methylation spreading is
156 unlikely to be involved as our hierarchical clustering shows that spreading varies from
157 long (RLG_Bdis004, RLG_Bdis180) to intermediate (RLG_Bdis0039) and short
158 (RLG_Bdis022) around these four families (Figure 3B). In the same manner, there is
159 no clear association neither with the age of the family (Figure 3B) nor with the
160 genetic background as all accessions show comparable gene expression patterns
161 (Figure 3A). It is nonetheless worth noting that RLG_Bdis004 and RLC_Bdis022 are
162 two of the most recently active families in the genome of *B. distachyon* (Stritt et al.
163 2019), suggesting a possible scenario where *de novo* silencing of these two elements
164 may influence nearby-gene regulation.

165 We eventually investigated the impact of TE insertion polymorphisms (TIPs)
166 on gene expression. Overall, we identified 1,833 high-confidence TIPs, i.e. insertions
167 present in at least one of the ten natural accessions but absent from the reference
168 genome Bd21 (Table S6). Due to the relative small number of TIPs located nearby
169 genes, we lacked power to investigate their impact on gene expression at the family
170 level. To circumvent this difficulty, we measured the log-fold change in gene
171 expression for loci harbouring a TIP in their vicinity (<1 kb) compared to the same

172 loci without a TIP in a different accession. We found that the majority of the TIPs
173 have no effect on adjacent gene expression (Figure S3), consistent with previous
174 findings that TE polymorphisms have little impact on methylation spreading in
175 *B. distachyon* (Eichten et al. 2016). These results contrast again with what is found in
176 *A. thaliana,* where polymorphic TEs are associated with differential expression of
177 nearby-genes among natural accessions (Stuart et al. 2016). A likely explanation for
178 this difference may be the different types of TIPs in the two species, notably the large
179 number of non-functional solo-LTR copies in *B. distachyon* (Stritt et al. 2019).
180 Despite the lack of global effect, a subset of 32 TIPs show a significant effect on
181 nearby gene expression at high confidence (Table S7), providing candidates for
182 further functional validation.

183

184 *Conclusions*

185 In contrast to the situation in *A. thaliana*, an eudicot of similar genome size and
186 complexity, TEs tend to have a limited impact on methylation spreading and only
187 punctual effects on gene expression in the model grass *B. distachyon*. Our findings
188 yet align to what is described in rice where TEs also have little impact on gene
189 expression. Methylation spreading has been suggested to be under purifying selection
190 and such a mechanism may largely constrain the impact of TEs on adjacent genes.
191 Compared to *A. thaliana*, TEs and genes are largely interspersed in *B. distachyon*. We
192 thus suggest that the differences we observed between the two species regarding the
193 impact of TEs are largely inherent to the contrasting structure of their genomes.

194

195

196 **Materials and Methods**

197 *Plant material*

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

204

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

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

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

226

227

228 *Methylation spreading*

229 We estimated methylation spreading by calculating methylation levels in 100 bp
230 windows spanning 2.5 kb upstream and downstream of each TE insertion. TE
231 insertions were categorized with regard to their position in
232 centromeric/pericentromeric regions or on chromosome arms and aggregated at the
233 family- (for the ten most active TE families) or superfamily levels. Heatmaps
234 depicting methylation levels around TEs were drawn in R using ggplot2 (Wickham
235 2009). We also used the levels of methylation spreading around TEs to calculate
236 Euclidean distances between the ten most active TE families and to perform a
237 hierarchical clustering using hclust in R (R development core team 2013).
238

239 *Impact of TEs on gene expression*

240 We assessed the potential link between the presence of a TE in the vicinity of a gene
241 and expression by assigning genes to different categories depending on whether they
242 harbor a TE in their sequence or within windows ranging from near vicinity (0-200
243 bp) to up to 2 kb. Average gene TPM and standard errors were calculated per distance
244 class with all TEs pooled together as well as for the ten most active families
245 separately and plotted using the R package plotrix (Lemon 2006). For a subset of
246 candidate families, we also compared the expression level (log1p transformed) of
247 genes displaying a TE in their vicinity (0-400 bp for RLG_Bdis0039, RLG_Bdis004
248 and RLC_Bdis022; 0-1 kb for RLG_Bdis180) to the expression level of genes
249 displaying a TE further away (between 400 bp-2 kb and 1 kb-2 kb respectively) with
250 a Wilcoxon test.
251

252 *Transposable element insertion polymorphism analysis*

253 Detection of transposable element insertion polymorphisms (TIP) was performed with
254 detettore (github.com/cstritt/detettore) using sequencing data from (Gordon et al.
255 2017). The expression of genes harboring a TIP in their vicinity (<1 kb) in at least one
256 of the accessions was compared to the corresponding genes in the other accessions.
257 For each single gene, difference in expression between groups was tested using a t-
258 test with a confidence interval of 0.95. Change in gene expression was considered
259 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 analysis will be made publically available upon acceptance of the manuscript. Scripts used in this study are available at https://github.com/mwylerCH.

273 **Literature**

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

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

278 Connors J, Krzywinski M, Schein J, Gascoyne R, Horsman D, Jones SJ, Marra MA.
279 2009. Circos : An information aesthetic for comparative genomics. Genome Res.
280 19:1639–1645.

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

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

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

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

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

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

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

300 Hollister JD, Smith LM, Guo Y, Ott F, Weigel D, Gaut BS. 2010. Transposable
301 elements and small RNAs contribute to gene expression divergence between
302 Arabidopsis thaliana and Arabidopsis lyrata. Proc Natl Acad Sci U S A 108:
303 2322–2327.

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

306 Horváth V, Merenciano M, González J. 2017. Revisiting the Relationship between
307 Transposable Elements and the Eukaryotic Stress Response. Trends in Genetics
308 33:832–841.

309 Liaw and M. Wiener. 2002. Classification and Regression by randomForest. R News
310 2:18-22.

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

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

315 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.

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.

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

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

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

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

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

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

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

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

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

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

376 Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, Henderson IR, Shinn
377 P, Pellegrini M, Jacobsen SE, et al. 2006. Genome-wide high-resolution
378 mapping and functional analysis of DNA methylation in arabidopsis. Cell
379 126:1189–1201.

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

382

383 **Figures and Tables**

384 **Figure 1: Genome organization and methylation distribution**

385 Circos plot displaying methylation levels as well as gene and TE density in 100 kb
386 windows in A) *A. thaliana* and B) *B. distachyon.* Centromeric and pericentromeric
387 regions are highlighted in grey. Panel C and D display CpG methylation levels
388 (expressed as % of methylated cytosines) for the most common TE superfamilies and
389 for genes in *A. thaliana* and *B. distachyon* respectively.

390

391 **Figure 2: Methylation spreading around transposable elements**

392 Heatmaps representing methylation (CpG) spreading around TEs located in the
393 pericentromere/centromere regions or on the arm of the chromosomes for *A. thaliana*
394 and *B. distachyon.*

395

396 **Figure 3: Impact of transposable elements on gene expression**

397 A) Average gene expression as a function of the distance to the nearest TE belonging
398 to the four LTR-retrotransposons RLG_BdisC039, RLG_BdisC004, RLG_BdisC180
399 and RLC_BdisC022 in the natural accessions of *B. distachyon*. Each coloured line
400 represents a different accession. The dashed line represents the genome-wide level of
401 expression. Boxplots display the expression of genes harbouring a TE in their vicinity
402 for two distance classes only, all accessions pooled together. Stars indicate Wilcoxon
403 test levels of significance: $^{***}<0.001$ B) Heatmap representing methylation (CpG)
404 spreading around the 10 most active LTR retrotransposon families. Families were
405 clustered based on the level of methylation spreading (long, intermediate and short)
406 around each insertion. Names on the right of the heatmap describe the TE lineage to
407 which each family belongs. Boxplots display the age of each families estimated by
408 (Stritt et al. 2019).

409

410

411 **Supplementary Materials**

412 **Figure S1: Relationship between gene or transposable element density and**
413 **methylation levels.**
414
415 **Figure S2: Impact of TEs on gene expression.** Average gene expression as a
416 function of the distance to the nearest TE in *A. thaliana* and in the natural accessions
417 of *B. distachyon,* all TEs pooled together.
418
419 **Figure S3: Transposable element insertion polymorphisms.** Volcano plot
420 displaying expression changes in genes harbouring a TIP in their vicinity. Significant
421 changes are displayed in red.
422
423 **Table S1: WGB- and RNA-sequencing output information**

424 **Table S2: Within-species glm outputs (CpG context)**

425 **Table S3: Within-species glm outputs (CHG context)**

426 **Table S4: Within-species glm outputs (CHH context)**

427 **Table S5: glm outputs - comparison *A. thaliana* vs. *B. distachyon***

428 **Table S6: Information about the 1,833 TIPs identified across the natural**
429 **accession of *B. distachyon.***

430 **Table S7: Information about the 32 high-confidence TIPs**

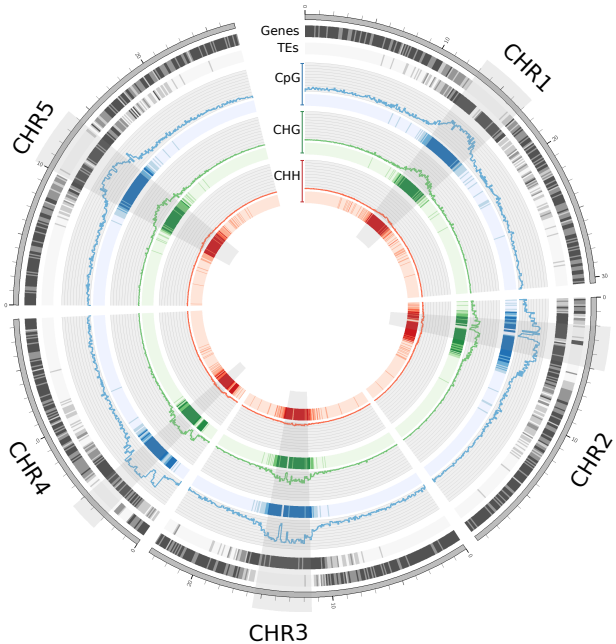
431 **Method S1: Extended Materials and Methods**

432

Figure 1

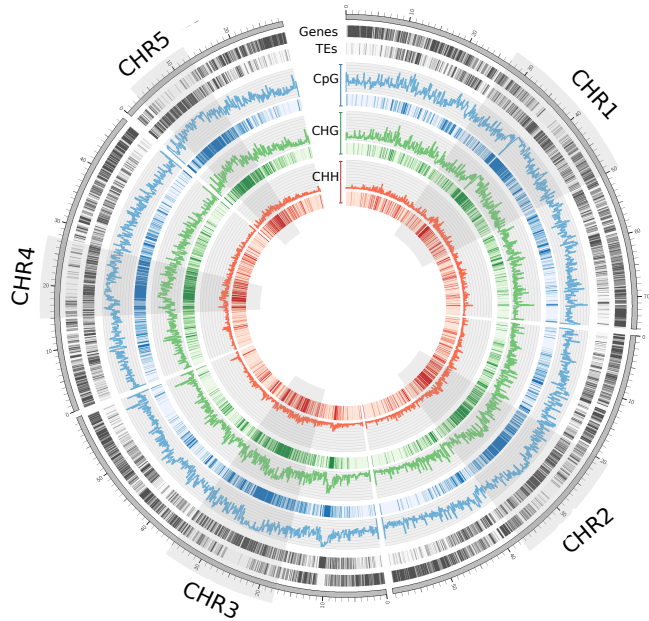
A)

A. thaliana

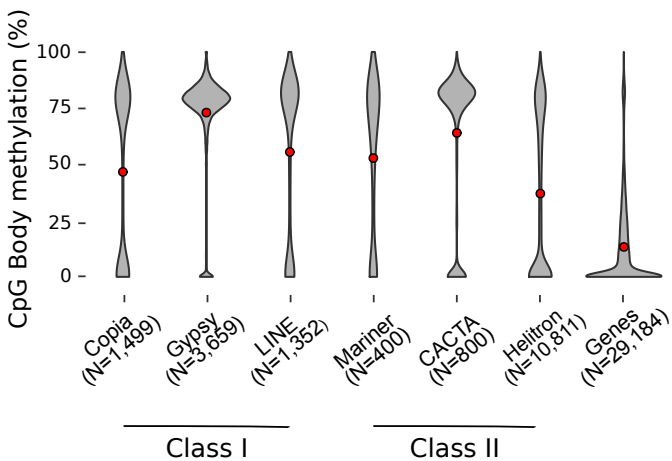


B)

B. distachyon



C)



D)

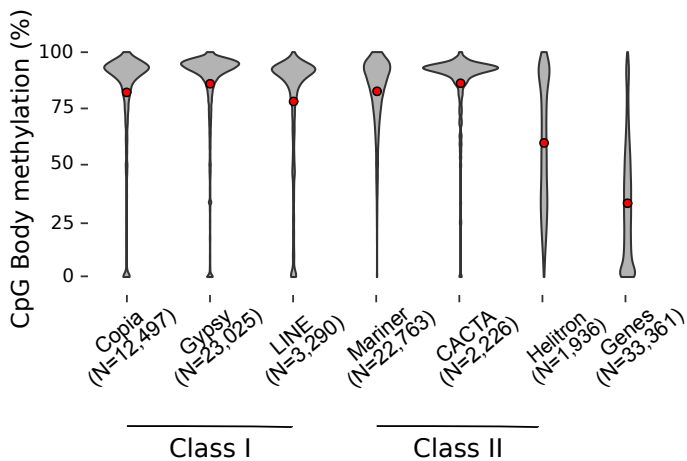


Figure 2

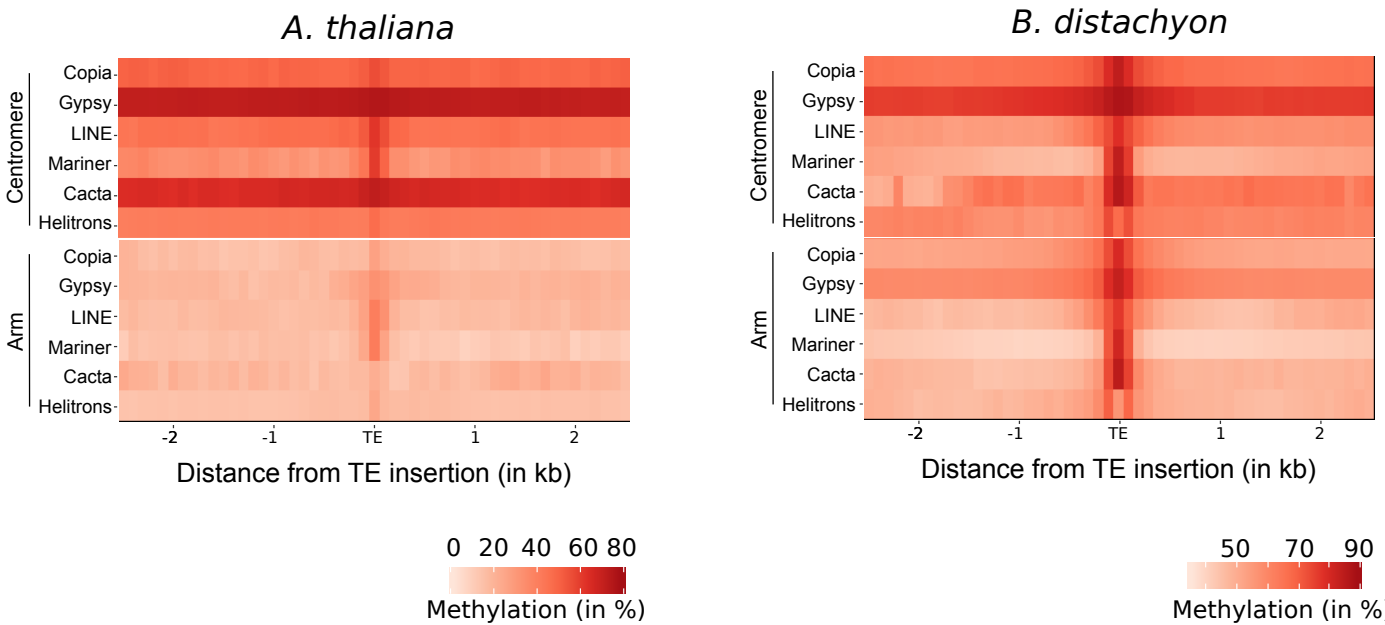
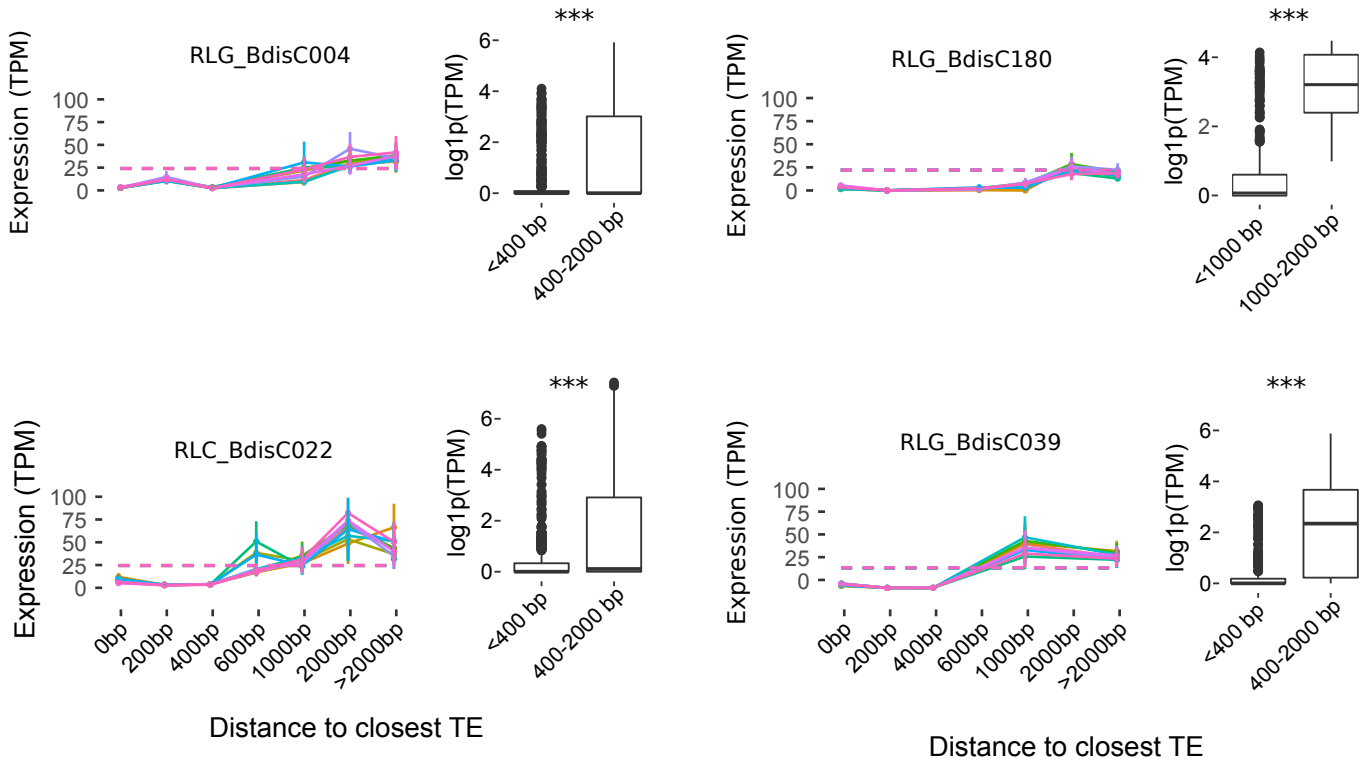

A. thaliana
B. distachyon
Centromere
Arm
Copia
Gypsy
LINE
Mariner
Cacta
Helitrons
-2
-1
TE
1
2
Distance from TE insertion (in kb)
0 20 40 60 80
50 70 90
Methylation (in %)

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)

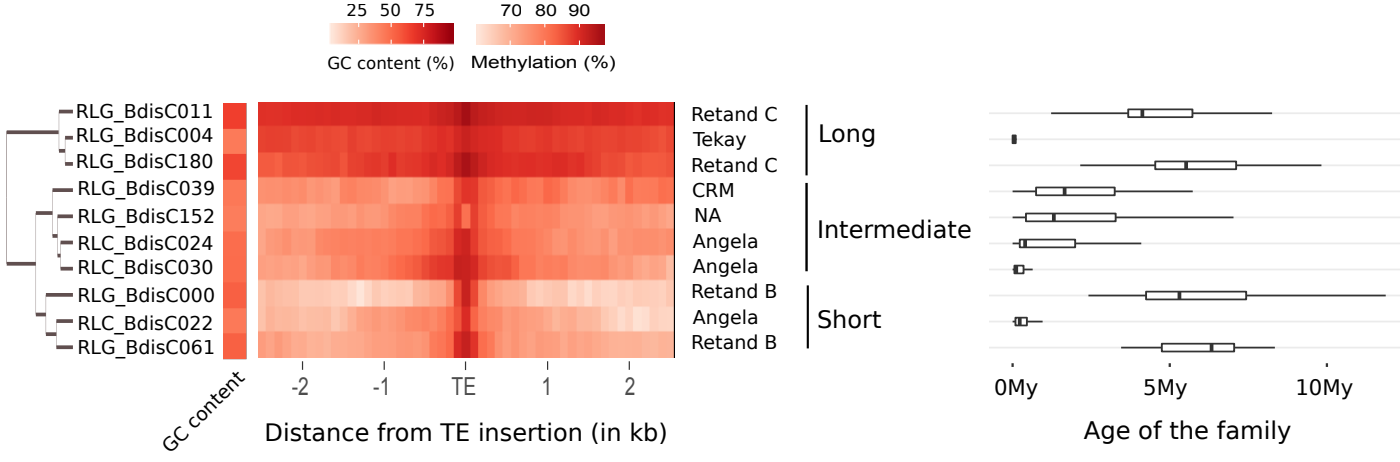

25 50 75
GC content (%)
70 80 90
Methylation (%)
RLG_BdisC011
RLG_BdisC004
RLG_BdisC180
RLG_BdisC039
RLG_BdisC152
RLC_BdisC024
RLC_BdisC030
RLG_BdisC000
RLC_BdisC022
RLG_BdisC061
Retand C
Tekay
Retand C
CRM
NA
Angela
Angela
Retand B
Angela
Retand B
Long
Intermediate
Short
GC content
-2
-1
TE
1
2
Distance from TE insertion (in kb)
0My
5My
10My
Age of the family

Figure S1

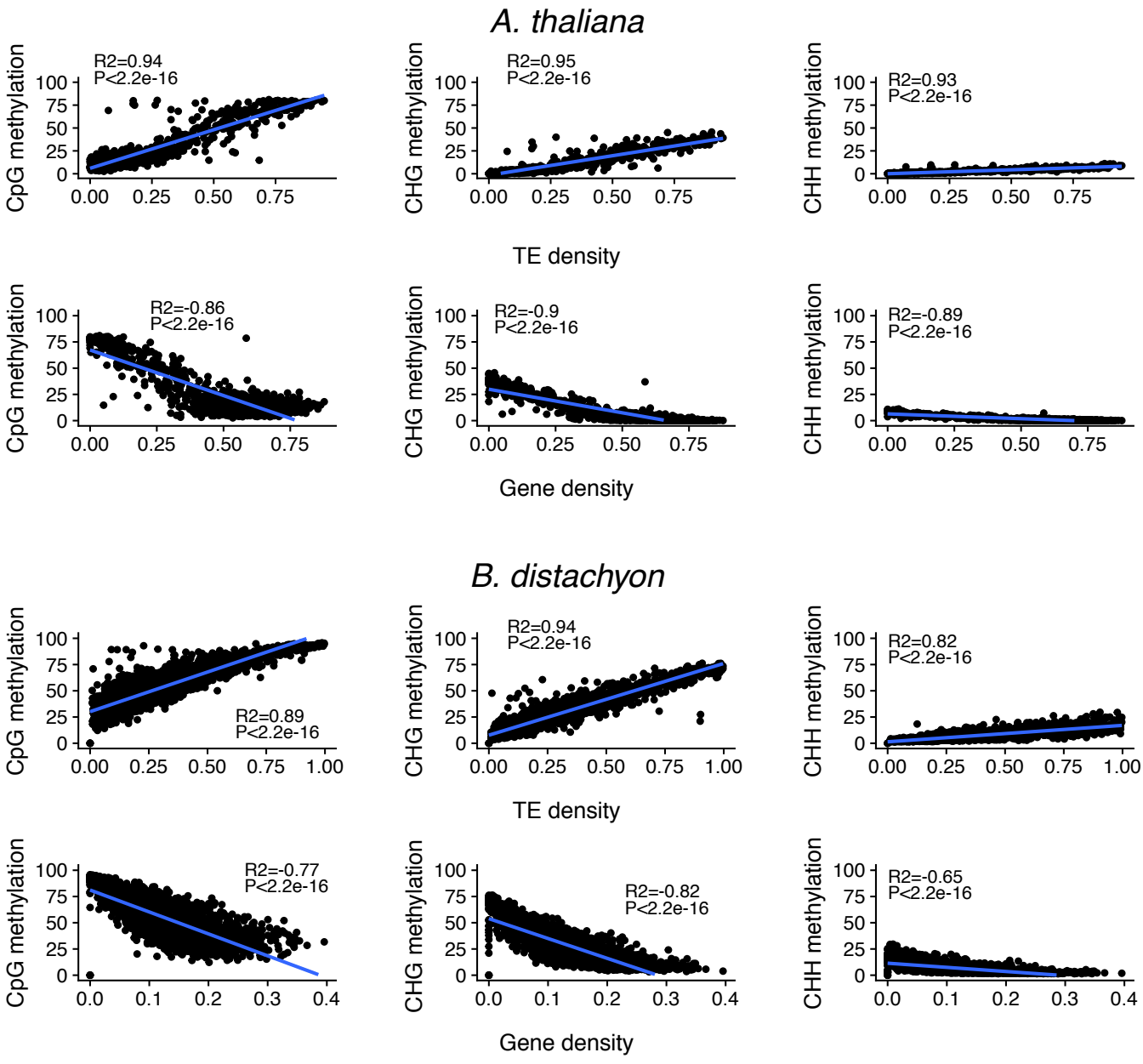

A. thaliana
CpG methylation
CHG methylation
CHH methylation
R2=0.94
P<2.2e-16
R2=0.95
P<2.2e-16
R2=0.93
P<2.2e-16
TE density
R2=-0.86
P<2.2e-16
R2=-0.9
P<2.2e-16
R2=-0.89
P<2.2e-16
Gene density
B. distachyon
R2=0.89
P<2.2e-16
R2=0.94
P<2.2e-16
R2=0.82
P<2.2e-16
TE density
R2=-0.77
P<2.2e-16
R2=-0.82
P<2.2e-16
R2=-0.65
P<2.2e-16
Gene density

Figure S2

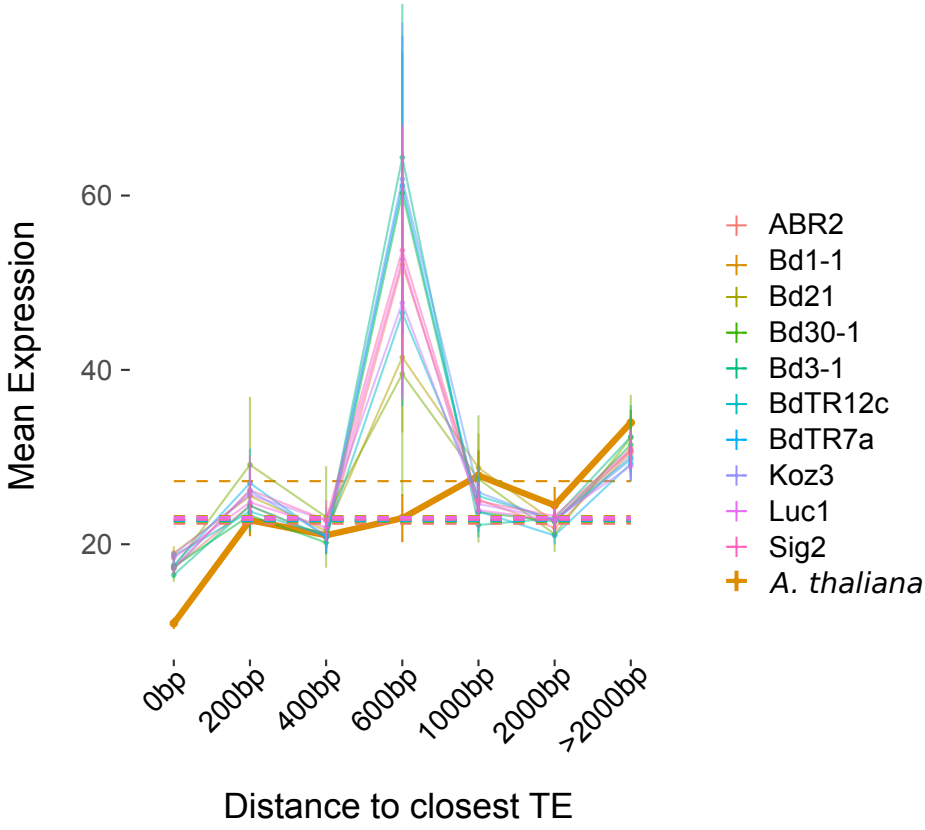
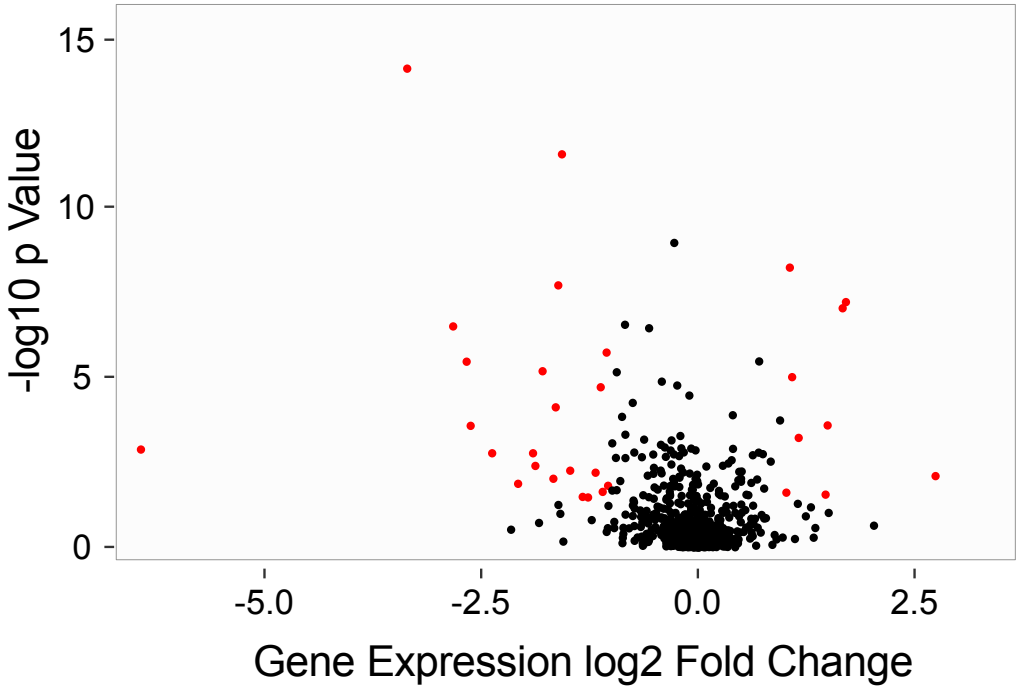

Mean Expression
60
40
20
0bp
200bp
400bp
600bp
1000bp
2000bp
>2000bp
Distance to closest TE
ABR2
Bd1-1
Bd21
Bd30-1
Bd3-1
BdTR12c
BdTR7a
Koz3
Luc1
Sig2
A. thaliana

Figure S3


15
10
5
0
-log10 p Value
-5.0
-2.5
0.0
2.5
Gene Expression log2 Fold Change