



HAL
open science

Transposition of HOPPLA in siRNA-deficient plants suggests a limited effect of the environment on retrotransposon mobility in *Brachypodium distachyon*

Michael Thieme, Nikolaos Minadakis, Christophe Himber, Bettina Keller, Wenbo Xu, Kinga Rutowicz, Calvin Matteoli, Marcel Böhrrer, Bart Rymen, Debbie Laudencia-Chingcuanco, et al.

► To cite this version:

Michael Thieme, Nikolaos Minadakis, Christophe Himber, Bettina Keller, Wenbo Xu, et al.. Transposition of HOPPLA in siRNA-deficient plants suggests a limited effect of the environment on retrotransposon mobility in *Brachypodium distachyon*. 2023. hal-04311886

HAL Id: hal-04311886

<https://hal.science/hal-04311886>

Preprint submitted on 28 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

2 Transposition of *HOPPLA* in siRNA-deficient plants suggests a 3 limited effect of the environment on retrotransposon mobility in 4 *Brachypodium distachyon*

5 Michael Thieme^{1*}, Nikolaos Minadakis¹, Christophe Himber², Bettina Keller¹, Wenbo Xu¹, Kinga
6 Rutowicz¹, Calvin Matteoli², Marcel Böhner², Bart Rymen², Debbie Laudencia-Chingcuanco³, John
7 Vogel⁴, Richard Sibout⁵, Christoph Stritt⁶, Todd Blevins² and Anne C. Roulin^{1*}

8 ¹ Department of Plant and Microbial Biology, University of Zurich, Zurich, Switzerland.

² Institut de Biologie Moléculaire des Plantes, CNRS, Université de Strasbourg, Strasbourg, France.

10 ³ USDA ARS Western Regional Research Center, Albany, CA, USA.

⁴ DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA.

12 ⁵ INRAE Unité BIA - 1268 Biopolymères Interactions Assemblages Equipe Paroi Végétale et Polymères Pariétaux
(PVPP), Nantes, France.

14 ⁶ Swiss Tropical and Public Health Institute (Swiss TPH), Allschwil, Switzerland.

16 * Corresponding authors: E-Mail

michael.thieme@botinst.uzh.ch (MT), anne.roulin@botinst.uzh.ch (ACR)

18

20 Abstract

22 Long terminal repeat retrotransposons (LTR-RTs) are powerful mutagens regarded as a major source
24 of genetic novelty and important drivers of evolution. Yet, the uncontrolled and potentially selfish
26 proliferation of LTR-RTs can lead to deleterious mutations and genome instability, with large fitness
28 costs for their host. While population genomics data suggest that an ongoing LTR-RT mobility is
30 common in many species, the understanding of their dual roles in evolution is limited. Here, we
32 harness the genetic diversity of 320 sequenced natural accessions of the Mediterranean grass
Brachypodium distachyon to characterize how genetic and environmental factors influence plant
34 LTR-RT dynamics in the wild. When combining a coverage-based approach to estimate global LTR-RT
36 copy number variations with mobilome-sequencing of nine accessions exposed to eight different
stresses, we find little evidence for a major role of environmental factors in LTR-RT accumulations in
B. distachyon natural accessions. Instead, we show that loss of RNA polymerase IV (Pol IV), which
mediates RNA-directed DNA methylation in plants, results in high transcriptional and transposition
activities of RLC_BdisC024 (*HOPPLA*) LTR-RT family elements, and that these effects are not stress-
specific. This work supports findings indicating an ongoing mobility in *B. distachyon* and reveals that
host RNA-directed DNA methylation rather than environmental factors controls their mobility in this
wild grass model.

Introduction

38 Transposable elements (TEs) are DNA sequences with the ability to form extrachromosomal copies
and to reintegrate elsewhere into the host genome. In plants, TE-derived sequences are ubiquitous
40 and can constitute more than 80 % of the genome (Mirouze & Vitte, 2014). In addition to playing a
major role in genome size variation (e.g. Hawkins et al., 2006; Piegu et al., 2006; Wang et al., 2021;
42 Yang et al., 2023), TEs can alter gene expression by acting as promoters or by providing *cis*-regulatory
elements to flanking regions (Butelli *et al.*, 2012; Makarevitch *et al.*, 2015; Roquis *et al.*, 2021). TEs
44 are therefore a major source of genetic change. Given that they are more likely than classic point
mutations to cause extreme changes in gene expression and phenotypes (Uzunović *et al.*, 2019;
46 Thieme *et al.*, 2022; Raúl *et al.*, 2023), they might be especially useful when the survival of an
organism or its descendants depends on a quick response to new or challenging environmental
48 conditions (for review Kawakatsu et al., 2016; Rey et al., 2016; Lanciano & Mirouze, 2018; Dubin et
al., 2018). Paradoxically, while population genomics data have revealed ongoing TE activity in natural
50 plant populations (e.g. (Stuart *et al.*, 2016; Stritt *et al.*, 2018; Baduel *et al.*, 2021), only a handful of
TE families have been reported to transpose in real-time. Therefore, how often or under which
52 natural conditions TEs are activated in the wild remain open questions. In addition, while ongoing
transposition is essential for TEs to survive, the presence of mobile and potentially ‘selfish’ DNA
54 sequences requires the host to evolve robust silencing mechanisms to prevent an uncontrolled TE
proliferation. TE activity thus remains a major puzzle in the field of evolutionary genomics.

56 In plants, the defence against TEs is multi-layered, comprising repressive histone modifications,
DNA methylation and RNA interference (Saze *et al.*, 2012; Zhang *et al.*, 2018; Liu *et al.*, 2022). One
58 of the main players of TE silencing is the RNA-directed DNA methylation (RdDM) pathway, which
involves two plant specific RNA-polymerases derived from Pol II, namely Pol IV and Pol V. The largest

60 subunits of each polymerase (NRPB1, NRPD1 and NRPE1, respectively) assemble with other proteins
into enzymes with distinct RNA products and functions (Ream et al., 2009; for review Rymen et al.,
62 2020). As a core component of RdDM, Pol IV (including NRPD1) transcribes TE regions into the
precursors of functionally specialized 24 nt small interfering RNAs (siRNAs) (Sigman & Slotkin, 2016;
64 Rymen *et al.*, 2020; Liu *et al.*, 2022). Upon the base pairing of 24 nt siRNAs to scaffold transcripts
produced by Pol V, the *de novo* DNA methyltransferase DRM2 (Zhong *et al.*, 2014) is recruited to
66 mediate the methylation and subsequent transcriptional repression of TEs. The essential role of
RdDM in TE silencing has been shown in *A. thaliana*, where the knockout of Pol IV and resulting
68 depletion of 24 nt siRNAs leads to a drastically increased heat-dependent transposition of the
ONSEN family (Tittel-Elmer *et al.*, 2010; Ito *et al.*, 2011).

70 The case of the heat-responsive *ONSEN* elements not only illustrates the importance of
epigenetic silencing in regulating TEs but also demonstrates that environmental factors may
72 modulate the dynamics of TEs in plants. Since their discovery by Barbara McClintock, who linked the
mobility of AC/DS elements in maize to the occurrence of a ‘genomic shock’ (McClintock, 1984), the
74 activity of TEs has been frequently associated with the presence of biotic and abiotic stressors. In
fact, certain TEs can sense specific physiological states of their host and use them to initiate their
76 own life cycle (Negi *et al.*, 2016). Besides *ONSEN* in *A. thaliana* (Cavrak *et al.*, 2014), the cold
inducible *Tcs1* element in blood oranges (Butelli *et al.*, 2012) or *Tos17* that gets activated during
78 tissue culture in rice (Hirochika *et al.*, 1996) are two other prominent cases of stress responsive TEs
in plants. Mechanistically, stress can activate TEs via specific motifs allowing the binding of
80 transcription factors and the subsequent recruitment of the transcription machinery to their
promoter-like sequences (Cavrak *et al.*, 2014; Grandbastien, 2015; Baduel *et al.*, 2021; Zhang *et al.*,
82 2022). The small window of increased activity during well-defined physiological states suggests that

some TEs have evolved a distinct lifestyle or ‘niche’ to successfully reproduce (Kidwell & Lisch, 1997;
84 Venner *et al.*, 2009; Stritt *et al.*, 2021; Stitzer *et al.*, 2021).

Since transcription constitutes the initial step to transposition, the abundance of TE
86 transcripts is often used as a proxy for TE activity (Lanciano & Cristofari, 2020). However, the life
cycle of TEs is complex (Schulman, 2013) and the fate of TE transcripts depends on many factors. For
88 instance, several transcriptionally active TEs have accumulated mutations that prevent the
production of enzymes needed for their autonomous transposition (Tanskanen *et al.*, 2007). To
90 selectively capture TEs that are not only transcriptionally active but also capable of transposing,
several protocols have been developed, including ALE-seq (Cho *et al.*, 2019), VLP-seq (Lee *et al.*,
92 2020) and mobilome-seq (Lanciano *et al.*, 2017). These recent approaches have been particularly
successful when targeting long terminal repeat retrotransposons (LTR-RTs), which represent the
94 largest fraction of TE-derived sequences in plant genomes (Vitte *et al.*, 2014). Indeed, LTR-RTs
transpose through a copy-and-paste mechanism which involves the reverse transcription of a full-
96 length RNA intermediate (Wicker *et al.*, 2007; Schulman, 2013). As part of their life cycle and
presumably through auto-integration, non-homologous and alternative end-joining, active LTR-RTs
98 can also form extrachromosomal circular DNA (eccDNA) intermediates (Flavell & Ish-horowicz, 1981;
Flavell, 1984; Garfinkel *et al.*, 2006; Wicker *et al.*, 2007; Lee *et al.*, 2020; Yang *et al.*, 2023a), whose
100 detection by mobilome-seq can be used as a proxy for their mobility (Lanciano *et al.*, 2017). For
instance, mobilome-seq has been successfully used to track full-length eccDNA of mobilized
102 autonomous RTs, containing both LTRs (2-LTR circles), in plants such as *A. thaliana* and rice
(Lanciano *et al.*, 2017; Thieme *et al.*, 2017).

104 While the activity of LTR-RTs has been extensively studied in the model *A. thaliana* (Baduel *et al.*, 2021), the interplay between genetic and environmental factors in other wild plant species
106 remains poorly investigated. To clarify these questions, we exploit here the *Brachypodium distachyon* diversity panel (Minadakis *et al.*, 2023a) and explore LTR-RT activity in a wild monocot.
108 *B. distachyon* is a Mediterranean grass with a compact diploid genome of ~272 Mb (International Brachypodium Initiative, 2010; Hasterok *et al.*, 2022) harboring 40 LTR-RT families (Stritt *et al.*, 2020)
110 that constitute about 30 % of the genome (International Brachypodium Initiative, 2010). In *B. distachyon*, LTR-RTs not only evolved varying insertion site preferences (Stritt *et al.*, 2020) but also
112 significantly differ in terms of transposition dynamics (Stritt *et al.*, 2018, 2020). While we previously suggested an ongoing transposition of LTR-RTs based on population genomics data (Stritt *et al.*, 2018; Horvath *et al.*, 2023), here we aimed to clarify the contribution of genetic and environmental factors
114 to LTR-RT activity in *B. distachyon*. To that end, we combined population genomics data available for 320 natural accessions with mobilome-seq under different stress conditions and asked: (i) does the
116 accumulation of LTR-RTs in these natural accessions correlate with environmental variables, (ii) is LTR-RT mobility induced by specific stresses, and (iii) which genetic factors influence the
118 accumulation of LTR-RTs ?

120

Results

122 **Abundances of LTR-RT families differ but show limited association with bioclimatic variables**
B. distachyon naturally occurs around the Mediterranean rim (**Fig 1A**) and groups into three main
124 genetic lineages (A, B and C) that further split into five genetic clades (Stritt *et al.*, 2022; Minadakis *et al.*, 2023a): an early diverged C clade and four clades found in Spain and France (B_West), Italy

126 (A_Italia), the Balkans and coastal Turkey (A_East) and inland Turkey and Lesser Caucasus (B_East).
We first aligned genomic reads of 320 accessions to the LTR-RT consensus sequences of
128 *B. distachyon* obtained from the TRansposable Elements Platform (TREP). We then computed the
abundance of TE-derived sequences, hence a proxy for copy number variation (pCNV), for the 40
130 annotated LTR-RT families using a coverage-based approach accounting for sample sequencing
depth (see Materials and Methods). We favored this approach over an analysis based on transposon
132 insertion polymorphisms (TIPs) because estimates based on TIPs are reference genome-dependent
and biased by the phylogeny in our study system. We have, for instance, previously shown that
134 accessions from the B_East clade harbor significantly less TIPs than accessions from the A_East clade
due to the fact that the reference genome Bd21 belongs the B_East clade (Stritt *et al.*, 2018). In
136 addition, whole-genome *de novo* assembly of 54 *B. distachyon* natural accessions and the
subsequent pangenome analysis revealed that non-reference accessions display large genomic
138 variations (Gordon *et al.*, 2017), which may further bias the estimates of TIP abundances.

The heatmap produced based on pCNV (**Fig 1B**) showed that LTR-RTs underwent different
140 transposon accumulations. We found that *Gypsy* elements (RLG) harbor a higher pCNV than *Copia*
(RLC) elements (Wilcoxon test, $W = 6517781$, $p\text{-value} < 2.2e\text{-}16$; **Fig 1B, C**). Furthermore, a PCA based
142 on RLG pCNVs did not allow us to discriminate accessions based on their phylogenetic relationship,
while a PCA performed with RLC elements separated the samples by genetic lineage (**Fig 1D**). The
144 strongest result was found with a PCA performed with the five youngest and putatively most recently
active LTR-RT families found in the pangenome of *B. distachyon* (the Angela families RLC_BdisC022,
146 RLC_BdisC024, RLC_BdisC030, RLC_BdisC209 and the Tekay family RLG_BdisC004), with the first two
axes together explaining more than 77.8% of the variance. Finally, with the exception of samples

148 from the most recently diverged clades A_East and A_Italia (13 kya), they further allowed us to discriminate samples based on the genetic clade of origin (Fig 1 B,D).

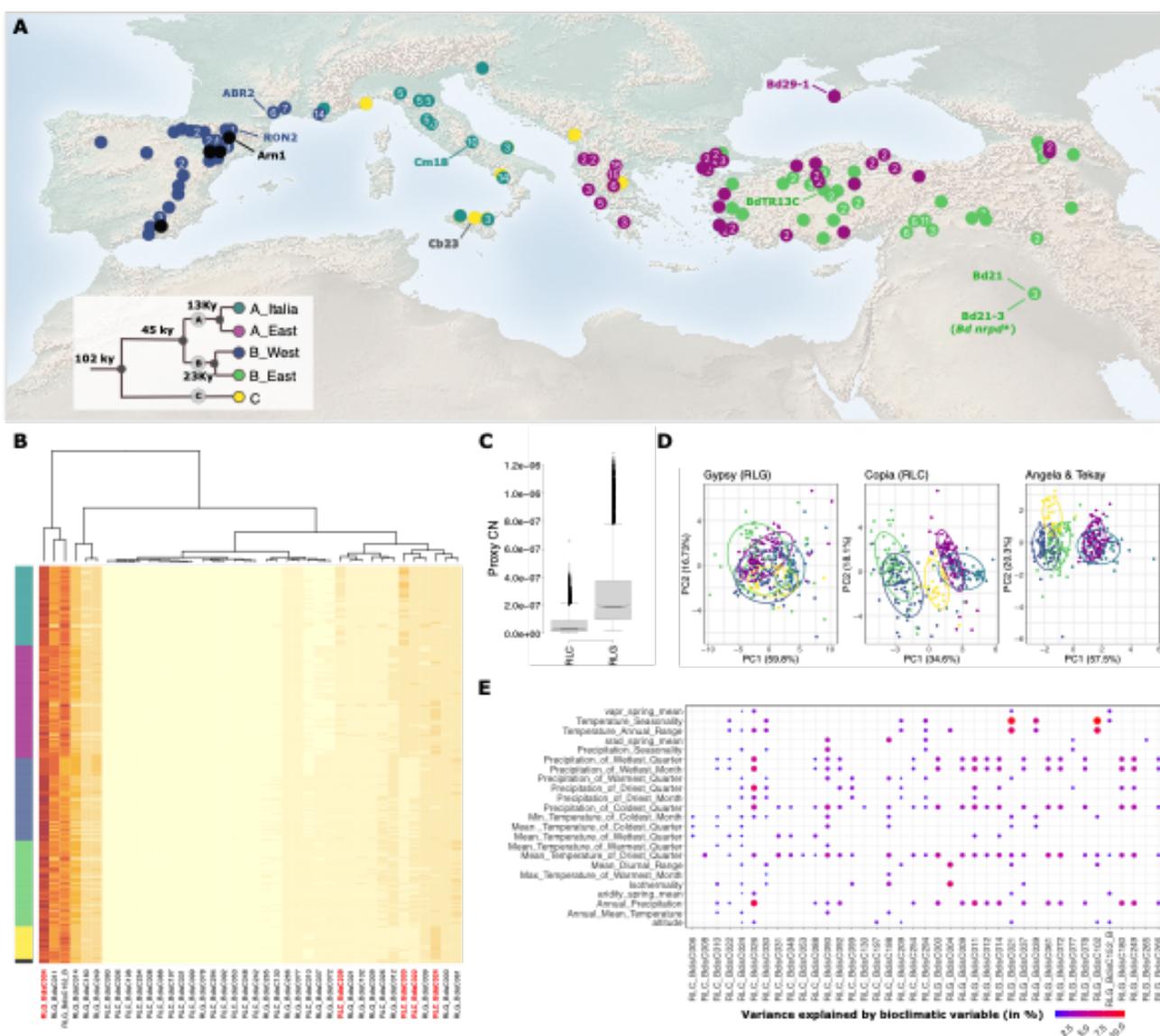


Fig 1. Natural diversity of proximal copy number variation (pCNV) of LTR-RTs in *B. distachyon*. (A) Origin of the 320 natural accessions included in this study. Accessions that were used for the mobilome-seq are labelled in the map and a PCA with the bioclimatic variables of their place of origin is depicted. Colors of points correspond to the genetic clades whose estimated split is shown in the phylogenetic tree. Black points indicate that the accession cannot be clearly assigned to one genetic clade. (B) Heatmap with read counts of TE-derived sequences (proxy for the copy number variations, pCNVs) of all 40 annotated LTR-RTs in 320 natural accessions of *B. distachyon*. Accessions are sorted according to their phylogeny and names of recently active TE-families are highlighted in red. (C) Overall estimates of the copy numbers of Copia and Gypsy-type LTR-RTs in 320 natural accessions. (D) PCAs of pCNVs of all Gypsy, Copia and all recently active LTR-RT families, highlighted in (B) belonging to the Angela & Tekay families (RLG_Bdis004, RLC_BdisC030, RLC_BdisC209, RLC_BdisC024 and RLC_BdisC022). Colors of points indicate the genetic clade of accessions. (E) Output of the LMM analyses between pCNVs of LTR-RTs and bioclimatic variables at the accessions' origin. Bubbles indicate a significant association (P -value < 0.05). Colors and sizes of bubbles show the part of the variance (marginal R^2) explained by the bioclimatic variables in %.

150 To test whether the accumulation of LTR-RT sequences correlated with environmental
factors, we retrieved bioclimatic variables comprising precipitation, temperature, aridity levels, solar
152 radiation and atmospheric pressure at each locality. We then ran linear mixed models (LMM) where
pCNV per LTR-RT family was entered as the response variable, the bioclimatic variables entered
154 separately as fixed factors and the clade of origin as random factors to account for population
structure. Marginal R^2 extracted for each LMM did not exceed 10% even for the putatively most
156 recently active LTR-RT families (**Fig 1E**), indicating that albeit significant, the association between
pCNV per family and the environment was mild in our study system. We observed similar
158 associations between pCNVs and bioclimatic variables when not accounting for population structure
and running classical linear model analyses (**S1 Fig**). With the exception of RLC_BdisC010 and
160 RLG_BdisC265, for which more than 40% of the variance in pCNVs was explained by environmental
factors, the LTR-RT families showed non-significant to mild associations with environmental variables
162 (**S1 Fig**).

164 **Recently active LTR-RT families produce eccDNAs**

The coverage-based approach for estimating LTR-RT CNVs considers all TE-derived sequences and
166 does not take into account that individual families differ in their age structures, turnover times and
proportion of full-length, potentially autonomous mobile elements (Stritt *et al.*, 2020). We therefore
168 complemented our *in silico* analysis by experimentally testing whether LTR-RT families were indeed
still active *in planta* and whether the mild but significant correlation we observed between global
170 pCNVs and bioclimatic variables may be due to a stress-specific activity.

To cover a wide range of the genetic, geographic and bioclimatic diversity of *B. distachyon*, we
172 selected nine natural accessions belonging to the five genetic clades and originating from contrasting

habitats (**Fig 1A**). Considering the role of Pol IV in the silencing of TEs in plants, we also included two
174 independent Pol IV mutant lines, a sodium azide mutagenized line (hereafter *Bd nrpd1-1*) and a T-
DNA line (hereafter *Bd nrpd1-2*), both carrying a homozygous mutation in the largest subunit of
176 Pol IV (NRPD1; Bradi2g34876) in the Bd21-3 background. We exposed plants to eight different
stresses, namely cold, drought, heat, salt, submergence, infection with *Magnaporthe oryzae*,
178 treatment with glyphosate and chemical de-methylation and performed mobilome-seq on all
resulting 105 samples (see Material and Methods).

180 Following the removal of organelle-derived reads, we first assembled mobilome reads and
aligned the resulting ten longest contigs of each sample to the reference assembly of *B. distachyon*.
182 We then screened genomic regions for which at least three assembled mobilome contigs of different
samples aligned and further assessed in which genotypes or stresses those contigs occurred. We
184 only retained circle-forming regions with a specificity above 50 % (i.e., regions for which more than
half of the contigs belonged to a certain stress or genotype). In addition, we also kept recurrently
186 active regions, present at a high frequency independently of the stress or the genotype in at least
ten samples. In total, we retained 15 circle-forming regions, all of which contained TE sequences
188 (**Fig 2**). Eight of these corresponded to the Angela family (RLC_BdisC024, RLC_BdisC022,
RLC_BdisC030, RLC_BdisC209), four to CRM elements (RLG_BdisC039, RLG_BdisC102), and one each
190 to the SIRE (RLC_BdisC026), the Alesia (RLC_BdisC010) and the non-autonomous and unclassified
RLG_BdisC152 family. We hereafter refer to RLC_BdisC024 as *HOPPLA* (German allusion for the the
192 surprising finding of a jumping element). We did not find stress specificity in the formation of
eccDNAs (**Fig 2A**). However, our results pointed to a genotype-dependent formation of eccDNAs for
194 the RLC_BdisC209, RLC_BdisC026 and *HOPPLA* families (**Fig 2B**). In particular, two contigs containing
HOPPLA elements were exclusively detected in the two *pol IV* mutants (**Fig 2B**).

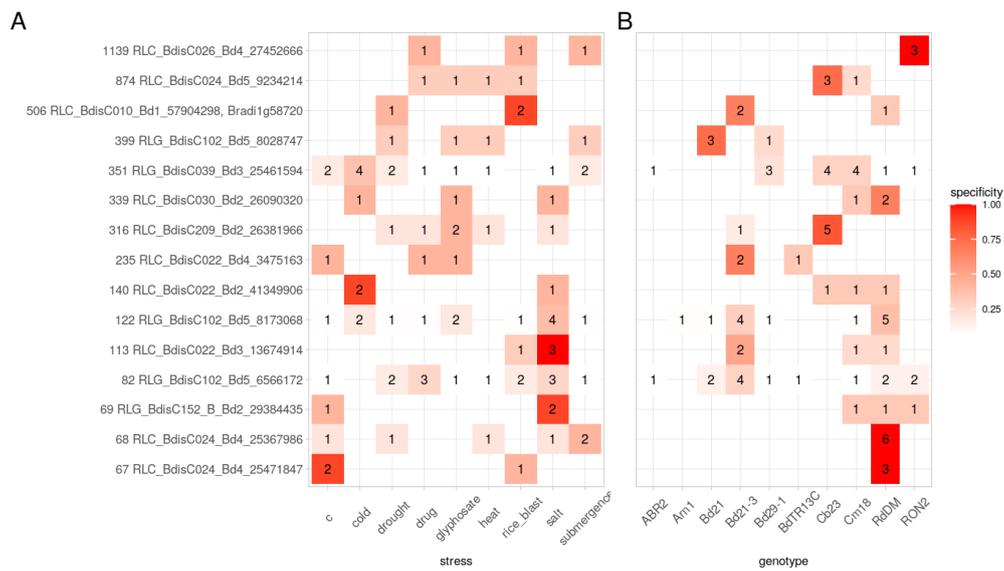


Fig 2 Assessment of LTR-RT mobility in *B. distachyon*. Stress (A) and genotype (B) specificity of the formation of eccDNA as determined by the alignment of assembled mobilome-seq reads. The color represents the degree of specificity and numbers indicate the count of samples from which one of their ten longest contigs aligns to each of the circle-forming regions. Annotations of regions are indicated on the y-axis. Multiple annotations in the same circle-forming region were concatenated. The two *pol IV* mutants and the controls Bd21-3 and the outcrossed line BdNRPD1 (+/+) are summarized as RdDM and Bd21-3, respectively

198 **HOPPLA activity is increased in the *pol IV* mutants regardless of the stress applied**

Fragmented eccDNAs or circles containing only of the two LTRs (1-LTR circles) can be formed following reverse transcription by auto-integration, alternative end-joining in the virus-like particles (Garfinkel *et al.*, 2006; Yang *et al.*, 2023a) or by a recombination of the two LTRs of genomic copies (Smith & Vinograd, 1972; Gaubatz, 1990). Hence, 1-LTR circles do not necessarily imply LTR-RT mobility. In contrast, recent work indicates that 2-LTR circles are formed following the complete reverse transcription by non-homologous end-joining of an intact full-length linear RT copy that is capable of integrating into the genome (Yang *et al.*, 2023a). Indeed, the detection of full-length 2-LTR circles of well characterized autonomous LTR-RTs such as *EVD (ATCOPIA93)* has been directly linked to their actual transposition (Lanciano *et al.*, 2017). As a complement to the assembly-based

208 analysis, we thus followed a stringent approach to analyse our mobilome-seq data. We aligned reads
to a library comprising artificial 3'LTR-5'LTR fusions of all full-length LTR-RTs annotated in the
210 *B. distachyon* reference assembly (Stritt *et al.*, 2020). Because such reads are not expected to be
present in genomic DNA, this allowed us to specifically detect intact 2-LTR circles of
212 extrachromosomal LTR-RTs capable of integrating into the genome. To control for possible traces of
undigested genomic DNA that may subsequently be amplified by the Phi29 enzyme during
214 mobilome-seq (Silander & Saarela, 2008), we also included publicly available genomic reads of all
nine accessions in our analysis.

216 We found that several LTR-RTs formed eccDNA with 2-LTR junctions. Yet, most of them
occurred sporadically and we did not observe a recurring stress-specific formation of 2-LTR circles
218 for any of the 37 LTR-RT families with annotated full-length copies (**Fig 3A**). For instance, we found
a very strong signal for RLC_BdisC031 that was solely detected in glyphosate-treated Bd21-3 plants
220 and therefore not further considered in the analysis. In contrast, and in accordance with the
assembly-based approach, the Angela element *HOPPLA* showed a recurrent formation of 2-LTR
222 circles. However, this formation was not triggered by a specific stress and only occurred in the two
independent *pol IV* mutants (**Fig 3B**). Finally, our attempt to transiently inhibit LTR-RT silencing using
224 alpha-amanitin and Zebularine (a combination of inhibitors shown to increase the activity of LTR-RTs
in *A. thaliana* and rice), did not result in the consistent activation of *HOPPLA* or other LTR-RTs in
226 multiple accessions (**Fig 3A**). The presence of 2-LTR circles of *HOPPLA* in *Bd nrpd1-2 (-/-)* was
confirmed by an inverse PCR on total DNA that was not subjected to a rolling circle amplification,
228 with outward facing primers specific to the two LTRs (**Fig 3C**). Notably, we also detected a faint signal
for the outcrossed line *Bd NRPD1 (+/+)* suggesting a weak activity of *HOPPLA* in wild-type plants.

230 Since the stress- or tissue-dependent activity of LTR-RTs is mediated by the specific binding
of transcription factor (TF), we screened the consensus sequence of *HOPPLA* for motifs of known TF
232 binding sites. First, we validated this approach by analyzing one of most active copies (AT1G11265)
of the heat-responsive *ONSEN* family of *A. thaliana*. As expected, a GO term analysis indicated a

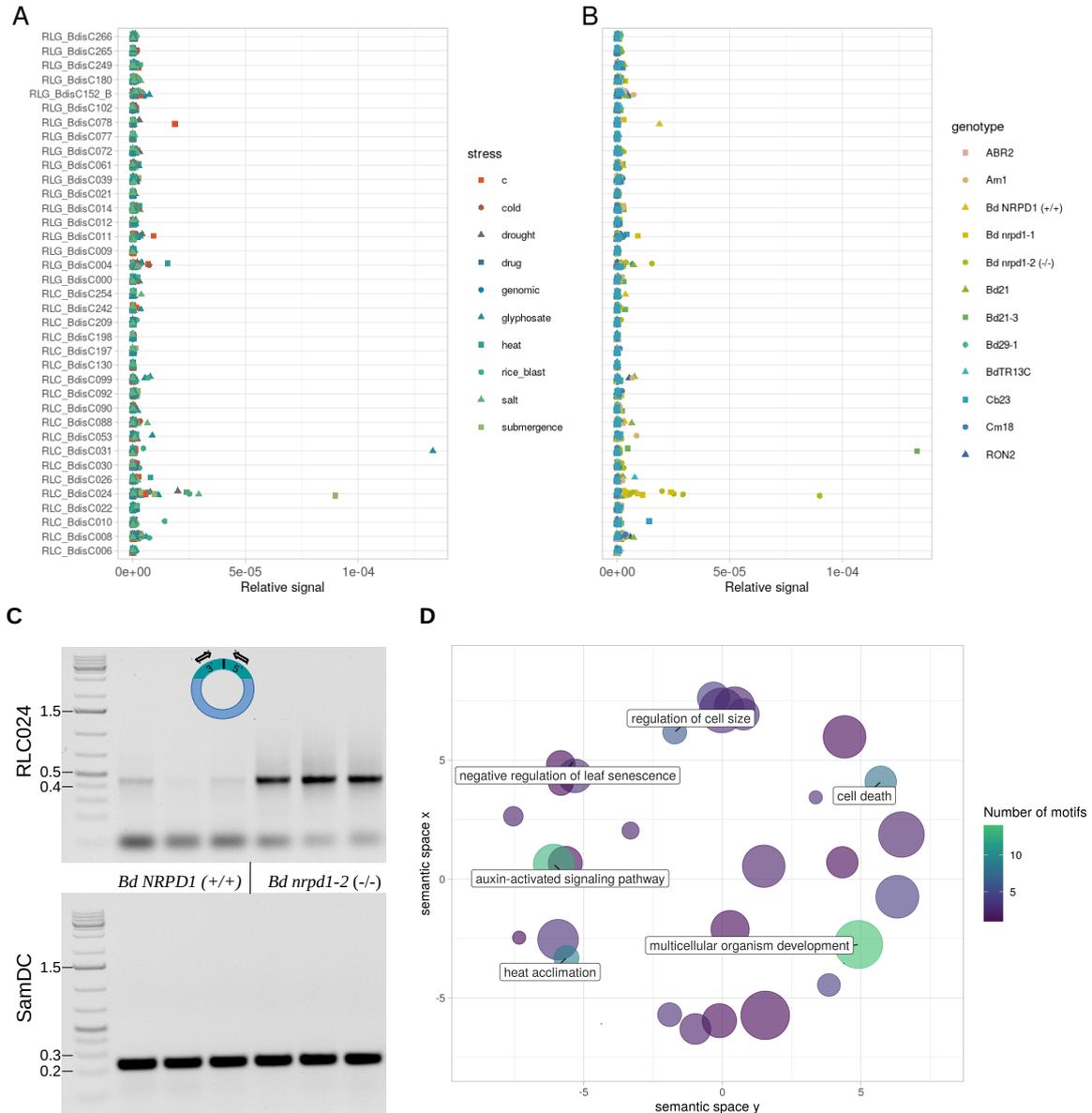


Fig 3 *HOPPLA* forms 2-LTR circles in the *pol IV* mutants. Relative abundance of 2-LTR-junction spanning reads depending on the stress (A) and the genotype (B) of individual mobilome-seq samples. (C) Inverse PCR using total DNA not subjected to a rolling circle amplification for the confirmation of an increased amount of extrachromosomal 2-LTR circles of *HOPPLA* in the *Bd nrpd1-2* (-/-) mutant compared to the *Bd NRPD1* (+/+) outcrossed line. Three biological replicates are shown. (D) GO enrichment analysis of transcription factors for which binding sites have been detected in the consensus sequence of *HOPPLA*. Colors indicate number of TF-binding sites found. GO-terms that occur at least six times are highlighted in the plot. All GO-terms and their number of occurrences are listed in S1 Table.

234 strong enrichment of heat-responsive TFs for this element (**S2 Fig**). In contrast to the well-known,
stress-responsive *ONSEN* LTR-RT, the GO terms of TFs that could bind to *HOPPLA* indicated that
236 developmental processes and auxin-activated signaling pathways played a role in its activity, rather
than specific stresses (**Fig 3D**).

238

Members of the *HOPPLA* family differ in activity

240 Because individual copies of the same LTR-RT family can differ in their activity (Cavrak *et al.*, 2014),
we also *anad* the relative abundance of 2-LTR-spanning reads for each annotated full-length copy of
242 *HOPPLA*. This analysis revealed a great diversity of eccDNA formation among individual copies of the
HOPPLA family and confirmed the strongest activity of *HOPPLA* in the two *pol IV* mutants (**Fig 4A**).
244 We also obtained a few reads spanning the 2-LTR-junction of the two most active *HOPPLA* copies
(Bd3_22992889 and Bd4_25471847) in the *Bd NRPD1 (+/+)* control line (**Fig 4A**), which confirmed
246 the weak but detectable band for the inverse 2-LTR PCR for *Bd NRPD1 (+/+)* (**Fig 3C**).

Using the meta information of individual *HOPPLA* copies described previously (Stritt *et al.*,
248 2020), we further assessed which genomic factors (DNA methylation, CG content, distance to the
closest gene or age of the copy) were linked to the 2-LTR circle formation for individual *HOPPLA*
250 copies. While no clear pattern emerged from this analysis, the most active copies of *HOPPLA* tend
to be rather young (**Fig 4B**).

252

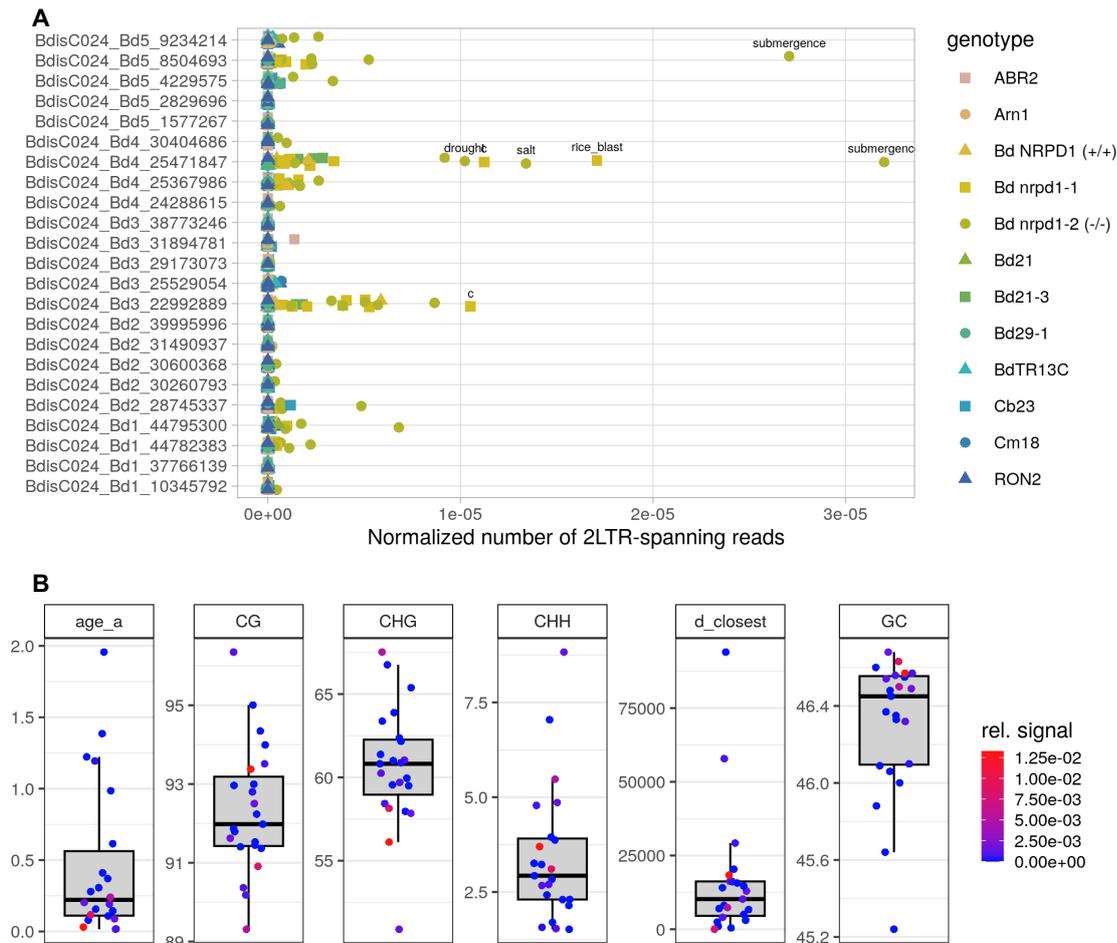


Fig 4 Members of the HOPPLA family differ in activity. (A) Relative abundance of 2-LTR-junction spanning reads of individual full-length copies of the HOPPLA family. **(B)** Age (age_a), closest distance to gene (d_closest), GC content (GC), and methylation levels in CG, CHG and CHH contexts of all individual genomic full-length copies of HOPPLA in percent. The color indicates relative abundance of 2-LTR-junction spanning reads from the mobilome-seq in (A).

254 **HOPPLA is targeted by Pol IV-dependent 24 nt siRNAs in the wild type and transposes in *pol IV* mutant plants**

256 The pivotal role of Pol IV in producing TE-specific 24-nt siRNAs for RNA-directed DNA methylation has been demonstrated in many plant species including *A. thaliana* (Ito *et al.*, 2011), rice (Xu *et al.*, 2020) and for the Alesia family (RLC_BdisC010) in *B. distachyon* (Böhrer *et al.*, 2020). To confirm that the increased production of 2-LTR eccDNA circles of HOPPLA in mutants deficient in *B. distachyon* NRPD1 is correlated with a depletion of 24 nt siRNAs, we performed a small RNA blot including

260

samples from the two *pol IV* mutants and their respective wild-type controls. Using a hybridization
262 probe specific to the *HOPPLA* LTRs, 24 nt siRNAs were detected in the control lines Bd21-3 and
Bd NRPD1 (+/+), but not in either of the *pol IV* mutant lines (**Fig 5A**). This finding strongly suggests
264 that *HOPPLA* is under control of the Pol IV-RdDM pathway, and that the absence of 24 nt siRNAs
results in the upregulation and increased production of 2-LTR eccDNAs from *HOPPLA*. Furthermore,
266 RNA-seq data from part of the same mutant panel shows that *HOPPLA* is the most upregulated LTR-
RT family in the *Bd nrpd1-2 (-/-)* background compared to the *Bd NRPD1 (+/+)* control line, indicating
268 that the reduction of 24 nt siRNAs is likely associated with an increased expression and subsequent
formation of *HOPPLA* eccDNAs in both *pol IV* mutants (**Fig 5B**).

270 To complete their life cycle, reverse transcribed extrachromosomal copies of LTR-RTs have to
integrate into the host genome (Schulman, 2013). Because all our analyses congruently pointed to
272 the activity of *HOPPLA* in the *Bd nrpd1-2 (-/-)* mutant, we sequenced the genome of seven

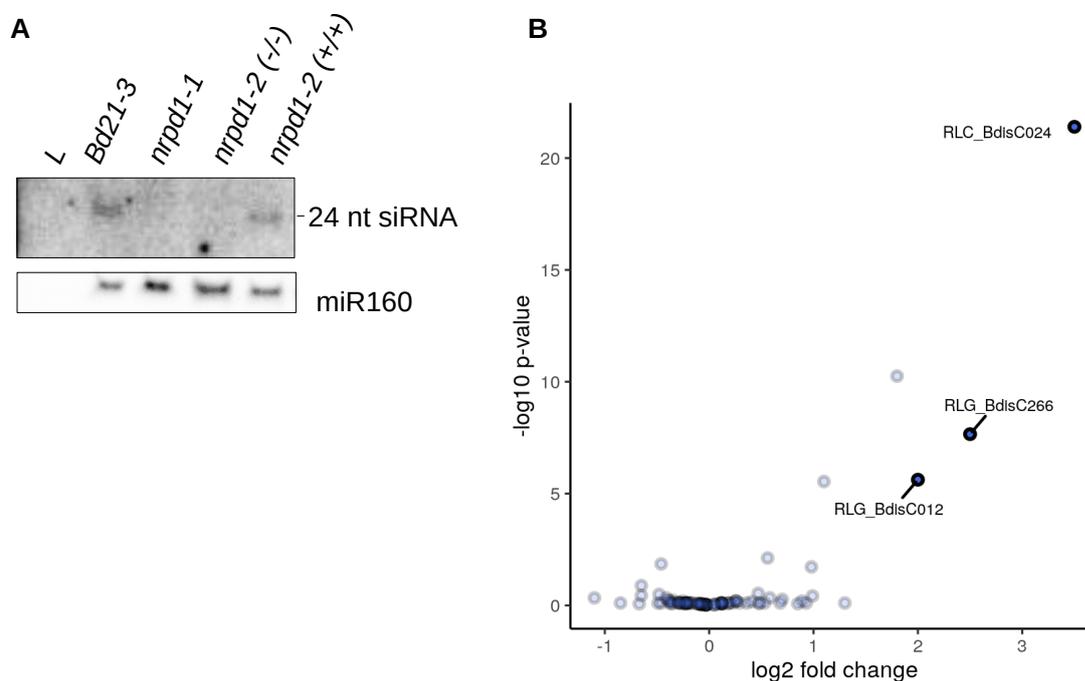


Fig 5 Loss of 24-nt siRNAs leads to an increased activity of *HOPPLA* in the Pol IV mutants (A) Northern plot for the detection of 24-nt siRNAs specific to the 3' LTR of *HOPPLA* in the *pol IV* mutants *Bd nrpd1-1* and *Bd nrpd1-2 (-/-)* and their control lines Bd21-3 and the outcrossed line *Bd NRPD1 (+/+)*. (B) SalmonTE analysis of the expression of LTR-RTs in *Bd nrpd1-2 (-/-)* relative to the outcrossed control line *Bd NRPD1 (+/+)*. LTR-RTs with a log₂ fold change of at least two are labeled, three biological replicates were analysed.

individuals of *Bd nrpd1-2* (-/-), six *Bd NRPD1* (+/+) plants and one wild-type Bd21-3 plant to detect
274 new *HOPPLA* insertions. As TIPs were detected relative to the reference genome Bd21 (an accession
closely related to Bd21-3 but not genetically identical), we first removed all conserved Bd21-3-
276 specific TIPs detected in multiple lines. We manually curated all filtered candidate TIPs and showed
that *HOPPLA* was the only family for which validated TIPs were identified in one of the re-sequenced
278 *Bd nrpd1-2* (-/-) plants (Bd1 38798495, Bd1 42205987, Bd4 28119639) (**S3-5 Figs**). This confirmed
that the loss of Pol IV function led to an increased production of eccDNA, as well as actual
280 transposition and accumulation of novel *HOPPLA* copies in the tested *Bd nrpd1-2* (-/-) mutant. The
presence of reads spanning the insertion site indicated that the detected *HOPPLA* insertions were
282 heterozygous or probably somatic for the insertion Bd4 28119639, which exhibited a specifically low
proportion of clipped reads. No TIPs were detected for any other LTR-RT family.

284

Genome-wide association studies for pCNVs do not recover known components of RdDM

286 To decipher the genetic basis of *HOPPLA* accumulations in natural populations, we first performed a
genome-wide association study (GWAS) using *HOPPLA* pCNVs in the diversity panel of 320 natural
288 accessions (**Fig 1B**) as a phenotype. We identified only one region with two significant peaks (FDR-
adjusted p-value < 0.05, Bd5 6920000-6960000 and 7210000-7240000) obtained for the mobile
290 *HOPPLA* family (**Fig 6**). Because inserted copies of *HOPPLA* may themselves lead to significantly
associated regions in the GWAS as shown in *A. thaliana* (Quadrana *et al.*, 2016), we first verified that
292 there were neither TIPs (Horvath *et al.*, 2023) nor annotated reference insertions of *HOPPLA* in that
region (**Fig 6**). As described above, our data suggested that the loss of 24-nt siRNAs in the Pol IV
294 mutants was sufficient to mobilize *HOPPLA*. We therefore further tested whether any of the genes
encoding subunits of Pol IV or Pol V would be localized in or near this region (window size 50 kb)

296 **(S3 Table)**. We did not detect any known Pol IV or Pol V- related genes, but instead found
 Bradi5g05225, an ortholog of the *A. thaliana* ROS1-associated methyl-DNA binding protein 1 (RMB1,
 298 AT1g63240) (Liu *et al.*, 2021a), to be co-localized with the peak.

To test whether genomic regions might be recurrently associated with their pCNVs, we finally
 300 extended the GWAS analyses to the four other most recently active families (RLC_BdisC022,
 RLC_BdisC030, RLC_BdisC209 and RLG_BdisC004) (**S6 Fig**). We extracted the candidate genes for

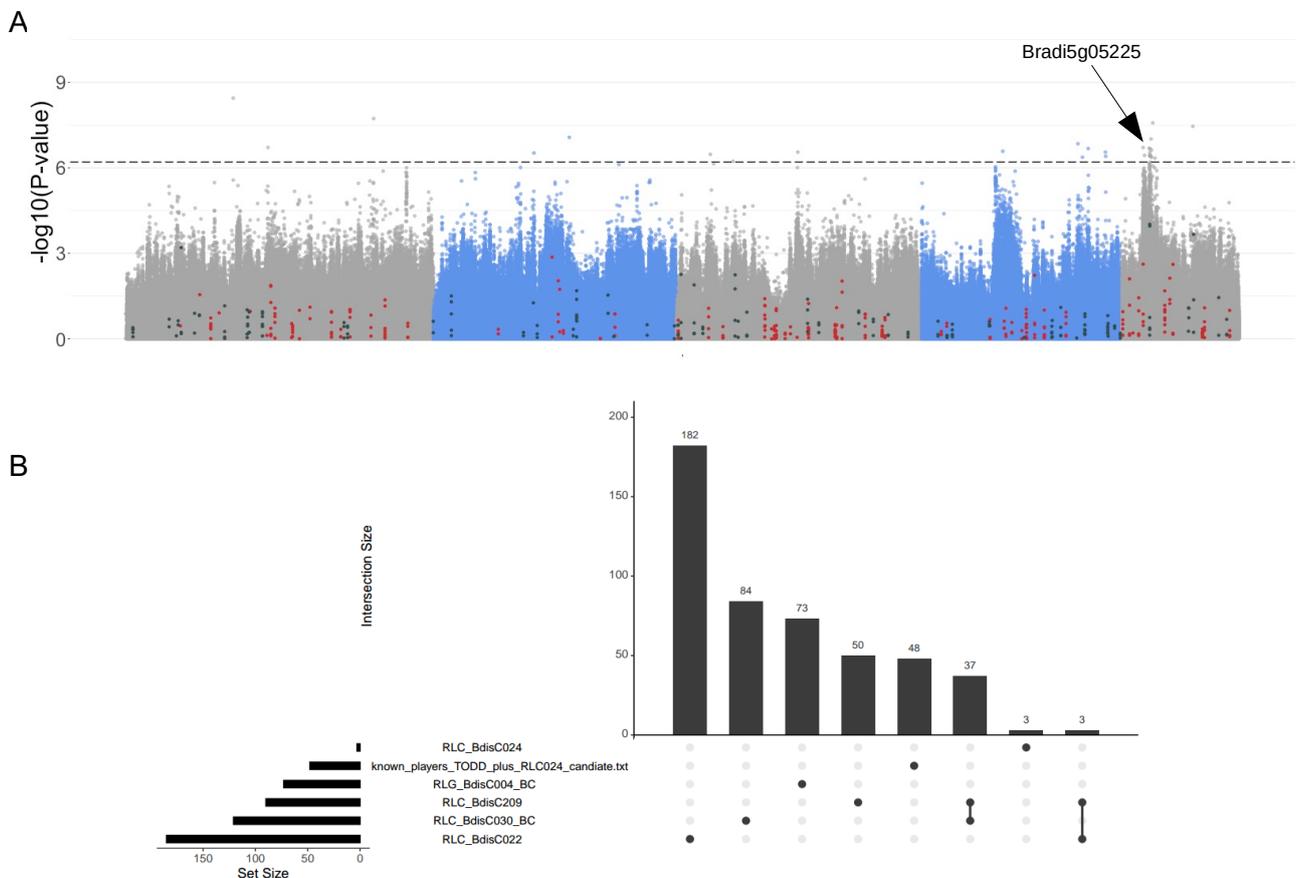


Fig 6 Genetic regions associated with pCNVs of recently active LTR-RTs (A) Manhattan plot depicting the GWAS results of pCNV of *HOPPLA* in 320 accessions of *B. distachyon*. Colored points indicate SNPs linked to TIPs (blue), TAPs (red) or known components of the Pol IV and V holoenzymes (asterisks). Threshold of significance (false discovery rate adjusted p-value <0.05) is marked with dashed lines. A significant region containing the candidate gene Bradi5g05225 (window size 50 kb) is highlighted. **(B)** UpSet plot of genes in 20 kb windows surrounding significant regions with at least two SNPs above the threshold of significance (FDR-adjusted p-value <0.05 for *HOPPLA*, RLC_BdisC209 and RLC_BdisC022 and Bonferroni correction for RLC_BdisC030 and RLG_BdisC004) of the five most recently active LTR-RT families in *B. distachyon*. To visualize potential overlaps, a list of the componentlangts of the Pol IV and Pol V holoenzymes is included in the UpSet plot.

302 each of the five families (see **S3 Table** and Material and Methods). Apart from the two closely related
families RLC_BdisC030 and RLC_BdisC209 that shared the majority of their GWAS candidates, and
304 RLC_BdisC022 and RLC_BdisC209 that shared three genes, we found no overlap of annotated loci
potentially contributing to the pCNVs of recently active families (**Fig 6**).

306

Discussion

308 Understanding the dynamics of TEs and their role in adaptation is currently one of the major
challenges in the field of evolutionary genomics. The fact that mobile TEs are a source of epi/genetic
310 diversity and potential drivers of evolution has been demonstrated in many organisms including
fungi (Muszewska *et al.*, 2019), insects (Gilbert *et al.*, 2021), mammals (Senft & Macfarlan, 2021)
312 and plants (Lisch, 2013). However, while there are a number of examples showing that certain TE
insertions facilitated the adaptation to changing environments (for review Bourgeois & Boissinot,
314 2019), TEs are generally harmful (Bourgeois *et al.*, 2020; Horvath *et al.*, 2023; Langmüller *et al.*, 2023)
and therefore controlled by complex silencing mechanisms. To foster our understanding of TE
316 activity, we investigated the environmental conditions and genetic factors associated with the
accumulation and mobility of LTR-RTs in plant genomes. By measuring LTR-RT pCNVs in a panel of
318 320 *B. distachyon* natural accessions, we show that the intra-specific variations of pCNVs of RLC
elements, but not the pCNVs of the generally older and more abundant RLG elements (Stritt *et al.*,
320 2020), separate accessions according to their genetic cluster of origin. This is even more striking for
members of the Angela (RLC_BdisC022, RLC_BdisC024, RLC_BdisC030, RLC_BdisC209) and the Tekay
322 (RLG_BdisC004) family, which are the youngest families in *B. distachyon* (Stritt *et al.*, 2020). Highly
polymorphic among natural accessions of *B. distachyon* (Stritt *et al.*, 2018, 2020), they are

324 expectedly the main drivers of lineage-specific expansions of pCNVs in our study system. Hence, we
do not only confirm that LTR-RT families in *B. distachyon* globally differ in size (Stritt *et al.*, 2020) but
326 also demonstrate that the accumulation of genomic sequences derived from specific families varies
significantly among natural accessions.

328 The transcriptional activity of LTR-RTs can be triggered by specific environmental stresses
(Grandbastien, 2015; Negi *et al.*, 2016). Given that *B. distachyon* occurs in a wide range of different
330 habitats in the Mediterranean area (Minadakis *et al.*, 2023b), this characteristic feature of LTR-RTs
provides a potential explanation for the pCNVs we observed across natural accessions (Quadrana *et*
332 *al.*, 2016; Baduel *et al.*, 2021). Yet, for all LTR-RT families, pCNVs correlate only moderately with
environmental factors. Consequently, our genomic data do not support a large effect of the
334 environment on LTR-RT activity in *B. distachyon*. While this result could seem startling, it is not
completely surprising. Indeed, many LTR-RT families, and especially the old RLG elements, do not
336 show signs of increased activity in the recent past in *B. distachyon* (Stritt *et al.*, 2020). Considering
that their copy number expansions took place in a climate that has drastically changed following the
338 last glacial maximum (Minadakis *et al.*, 2023b), a limited link between their activity and the current
environmental conditions is actually expected for most families. In contrast, the lack of correlation
340 between current bioclimatic variables and copy number variation for families with ongoing activity
(RLC_BdisC022, RLC_BdisC024, RLC_BdisC030, RLC_BdisC209 and RLG_BdisC004; Stritt *et al.*, 2020),
342 suggests a more complex mechanism than their pure dependence on specific stresses in certain
environments. This hypothesis is supported by previous findings in *A. thaliana*. Indeed, a minor
344 impact of the environment on transpositional activity was also found in this species, where the two
most associated environmental variables ('seasonality of precipitation' and 'diurnal temperature
346 range') only explained about 9 % of the observed variation (Baduel *et al.*, 2021).

Genetic factors are well-known to be essential in regulating LTR-RT activity (Miura *et al.*, 2001; 348 Tsukahara *et al.*, 2009; Mirouze *et al.*, 2009; Bourque *et al.*, 2018). As the loss of main players of the RdDM silencing pathway leads to increased TE activity (Ito *et al.*, 2011; Benoit *et al.*, 2019; Xu *et al.*, 350 2020; Baduel *et al.*, 2021), the two *B. distachyon* Pol IV (NRPD1) mutants provided an ideal functional tool to experimentally validate our *in silico* analysis. Since transcriptionally active LTR-RTs are not 352 necessarily able to transpose (Bajus *et al.*, 2022), we used a mobilome-seq approach to detect TE-derived eccDNAs and transpositionally active LTR-RT families. We deliberately followed a very 354 stringent approach for analysing the data and by doing so, identified *HOPPLA* as the only highly active LTR-RT family in *B. distachyon*. Indeed, *HOPPLA* is the only family for which we further detect 356 newly inserted copies in the Pol IV mutant.

The non-stress-specific activity of *HOPPLA* in the two independent Pol IV mutants supported 358 our *in silico* approach and strengthened the idea that genetic, rather than environmental stresses, are major drivers of LTR-RT activity in *B. distachyon*. These results are also in line with our TF binding 360 sites analysis. In contrast to the heat-responsive *A. thaliana* element *ONSEN*, for which we predominantly recovered TF-binding sites associated with heat response, *HOPPLA* seemed to be 362 targeted by TFs involved in developmental processes and auxin signaling (Leyser, 2018). Follow-up studies should therefore address the question of whether the activity of *HOPPLA* or other families 364 differs between tissues or developmental stages, as observed for the endosperm-specific mobility of *PopRice* in rice, for example (Lanciano *et al.*, 2017). Strikingly, despite the central role of Pol IV in 366 the RdDM pathway, we did not observe bursts of multiple LTR-RT families but instead found that the loss of 24 nt siRNAs specifically activated individual copies of the *HOPPLA* family. Interestingly, we 368 also detected a weak signal for 2-LTR eccDNAs in the Bd21-3 wt and the outcrossed line *Bd NRPD1 (+/+)* but not in other natural accessions. This suggests that the accession-specific composition of

370 the mobilome, and hence the genetic background of the *pol IV* mutant line, plays an important role
in LTR-RT activity. Related to this, we sporadically observed very strong signals for individual samples,
372 which could indicate an accession-specific response of the mobilome to certain triggers.

Given that pCNVs vary greatly among genetic clades, assessing the effect of a genetic
374 mutation of major components of the RdDM pathway in a set of genetically diverse natural
accessions would be timely, yet labor-intensive as transformation works more efficiently in the Bd21-
376 3 background than in the other accession tested. Our attempt to transiently reduce LTR-RT silencing
in multiple accessions from different genetic clades using the chemical inhibition of Pol II and DNA
378 methyltransferases (Thieme *et al.*, 2017) did not result in an increased activity of *HOPPLA* or
members of any other LTR-RT family. In addition, and despite the differences of activities observed
380 among individual *HOPPLA* copies, we could not detect, in the present study, a clear link between
their activity and GC contents or methylation states. Taken together, these findings suggest that the
382 specific function of the canonical RdDM with Pol IV, rather than generic DNA methylation states are
regulating *HOPPLA* activity. Yet, our GWAS failed to recover major components of the RdDM
384 pathway. Instead, the diversity of activity within the *HOPPLA* family may suggest that the presence
of single active copies could determine the fate of an entire family. In addition, pCNV is also
386 dependent on the removal rate of LTR-RT families, which varies greatly in *B. distachyon* (Stritt *et al.*,
2020). This complexity of parameters affecting the dynamics of LTR-RT might explain why none of
388 the genes known to be involved in silencing LTR-RTs are associated with pCNV for *HOPPLA* or any
recently active family. Our candidate locus containing Bradi5g05225, a gene related to RMB1 whose
390 loss of function has been shown to result in DNA hypermethylation (Liu *et al.*, 2021b), remains
nonetheless a great candidate for functional validation.

392 Altogether, our work confirms that LTR-RTs in *B. distachyon* are ‘well-behaved’ (Stritt *et al.*,
2018) and that the evolutionary consequences of their mobility are hard to study in real-time.
394 Indeed, while mobilome-seq revealed a sporadic activity for other families, we only found recurring
activity and new insertions of *HOPPLA* in the *pol IV* mutant. These results somewhat contrast with
396 our population genomics analyses which clearly indicate an ongoing activity of several LTR-RT
families in natural accessions. We propose that the activity of LTR-RTs is relatively low and might
398 depend on a complex interaction between genetic factors, developmental stages and, more
marginally, the punctual occurrence of stresses.

400

402 **Material and methods**

Estimation of LTR-RT pCNVs

404 We used publicly available genomic reads of 320 sequenced natural accessions of *B. distachyon*
(Gordon *et al.*, 2017, 2020; Skalska *et al.*, 2020; Stritt *et al.*, 2022; Minadakis *et al.*, 2023a) to assess
406 the natural variation of copy numbers of LTR-RTs. Because sequencing depth differed substantially
between accessions (Minadakis *et al.*, 2023a), we first downsampled all fastq files to the read
408 number of the sample with the lowest number of reads (4.230.721 reads) using the `reformat.sh`
function of BBtools (v 38.75, BBMap Bushnell B., sourceforge.net/projects/bbmap/). Downsampled
410 reads were aligned to the TREP consensus sequences of LTR-RTs and the reference assembly of Bd21
(v 3.0) using BWA-MEM (v 0.7.17-r1188) (Li & Durbin, 2009) with the `-M` and the `-a` options set,
412 hence outputting all alignments found. Coverages of LTR-RTs and the reference assembly were
assessed using `bedtools` (v 2.30.0) (Quinlan & Hall, 2010) `genomecov` with the `-d` and `-split`

414 parameters set. For each of the LTR-RT families, a proxy for copy number was obtained by
normalizing the coverage signals by the coverage of the entire reference assembly and by correcting
416 for the length of the consensus sequences. We favoured this coverage-based approach over an
analysis based on transposon insertion polymorphisms (TIPs) as estimates based on TIPs are
418 reference-dependent and biased by the phylogeny in our study system. We have for instance
previously shown that accessions from the B_East clade harbor significantly less TIPs than accessions
420 from the A_East clade due to the fact that the reference genome Bd21 belongs the B_East clade
(Stritt *et al.*, 2018). In addition, whole-genome *de novo* assembly of 54 *B. distachyon* natural
422 accessions and the subsequent pangenome analysis revealed that non-reference accessions display
large genomic variations (Gordon *et al.*, 2017) which may further bias the estimates of TIP
424 abundances. pCNV raw data were processed using R (v 3.6.3 and 4.0.2) (R Core Team, 2020) in
Rstudio(RStudio Team, 2016).

426 Variation in pCNVs across the 320 natural accessions was visualized with a heatmap drawn
with the heatmap() function natively provided in R version 4.0.2. We computed pairwise genetic
428 distances between accessions with the R package pvclust v 2.2.0 (Suzuki & Shimodaira, 2006). The
resulting tree was used to order accessions phylogenetically on the heatmap. PCAs based on pCNVs
430 were obtained with the R package ggbiplot v 0.55 (Vu, 2011).

To test for an association between pCNVs and environmental variables, we retrieved information
432 about climatic variables at each local site from (Minadakis *et al.*, 2023b). Linear mixed model
analyses where pCNV per LTR-RT family was entered as the response variable, the bioclimatic
434 variables entered separately as fixed factors and the clade of origin as random factors to account for
population structure were ran with the R package lme4 (Bates *et al.*, 2015). The part of the variance
436 explained by the fixed- (marginal R^2) were computed following (Nakagawa & Schielzeth, 2013) and

visualized as bubble plot with the R package ggplot2 (Wickham, 2016). Classical linear models were
438 run in base R.

440 **Plant material, growth conditions and stresses for mobilome-seq**

Brachypodium distachyon natural accessions used in this study comprised Bd21, Bd21-3, Cm18,
442 Cb23, ABR2, Bd29-1 BdTR13c, RON2 and Arn1. Because Pol IV is known to play an important role in
LTR-RT silencing in plants (Stonaker *et al.*, 2009; Ito *et al.*, 2011; Ferrafiat *et al.*, 2019), we also
444 included the sodium azide mutagenized *pol IV* mutant line NaN74 (*Bd nrpd1-1*) (Dalmais *et al.*, 2013;
Böhrer *et al.*, 2020), the T-DNA insertion *pol IV* mutant line JJJ18557 Nr31 (Bragg *et al.*, 2012)
446 *Bd nrpd1-2* (-/-) and a corresponding sibling, outcrossed control line *Bd NRPD1* (+/+) in the
background of the natural accession Bd21-3. For *in vitro* experiments, seeds were soaked for 4 h in
448 tap water and, without damaging the embryo, the lemma was carefully peeled off. Seeds were then
surface-sterilized for 30 seconds in 100% ethanol and immediately rinsed three times with sterile
450 tap water. Surface-sterilized seeds were placed with the embryo facing down and at an angle of
about 30° towards the side, onto solid ½ MS-medium (2.15 g/L MS basal salt without vitamins
452 (Duchefa Biochemie, Haarlem, NL), 0.5 g/L MES-Monohydrate, 10 g/L sucrose, pH 5.8 (KOH), 0.25 %
Phytagel (Sigma-Aldrich, St. Louis, USA) in 'De Wit' culture tubes (Duchefa Biochemie, Haarlem, NL).
454 Plants were grown at 24 °C (day) / 22 °C (night), 16 h light under controlled conditions in an Aralab
600 growth chamber (Rio de Mouro, PT) for 25 to 29 days until the onset of stresses. For salt stress,
456 seedlings were transplanted to solid ½ MS-medium supplied with 300 mM NaCl and grown for five
days at 24/22 °C, 16h light. A solution of sterile-filtrated Glyphosate (Sintagro AG, Härkingen, CH) (20
458 mM, diluted in water) was applied to leaves using a piece of soaked sterile filter paper and plants
were incubated for four days at 24/22 °C, 16h light. Drought stress was induced by uprooting plants

460 from the medium and incubating them for 2:15 h at 24 °C in the light. Before sampling, plants were
allowed to recover for two hours on fresh ½ MS-medium. For the infection with *Magnaporthe oryzae*
462 (rice blast) six isolates (FR13, Mo15-27, 9475-1-3, IK81, M64 and Mo15-19) with spore
concentrations between 130'000- 200'000 K spores per isolate per mL sterile water, supplied with
464 0.2 % Tween 20 were mixed and applied with a cotton swab to plant leaves. Plants were incubated
for 24 h in the dark (24/22 °C) and then grown for another three days at 24/22 °C, 16h light. For heat
466 stress, plants were incubated for 8 h at 42 °C. Before sampling, heat-stressed plants were allowed to
recover for 16 h at 24/22 °C. Cold stress was induced by incubating plants for 24 h at 2 °C on ice at
468 16 h light. Prior to sampling, plants were allowed to recover for two hours at 24 °C in the light. For
submergence stress, two small holes were drilled just above the growth medium and at the top
470 through the wall of the culture tubes. Tubes were then inverted and submerged upside down for 48
h at 24/22 °C, 16h light using a custom rack in a plastic beaker filled with 2.5 liters of 24 °C tap water.
472 In this way, it is possible to submerge plant leaves without the medium coming into contact with the
water. Chemical de-methylation of DNA was conducted according to (Thieme *et al.*, 2017) by
474 germinating and growing plants for 28 days on ½ MS-medium supplied with a mixture of Zebularine
(Sigma-Aldrich, St. Louis, USA) and alpha-amanitin (Sigma-Aldrich, St. Louis, USA). Because the drug
476 treatment severely affected the growth of seedlings, we omitted a treatment of mutant plants and
used reduced concentrations of 20 uM (Zebularine) and 2.5 mg/ml (alpha-amanitin), respectively
478 for all natural accessions.

480 **Mobilome sequencing and validation of eccDNAs**

DNA was extracted using the DNeasy plant kit (Qiagen, Venlo, Netherlands) according to the protocol
482 of the manufacturer. DNA concentration was measured using the Qubit high sensitivity kit

(Invitrogen, Waltham, USA). Mobilome sequencing was performed according to (Lanciano *et al.*,
484 2017) using pooled DNA of two biological replicates per sample. For this, 50 ng of DNA from both
biological replicates were pooled and diluted to a volume of 58 μ L. To enrich eccDNA, DNA was first
486 purified using the GENE CLEAN kit (MP Biomedicals, Santa Ana, USA) according to manufactures
recommendations using 5 μ L glass milk with an elution volume of 35 μ L. Thirty μ L of the eluate were
488 digested using the Plasmid-Safe ATP-dependent DNase (Biosearch Technologies, Hoddesdon, UK) for
17 h at 37 °C. The digestion product was then subjected to an ethanolic precipitation and the
490 precipitated eccDNA amplified using the illustraTempliPhi Amplification Kit (Cytiva, Marlborough,
USA) according to (Lanciano *et al.*, 2017) with an extended incubation time of 65 h at 28 °C. The
492 templphi product was diluted 1:10, quantified using the Qubit high sensitivity kit and 120 ng per
sample were used for library preparation. Sequencing libraries were prepared using the Nextera DNA
494 Flex Library Prep and the Nextera DNA CD Indexes (Illumina, San Diego, USA). Quality of libraries
were assessed using the Tape Station (Agilent Technologies, Santa Clara, USA) with High Sensitivity
496 D1000 screen tapes and concentrations were measured using the Qubit high sensitivity kit. Up to 12
indexed libraries were pooled and sequenced with an Illumina MiSeq sequencer using the MiSeq
498 reagent kit v3 (600 cycles). Raw reads have been uploaded to ENA (accession number PRJEB58186).

The presence of extrachromosomal circular copies of *HOPPLA* (RLC_BdisC024) was validated
500 by an inverse PCR using 7 ng/ μ L total DNA. Input quantities of DNA were controlled using primers
specific to the S-adenosylmethionine decarboxylase (SamDC) gene (Hong *et al.*, 2008). Sequences of
502 primers are listed in **S5 Table**.

504

506 **Analysis of Mobilome-seq**

507 Reads were trimmed using the BBDuk tool of BBtools (BBMap (v 38.75, Bushnell B.,
508 sourceforge.net/projects/bbmap/) with the parameters qtrim = rl and trimq = 20. Reads originating
509 from organelles were removed by aligning reads to the chloroplast genome (NC_011032.1) (Bortiri
510 *et al.*, 2008) and the mitochondrion genome (v 1.0.0) of *B. distachyon* using BWA-MEM (v 0.7.17-
511 r1188) (Li & Durbin, 2009) with the -M parameter set. Unmapped reads were isolated using samtools
512 (v 1.13) (Danecek *et al.*, 2021) view -b -f 4 and bedtools (v 2.30.0) (Quinlan & Hall, 2010) bamtofastq.

513 Organelle-filtered mobilome reads were assembled using the SPAdes genome assembler (v
514 3.13.0) (Prijbelski *et al.*, 2020). From each assembly, the top ten contigs were extracted and jointly
515 aligned to the reference assembly of Bd21 (v 3.0) using BWA-MEM (v 0.7.17-r1188) with the -M
516 parameter set. Bam files were converted into bed files using bedtools (v 2.30.0) bamtoBED with the
517 -split option set and overlapping contigs were merged using bedtools merge with the -o distinct,
518 count, count_distinct and -c 4 parameters set. Assembled, circle-forming regions were annotated
519 with bedtools intersect using the version 3.1 annotation of the reference assembly and the
520 annotation of all full-length LTR-RTs (Stritt *et al.*, 2020) of the reference assembly. Annotated regions
521 were extracted with bedtools getfasta and all sequences longer than 2 kb were isolated using SeqKit
522 seq (v 0.11.0) (Shen *et al.*, 2016). Circle-forming regions that occurred in less than three samples
523 were not included in the analysis.

524 To specifically detect mobilized LTR-RTs, we first extracted all annotated full-length LTR-RTs
525 of the Bd21 reference assembly (Stritt *et al.*, 2020). Using a custom python script, we then merged
526 the last 300 bp of the 3' to the first 300 bp of the 5' LTR to obtain a 'tail-to-head' library containing
527 all annotated full-length LTR-RT copies annotated in Bd21. We then aligned organelle-filtered
528 mobilome reads to the tail-to-head library of LTR-RTs and used bedtools (v 2.30.0) intersect to extract

aligned reads that were spanning the 2-LTR junction and that aligned to at least 5 bp of both LTRs.

530 The coverage of the junction-spanning reads was calculated using deeptools (v 3.5.1) (Ramírez *et al.*,
2016) with the parameters `-bs 1, --ignoreDuplicates --outRawCounts` set. To account for differences
532 in sequencing depth, the obtained coverage for 2-LTR-junction spanning reads was normalized with
the total coverage obtained with bedtools (v 2.30.0) `genomecov` with the `-d` and `-split` parameters
534 set, from the alignments of filtered reads to the reference assembly of Bd21 (v 3.0) obtained from
Phytozome 12 (International Brachypodium Initiative, 2010) generated by BWA-MEM (v 0.7.17-
536 r1188) (Li & Durbin, 2009). To plot the overall activity per family, normalized signals were summed
up for every individual TE family.

538

mRNA-sequencing and small RNA northern blotting

540 Leaves of 4-week-old *B. distachyon* plants were ground in liquid nitrogen and 500 µL of this powder
was subjected to TRIzol extraction following the supplier instructions (Invitrogen, CA, USA). 20 µg of
542 total RNA was treated with DNase I for 30 min., then repurified via phenol-chloroform extraction
and ethanol precipitation. DNase-treated total RNA samples were sent to Fasteris/Genesupport
544 (Plan-les-Ouates, Switzerland), subjected to poly(A)-tail selection, and then aliquoted for library
construction via the Illumina TruSeq Stranded mRNA Library Prep kit. Resulting stranded polyA+
546 RNA-seq (mRNA-seq) libraries were sequenced on an Illumina NovaSeq 6000. The raw paired-end
read data were deposited at the NCBI Gene Expression Omnibus (GEO accession: GSE243693).

548 For the small RNA blot analysis, 200 µg of each total RNA were size-fractionated using the RNeasy
Midi Kit (QIAGEN), as described previously (Böhrer *et al.*, 2020). Low molecular weight (LMW, <200
550 nt) RNAs are not bound by the silica membrane of the columns and were isolated from the collected
flow-through and wash aliquots. LMW RNAs were precipitated overnight using isopropanol.

552 Following a centrifugation step (45 min. at 24000 x g, 4°C) and the removal of the supernatant, the
pellet was washed with 75% ethanol, centrifuged (15 min. at 24000 x g, 4°C), dried at RT for 20 min.
554 then at 65°C for 5 min., and resuspended in 41 µL of DEPC-treated MilliQ water. LMW RNAs were
quantified using a Nanodrop device and 12.3 µg of LMW RNAs from each sample were loaded into
556 the 16% polyacrylamide gel (Böhrer *et al.*, 2020). After running, transfer and UV crosslinking,
membrane was prehybridized in PerfectHyb Plus buffer (Merck, Darmstadt, Germany) at 35°C and
558 then and hybridized at 35°C with the Klenow internally-labeled probe (*HOPPLA*), or with the 5'-end
labeled probe (miR160) (Böhrer *et al.*, 2020). After overnight hybridization, washing was performed
560 at 37°C. Signal detection requires 5-7 days exposure for *HOPPLA* and 1-2 days for miR160.
Oligonucleotide sequences for the probes are listed in **S5 Table**.

562

LTR-RT expression analysis

564 RNA-seq raw reads of *Bd nrpd1-2 (-/-)* and *Bd nrpd1-2 (+/+)* were trimmed for adapters using fastp
(v 0.23.2) (Chen, 2023) with the following options: `--qualified_quality_phred 15 --`
566 `unqualified_percent_limit 4 --n_base_limit 20 --low_complexity_filter --`
`overrepresentation_analysis --correction --detect_adapter_for_pe`. Cleaned reads were then
568 analysed using SalmonTE (v 0.4) (Jeong *et al.*, 2018) to measure global expression of LTR-RTs. LTR-RT
consensus sequences of *B. distachyon* obtained from the TRansposable Elements Platform (TREP,
570 <https://trep-db.uzh.ch/>) were used to generate the custom library for SalmonTE. Default options of
SalmonTE quant and test function were used to quantify expression and to perform statistical
572 analysis. Expression data were plotted using R (v 3.6.3) in RStudio (v 7d165dcf).

574

Motif analysis

576 The consensus sequence of *HOPPLA* was screened for known transcription factor binding sites
obtained from the PlantTFDB (Jin *et al.*, 2017) using FIMO (v 5.1.1) (Grant *et al.*, 2011). To
578 functionally annotate transcription factors that could bind to *HOPPLA*, we used GO-terms of the
Gramene (release 50) database (Tello-Ruiz *et al.*, 2021) downloaded from the platform agriGO (v 2.0)
580 (Tian *et al.*, 2017). Generic, TF specific GO terms (GO:0003700, GO:0006355, GO:0005634,
GO:0003677, GO:0043565, GO:0046983, GO:0003682 GO:0045893) such as ‘positive regulation of
582 DNA-templated transcription’ were removed from the list of GO terms as they would interfere with
the downstream analysis. The remaining GO terms of transcription factors potentially binding to
584 *HOPPLA* were visualized with REVIGO (Supek *et al.*, 2011) using the ‘SimRel’ semantic similarity
measure, the option ‘small’ and the GO terms of the *Oryza sativa* Japonica Group. The total number
586 of occurrences of individual GO terms was taken into account with the option ‘higher value is better’.
GO terms occurring more than five times were labelled in the plots. As a proof of concept, we
588 followed the exact same approach using the sequence of one of the most active *ONSEN* copies
(AT1G11265) and the GO terms of *Arabidopsis thaliana*.

590

Detection of novel *HOPPLA* insertions in *Bd nrpd1-2 (-/-)*

592 DNA of adult plants was extracted using the DNeasy plant kit (Qiagen, Venlo, Netherlands) according
to the protocol of the manufacturer subjected to whole genome sequencing. Reads were trimmed
594 using the BBDuk tool of BBtools (BBMap (v 38.75, Bushnell B., sourceforge.net/projects/bbmap/)
with the parameters qtrim = rl and trimq = 20. Trimmed reads were aligned to the reference
596 assembly of Bd21 (v 3.0) using BWA-MEM (v 0.7.17-r1188) (Li & Durbin, 2009) with the -M
parameter set. Samtools (v 1.13) (Danecek *et al.*, 2021) was used to obtain sorted and indexed bam

598 files. TIPs were detected with detettore (v 2.0.3) (<https://github.com/cstritt/detettore>) with the
options `-require_split, -q 30` and using the consensus sequences of LTR-RTs of *B. distachyon* (TREP-
600 database) and the annotation of all full-length LTR-RTs of Bd21 (Stritt *et al.*, 2020). Because both
Bd nrpd1-2 lines were in the Bd21-3 background we were able to exclude all Bd21-3 specific TIPs by
602 removing those insertions that were detected in multiple individuals with more than one genetic
background. Remaining TIPs were manually curated using the genome browser IGV (v 2.15.4.12)
604 (Robinson *et al.*, 2011). HOPPLA TIPs were visualized with JBrowse 2 (v 2.6.1) (Diesh *et al.*, 2023) .
Raw genomic reads of the re-sequencing of *Bd nrpd1-2* (-/-), *Bd NRPD1* (+/+) and Bd21-3 have been
606 uploaded to ENA (accession number PRJEB64053).

608 **GWAS for pCNV**

GEMMA 0.98.5 (Zhou & Stephens, 2012) was used to test for associations between SNPs
610 (Minadakis *et al.*, 2023a) and the LTR-RT families pCVNs, while correcting for population
structure (Stritt *et al.*, 2022; Minadakis *et al.*, 2023a). A centered relatedness matrix was first
612 created with the option `-gk 1` and association tests were performed using the option `-maf 0.05`
to exclude rare alleles, and the default SNP missingness threshold applied by GEMMA that
614 excludes SNPs with missing data in more than 5% of the accessions. We selected 20 kb genomic
regions with a 10 kb overlap that contained at least two SNPs above the False Discovery Rate of
616 0.05 or Bonferroni correction threshold as candidate region using the R package *rehh* (v 3.2.2)
(Gautier & Vitalis, 2012). Genes overlapping with candidate regions were selected with the
618 BEDTOOLS (v 2.26.0) (Quinlan & Hall, 2010) `intersect` command using the version 3.1 of the *B.*
distachyon annotation file (<https://phytozome-next.jgi.doe.gov>) that are contained in the
620 significant regions. The UpSetR (Conway *et al.*, 2017) R package was used to visualize the

intersections of significant genes between the variables. Protein constituents of the Pol IV and Pol V
622 enzymes (see **S4 Table**) were downloaded from the plant RNA polymerase database
<http://rna.polymerase.eu/>.

624

Acknowledgements and author contributions

626 We thank the Genetic Diversity Center Zürich for providing the infrastructure for the mobilome-seq.
M.T. and A.C.R. conceived the study and planned experiments. M.T., B.K., K.R., C.H., B.R. and M.B.
628 generated data. M.T., A.C.R., N.M. and W.X. analysed data. D.L., J.V. and R.S. provided mutagenized
germplasm and online TILLING/T-DNA browsing tools. C.H., B.R. and T.B. identified and screened for
630 the *pol IV* null mutants. M.T. and A.C.R. wrote the paper with contributions from T.B., C.S. All authors
read the manuscript. The authors declare that they have no conflict of interest. Professional editing
632 performed by Emanuelle Botté.

634 Funding

University of Zurich Research Priority Programs (URPP) Evolution in Action to M.T. and A.C.R.;
636 Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung. Grant Number:
31003A_182785 to A.C.R., W.X., N.M. and B.K.; Schweizerischer Nationalfonds zur Förderung der
638 Wissenschaftlichen Forschung. Grant Number: PZ00P3_154724 to C.S.; T.B. was supported by the
Interdisciplinary Thematic Institute IMCBio (ITI 2021-2028 program), including funds from IdEx
640 Unistra (ANR-10-IDEX-0002), SFRI-STRAT'US (ANR 20-SFRI-0012) and EUR IMCBio (ANR-17-EURE-
0023) in the framework of the French Investments for the Future Program.

642 References

- 644 **Baduel P, Leduque B, Ignace A, Gy I, Gil J, Loudet O, Colot V, Quadrana L. 2021.** Genetic and environmental modulation of transposition shapes the evolutionary potential of *Arabidopsis thaliana*. *Genome Biology* **22**: 138.
- 646 **Bajus M, Macko-Podgórní A, Grzebelus D, Baránek M. 2022.** A review of strategies used to identify transposition events in plant genomes. *Frontiers in Plant Science* **13**: 1080993.
- 648 **Bates D, Mächler M, Bolker BM, Walker SC. 2015.** Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67**: 1–48.
- 650 **Benoit M, Drost HG, Catoni M, Gouil Q, Lopez-Gomollon S, Baulcombe D, Paszkowski J. 2019.** Environmental and epigenetic regulation of Rider retrotransposons in tomato. *PLoS Genetics* **15**:
652 e1008370.
- 654 **Böhrer M, Rymen B, Himer C, Gerbaud A, Pflieger D, Laudencia-Chingcuanco D, Cartwright
656 A, Vogel J, Sibout R, Blevins T. 2020.** Integrated Genome-Scale Analysis and Northern Blot Detection of Retrotransposon siRNAs Across Plant Species. *Methods in molecular biology (Clifton, N.J.)* **2166**: 387–411.
- 658 **Bortiri E, Coleman-Derr D, Lazo GR, Anderson OD, Gu YQ. 2008.** The complete chloroplast genome sequence of *Brachypodium distachyon*: Sequence comparison and phylogenetic analysis of eight grass plastomes. *BMC Research Notes* **1**: 61.
- 660 **Bourgeois Y, Boissinot S. 2019.** On the Population Dynamics of Junk: A Review on the Population Genomics of Transposable Elements. *Genes* **10**.
- 662 **Bourgeois Y, Ruggiero RP, Hariyani I, Boissinot S. 2020.** Disentangling the determinants of transposable elements dynamics in vertebrate genomes using empirical evidences and simulations.
664 *PLoS genetics* **16**.
- 666 **Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS, et al. 2018.** Ten things you should know about transposable elements. *Genome Biology* **19**:1 19: 1–12.
- 668 **Bragg JN, Wu J, Gordon SP, Guttman ME, Thilmony R, Lazo GR, Gu YQ, Vogel JP. 2012.** Generation and Characterization of the Western Regional Research Center *Brachypodium* T-DNA
670 Insertional Mutant Collection. *PLOS ONE* **7**: e41916.
- 672 **Butelli E, Licciardello C, Zhang Y, Liu J, Mackay S, Bailey P, Reforgiato-Recupero G, Martin C. 2012.** Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *Plant Cell* **24**: 1242–1255.
- 674 **Cavrak V V., Lettner N, Jamge S, Kosarewicz A, Bayer LM, Mittelsten Scheid O. 2014.** How a Retrotransposon Exploits the Plant's Heat Stress Response for Its Activation. *PLOS Genetics* **10**:
676 e1004115.
- 678 **Chen S. 2023.** Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta* **2**: e107.
- 680 **Cho J, Benoit M, Catoni M, Drost H-G, Brestovitsky A, Oosterbeek M, Paszkowski J. 2019.** Sensitive detection of pre-integration intermediates of long terminal repeat retrotransposons in crop plants. *Nature Plants* **5**: 26–33.

- 682 **Conway JR, Lex A, Gehlenborg N. 2017.** UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**: 2938–2940.
- 684 **Dalmais M, Antelme S, Ho-Yue-Kuang S, Wang Y, Darracq O, d’Yvoire MB, Cézard L, Légée F, Blondet E, Oria N, et al. 2013.** A TILLING Platform for Functional Genomics in *Brachypodium distachyon*. *PLoS ONE* **8**: 65503.
- 686 **Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM. 2021.** Twelve years of SAMtools and BCFtools. *GigaScience* **10**: 1–4.
- 688 **Diesh C, Stevens GJ, Xie P, De Jesus Martinez T, Hershberg EA, Leung A, Guo E, Dider S, Zhang J, Bridge C, et al. 2023.** JBrowse 2: a modular genome browser with views of synteny and structural variation. *Genome Biology* **24**: 1–21.
- 690 **Dubin MJ, Mittelsten Scheid O, Becker C. 2018.** Transposons: a blessing curse. *Current Opinion in Plant Biology* **42**: 23–29.
- 692 **Ferrafiat L, Pflieger D, Singh J, Thieme M, Böhrer M, Himber C, Gerbaud A, Bucher E, Pikaard CS, Blevins T. 2019.** The NRPD1 N-terminus contains a Pol IV-specific motif that is critical for genome surveillance in *Arabidopsis*. *Nucleic acids research* **47**: 9037–9052.
- 696 **Flavell AJ. 1984.** Role of reverse transcription in the generation of extrachromosomal copia mobile genetic elements. *Nature* **310**: 514–516.
- 698 **Flavell AJ, Ish-horowicz D. 1981.** Extrachromosomal circular copies of the eukaryotic transposable element copia in cultured *Drosophila* cells. *Nature* **1981** 292:5824 **292**: 591–595.
- 700 **Garfinkel DJ, Stefanisko KM, Nyswaner KM, Moore SP, Oh J, Hughes SH. 2006.** Retrotransposon Suicide: Formation of Ty1 Circles and Autointegration via a Central DNA Flap. *Journal of Virology* **80**: 11920.
- 702 **Gaubatz JW. 1990.** Extrachromosomal circular DNAs and genomic sequence plasticity in eukaryotic cells *. *Mutation Research* **237**: 271–292.
- 704 **Gautier M, Vitalis R. 2012.** rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics (Oxford, England)* **28**: 1176–1177.
- 706 **Gilbert C, Peccoud J, Cordaux R. 2021.** Transposable Elements and the Evolution of Insects. <https://doi.org/10.1146/annurev-ento-070720-074650> **66**: 355–372.
- 708 **Gordon SP, Contreras-Moreira B, Levy JJ, Djamei A, Czedik-Eysenberg A, Tartaglio VS, Session A, Martin J, Cartwright A, Katz A, et al. 2020.** Gradual polyploid genome evolution revealed by pan-genomic analysis of *Brachypodium hybridum* and its diploid progenitors. *Nature Communications* **11**.
- 710 **Gordon SP, Contreras-Moreira B, Woods DP, Des Marais DL, Burgess D, Shu S, Stritt C, Roulin AC, Schackwitz W, Tyler L, et al. 2017.** Extensive gene content variation in the *Brachypodium distachyon* pan-genome correlates with population structure. *Nature Communications* **8**: 1–13.
- 714 **Grandbastien MA. 2015.** LTR retrotransposons, handy hitchhikers of plant regulation and stress response. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1849**: 403–416.
- 716 **Grant CE, Bailey TL, Noble WS. 2011.** FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**: 1017–1018.
- 718 **Hasterok R, Catalan P, Hazen SP, Roulin AC, Vogel JP, Wang K, Mur LAJ. 2022.** *Brachypodium*: 20 years as a grass biology model system; the way forward? *Trends in Plant Science* **27**: 1002–1016.

- 724 **Hawkins JS, Kim HR, Nason JD, Wing RA, Wendel JF. 2006.** Differential lineage-specific
726 amplification of transposable elements is responsible for genome size variation in *Gossypium*.
Genome Research **16**: 1252.
- Hirochika H, Sugimoto K, Otsuki Y, Tsugawa H, Kanda M. 1996.** Retrotransposons of rice
728 involved in mutations induced by tissue culture. *Proceedings of the National Academy of Sciences of
the United States of America* **93**: 7783–7788.
- 730 **Hong SY, Seo PJ, Yang MS, Xiang F, Park CM. 2008.** Exploring valid reference genes for gene
expression studies in *Brachypodium distachyon* by real-time PCR. *BMC Plant Biology* **8**: 112.
- 732 **Horvath R, Minadakis N, Bourgeois Y, Roulin AC. 2023.** The evolution of transposable elements
734 in *Brachypodium distachyon* is governed by purifying selection, while neutral and adaptive processes
play a minor role. *bioRxiv*: 2023.09.15.557873.
- International Brachypodium Initiative. 2010.** Genome sequencing and analysis of the model grass
736 *Brachypodium distachyon*. *Nature* **463**: 763–768.
- Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, Paszkowski J. 2011.** An siRNA pathway
738 prevents transgenerational retrotransposition in plants subjected to stress. *Nature* **472**: 115–120.
- Jeong H-H, Yalamanchili HK, Guo C, Shulman JM, Liu Z. 2018.** An ultra-fast and scalable
740 quantification pipeline for transposable elements from next generation sequencing data.
Biocomputing 2018: 168–179.
- 742 **Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, Gao G. 2017.** PlantTFDB 4.0: Toward a central
744 hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research* **45**: D1040–
D1045.
- Kawakatsu T, Huang SS, Jupe F, Sasaki E, Schmitz RJ, Urich MA, Castanon R, Nery JR,
746 Barragan C, He Y, et al. 2016.** Epigenomic Diversity in a Global Collection of *Arabidopsis thaliana*
Accessions. *Cell* **166**: 492–505.
- 748 **Kidwell MG, Lisch D. 1997.** Transposable elements as sources of variation in animals and plants.
Proceedings of the National Academy of Sciences of the United States of America **94**: 7704–7711.
- 750 **Lanciano S, Carpentier MC, Llauro C, Jobet E, Robakowska-Hyzorek D, Lasserre E,
Ghesquiere A, Panaud O, Mirouze M. 2017.** Sequencing the extrachromosomal circular mobilome
752 reveals retrotransposon activity in plants. *PLoS Genet* **13**: e1006630.
- Lanciano S, Cristofari G. 2020.** Measuring and interpreting transposable element expression.
754 *Nature Reviews Genetics* **21**: 721–736.
- Lanciano S, Mirouze M. 2018.** Transposable elements: all mobile, all different, some stress
756 responsive, some adaptive? *Current Opinion in Genetics & Development* **49**: 106–114.
- Langmüller AM, Langmüller L, Nolte V, Dolezal M, Schi Otterer C. 2023.** The genomic
758 distribution of transposable elements is driven by spatially variable purifying selection. *Nucleic Acids
Research* **51**: 9203–9213.
- 760 **Lee SC, Ernst E, Berube B, Borges F, Parent JS, Ledon P, Schorn A, Martienssen RA. 2020.**
762 *Arabidopsis* retrotransposon virus-like particles and their regulation by epigenetically activated small
RNA. *Genome Research* **30**: 576–588.
- Leyser O. 2018.** Auxin Signaling. *Plant Physiology* **176**: 465–479.
- 764 **Li H, Durbin R. 2009.** Fast and accurate short read alignment with Burrows-Wheeler transform.
Bioinformatics **25**: 1754–1760.

- 766 **Lisch D. 2013.** How important are transposons for plant evolution? *Nat Rev Genet* **14**: 49–61.
- Liu P, Cuerda-Gil D, Shahid S, Slotkin RK. 2022.** The Epigenetic Control of the Transposable
768 Element Life Cycle in Plant Genomes and Beyond. <https://doi.org/10.1146/annurev-genet-072920-015534> **56**: 63–87.
- 770 **Liu P, Nie WF, Xiong X, Wang Y, Jiang Y, Huang P, Lin X, Qin G, Huang H, Niu Q, et al. 2021a.**
772 A novel protein complex that regulates active DNA demethylation in Arabidopsis. *Journal of Integrative Plant Biology* **63**: 772–786.
- Liu P, Nie WF, Xiong X, Wang Y, Jiang Y, Huang P, Lin X, Qin G, Huang H, Niu Q, et al. 2021b.**
774 A novel protein complex that regulates active DNA demethylation in Arabidopsis. *Journal of Integrative Plant Biology* **63**: 772–786.
- 776 **Makarevitch I, Waters AJ, West PT, Stitzer M, Hirsch CN, Ross-Ibarra J, Springer NM. 2015.**
778 Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLoS Genet* **11**: e1004915.
- McClintock B. 1984.** The significance of responses of the genome to challenge. *Science* **226**: 792–
780 801.
- Minadakis N, Williams H, Horvath R, Caković D, Stritt C, Thieme M, Bourgeois Y, Roulin AC. 2023a.**
782 New resources for environmental genomics in the wild Mediterranean grass *B. distachyon*. *bioRxiv*: 2023.06.01.543285.
- 784 **Minadakis N, Williams H, Horvath R, Caković D, Stritt C, Thieme M, Bourgeois Y, Roulin AC. 2023b.**
786 The demographic history of the wild crop relative *Brachypodium distachyon* is shaped by distinct past and present ecological niches. *bioRxiv*: 2023.06.01.543285.
- Mirouze M, Reinders J, Bucher E, Nishimura T, Schneeberger K, Ossowski S, Cao J, Weigel D, Paszkowski J, Mathieu O. 2009.** Selective epigenetic control of retrotransposition in Arabidopsis. *Nature* **461**: 427–430.
- 790 **Mirouze M, Vitte C. 2014.** Transposable elements, a treasure trove to decipher epigenetic variation: insights from Arabidopsis and crop epigenomes. *J Exp Bot* **65**: 2801–2812.
- 792 **Miura A, Yonebayashi S, Watanabe K, Toyama T, Shimada H, Kakutani T. 2001.** Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. *Nature* **411**: 212–214.
- 794 **Muszewska A, Steczkiewicz K, Stepniewska-Dziubinska M, Ginalska K. 2019.** Transposable elements contribute to fungal genes and impact fungal lifestyle. *Scientific Reports* **9**: 1–10.
- 796 **Nakagawa S, Schielzeth H. 2013.** A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods in Ecology and Evolution* **4**: 133–142.
- 798 **Negi P, Rai AN, Suprasanna P. 2016.** Moving through the stressed genome: Emerging regulatory roles for transposons in plant stress response. *Frontiers in Plant Science* **7**: 1448.
- 800 **Piegu B, Guyot R, Picault N, Roulin A, Sanyal A, Kim H, Collura K, Brar DS, Jackson S, Wing RA, et al. 2006.** Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Res* **16**: 1262–
802 1269.
- 804 **Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020.** Using SPAdes De Novo Assembler. *Current Protocols in Bioinformatics* **70**: e102.
- 806 **Quadrana L, Silveira AB, Mayhew GF, Leblanc C, Martienssen RA, Jeddloh JA, Colot V. 2016.** The Arabidopsis thaliana mobilome and its impact at the species level.

- 808 **Quinlan AR, Hall IM. 2010.** BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.
- 810 **R Core Team. 2020.** R: A Language and Environment for Statistical Computing.
- Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke**
812 **T. 2016.** deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research* **44**: W160–W165.
- 814 **Raúl C, Noemia M-D, Sonal G, Michael P, M. CJ. 2023.** Transposons are important contributors to gene expression variability under selection in rice populations. *eLife* **12**.
- 816 **Ream TS, Haag JR, Wierzbicki AT, Nicora CD, Norbeck AD, Zhu JK, Hagen G, Guilfoyle TJ, Pasa-Tolic L, Pikaard CS. 2009.** Subunit compositions of the RNA-silencing enzymes Pol IV and
818 Pol V reveal their origins as specialized forms of RNA polymerase II. *Mol Cell* **33**: 192–203.
- Rey O, Danchin E, Mirouze M, Loot C, Blanchet S. 2016.** Adaptation to Global Change: A
820 Transposable Element-Epigenetics Perspective. *Trends in Ecology and Evolution* **31**: 514–526.
- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP.**
822 **2011.** Integrative genomics viewer. *Nature Publishing Group*.
- Roquis D, Robertson M, Yu L, Thieme M, Julkowska M, Bucher E. 2021.** Genomic impact of
824 stress-induced transposable element mobility in Arabidopsis. *Nucleic Acids Research*.
- RStudio Team. 2016.** RStudio: Integrated Development Environment for R.
- 826 **Rymen B, Ferrafiat L, Blevins T. 2020.** Non-coding RNA polymerases that silence transposable elements and reprogram gene expression in plants. *Transcription* **11**: 172–191.
- 828 **Saze H, Tsugane K, Kanno T, Nishimura T. 2012.** DNA Methylation in Plants: Relationship to Small RNAs and Histone Modifications, and Functions in Transposon Inactivation. *Plant and Cell*
830 *Physiology* **53**: 766–784.
- Schulman AH. 2013.** Retrotransposon replication in plants. *Current Opinion in Virology* **3**: 604–614.
- 832 **Senft AD, Macfarlan TS. 2021.** Transposable elements shape the evolution of mammalian development. *Nature Reviews Genetics* **22**: 691–711.
- 834 **Shen W, Le S, Li Y, Hu F. 2016.** SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation.
- 836 **Sigman MJ, Slotkin RK. 2016.** The First Rule of Plant Transposable Element Silencing: Location, Location, Location. *The Plant Cell* **28**: 304–313.
- 838 **Silander K, Saarela J. 2008.** Whole genome amplification with Phi29 DNA polymerase to enable genetic or genomic analysis of samples of low DNA yield. *Methods in molecular biology (Clifton, N.J.)* **439**: 1–18.
- 840 **Skalska A, Stritt C, Wyler M, Williams HW, Vickers M, Han J, Tuna M, Tuna GS, Susek K, Swain M, et al. 2020.** Genetic and methylome variation in Turkish brachypodium distachyon accessions differentiate two geographically distinct subpopulations. *International Journal of*
844 *Molecular Sciences* **21**: 1–17.
- Smith CA, Vinograd J. 1972.** Small polydisperse circular DNA of HeLa cells. *Journal of molecular biology* **69**.
- 846 **Stitzer MC, Anderson SN, Springer NM, RossIbarra J. 2021.** The genomic ecosystem of
848 transposable elements in maize. *PLoS Genetics* **17**: e1009768.

- 850 **Stonaker JL, Lim JP, Erhard KF, Hollick JB. 2009.** Diversity of Pol IV Function Is Defined by
Mutations at the Maize *rrm7* Locus. *PLOS Genetics* **5**: e1000706.
- 852 **Stritt C, Gimmi EL, Wyler M, Bakali AH, Skalska A, Hasterok R, Mur LAJ, Pecchioni N,
Roulin AC. 2022.** Migration without interbreeding: Evolutionary history of a highly selfing
Mediterranean grass inferred from whole genomes. *Molecular Ecology* **31**: 70–85.
- 854 **Stritt C, Gordon SP, Wicker T, Vogel JP, Roulin AC. 2018.** Recent Activity in Expanding
Populations and Purifying Selection Have Shaped Transposable Element Landscapes across Natural
856 Accessions of the Mediterranean Grass *Brachypodium distachyon*. *Genome Biology and Evolution*
10: 304–318.
- 858 **Stritt C, Thieme M, Roulin AC. 2021.** Rare transposable elements challenge the prevailing view of
transposition dynamics in plants. *American journal of botany* **108**: 1310–1314.
- 860 **Stritt C, Wyler M, Gimmi EL, Pippel M, Roulin AC. 2020.** Diversity, dynamics and effects of long
terminal repeat retrotransposons in the model grass *Brachypodium distachyon*. *New Phytologist* **227**:
862 1736–1748.
- 864 **Stuart T, Eichten SR, Cahn J, Karpievitch Y V, Borevitz JO, Lister R. 2016.** Population scale
mapping of transposable element diversity reveals links to gene regulation and epigenomic variation.
Elife **5**.
- 866 **Supek F, Bošnjak M, Škunca N, Šmuc T. 2011.** REVIGO Summarizes and Visualizes Long Lists
of Gene Ontology Terms (C Gibas, Ed.). *PLoS ONE* **6**: e21800.
- 868 **Suzuki R, Shimodaira H. 2006.** Pvcust: an R package for assessing the uncertainty in hierarchical
clustering. *Bioinformatics (Oxford, England)* **22**: 1540–1542.
- 870 **Tanskanen JA, Sabot F, Vicient C, Schulman AH. 2007.** Life without GAG: the BARE-2
retrotransposon as a parasite’s parasite. *Gene* **390**: 166–174.
- 872 **Tello-Ruiz MK, Naithani S, Gupta P, Olson A, Wei S, Preece J, Jiao Y, Wang B, Chougule K,
Garg P, et al. 2021.** Gramene 2021: Harnessing the power of comparative genomics and pathways
874 for plant research. *Nucleic Acids Research* **49**: D1452–D1463.
- 876 **Thieme M, Brêchet A, Bourgeois Y, Keller B, Bucher E, Roulin AC. 2022.** Experimentally heat-
induced transposition increases drought tolerance in *Arabidopsis thaliana*. *New Phytologist* **236**: 182–
194.
- 878 **Thieme M, Lanciano S, Balzergue S, Daccord N, Mirouze M, Bucher E. 2017.** Inhibition of RNA
polymerase II allows controlled mobilisation of retrotransposons for plant breeding. *Genome Biol* **18**:
880 134.
- 882 **Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, Xu W, Su Z. 2017.** AgriGO v2.0: A GO analysis toolkit
for the agricultural community, 2017 update. *Nucleic Acids Research* **45**: W122–W129.
- 884 **Tittel-Elmer M, Bucher E, Broger L, Mathieu O, Paszkowski J, Vaillant I. 2010.** Stress-Induced
Activation of Heterochromatic Transcription. *PLOS Genetics* **6**: e1001175.
- 886 **Tsukahara S, Kobayashi A, Kawabe A, Mathieu O, Miura A, Kakutani T. 2009.** Bursts of
retrotransposition reproduced in *Arabidopsis*. *Nature* **461**: 423–426.
- 888 **Uzunović J, Josephs EB, Stinchcombe JR, Wright SI, Parsch J. 2019.** Transposable Elements Are
Important Contributors to Standing Variation in Gene Expression in *Capsella Grandiflora*. *Molecular
Biology and Evolution* **36**: 1734–1745.

- 890 **Venner S, Feschotte C, Biéumont C. 2009.** Dynamics of transposable elements: towards a community
ecology of the genome. *Trends in Genetics* **25**: 317–323.
- 892 **Vitte C, Fustier MA, Alix K, Tenaillon MI. 2014.** The bright side of transposons in crop evolution.
Briefings in Functional Genomics and Proteomics **13**: 276–295.
- 894 **Vu VQ. 2011.** ggbiplot: A ggplot2 based biplot. R package version 0.55.
- Wang D, Zheng Z, Li Y, Hu H, Wang Z, Du X, Zhang S, Zhu M, Dong L, Ren G, et al. 2021.**
896 Which factors contribute most to genome size variation within angiosperms? *Ecology and Evolution*
11: 2660–2668.
- 898 **Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P,
Morgante M, Panaud O, et al. 2007.** A unified classification system for eukaryotic transposable
900 elements. *Nature Reviews Genetics* **8**: 973–982.
- Wickham H. 2016.** ggplot2 Elegant Graphics for Data Analysis. *Use R! series*: 211.
- 902 **Xu L, Yuan K, Yuan M, Meng X, Chen M, Wu J, Li J, Qi Y. 2020.** Regulation of Rice Tillering
by RNA-Directed DNA Methylation at Miniature Inverted-Repeat Transposable Elements. *Molecular*
904 *plant* **13**: 851–863.
- Yang F, Su W, Chung OW, Tracy L, Wang L, Ramsden DA, Zhang ZZ. 2023a.** Retrotransposons
906 hijack alt-EJ for DNA replication and eccDNA biogenesis. *Nature* **2023**: 1–8.
- Yang L-L, Zhang X-Y, Wang L-Y, Li Y-G, Li X-T, Yang Y, Su Q, Chen N, Zhang Y-L, Li N, et**
908 **al. 2023b.** Lineage-specific amplification and epigenetic regulation of LTR-retrotransposons
contribute to the structure, evolution, and function of Fabaceae species. *BMC Genomics* **2023** *24*:1
910 **24**: 1–15.
- Zhang H, Lang Z, Zhu JK. 2018.** Dynamics and function of DNA methylation in plants. *Nature*
912 *Reviews Molecular Cell Biology* **2018** *19*:8 **19**: 489–506.
- Zhang Y, Li Z, Liu J, Zhang Y, Ye L, Peng Y, Wang H, Diao H, Ma Y, Wang M, et al. 2022.**
914 Transposable elements orchestrate subgenome-convergent and -divergent transcription in common
wheat. *Nature Communications* **13**.
- 916 **Zhong X, Du J, Hale CJ, Gallego-Bartolome J, Feng S, Vashisht AA, Chory J, Wohlschlegel JA,
Patel DJ, Jacobsen SE. 2014.** Molecular Mechanism of Action of Plant DRM De Novo DNA
918 Methyltransferases. *Cell* **157**: 1050–1060.
- Zhou X, Stephens M. 2012.** Genome-wide efficient mixed-model analysis for association studies.
920 *Nature Genetics* **2012** *44*:7 **44**: 821–824.

922

924

926

Supporting Information

928

930

932

934

936



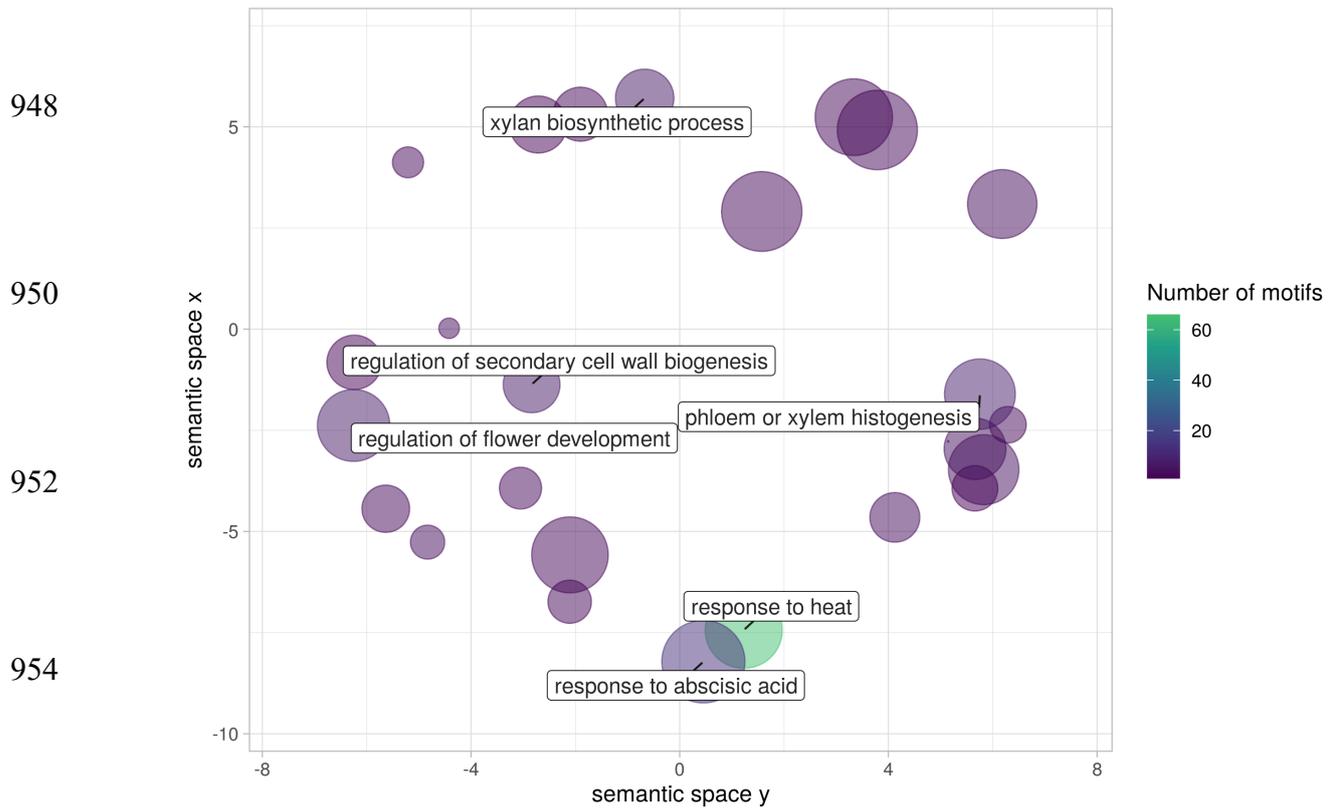
938 **S1 Fig. Correlation of bioclimatic variables with pCNV when not correcting for population structure.** Colors and sizes of bubbles show the part of the variance (R^2) explained by the bioclimatic variables in %

940

942

944

946



956 **S2 Fig. TFs binding to *ONSEN* are heat-stress inducible.** GO-enrichment analysis of transcription factors for which
957 binding sites have been detected in AT1G11265, a member of the heat-responsive *ONSEN* (*ATCOPIA78*) LTR-RT family in
958 *A. thaliana*. Colors indicate number of TF-binding sites found. GO terms that occur at least six times are highlighted in
959 the plot. All GO-terms and their number of occurrences is listed in **S1 Table**.

960

962



964 **S3 Fig**



966

S4 Fig

968

970



972

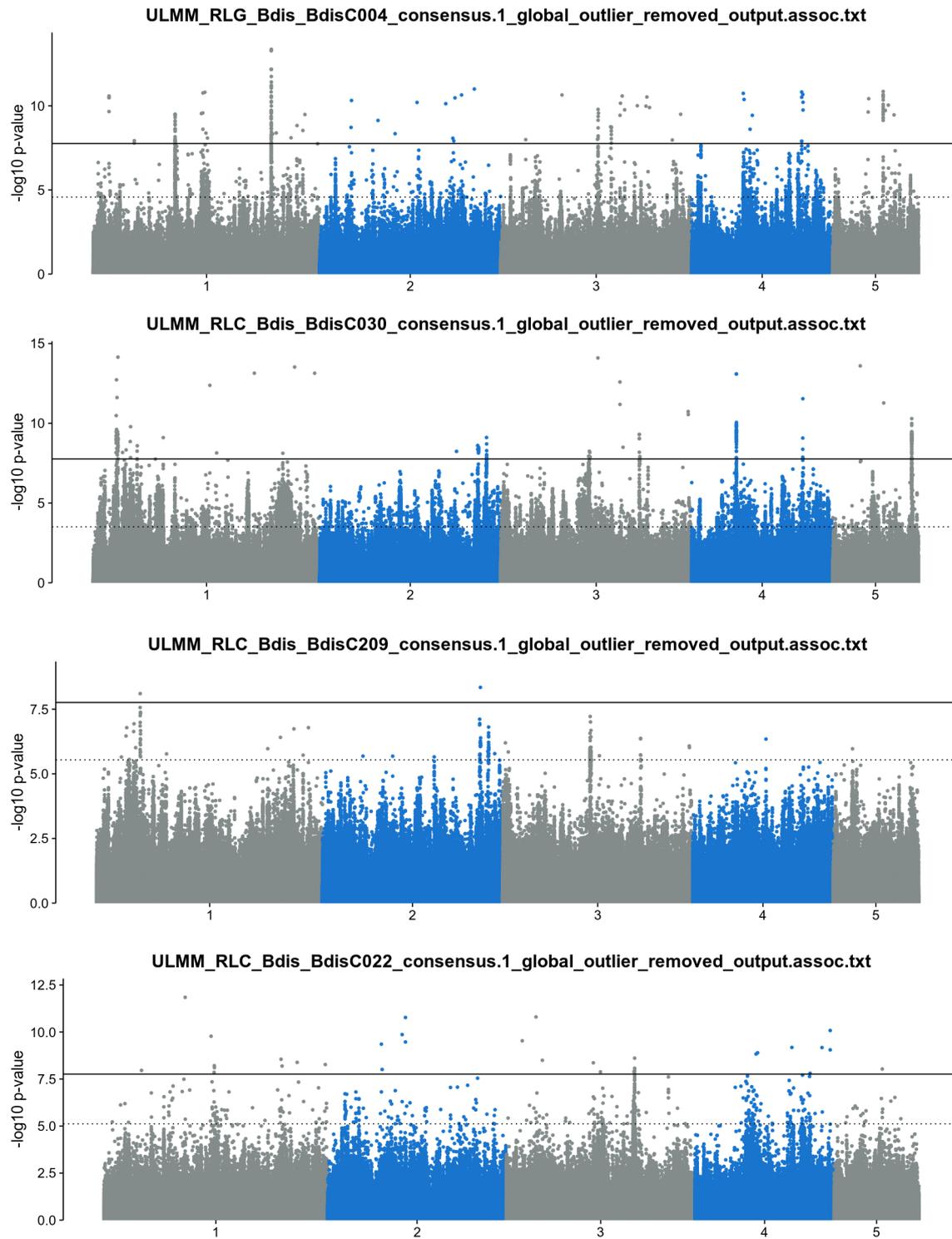
974 **S3-5 Figs. TIPS of HOPPLA in one of the resequenced *Bd nrpd1-2 (-/-)* plants.** JBrowse screenshot of the insertion site
976 in *Bd nrpd1-2 (-/-)* (bottom) compared to the Bd21-3 wt (top). The target side duplication (TSD) is annotated and soft
clipped parts of reads are coloured.

978

980

982

984



986 **S6 Fig. Manhattan plots of the GWASs of pCNVs of four recently active LTR-RT families.** From top: RLG_BdisC004,
988 RLC_BdisC030 RLC_BdisC209 and RLC_BdisC022. The two significance levels, false discovery rate < 0.05 (dashed line)
and Bonferroni correction (solid line) are depicted

988

- 990 **S1 Table. REVIGO output of the processing of GO terms of transcription factors for which binding sites have been detected in the *HOPPLA* and *ONSEN* consensus sequences.**
- 992 **S2 Table. Normalized pCNV of LTR-RTs and bioclimatic variables of the 320 natural accessions of *B. distachyon***
- 994 **S3 Table. Gene list of pCNV GWAS with different levels of significance (FDR < 0.05, BC) and window sizes (20 kb, 50 kb)**
- 996 **S4 Table. Components of the Pol IV and Pol IV holoenzymes in *B. distachyon***
- S5 Table. Sequences of oligos used in this study**