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

Michael Thieme[1*], Nikolaos Minadakis[1], Christophe Himber[2], Bettina Keller[1], Wenbo Xu[1], Kinga Rutowicz[1], Calvin Matteoli[2], Marcel Böhrer[2], Bart Rymen[2], Debbie Laudencia-Chingcuanco[3], John Vogel[4], Richard Sibout[5], Christoph Stritt[6], Todd Blevins[2] and Anne C. Roulin[1*]

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

[2] Institut de Biologie Moléculaire des Plantes, Centre National de la Recherche Scientifique, Université de Strasbourg, Strasbourg, France.

[3] United States Department of Agriculture Agricultural Research Service Western Regional Research Center, Albany, California, United States of America

[4] United States Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, California, United States of America

[5] Institut National de la Recherche Agronomique Unité BIA- 1268 Biopolymères Interactions Assemblages Equipe Paroi Végétale et Polymères Pariétaux (PVPP), Nantes, France.

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

*** Corresponding authors**: michael.thieme@botinst.uzh.ch, anne.roulin@botinst.uzh.ch

Abstract

Long terminal repeat retrotransposons (LTR-RTs) are powerful mutagens regarded as a major source of genetic novelty and important drivers of evolution. Yet, the uncontrolled and potentially selfish proliferation of LTR-RTs can lead to deleterious mutations and genome instability, with large fitness costs for their host. While population genomics data suggest that an ongoing LTR-RT mobility is common in many species, the understanding of their dual role in evolution is limited. Here, we harness the genetic diversity of 320 sequenced natural accessions of the Mediterranean grass *Brachypodium distachyon* to characterize how genetic and environmental factors influence plant LTR-RT dynamics in the wild. When combining a coverage-based approach to estimate global LTR-RT 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 transpositional 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.

Author summary

Long terminal repeat retrotransposons (LTR-RTs) are major components of plant genomes. Their 'copy-and-paste' replication mechanism allows them to rapidly increase in copy number, with potentially negative effects on host fitness. On the other hand, because they can rewire transcriptional networks and alter phenotypes, their mobility is an important driver of evolution. Ever since their discovery, LTR-RT activity has been linked to stress exposure, suggesting that LTR-RTs modulate the pace of evolution in response to the environment. In this study, we test this hypothesis by harnessing the genetic variation in a set of 320 natural accessions of the Mediterranean grass *Brachypodium distachyon* originating from diverse habitats. We find little evidence for the importance of stresses in activating *B. distachyon* LTR-RTs. Instead, we show that the loss of RNA polymerase IV, a component of plant retrotransposon silencing, leads to the activation and transposition of an LTR-RT family that we name *HOPPLA*. *HOPPLA* is the first LTR-RT family in *B. distachyon* shown to transpose in real-time. These findings open up new avenues for studying retrotransposon-mediated evolution in this close relative of staple crops, such as rice and wheat.

Introduction

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 and can constitute more than 80 % of the genome [1]. In addition to playing a major role in genome size variation (e.g. [2–5]), TEs can alter gene expression by acting as promoters or by providing *cis*-regulatory elements to flanking regions [6–8]. TEs 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 [9–11], they might be especially useful when the survival of an organism or its descendants depends on a quick response to new or challenging environmental conditions (for review [12–15]). Paradoxically, while population genomics data have revealed ongoing TE activity in natural plant populations (e.g. [16–18]), only a handful of TE families have been experimentally shown to transpose. Therefore, how often or under which 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 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.

In plants, the defence against TEs is multi-layered, comprising repressive histone modifications, DNA methylation and RNA interference [19–21]. One 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 subunits of each polymerase (NRPB1, NRPD1 and NRPE1, respectively) assemble with other proteins into enzymes with distinct RNA products and functions [22,23]. 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) [21,23,24]. Upon the base pairing of 24 nt siRNAs to scaffold transcripts produced by Pol V, the *de novo* DNA methyltransferase DRM2 [25] is recruited to 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 depletion of 24 nt siRNAs leads to a drastically increased heat-dependent transposition of the *ONSEN* family [26,27].

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 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' [28], the 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 own life cycle [29]. Besides *ONSEN* in *A. thaliana* [30], the cold inducible *Tcs1* element in blood oranges [8], *Tos17* that gets activated during tissue culture in rice [31], *Tnt1* of tobacco that reacts to wounding [32] or the drought responsive *Rider* retrotransposon in tomato [33] are other prominent cases of stress-responsive TEs in plants. Mechanistically, stress can activate TEs via specific motifs allowing the binding of transcription factors and the subsequent recruitment of the transcription machinery to their promoter-like sequences [18,30,34,35]. 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 [36–39].

Since transcription constitutes the initial step to transposition, the abundance of TE transcripts is often used as a proxy for TE activity [40]. However, the life cycle of TEs is complex [41] and the fate of TE transcripts depends on many factors. For instance, several transcriptionally active TEs have accumulated mutations that prevent the production of enzymes needed for their autonomous transposition [42]. To selectively capture TEs that are not only transcriptionally active but also capable of transposing, several protocols have been developed, including ALE-seq [43], VLP-seq [44] and mobilome-seq [45]. These recent approaches have been particularly successful when targeting long terminal repeat retrotransposons (LTR-RTs), which represent the largest fraction of TE-derived sequences in plant genomes [46]. Indeed, LTR-RTs transpose through a copy-and-paste mechanism which involves the reverse transcription of a full-length RNA intermediate [41,47]. As part of their life cycle and presumably through auto-integration, non-homologous and alternative end-joining, active LTR-RTs can also form extrachromosomal circular DNA (eccDNA) intermediates [44,47–51], whose detection by mobilome-seq can be used as a proxy for their mobility [45]. For instance, mobilome-seq has been successfully used to track full-length eccDNA of mobilized autonomous RTs, containing both LTRs (2-LTR circles), in plants such as *A. thaliana* and rice [45,52].

While the activity of LTR-RTs has been extensively studied in the model *A. thaliana* [18], the interplay between genetic and environmental factors in other wild plant species remains poorly investigated. To clarify these questions, we exploit here the *Brachypodium distachyon* diversity panel [53] and explore

LTR-RT activity in a wild monocot. *B. distachyon* is a Mediterranean grass with a compact diploid genome of ~272 Mb [54,55] harboring 40 LTR-RT families [56] that constitute about 30 % of the genome [55]. In *B. distachyon*, LTR-RTs not only evolved varying insertion site preferences [56] but also significantly differ in terms of transposition dynamics [17,56]. While we previously suggested an ongoing transposition of LTR-RTs based on population genomics data [17,57], here we aimed to clarify the contribution of genetic and environmental factors 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 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 accumulation of LTR-RTs?

Results

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 genetic lineages (A, B and C) that further split into five genetic clades [53,58]: an early diverged C clade and four clades found in Spain and France (B_West), Italy (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 *B. distachyon* obtained from the TRansposable Elements Platform (TREP). We then computed the abundance of TE-derived sequences, hence a proxy for copy numbers (pCNs), for the 40 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 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 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 to the B_East clade [17]. In 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 variations [59], which may further bias the estimates of TIP abundances.

The heatmap produced based on pCN variation (**Fig 1B**) showed that LTR-RTs underwent different transposon accumulations. We found that *Ty3* elements (RLG) harbor higher pCNs than *Ty1/Copia* (RLC)

elements (Wilcoxon test, W = 6517781, p-value < 2.2e-16; **Fig 1B and 1C**). Furthermore, a PCA based on RLG pCNs 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 strongest

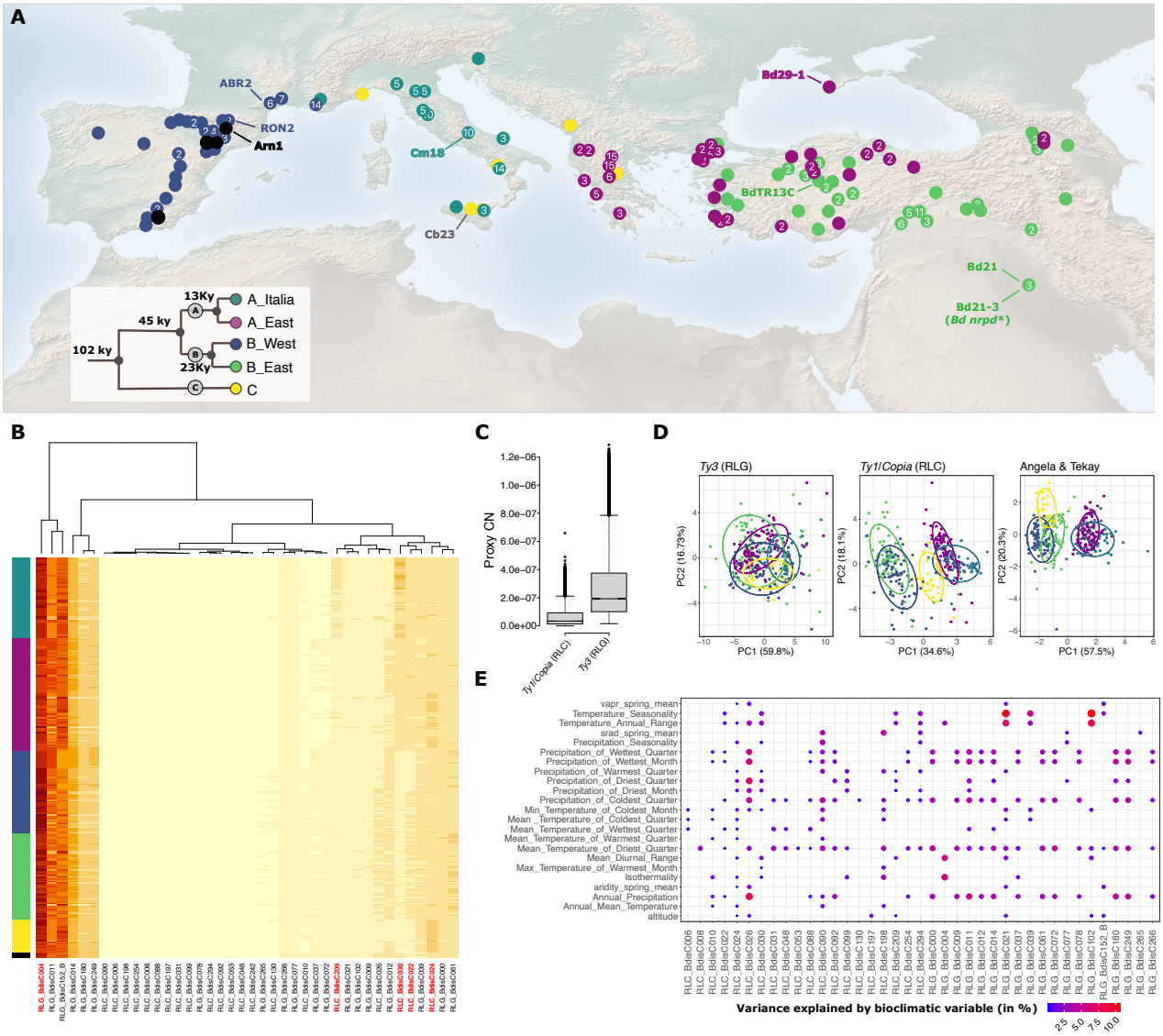


Fig 1. Natural diversity of proximal copy number (pCN) variation of LTR-RTs in *B. distachyon*. (**A**) Origin of the 320 natural accessions included in this study. Accessions that were used for the mobilome-seq are labelled in the map and a PCA with the bioclimatic variables of their place of origin 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. Numbers in dots indicate how many sequenced natural accessions were sampled in the marked area. The map has been obtained from (https://www.naturalearthdata.com/downloads/50m-cross-blend-hypso/50m-cross-blended-hypso-with-shaded-relief-and-water/). (**B**) Heatmap with read counts of TE-derived sequences (proxy for the copy number (pCN) variation) of all 40 annotated LTR-RTs in 320 natural accessions of *B. distachyon.* TE-families are clustered by their pCNs (dendrogram above the heatmap) and accessions are sorted according to their phylogeny. Names of recently active TE-families are highlighted in red. (**C**) Overall estimates of the copy numbers of *Ty1/Copia* and *Ty3*-type LTR-RTs in 320 natural accessions. (**D**) PCAs of pCNs of all *Ty3*, *Ty1/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 pCNs 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 %.

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, RLC_BdisC024, RLC_BdisC030, RLC_BdisC209 and the Tekay family RLG_BdisC004) [56]), with the first and the second principal components together explaining more than 77.8% of the variance. Finally, with the exception of samples 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 1B and 1D**).

To test whether the accumulation of LTR-RT sequences correlated with environmental factors, we retrieved bioclimatic variables comprising precipitation, temperature, aridity levels, solar radiation and atmospheric pressure at each locality. We then ran linear mixed models (LMM) where pCNs per LTR-RT family was entered as the response variable, the bioclimatic variables entered 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 recently active LTR-RT families (**Fig 1E**), indicating that albeit significant, the association between pCNs per family and the environment was mild in our study system. We observed similar associations between pCNs and bioclimatic variables when not accounting for population structure and running classical linear model analyses (**S1 Fig**). With the exception of the two relatively young and low-copy families RLC_BdisC010 and RLG_BdisC265 [56], for which more than 40 % of the variance in pCNs was explained by environmental factors, the LTR-RT families showed non-significant to mild associations with environmental variables (**S1 Fig**).

Recently active LTR-RT families produce eccDNAs

The coverage-based approach for estimating LTR-RT pCNs considers all TE-derived sequences and does not take into account that individual families differ in their age structures, turnover times and proportion of full-length, potentially autonomous mobile elements [56]. We therefore 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 variations in pCNs 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 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 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 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*, treatment with glyphosate and chemical de-methylation and performed mobilome-seq on all resulting 105 samples (S6 **Table)**. Mobilome-seq was developed to specifically capture circular extrachromosomal DNA that is formed as an intermediate during retrotransposon mobility [45] (see Material and Methods).

The rolling circle amplification of eccDNAs during mobilome-seq results in a high coverage of mobile elements allowing to reconstruct near complete mobile LTR-RTs from short reads [45]. Following the removal of organelle-derived reads, we thus first assembled mobilome reads and only aligned the resulting ten longest contigs of each sample to the reference assembly of *B. distachyon*. The size selection of contigs allowed us to only retain the most relevant eccDNA candidates of each sample, potentially representing full-length mobile LTR-RTs. 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. In a next step, we attempted to identify those genomic regions that were enriched in contigs assembled from mobilome reads from certain genotypes or stress conditions. We hence 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 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 (**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 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 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 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**).

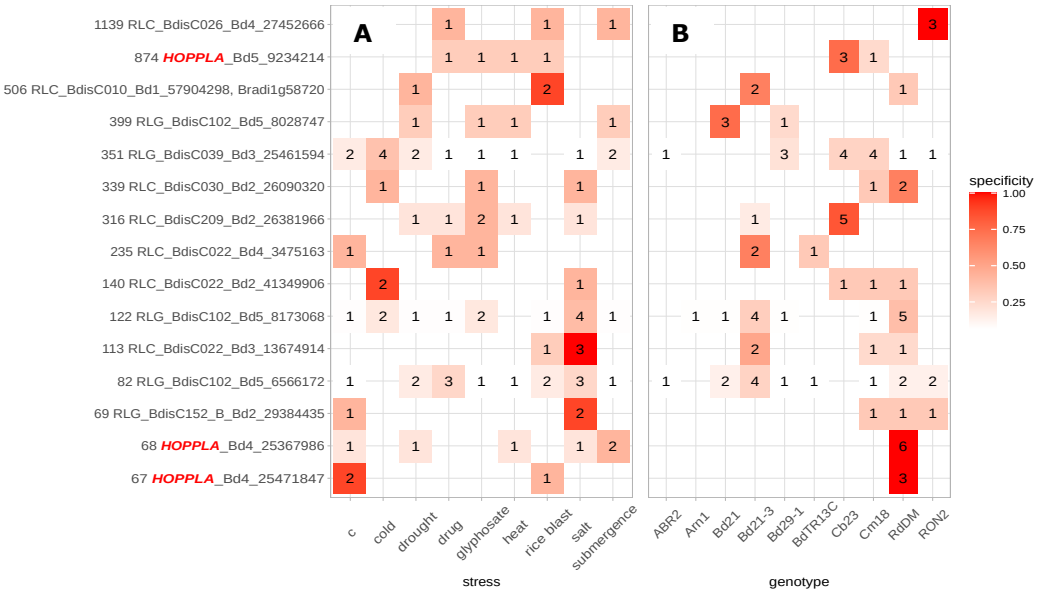


Fig 2. Assessment of LTR-RT mobility in *B. distachyon*. Stress (**A**) and genotype (**B**) specificity of the formation of eccDNA as determined by the alignment of assembled mobilome-seq reads. The color represents the degree of specificity and numbers indicate the count of samples from which one of their ten longest contigs aligns to each of the circle-forming regions. Annotations of regions are indicated on the y-axis. Loci containing *HOPPLA* (RLC_BdisC024) are highlighted in red. Multiple annotations in the same circle-forming region were concatenated. The two *pol IV* mutants and the controls Bd21-3 and the outcrossed line *Bd NRPD1 (+/+)* are summarized as RdDM and Bd21-3, respectively. The following stresses were applied: c (control conditions), cold (2 °C on ice, 24h), drought (uprooting, 2:15 h), drug (chemical de-methylation with Zebularine (20 uM) and alpha-amanitin (2.5 mg/ml), 28 days), glyphosate (20 mM, four days), heat (42 °C, 8h), rice blast (*Magnaporthe oryzae* infection, four days), salt (300 mM NaCl, five days) and submergence (48 h). See material and methods for details.

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

Fragmented eccDNAs or circles containing only one of the two LTRs (1-LTR circles) can be formed following reverse transcription by auto-integration, alternative end-joining in the virus-like particles [49,50] or by a recombination of the two LTRs of genomic copies [60,61]. 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 [50]. 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 [45]. As a complement to the assembly-based 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 *B. distachyon* reference assembly [56]. This allowed us to specifically detect intact 2-LTR circles of extrachromosomal LTR-RTs capable of integrating into the genome. To control for possible traces of undigested genomic

DNA that may also contain regions resembling LTR-LTR junctions and that may subsequently be amplified by the Phi29 enzyme during mobilome-seq [62], we also included publicly available genomic reads of all nine accessions in our analysis.

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 for any of the 37 LTR-RT families with annotated full-length copies (**Fig 3A**). For instance, we found a very strong signal

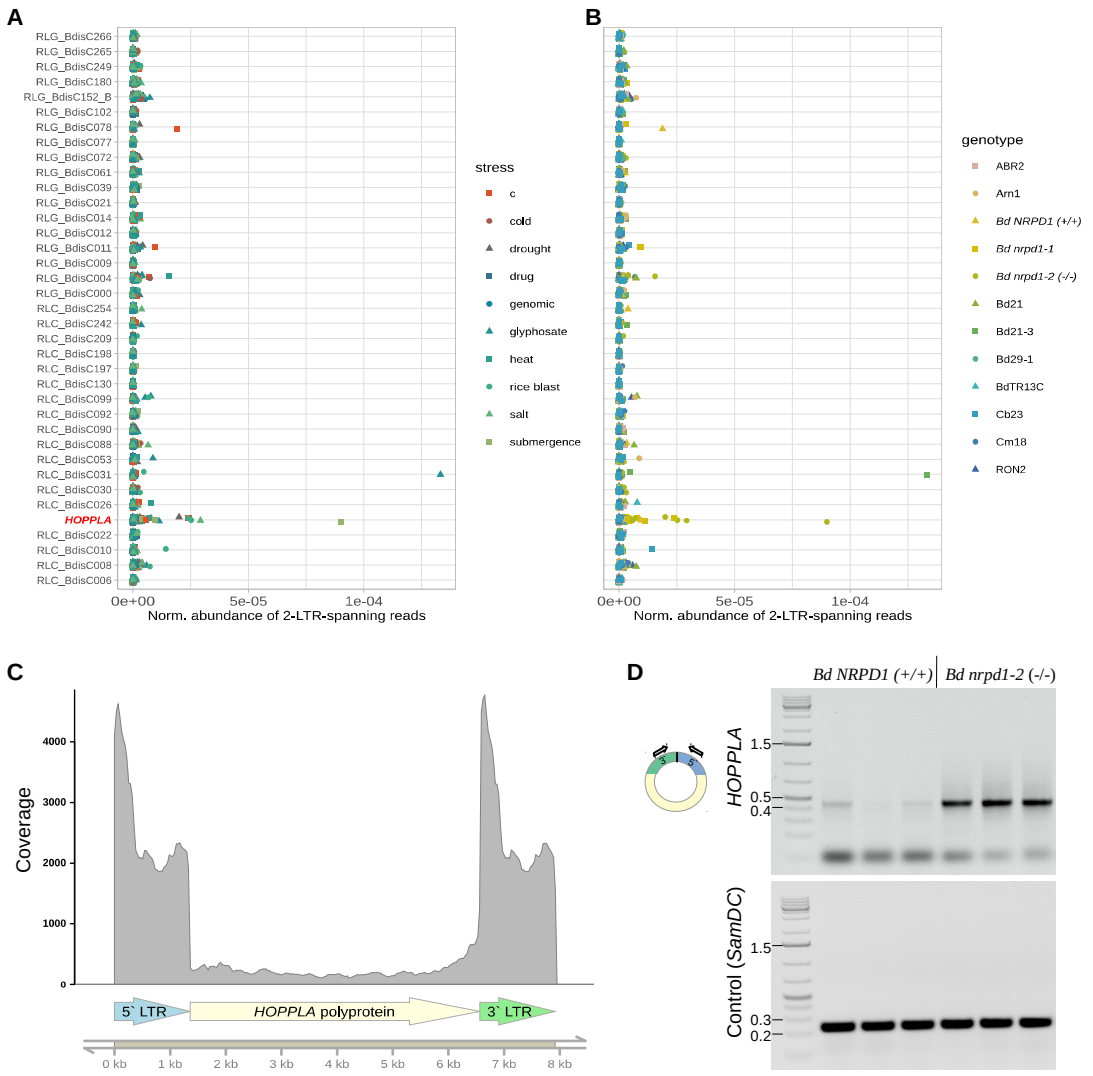


Fig 3. *HOPPLA* forms 2-LTR circles in the *pol IV* mutants. Normalized abundance of 2-LTR-junction spanning reads depending on the stress (**A**) and the genotype (**B**) of individual mobilome-seq samples. *HOPPLA* (RLC_BdisC024) is highlighted in red. The alignment of publicly available genomic reads (genomic) served as a control for the presence of genomic reads aligning to 2-LTR-junctions. See Fig 2 and methods section for details about the applied c (control) conditions and stresses. (**C**) Coverage of mobilome-seq reads of the *HOPPLA* consensus sequence of sample *Bd nrpd1-2 (-/-)* submergence stress. The structure of *HOPPLA* element is depicted. (**D**) 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. Loaded PCR reactions with primers specific to the *HOPPLA* LTRs (top gel) and to the genomic control gene *SamDC* (Bradi5g14640) (bottom gel) are shown. Three biological replicates are depicted. Schematic representation of primer design for the inverse PCR is shown.

for RLC_BdisC031 that was solely detected in glyphosate-treated Bd21-3 plants 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 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 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 multiple accessions (**Fig 3A**).

The mobility of *HOPPLA* in the *Bd nrpd1-2 (-/-)* was further confirmed by the fact that the alignment of mobilome-reads to the consensus sequence of *HOPPLA* resulted in a high coverage of the entire element (**Fig 3C)**. Finally, 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, with outward facing primers specific to the two LTRs (**Fig 3D**). Notably, in the PCR, we also detected a faint signal for the outcrossed line *Bd NRPD1 (+/+)* suggesting a weak activity of *HOPPLA* in wild-type plants.

Members of the *HOPPLA* family differ in activity

Because individual copies of the same LTR-RT family can differ in their activity [30], we also analysed the relative abundance of 2-LTR-spanning reads for each annotated full-length copy of *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**). 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 the weak but detectable band for the inverse 2-LTR PCR for *Bd NRPD1 (+/+)* (**Fig 3D**).

Using the meta information of individual *HOPPLA* copies described previously [56], 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* copies. While no clear pattern emerged from this analysis, the most active copies of *HOPPLA* tend to be rather young (**Fig 4B**).

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 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 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* may indicate that developmental processes and auxin-activated signaling pathways played a role in its activity, rather than specific stresses (**Fig 4C**).

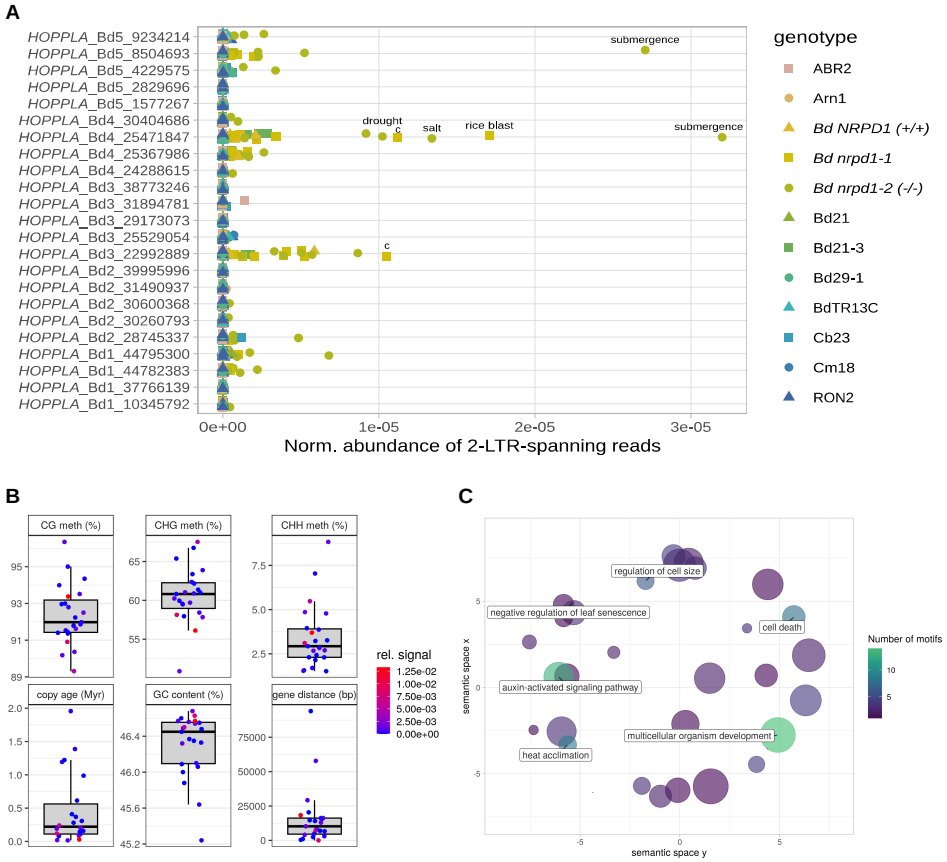


Fig 4. Members of the *HOPPLA* family differ in activity. (**A**) Normalized abundance of 2-LTR-junction spanning reads of individual full-length copies of the *HOPPLA* family. Samples with a high signal are labelled with c (control conditions) or the respective stress applied. See Fig 2 and material and methods for details. (**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**). (**C**) 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.

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

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* [27], rice [63] and for the Alesia family (RLC_BdisC010) in *B. distachyon* [64]. 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 samples from the two *pol IV* mutants and their respective wild-type controls. Using a hybridization 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 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, 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 that the reduction of 24 nt siRNAs is likely associated with an increased expression and subsequent formation of *HOPPLA* eccDNAs in both *polIV* mutants (**Fig 5B**).

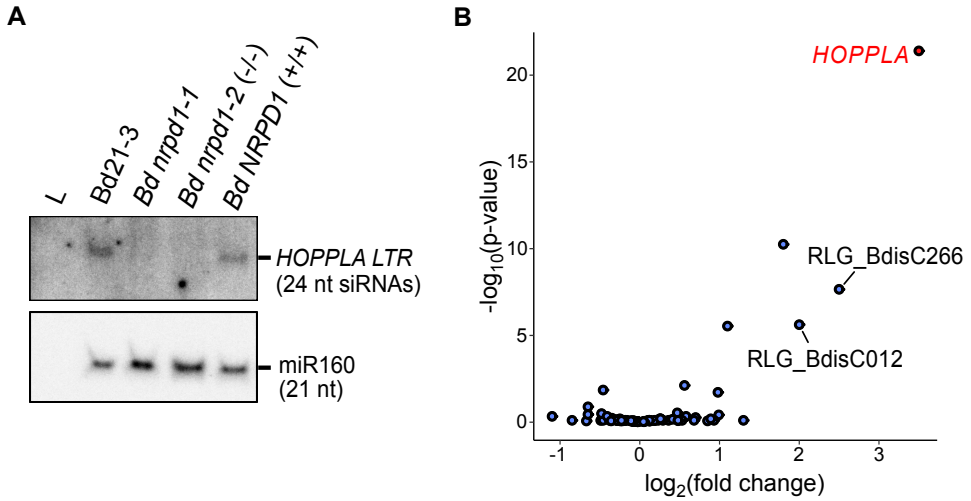


Fig 5. Loss of 24-nt siRNAs leads to an increased activity of *HOPPLA* in the Pol IV mutants. (**A**) Northern plot for the detection of 24-nt siRNAs specific to the 3` LTR of *HOPPLA* in the *pol IV* mutants *Bd nrpd1-1* and *Bd nrpd1-2* (-/-) and their control lines Bd21-3 and the outcrossed line *Bd 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, *HOPPLA* is highlighted in red, three biological replicates were analysed.

To complete their life cycle, reverse transcribed extrachromosomal copies of LTR-RTs have to integrate into the host genome [41]. Because all our analyses congruently pointed to the activity of *HOPPLA* in the *Bd nrpd1-2 (-/-)* mutant, we sequenced the genome of seven individuals of *Bd nrpd1-2 (-/-)*, six *Bd NRPD1 (+/+)* plants and one wild-type Bd21-3 plant to detect 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-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 *Bd nrpd1-2 (-/-)* plants (Bd1 38798495, Bd1

42205987, Bd4 28119639) (**S3A-S3C Fig**). This confirmed that the loss of Pol IV function led to an increased production of eccDNA, as well as actual 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 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.

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

To decipher the genetic basis of *HOPPLA* accumulations in natural populations, we first performed a genome-wide association study (GWAS) using *HOPPLA* pCNs in the diversity panel of 320 natural 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 *HOPPLA* family (**Fig 6**). Because inserted copies of *HOPPLA* may themselves lead to significantly associated regions in the GWAS as shown in *A. thaliana* [65], we first verified that there were neither TIPs [57] 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 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) (**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, AT1g63240) [66], to be co-localized with the peak.

To test whether genomic regions might be recurrently associated with their pCNs, we finally extended the GWAS analyses to the four other most recently active families (RLC_BdisC022, RLC_BdisC030, RLC_BdisC209 and RLG_BdisC004) (**S4 Fig**). We extracted the candidate genes for 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 RLC_BdisC022 and RLC_BdisC209 that shared three genes, we found no overlap of annotated loci potentially contributing to the pCN variations of recently active families (**Fig 6**).

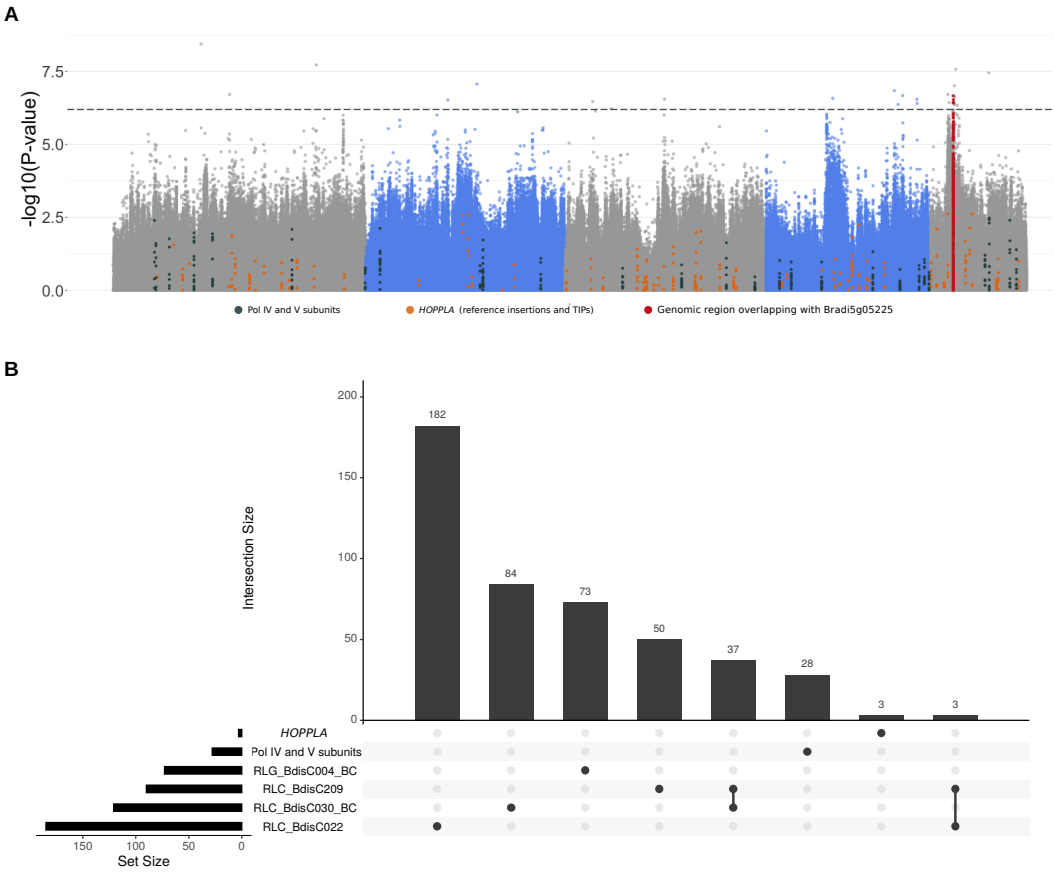


Fig 6. Genetic regions associated with pCN variations of recently active LTR-RTs. (**A**) Manhattan plot depicting the GWAS results of pCN variation of *HOPPLA* in 320 accessions of *B. distachyon.* Colored points indicate SNPs linked (+/- 10 kb) to known Pol IV and V subunits (dark grey), *HOPPLA* reference insertions and TIPs (orange) and the genomic region containing the candidate gene Bradi5g05225 (red). 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 components of the Pol IV and Pol V holoenzymes is included in the UpSet plot.

Discussion

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 diversity and potential drivers of evolution has been demonstrated in many organisms including fungi [67], insects [68], mammals [69] and plants [70]. However, while there are a number of examples showing that certain TE insertions facilitated the adaptation to changing environments (for review [71]), TEs are generally harmful and therefore controlled by complex silencing mechanisms. To foster our understanding of TE 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 pCNs in a panel of 320 *B. distachyon* natural accessions, we show that the intra-specific variations of pCNs of RLC elements, but not the pCNs of the generally older and more abundant RLG elements [56], 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 (RLG_BdisC004) family, which are the youngest families in *B. distachyon* [56]. Highly polymorphic among natural accessions of *B. distachyon* [17,56], they are expectedly the main drivers of lineage-specific expansions of pCNs in our study system. Hence, we do not only confirm that LTR-RT families in *B. distachyon* globally differ in size [56] but also demonstrate that the accumulation of genomic sequences derived from specific families varies significantly among natural accessions.

The transcriptional activity of LTR-RTs can be triggered by specific environmental stresses [29,34]. Given that *B. distachyon* occurs in a wide range of different habitats in the Mediterranean area [53], this characteristic feature of LTR-RTs provides a potential explanation for the pCN variation we observed across natural accessions [18,65]. Yet, for the large majority of LTR-RT families, pCNs correlate only moderately with environmental factors. Consequently, our genomic data do not support a large effect of the 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 show signs of increased activity in the recent past in *B. distachyon* [56]. Considering that their copy number expansions took place in a climate that has drastically changed following the last glacial maximum [53], a limited link between their activity and the current environmental conditions is actually expected for most families. In contrast, the lack of correlation between current bioclimatic variables and copy number variation for families with ongoing activity (RLC_BdisC022, RLC_BdisC024, RLC_BdisC030, RLC_BdisC209 and RLG_BdisC004; [56]), 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 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 range') only explained about 9 % of the observed variation [18].

Genetic factors are well-known to be essential in regulating LTR-RT activity [72–75]. As the loss of main players of the RdDM silencing pathway leads to increased TE activity [18,27,33,63], 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 necessarily able to transpose [76], we used a mobilome-seq approach to detect TE-derived eccDNAs and transpositionally active LTR-RT families. We deliberately followed a very 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 newly inserted copies in the Pol IV mutant.

The non-stress-specific activity of *HOPPLA* in the two independent Pol IV mutants and limited activity of other elements supported our *in silico* approach and strengthened the idea that genetic, rather than environmental stresses, are major drivers of LTR-RT activity in *B. distachyon*. While we cannot exclude the possibility that other specific stress conditions may trigger the eccDNA formation of other young autonomous LTR-RTs, specifically for *HOPPLA*, these results are also in line with our TF binding 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 targeted by TFs involved in developmental processes and auxin signaling [77]. As this study already covers a broad range of (a)biotic stresses, follow-up studies should therefore address the question of whether the activity of *HOPPLA* or other families differs between tissues or developmental stages, as also observed for the endosperm-specific mobility of *PopRice* in rice, for example [45]. Strikingly, despite the central role of Pol IV in 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 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 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, which could indicate an accession-specific response of the mobilome to certain triggers.

Given that pCNs vary greatly among genetic clades, assessing the effect of a genetic 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 Bd21-3 background than in the other accessions 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 methyltransferases [52] 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 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 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 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, pCNs are also dependent on the removal rate of LTR-RT families, resulting in the formation of TE-fragments and soloLTRs, which varies greatly in *B. distachyon* [56]. This complexity of parameters affecting the dynamics of LTR-RT might explain why none of the genes known to be involved in silencing LTR-RTs are associated with the pCN variation of *HOPPLA* or any recently active family. Our candidate locus containing Bradi5g05225, a gene related to RMB1 whose loss of function has been shown to result in DNA hypermethylation [78], remains nonetheless a great candidate for functional validation.

Altogether, our work confirms that LTR-RTs in *B. distachyon* are 'well-behaved' [17] and that the evolutionary consequences of their mobility are hard to study in real-time. 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 our previous 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 depend on a complex interaction between genetic factors, developmental stages and, more marginally, the punctual occurrence of stresses. This study not only elucidates fundamental mechanisms of LTR-RT-dynamics of an undomesticated grass in the wild, but may also be relevant to better understand the biology of mobile elements in more complex genomes.

Materials and methods

Estimation of LTR-RT pCNs

We used publicly available genomic reads of 320 sequenced natural accessions of *B. distachyon* [53,58,59,79,80] to assess the natural variation of copy numbers of LTR-RTs. Because sequencing depth differed substantially between accessions [53], we first downsampled all fastq files to the read 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 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) [81] with the -M and the -a options set, hence outputting all alignments

found. Coverages of LTR-RTs and the reference assembly were assessed using bedtools (v 2.30.0) [82] genomecov with the -d and -split parameters set. The assessment of the global coverage of the reference genome allowed us to compensate for potential differences in read length and/or quality. A proxy for copy numbers was thus obtained by normalizing the coverage signals of each of the LTR-RT consensus sequences by the coverage of the entire reference assembly and by correcting for the length of the consensus sequences. pCN raw data were processed using R (v 3.6.3 and 4.0.2) [83] in Rstudio [84].

Variation in pCNs 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 distances between accessions with the R package pvclust v 2.2.0 [85]. The resulting tree was used to order accessions phylogenetically on the heatmap. PCAs based on pCNs were obtained with the R package ggbiplot v 0.55 [86].

To test for an association between pCNs and environmental variables, we retrieved information about climatic variables at each local site from [53]. Linear mixed model analyses where the pCN per LTR-RT family was entered as the response variable, the bioclimatic 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 [87]. The part of the variance explained by the fixed- (marginal R^2) were computed following [88] and visualized as bubble plot with the R package ggplot2 [89]. Classical linear models were run in base R.

Plant material, growth conditions and stresses for mobilome-seq

Brachypodium distachyon natural accessions used in this study comprised Bd21, Bd21-3, Cm18, Cb23, ABR2, Bd29-1 BdTR13c, RON2 and Arn1. Because Pol IV is known to play an important role in LTR-RT silencing in plants [27,90,91], we also included the sodium azide mutagenized *pol IV* mutant line NaN74 (*Bd nrpd1-1*) [64,92], the T-DNA insertion *pol IV* mutant line JJJ18557 Nr31 [93] *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 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 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 (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). 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, 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 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 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* (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 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 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 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 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. 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 [52] by 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 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 for all natural accessions.

Mobilome sequencing and validation of eccDNAs

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

Waltham, USA). Mobilome sequencing was performed according to [45] 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 µL. To enrich eccDNA, DNA was first purified using the GENECLEAN kit (MP Biomedicals, Santa Ana, USA) according to manufactures recommendations using 5 µL glass milk with an elution volume of 35 µL. Thirty µL of the eluate were 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 precipitated eccDNA amplified using the illustraTempliPhi Amplification Kit (Cytiva, Marlborough, USA) according to [45] with an extended incubation time of 65 h at 28 °C. The templiphi 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 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 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 reagent kit v3 (600 cycles) (S6 **Table)**. Raw reads have been uploaded to ENA (accession number PRJEB58186).

The presence of extrachromosomal circular copies of *HOPPLA* (RLC_BdisC024) was validated by an inverse PCR using 7 ng/µL total DNA. Input quantities of DNA were controlled using primers specific to the S-adenosylmethionine decarboxylase (SamDC) gene (Bradi5g14640) that have previously been shown to be efficient and are therefore also used for the control reaction amplifying reference genes in real-time PCRs in *B. distachyon* [94]. Sequences of primers are listed in **S5 Table**.

Analysis of Mobilome-seq

Reads were trimmed using the BBDuk tool of BBtools (BBMap (v 38.75, Bushnell B., sourceforge.net/projects/bbmap/) with the parameters qtrim = rl and trimq = 20. Reads originating from organelles were removed by aligning reads to the chloroplast genome (NC_011032.1) [95] and the mitochondrion genome (v 1.0.0) of *B. distachyon* using BWA-MEM (v 0.7.17-r1188) [81] with the -M parameter set. Unmapped reads were isolated using samtools (v 1.13) [96] view -b -f 4 and bedtools (v 2.30.0) [82] bamtofastq.

Organelle-filtered mobilome reads were assembled using the SPAdes genome assembler (v 3.13.0) [97]. From each assembly, the top ten contigs were extracted and jointly aligned to the reference assembly of Bd21 (v 3.0) using BWA-MEM (v 0.7.17-r1188) with the -M parameter set. Bam files were converted into bed files using bedtools (v 2.30.0) bamtobed with the -split option set and overlapping contigs were merged using bedtools merge with the -o distinct, count, count_distinct and -c 4 parameters set. Assembled, circle-forming regions were annotated with bedtools intersect using the version 3.1 annotation of the reference assembly and the annotation of all full-length LTR-RTs [56] of the reference assembly. Annotated regions were extracted with bedtools getfasta and all sequences longer than 2 kb were isolated using SeqKit seq (v 0.11.0) [98]. Circle-forming regions that occurred in less than three samples were not included in the analysis.

To specifically detect mobilized LTR-RTs, we first extracted all annotated full-length LTR-RTs of the Bd21 reference assembly [56]. Using a custom python script, we then merged the last 300 bp of the 3' to the first 300 bp of the 5' LTR to obtain a 'tail-to-head' library containing all annotated full-length LTR-RT copies annotated in Bd21. We then aligned organelle-filtered 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. The coverage of the junction-spanning reads was calculated using deeptools (v 3.5.1) [99] with the parameters -bs 1, --ignoreDuplicates --outRawCounts set. To account for differences 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 set, from the alignments of filtered reads to the reference assembly of Bd21 (v 3.0) obtained from Phytozome 12 [55] generated by BWA-MEM (v 0.7.17-r1188) [81]. To control for potential traces with undigested genomic DNA that may contain inserted LTR-LTR junctions, we also aligned publicly available reads for each of the accessions, and measured their relative abundance as described above. To plot the overall activity per family (**Fig 3A and 3B**), normalized signals were summed up for every individual TE family. Otherwise (**Fig 4A**) signals were plotted for each individual *HOPPLA* full-length copy.

To visualize the coverage of the *HOPPLA* TREP consensus sequence, reads were aligned using BWA-MEM with the -M and -a options set. Aligned reads were visualized using the packages GVIZ (v.1.28.3) [100] and RTRACKLAYER (v.1.44.4) [101] in R and Rstudio.

mRNA-sequencing and small RNA northern blotting

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 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 (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+ 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).

For the small RNA blot analysis, 200 µg of each total RNA were size-fractionated using the RNeasy Midi Kit (QIAGEN), as described previously [64]. Low molecular weight (LMW, <200 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. 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. 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 the 16% polyacrylamide gel [64]. After running, transfer and UV crosslinking, membrane was prehybridized in PerfectHyb Plus buffer (Merck, Darmstadt, Germany) at 35°C and then and hybridized at 35°C with the Klenow internally-labeled probe (*HOPPLA*), or with the 5'-end labeled probe (miR160) [64]. After overnight hybridization, washing was performed 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.**

LTR-RT expression analysis

RNA-seq raw reads of *Bd nrpd1-2 (-/-)* and *Bd NRPD1 (+/+)* were trimmed for adapters using fastp (v 0.23.2) [102] with the following options: --qualified_quality_phred 15 --unqualified_percent_limit 4 --n_base_limit 20 --low_complexity_filter --overrepresentation_analysis --correction --detect_adapter_for_pe. Cleaned reads were then analysed using SalmonTE (v 0.4) [103] to measure

global expression of LTR-RTs. LTR-RT consensus sequences of *B. distachyon* obtained from the TRansposable Elements Platform (TREP, 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 analysis. Expression data were plotted using R (v 3.6.3) in RStudio (v 7d165dcf).

Motif analysis

The consensus sequence of *HOPPLA* was screened for known transcription factor binding sites obtained from the PlantTFDB [104] using FIMO (v 5.1.1) [105]. To functionally annotate transcription factors that could bind to *HOPPLA,* we used GO-terms of the Gramene (release 50) database [106] downloaded from the platform agriGO (v 2.0) [107]. 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 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 *HOPPLA* were visualized with REVIGO [108] using the 'SimRel' semantic similarity measure, the option 'small' and the GO terms of the *Oryza sativa* Japonica Group. The total number 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 followed the exact same approach using the sequence of one of the most active *ONSEN* copies (AT1G11265) and the GO terms of *Arabidopsis thaliana*.

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

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 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 assembly of Bd21 (v 3.0) using BWA-MEM (v 0.7.17-r1188) [81] with the -M parameter set. Samtools (v 1.13) [96] was used to obtain sorted and indexed bam 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-database) and the annotation of all full-length LTR-RTs of

Bd21 [56]. Because both *Bd nrpd1-2* lines were in the Bd21-3 background we were able to exclude all Bd21-3 specific TIPs by 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) . *HOPPLA* TIPs were visualized with JBrowse 2 (v 2.6.1) [109]. Raw genomic reads of the re-sequencing of *Bd nrpd1-2 (-/-), Bd NRPD1 (+/+)* and Bd21-3 have been uploaded to ENA (accession number PRJEB73379).

GWAS for pCNs

GEMMA 0.98.5 [110] was used to test for associations between SNPs [53] and the LTR-RT families pCNs, while correcting for population structure [53,58]. A centered relatedness matrix was first 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 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 0.05 or Bonferroni correction threshold as candidate region using the R package rehh (v 3.2.2) [111]. Genes overlapping with candidate regions were selected with the BEDTOOLS (v 2.26.0) [82] intersect command using the version 3.1 of the *B. distachyon* annotation file (https://phytozome-next.jgi.doe.gov). The UpSetR [112] R package was used to visualize the intersections of significant genes between the variables. Protein constituents of the Pol IV and Pol V enzymes (see **S4 Table**) were downloaded from the plant RNA polymerase database http://rna.polymerase.eu/.

Acknowledgements

We thank the Genetic Diversity Center Zürich for providing the infrastructure for the mobilome-seq. This manuscript was professionally edited by Dr Emmanuelle Botté, from Manuscribe (https://manuscribe.com.au/).

Author contributions

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. generated data. M.T., A.C.R., N.M. and W.X. analysed data. C.M. handled sequencing 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 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.

Funding

This work was supported by the University of Zurich Research Priority Programs (URPP) Evolution in Action to MT and ACR; the Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (grant number 31003A_182785 to ACR, WX, NM, KR and BK); the Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (grant number PZ00P3_154724 to CS), and the Interdisciplinary Thematic Institute IMCBio (ITI 2021-2028 program to TB), including funds from IdEx Unistra (ANR-10-IDEX-0002 to TB), SFRI-STRAT'US (ANR 20-SFRI-0012 to TB) and EUR IMCBio (ANR-17-EURE-0023 to TB) in the framework of the French Investments for the Future Program. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Mirouze M, Vitte C. Transposable elements, a treasure trove to decipher epigenetic variation: insights from Arabidopsis and crop epigenomes. J Exp Bot. 2014;65: 2801–2812. doi:10.1093/jxb/eru120
2. Piegu B, Guyot R, Picault N, Roulin A, Sanyal A, Kim H, et al. Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in Oryza australiensis, a wild relative of rice. Genome Res. 2006;16: 1262–1269. doi:10.1101/gr.5290206
3. Hawkins JS, Kim HR, Nason JD, Wing RA, Wendel JF. Differential lineage-specific amplification of transposable elements is responsible for genome size variation in Gossypium. Genome Res. 2006;16: 1252. doi:10.1101/GR.5282906
4. Wang D, Zheng Z, Li Y, Hu H, Wang Z, Du X, et al. Which factors contribute most to genome size variation within angiosperms? Ecol Evol. 2021;11: 2660–2668. doi:10.1002/ECE3.7222
5. Yang L-L, Zhang X-Y, Wang L-Y, Li Y-G, Li X-T, Yang Y, et al. Lineage-specific amplification and epigenetic regulation of LTR-retrotransposons contribute to the structure, evolution, and

function of Fabaceae species. BMC Genomics 2023 241. 2023;24: 1–15. doi:10.1186/S12864-023-09530-Y

6. Roquis D, Robertson M, Yu L, Thieme M, Julkowska M, Bucher E. Genomic impact of stress-induced transposable element mobility in Arabidopsis. Nucleic Acids Res. 2021;49: 10431–10447. doi:10.1093/nar/gkab828

7. Makarevitch I, Waters AJ, West PT, Stitzer M, Hirsch CN, Ross-Ibarra J, et al. Transposable elements contribute to activation of maize genes in response to abiotic stress. PLoS Genet. 2015;11: e1004915. doi:10.1371/journal.pgen.1004915

8. Butelli E, Licciardello C, Zhang Y, Liu J, Mackay S, Bailey P, et al. Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. Plant Cell. 2012;24: 1242–1255. doi:10.1105/tpc.111.095232

9. Uzunović J, Josephs EB, Stinchcombe JR, Wright SI, Parsch J. Transposable Elements Are Important Contributors to Standing Variation in Gene Expression in Capsella Grandiflora. Mol Biol Evol. 2019;36: 1734–1745. doi:10.1093/molbev/msz098

10. Thieme M, Brêchet A, Bourgeois Y, Keller B, Bucher E, Roulin AC. Experimentally heat-induced transposition increases drought tolerance in Arabidopsis thaliana. New Phytol. 2022;236: 182–194. doi:10.1111/NPH.18322

11. Raúl C, Noemia M-D, Sonal G, Michael P, M. CJ. Transposons are important contributors to gene expression variability under selection in rice populations. Elife. 2023;12. doi:10.7554/ELIFE.86324

12. Kawakatsu T, Huang SS, Jupe F, Sasaki E, Schmitz RJ, Urich MA, et al. Epigenomic Diversity in a Global Collection of Arabidopsis thaliana Accessions. Cell. 2016;166: 492–505. doi:10.1016/j.cell.2016.06.044

13. Rey O, Danchin E, Mirouze M, Loot C, Blanchet S. Adaptation to Global Change: A Transposable Element-Epigenetics Perspective. Trends in Ecology and Evolution. Elsevier Ltd; 2016. pp. 514–526. doi:10.1016/j.tree.2016.03.013

14. Lanciano S, Mirouze M. Transposable elements: all mobile, all different, some stress responsive, some adaptive? Curr Opin Genet Dev. 2018;49: 106–114. doi:10.1016/j.gde.2018.04.002

15. Dubin MJ, Mittelsten Scheid O, Becker C. Transposons: a blessing curse. Curr Opin Plant Biol. 2018;42: 23–29. doi:10.1016/J.PBI.2018.01.003

16. Stuart T, Eichten SR, Cahn J, Karpievitch Y V., Borevitz JO, Lister R. Population scale mapping of transposable element diversity reveals links to gene regulation and epigenomic variation. Elife. 2016;5: e20777. doi:10.7554/ELIFE.20777

17. Stritt C, Gordon SP, Wicker T, Vogel JP, Roulin AC. Recent Activity in Expanding Populations and Purifying Selection Have Shaped Transposable Element Landscapes across Natural Accessions of the Mediterranean Grass Brachypodium distachyon. Genome Biol Evol. 2018;10: 304–318. doi:10.1093/gbe/evx276

18. Baduel P, Leduque B, Ignace A, Gy I, Gil J, Loudet O, et al. Genetic and environmental modulation of transposition shapes the evolutionary potential of Arabidopsis thaliana. Genome Biol. 2021;22: 138. doi:10.1186/s13059-021-02348-5

19. Saze H, Tsugane K, Kanno T, Nishimura T. DNA Methylation in Plants: Relationship to Small RNAs and Histone Modifications, and Functions in Transposon Inactivation. Plant Cell Physiol. 2012;53: 766–784. doi:10.1093/PCP/PCS008

20. Zhang H, Lang Z, Zhu JK. Dynamics and function of DNA methylation in plants. Nat Rev Mol Cell Biol 2018 198. 2018;19: 489–506. doi:10.1038/s41580-018-0016-z

21. Liu P, Cuerda-Gil D, Shahid S, Slotkin RK. The Epigenetic Control of the Transposable Element Life Cycle in Plant Genomes and Beyond. https://doi.org/101146/annurev-genet-072920-015534. 2022;56: 63–87. doi:10.1146/ANNUREV-GENET-072920-015534

22. Ream TS, Haag JR, Wierzbicki AT, Nicora CD, Norbeck AD, Zhu JK, et al. Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. Mol Cell. 2009;33: 192–203. doi:10.1016/j.molcel.2008.12.015

23. Rymen B, Ferrafiat L, Blevins T. Non-coding RNA polymerases that silence transposable elements and reprogram gene expression in plants. Transcription. 2020;11: 172–191. doi:10.1080/21541264.2020.1825906

24. Sigman MJ, Slotkin RK. The First Rule of Plant Transposable Element Silencing: Location, Location, Location. Plant Cell. 2016;28: 304–313. doi:10.1105/TPC.15.00869

25. Zhong X, Du J, Hale CJ, Gallego-Bartolome J, Feng S, Vashisht AA, et al. Molecular Mechanism of Action of Plant DRM De Novo DNA Methyltransferases. Cell. 2014;157: 1050–1060. doi:10.1016/J.CELL.2014.03.056

26. Tittel-Elmer M, Bucher E, Broger L, Mathieu O, Paszkowski J, Vaillant I. Stress-Induced Activation of Heterochromatic Transcription. PLOS Genet. 2010;6: e1001175. doi:10.1371/JOURNAL.PGEN.1001175

27. Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, Paszkowski J. An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. Nature. 2011;472: 115–120. doi:10.1038/NATURE09861

28. McClintock B. The significance of responses of the genome to challenge. Science (80-). 1984;226: 792–801. Available: internal-pdf://189.66.194.35/McClintock

29. Negi P, Rai AN, Suprasanna P. Moving through the stressed genome: Emerging regulatory roles for transposons in plant stress response. Front Plant Sci. 2016;7: 1448. doi:10.3389/fpls.2016.01448

30. Cavrak V V., Lettner N, Jamge S, Kosarewicz A, Bayer LM, Mittelsten Scheid O. How a Retrotransposon Exploits the Plant's Heat Stress Response for Its Activation. PLOS Genet. 2014;10: e1004115. doi:10.1371/JOURNAL.PGEN.1004115

31. Hirochika H, Sugimoto K, Otsuki Y, Tsugawa H, Kanda M. Retrotransposons of rice involved in mutations induced by tissue culture. Proc Natl Acad Sci U S A. 1996;93: 7783–7788. Available: internal-pdf://0387344111/Hirohiko_tissue culture rice tos.pdf

32. Mhiri C, Morel JB, Vernhettes S, Casacuberta JM, Lucas H, Grandbastien MA. The promoter of the tobacco Tnt1 retrotransposon is induced by wounding and by abiotic stress. Plant Mol Biol. 1997;33: 257–266. doi:10.1023/A:1005727132202

33. Benoit M, Drost HG, Catoni M, Gouil Q, Lopez-Gomollon S, Baulcombe D, et al. Environmental and epigenetic regulation of Rider retrotransposons in tomato. PLoS Genet. 2019;15: e1008370. doi:10.1371/journal.pgen.1008370

34. Grandbastien MA. LTR retrotransposons, handy hitchhikers of plant regulation and stress response. Biochim Biophys Acta - Gene Regul Mech. 2015;1849: 403–416. doi:10.1016/J.BBAGRM.2014.07.017

35. Zhang Y, Li Z, Liu J, Zhang Y, Ye L, Peng Y, et al. Transposable elements orchestrate subgenome-convergent and -divergent transcription in common wheat. Nat Commun. 2022;13. doi:10.1038/S41467-022-34290-W

36. Kidwell MG, Lisch D. Transposable elements as sources of variation in animals and plants. Proc Natl Acad Sci U S A. 1997;94: 7704–7711. doi:10.1073/pnas.94.15.7704

37. Venner S, Feschotte C, Biémont C. Dynamics of transposable elements: towards a community ecology of the genome. Trends Genet. 2009;25: 317–323. doi:10.1016/J.TIG.2009.05.003

38. Stritt C, Thieme M, Roulin AC. Rare transposable elements challenge the prevailing view of transposition dynamics in plants. Am J Bot. 2021;108: 1310–1314. doi:10.1002/AJB2.1709

39. Stitzer MC, Anderson SN, Springer NM, RossIbarra J. The genomic ecosystem of transposable elements in maize. PLoS Genet. 2021;17: e1009768. doi:10.1371/journal.pgen.1009768

40. Lanciano S, Cristofari G. Measuring and interpreting transposable element expression. Nat Rev Genet 2020 2112. 2020;21: 721–736. doi:10.1038/s41576-020-0251-y

41. Schulman AH. Retrotransposon replication in plants. Curr Opin Virol. 2013;3: 604–614. doi:10.1016/J.COVIRO.2013.08.009

42. Tanskanen JA, Sabot F, Vicient C, Schulman AH. Life without GAG: the BARE-2 retrotransposon as a parasite's parasite. Gene. 2007;390: 166–174. doi:10.1016/J.GENE.2006.09.009

43. Cho J, Benoit M, Catoni M, Drost H-G, Brestovitsky A, Oosterbeek M, et al. Sensitive detection of pre-integration intermediates of long terminal repeat retrotransposons in crop plants. Nat Plants. 2019;5: 26–33. doi:10.1038/s41477-018-0320-9

44. Lee SC, Ernst E, Berube B, Borges F, Parent JS, Ledon P, et al. Arabidopsis retrotransposon virus-like particles and their regulation by epigenetically activated small RNA. Genome Res. 2020;30: 576–588. doi:10.1101/GR.259044.119/-/DC1

45. Lanciano S, Carpentier MC, Llauro C, Jobet E, Robakowska-Hyzorek D, Lasserre E, et al. Sequencing the extrachromosomal circular mobilome reveals retrotransposon activity in plants. PLoS Genet. 2017;13: e1006630. doi:10.1371/journal.pgen.1006630

46. Vitte C, Fustier MA, Alix K, Tenaillon MI. The bright side of transposons in crop evolution. Briefings Funct Genomics Proteomics. 2014;13: 276–295. doi:10.1093/bfgp/elu002

47. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. Nature Reviews Genetics. Nature Publishing Group; 2007. pp. 973–982. doi:10.1038/nrg2165

48. Flavell AJ, Ish-horowicz D. Extrachromosomal circular copies of the eukaryotic transposable element copia in cultured Drosophila cells. Nat 1981 2925824. 1981;292: 591–595. doi:10.1038/292591a0

49. Garfinkel DJ, Stefanisko KM, Nyswaner KM, Moore SP, Oh J, Hughes SH. Retrotransposon Suicide: Formation of Ty1 Circles and Autointegration via a Central DNA Flap. J Virol. 2006;80: 11920. doi:10.1128/JVI.01483-06

50. Yang F, Su W, Chung OW, Tracy L, Wang L, Ramsden DA, et al. Retrotransposons hijack alt-EJ for DNA replication and eccDNA biogenesis. Nat 2023. 2023; 1–8. doi:10.1038/s41586-023-06327-7

51. Flavell AJ. Role of reverse transcription in the generation of extrachromosomal copia mobile genetic elements. Nature. 1984;310: 514–516. doi:10.1038/310514A0

52. Thieme M, Lanciano S, Balzergue S, Daccord N, Mirouze M, Bucher E. Inhibition of RNA polymerase II allows controlled mobilisation of retrotransposons for plant breeding. Genome Biol. 2017;18: 134. doi:10.1186/s13059-017-1265-4

53. Minadakis N, Williams H, Horvath R, Caković D, Stritt C, Thieme M, et al. The demographic history of the wild crop relative *Brachypodium distachyon* is shaped by distinct past and present ecological niches. Peer Community J. 2023;3. doi:10.24072/PCJOURNAL.319

54. Hasterok R, Catalan P, Hazen SP, Roulin AC, Vogel JP, Wang K, et al. Brachypodium: 20 years as a grass biology model system; the way forward? Trends Plant Sci. 2022;27: 1002–1016. doi:10.1016/J.TPLANTS.2022.04.008

55. International Brachypodium Initiative. Genome sequencing and analysis of the model grass Brachypodium distachyon. Nature. 2010;463: 763–768. doi:10.1038/nature08747

56. Stritt C, Wyler M, Gimmi EL, Pippel M, Roulin AC. Diversity, dynamics and effects of long terminal repeat retrotransposons in the model grass *Brachypodium distachyon*. New Phytol. 2020;227: 1736–1748. doi:10.1111/nph.16308

57. Horvath R, Minadakis N, Bourgeois Y, Roulin AC. The evolution of transposable elements in Brachypodium distachyon is governed by purifying selection, while neutral and adaptive processes play a minor role. Elife. 2023;12. doi:10.7554/ELIFE.93284.1

58. Stritt C, Gimmi EL, Wyler M, Bakali AH, Skalska A, Hasterok R, et al. Migration without interbreeding: Evolutionary history of a highly selfing Mediterranean grass inferred from whole genomes. Mol Ecol. 2022;31: 70–85. doi:10.1111/mec.16207

59. Gordon SP, Contreras-Moreira B, Woods DP, Des Marais DL, Burgess D, Shu S, et al. Extensive gene content variation in the Brachypodium distachyon pan-genome correlates with population structure. Nat Commun. 2017;8: 1–13. doi:10.1038/s41467-017-02292-8

60. Smith CA, Vinograd J. Small polydisperse circular DNA of HeLa cells. J Mol Biol. 1972;69: 163–178. doi:10.1016/0022-2836(72)90222-7

61. Gaubatz JW. Extrachromosomal circular DNAs and genomic sequence plasticity in eukaryotic cells *. Mutat Res. 1990;237: 271–292.

62. Silander K, Saarela J. Whole genome amplification with Phi29 DNA polymerase to enable genetic or genomic analysis of samples of low DNA yield. Methods Mol Biol. 2008;439: 1–18. doi:10.1007/978-1-59745-188-8_1

63. Xu L, Yuan K, Yuan M, Meng X, Chen M, Wu J, et al. Regulation of Rice Tillering by RNA-Directed DNA Methylation at Miniature Inverted-Repeat Transposable Elements. Mol Plant. 2020;13: 851–863. doi:10.1016/J.MOLP.2020.02.009

64. Böhrer M, Rymen B, Himber C, Gerbaud A, Pflieger D, Laudencia-Chingcuanco D, et al. Integrated Genome-Scale Analysis and Northern Blot Detection of Retrotransposon siRNAs Across Plant Species. Methods Mol Biol. 2020;2166: 387–411. doi:10.1007/978-1-0716-0712-1_23

65. Quadrana L, Silveira AB, Mayhew GF, LeBlanc C, Martienssen RA, Jeddeloh JA, et al. The Arabidopsis thaliana mobilome and its impact at the species level. Elife. 2016;5: e15716. doi:10.7554/ELIFE.15716

66. Liu P, Nie WF, Xiong X, Wang Y, Jiang Y, Huang P, et al. A novel protein complex that regulates active DNA demethylation in Arabidopsis. J Integr Plant Biol. 2021;63: 772–786. doi:10.1111/jipb.13045

67. Muszewska A, Steczkiewicz K, Stepniewska-Dziubinska M, Ginalski K. Transposable elements contribute to fungal genes and impact fungal lifestyle. Sci Reports 2019 91. 2019;9: 1–10. doi:10.1038/s41598-019-40965-0

68. Gilbert C, Peccoud J, Cordaux R. Transposable Elements and the Evolution of Insects. https://doi.org/101146/annurev-ento-070720-074650. 2021;66: 355–372. doi:10.1146/ANNUREV-ENTO-070720-074650

69. Senft AD, Macfarlan TS. Transposable elements shape the evolution of mammalian development. Nat Rev Genet 2021 2211. 2021;22: 691–711. doi:10.1038/s41576-021-00385-1

70. Lisch D. How important are transposons for plant evolution? Nat Rev Genet. 2013;14: 49–61. doi:10.1038/nrg3374

71. Bourgeois Y, Boissinot S. On the Population Dynamics of Junk: A Review on the Population Genomics of Transposable Elements. Genes (Basel). 2019;10. doi:10.3390/genes10060419

72. Tsukahara S, Kobayashi A, Kawabe A, Mathieu O, Miura A, Kakutani T. Bursts of retrotransposition reproduced in Arabidopsis. Nature. 2009/09/08. 2009;461: 423–426. doi:10.1038/nature08351

73. Miura A, Yonebayashi S, Watanabe K, Toyama T, Shimada H, Kakutani T. Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. Nature. 2001;411: 212–214. doi:http://www.nature.com/nature/journal/v411/n6834/suppinfo/411212a0_S1.html

74. Mirouze M, Reinders J, Bucher E, Nishimura T, Schneeberger K, Ossowski S, et al. Selective epigenetic control of retrotransposition in Arabidopsis. Nature. 2009;461: 427–430. doi:10.1038/nature08328

75. Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, et al. Ten things you should know about transposable elements. Genome Biol 2018 191. 2018;19: 1–12. doi:10.1186/S13059-018-1577-Z

76. Bajus M, Macko-Podgórni A, Grzebelus D, Baránek M. A review of strategies used to identify transposition events in plant genomes. Front Plant Sci. 2022;13: 1080993. doi:10.3389/FPLS.2022.1080993/BIBTEX

77. Leyser O. Auxin Signaling. Plant Physiol. 2018;176: 465–479. doi:10.1104/PP.17.00765

78. Liu P, Nie WF, Xiong X, Wang Y, Jiang Y, Huang P, et al. A novel protein complex that regulates active DNA demethylation in Arabidopsis. J Integr Plant Biol. 2021;63: 772–786. doi:10.1111/JIPB.13045/SUPPINFO

79. Gordon SP, Contreras-Moreira B, Levy JJ, Djamei A, Czedik-Eysenberg A, Tartaglio VS, et al. Gradual polyploid genome evolution revealed by pan-genomic analysis of Brachypodium

hybridum and its diploid progenitors. Nat Commun. 2020;11: 3670. doi:10.1038/s41467-020-17302-5

80. Skalska A, Stritt C, Wyler M, Williams HW, Vickers M, Han J, et al. Genetic and methylome variation in Turkish brachypodium distachyon accessions differentiate two geographically distinct subpopulations. Int J Mol Sci. 2020;21: 1–17. doi:10.3390/ijms21186700

81. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25: 1754–1760. doi:10.1093/bioinformatics/btp324

82. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26: 841–842. doi:10.1093/bioinformatics/btq033

83. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2020. Available: https://www.r-project.org/

84. RStudio Team. RStudio: Integrated Development Environment for R. Boston, MA; 2016. Available: http://www.rstudio.com/

85. Suzuki R, Shimodaira H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics. 2006;22: 1540–1542. doi:10.1093/BIOINFORMATICS/BTL117

86. Vu VQ. ggbiplot: A ggplot2 based biplot. R package version 0.55. 2011.

87. Bates D, Mächler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015;67: 1–48. doi:10.18637/jss.v067.i01

88. Nakagawa S, Schielzeth H. A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods Ecol Evol. 2013;4: 133–142. doi:10.1111/J.2041-210X.2012.00261.X

89. Wickham H. ggpolt2 Elegant Graphics for Data Analysis. Use R! Ser. 2016; 211.

90. Ferrafiat L, Pflieger D, Singh J, Thieme M, Böhrer M, Himber C, et al. The NRPD1 N-terminus contains a Pol IV-specific motif that is critical for genome surveillance in Arabidopsis. Nucleic Acids Res. 2019;47: 9037–9052. doi:10.1093/nar/gkz618

91. Stonaker JL, Lim JP, Erhard KF, Hollick JB. Diversity of Pol IV Function Is Defined by Mutations at the Maize rmr7 Locus. PLOS Genet. 2009;5: e1000706. doi:10.1371/JOURNAL.PGEN.1000706

92. Dalmais M, Antelme S, Ho-Yue-Kuang S, Wang Y, Darracq O, d'Yvoire MB, et al. A TILLING Platform for Functional Genomics in Brachypodium distachyon. PLoS One. 2013;8: 65503. doi:10.1371/JOURNAL.PONE.0065503

93. Bragg JN, Wu J, Gordon SP, Guttman ME, Thilmony R, Lazo GR, et al. Generation and Characterization of the Western Regional Research Center Brachypodium T-DNA Insertional Mutant Collection. PLoS One. 2012;7: e41916. doi:10.1371/JOURNAL.PONE.0041916

94. Hong SY, Seo PJ, Yang MS, Xiang F, Park CM. Exploring valid reference genes for gene expression studies in Brachypodium distachyon by real-time PCR. BMC Plant Biol. 2008;8: 112. doi:10.1186/1471-2229-8-112

95. Bortiri E, Coleman-Derr D, Lazo GR, Anderson OD, Gu YQ. The complete chloroplast genome sequence of Brachypodium distachyon: Sequence comparison and phylogenetic analysis of eight grass plastomes. BMC Res Notes. 2008;1: 61. doi:10.1186/1756-0500-1-61

96. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10: 1–4. doi:10.1093/gigascience/giab008

97. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes De Novo Assembler. Curr Protoc Bioinforma. 2020;70: e102. doi:10.1002/cpbi.102

98. Shen W, Le S, Li Y, Hu F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLoS One. 2016;11: e0163962. doi:10.1371/journal.pone.0163962

99. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016;44: W160–W165. doi:10.1093/NAR/GKW257

100. Hahne F, Ivanek R. Visualizing genomic data using Gviz and bioconductor. Methods in Molecular Biology. Humana Press Inc.; 2016. pp. 335–351. doi:10.1007/978-1-4939-3578-9_16

101. Lawrence M, Gentleman R, Carey V. rtracklayer: An R package for interfacing with genome browsers. Bioinformatics. 2009;25: 1841–1842. doi:10.1093/bioinformatics/btp328

102. Chen S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. iMeta. 2023;2: e107. doi:10.1002/IMT2.107

103. Jeong H-H, Yalamanchili HK, Guo C, Shulman JM, Liu Z. An ultra-fast and scalable quantification pipeline for transposable elements from next generation sequencing data. Biocomput 2018. 2018; 168–179. doi:10.1142/9789813235533_0016

104. Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, et al. PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res. 2017;45: D1040–D1045. doi:10.1093/nar/gkw982

105. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. Bioinformatics. 2011;27: 1017–1018. doi:10.1093/bioinformatics/btr064

106. Tello-Ruiz MK, Naithani S, Gupta P, Olson A, Wei S, Preece J, et al. Gramene 2021: Harnessing the power of comparative genomics and pathways for plant research. Nucleic Acids Res. 2021;49: D1452–D1463. doi:10.1093/nar/gkaa979

107. Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, et al. AgriGO v2.0: A GO analysis toolkit for the agricultural community, 2017 update. Nucleic Acids Res. 2017;45: W122–W129. doi:10.1093/nar/gkx382

108. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. Gibas C, editor. PLoS One. 2011;6: e21800. doi:10.1371/journal.pone.0021800

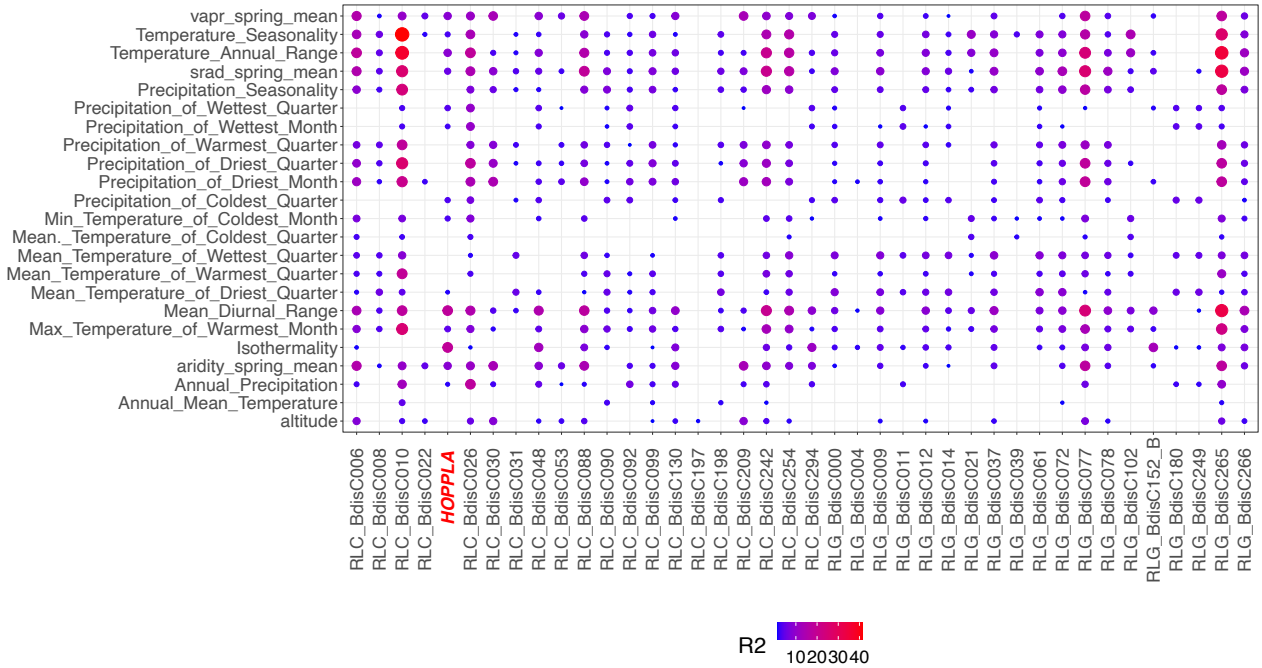
109. Diesh C, Stevens GJ, Xie P, De Jesus Martinez T, Hershberg EA, Leung A, et al. JBrowse 2: a modular genome browser with views of synteny and structural variation. Genome Biol. 2023;24: 1–21. doi:10.1186/S13059-023-02914-Z/FIGURES/11

110. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet 2012 447. 2012;44: 821–824. doi:10.1038/ng.2310

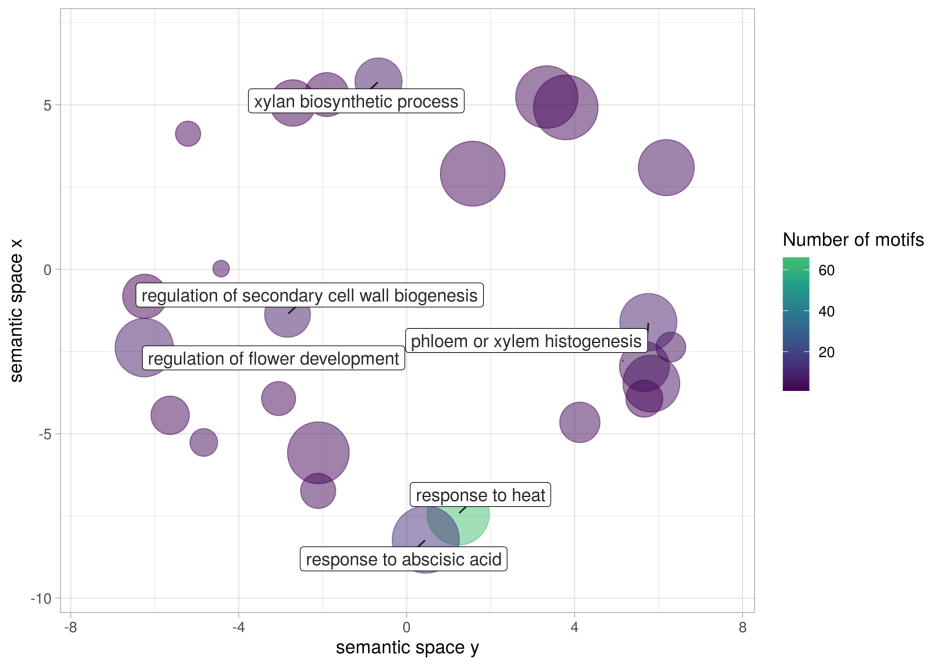
111. Gautier M, Vitalis R. rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. Bioinformatics. 2012;28: 1176–1177. doi:10.1093/BIOINFORMATICS/BTS115

112. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics. 2017;33: 2938–2940. doi:10.1093/BIOINFORMATICS/BTX364

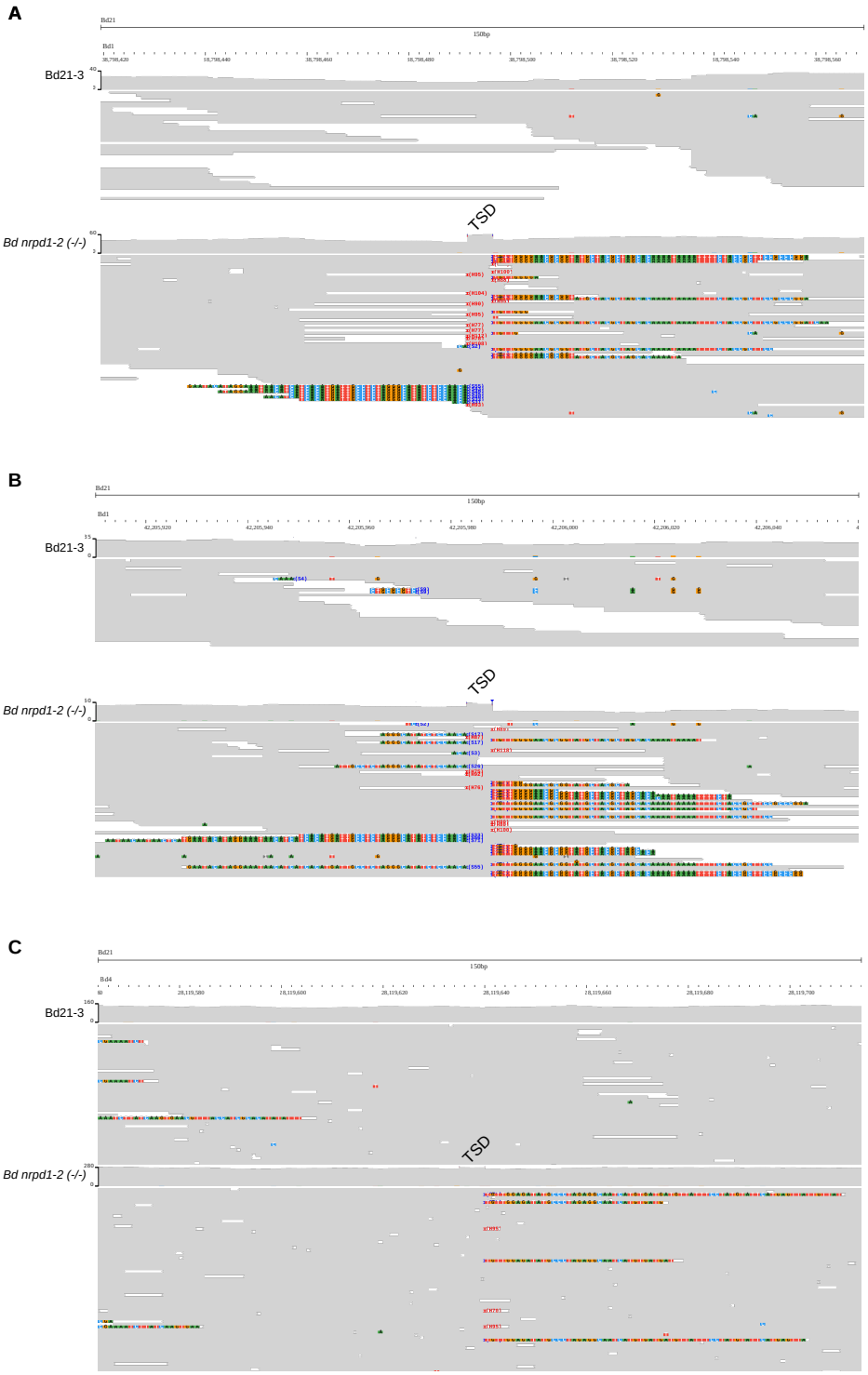
Supporting Information - Figures



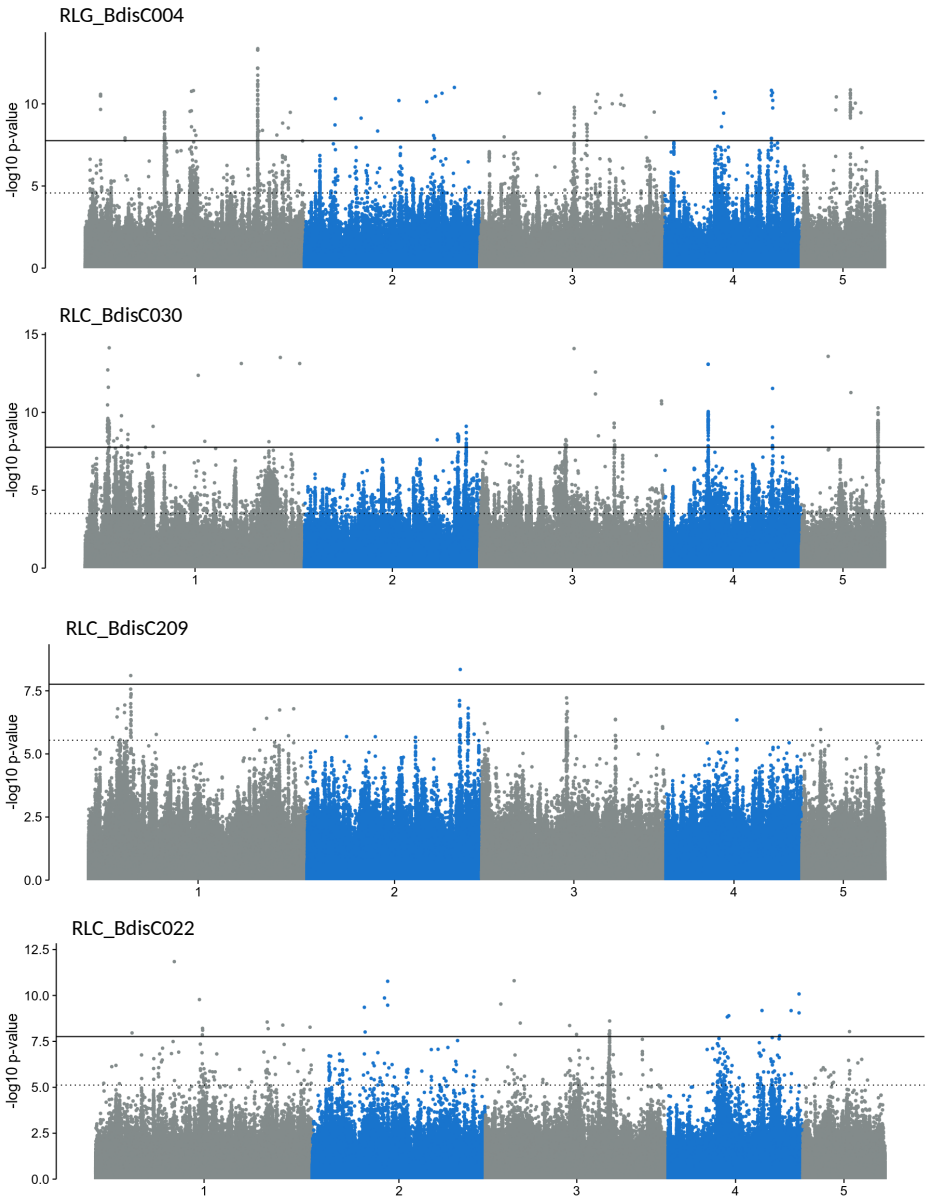
S1 Fig. Correlation of bioclimatic variables with pCNs when not correcting for population structure. Colors and sizes of bubbles show the part of the variance (R^2) explained by the bioclimatic variables in %



S2 Fig. TFs binding to *ONSEN* are heat-stress inducible. GO-enrichment analysis of transcription factors for which binding sites have been detected in AT1G11265, a member of the heat-responsive *ONSEN* (*ATCOPIA78*) LTR-RT family in *A. thaliana*. 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 is listed in **S1 Table**.



S3 Fig. TIPs of *HOPPLA* in one of the resequenced *Bd nrpd1-2 (-/-)* plants. JBrowse screenshot of three insertion sites (**A-C**) 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.



S4 Fig. Manhattan plots of the GWASs of pCNs of four recently active LTR-RT families. From top: RLG_BdisC004, 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

Supporting Information – Table legends

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.

S2 Table. Normalized pCNs of LTR-RTs and bioclimatic variables of the 320 natural accessions of *B. distachyon*

S3 Table. Gene list of pCNs GWAS with different levels of significance (FDR < 0.05, BC) and window sizes (20 kb, 50 kb)

S4 Table. Components of the Pol IV and Pol V holoenzymes in *B. distachyon*

S5 Table. Sequences of oligos used in this study

S6 Table. Meta information of mobilome-seq samples