Nucleic Acids Research, 2024, **52**, 1226–1242
https://doi.org/10.1093/nar/gkad1214
Advance access publication date: 24 December 2023
Genomics

Recent reactivation of a pathogenicity-associated transposable element is associated with major chromosomal rearrangements in a fungal wheat pathogen

Thomas Badet[1], Sabina Moser Tralamazza[1], Alice Feurtey[1,2] and Daniel Croll[1,*]

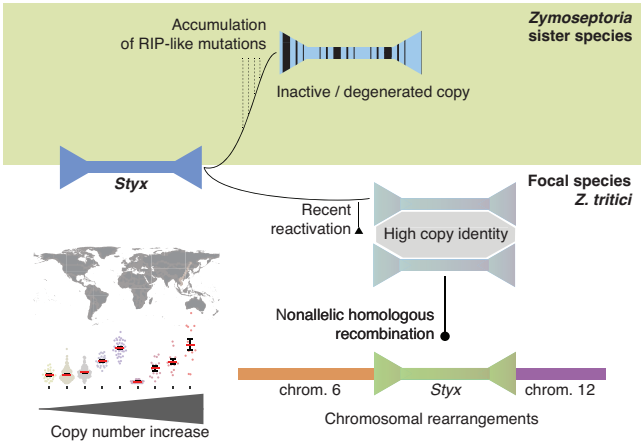
[1]Laboratory of Evolutionary Genetics, Institute of Biology, University of Neuchâtel, CH-2000 Neuchâtel, Switzerland
[2]Plant Pathology, D-USYS, ETH Zurich, CH-8092 Zurich, Switzerland

*To whom correspondence should be addressed. Tel: +41 32 718 23 30; Fax: +41 32 718 30 01; Email: daniel.croll@unine.ch

Abstract

Transposable elements (TEs) are key drivers of genomic variation contributing to recent adaptation in most species. Yet, the evolutionary origins and insertion dynamics within species remain poorly understood. We recapitulate the spread of the pathogenicity-associated *Styx* element across five species that last diverged ∼11 000 years ago. We show that the element likely originated in the *Zymoseptoria* fungal pathogen genus and underwent multiple independent reactivation events. Using a global 900-genome panel of the wheat pathogen *Zymoseptoria tritici,* we assess *Styx* copy number variation and identify renewed transposition activity in Oceania and South America. We show that the element can mobilize to create additional *Styx* copies in a four-generation pedigree. Importantly, we find that new copies of the element are not affected by genomic defenses suggesting minimal control against the element. *Styx* copies are preferentially located in recombination breakpoints and likely triggered multiple types of large chromosomal rearrangements. Taken together, we establish the origin, diversification and reactivation of a highly active TE with likely major consequences for chromosomal integrity and the expression of disease.

Graphical abstract



Introduction

Transposable elements (TEs) are major constituents of most eukaryotic genomes. Bursts of TE amplification are major drivers of genome evolution as they can create vast adaptive genomic variation (1). However, the repetitive nature of TEs also promotes genome instability through non-allelic homologous recombination potentially triggering chromosomal rearrangements (2). TE mobilization also causes deleterious mutations by the disruption or modification of coding regions that can affect host fitness (3–5). TE sequences relocate to novel loci directly through either DNA or RNA intermediates that often encode the proteins required for sequence excision and integration (1). Highly successful TEs include the short interspersed nuclear element *Alu*, which is present in >1 million copies in the human genome (6). *Alu*s are non-autonomous elements that hijack the retrotransposition machinery from another retrotransposon for their own mobilization. Similar mechanisms are associated with the recent expansion of the *mPing* miniature inverted-repeat TE in rice (7). Delegating the transposition catalysis to other transcriptionally active elements is an effective strategy for TEs to rapidly increase in copy number. However, retracing the precise movements of

Received: April 9, 2023. **Revised:** November 30, 2023. **Editorial Decision:** December 7, 2023. **Accepted:** December 11, 2023

TEs across the genome remains challenging due to low sequence complexity of insertion sites and the lack of fully resolved genome assemblies.

A major constraint on TE proliferation in the genome is the activity of the host epigenetic machinery, including histone modifications, DNA methylation and RNA interference suppressing active TE copies (8). In rice, both *Ping* and *Pong* can catalyze the transposition of *mPing in vitro*, yet expression of the *Pong* element is effectively repressed by the host silencing machinery so that only the *Ping* element contributes to mobilization (7,9). To counteract host epigenetic control, some TEs encode their own promoter sequences. The *Ty1/Copia*-like retrotransposon *ONSEN* mobilizes upon heat stress in Brassicaceae species and carries a heat-inducible promoter sequence (10,11). Stress-induced activation has been reported for other TEs across kingdoms and is often associated with transposition bursts (12–15). Activated *ONSEN* elements show an insertion preference for GC- and gene-rich regions (16). Insertion bias toward gene-rich sequences likely facilitates the escape from epigenetic control and promotes sustained transposition activity (17).

Independently of transcriptional control, mutations accumulating in TE sequences can dramatically alter transposition rates. Eliminating mutations in the *Sleeping Beauty* Tc1/mariner transposon reactivated the TE in salmon genomes suggesting that these mutations were deleterious to transposon activity (18). In contrast, introducing point mutations at the 3′ terminal inverted repeat (TIR) of the non-autonomous *mPing* element in the rice genome reduces excision events 10-fold (19). Ascomycete fungi have evolved a unique mechanism of mutation-driven TE control called repeat-induced point mutations (RIP) (20). Targeted mutagenesis of TE copies disrupts coding sequence integrity through very high rates of C→T transitions. Triggered by the presence of repeated sequences in the genome, RIP can efficiently eliminate the transposition potential of TEs. Understanding how epigenetic factors and mutation accumulation interact to counter TE invasions is key to our understanding of TE expansion dynamics.

In this work, we identify the factors driving the invasion of a highly active transposon in the fungal pathogen *Zymoseptoria tritici*. The pathogen attacks wheat and causes global losses in wheat production (21,22). The species experienced population-level bursts of TEs linked to the creation of genomic diversity and rapid adaptation (23–26). Despite active RIP and RNA interference, some *Gypsy* retroelements and class II TIR elements have recently accumulated high copy numbers (23). A substantial number of TEs undergo derepression when the pathogen colonizes the plant host suggesting that the pathogen lifestyle constitutes risks to effective TE control (14). Recent work has shown that transcription of a DNA class II transposon named *Styx* negatively impacts asexual reproduction (27). *Styx* was also linked to a large-scale chromosomal rearrangement and copy number variation among field isolates (27). The propensity to create new copies in the recent evolutionary history of the species makes *Styx* transposon an ideal element to investigate how TEs activate and propagate in genomes.

Here, we recapitulate the evolutionary history of the *Styx* element across recent speciation events using reference-quality genomes of the *Zymoseptoria* plant pathogen genus. We focus on *Z. tritici* showing the most recent *Styx* activity. *Styx* copy number estimates derived from a 900-genome panel reveal two substantial recent expansions in geographically restricted populations. Using a four-generation pedigree with completely assembled genomes, we track active transposition of *Styx*. We show that *Styx* triggered several distinct rearrangement types, including deletions and duplications of flanking regions, as well as chromosomal fusions. In conjunction, we establish how a recently reactivated TE escapes genomic defenses and triggers genomic rearrangements at observable rates.

Materials and methods

TE annotation and insertion genotyping

We accessed previously published complete genome assemblies of 20 reference-quality genomes of *Z. tritici* spanning the global distribution range of the pathogen (28,29) (Supplementary Table S1). In addition, we analyzed four complete genomes of sister species *Z. passerinii*, *Z. ardabiliae*, *Z. brevis* and *Z. pseudotritici* (28) (Supplementary Table S1). We also analyzed 10 complete genomes assembled from a four-generation pedigree started by the isolates 1A5 and 1E4 (30) (Supplementary Table S1). We annotated all genomes for TEs using the method and consensus sequences described in (29). In addition, for *Styx*, we implemented an additional annotation based on the 36-bp TIR as described in (27) to identify conserved copies across the 34 genome assemblies using the R package *packFinder* version 1.6.0 (31). We performed the *packSearch* in R version 4.1.3 using the following parameters: *tsdMismatch* = 6, *mismatch* = 8, *elementLength* = c(100, 30000) and *tsdLength* = 6. TIR positions were deduced using the *identifyTirMatches* function of *packFinder*. Single TIRs with no co-occurring pair within 30 kb distance were considered solo-TIRs.

For the TEs DTA_Vera, DTT_Tapputi and DXX_Birute, we first performed self-alignments of their consensus sequences to identify TIRs using blastn with *-word_size 4 -perc_identity 85 -dust no* as parameters (fasta sequences are provided in Supplementary Data S1). The TIR was then used to retrieve copies of each element in the parent and progeny genomes of the pedigree using packFinder with parameters adjusted to the element length [*mismatch* = round(*TIR_length* * 0.15), *tsdMismatch* = round(*TSD_length* * 0.4), *elementLength* = c(*TE_length* − 5, *TE_length* + 5) and *tsdLength* = c(2:6)]. Identified *Styx* loci and target site duplications were then used to identify homologous positions among *Styx* copies in genomes of the pedigree. For the high-copy TE RLC_Deimos, we used RepeatMasker to annotate progeny genomes retaining only annotations covering ≥95% of the element consensus sequence tolerating elements exceeding length by ≤5% (29). To identify copies at homologous positions, we considered five additional base pairs on each side of the Deimos sequences as putative TSD.

To investigate genomic characteristics near TE loci, we binned each genome of the pedigree into 10 000-bp windows to compute gene, TE and GC coverage. We used gene and TE annotations from the Badet *et al.* (29) study together with the bedtools *coverage* function. Similarly, we used the bedtools *nuc* function to compute GC content across all genomic 10-kb windows.

Sequence alignments and phylogenetic tree inference

All sequence alignments were performed with the software MAFFT v 7.475 using the –maxiterate 1000 and –leavegappyregion parameters (32). The TIR sequences were recovered based on the positions provided by the *identifyTirMatches* function of the *packFinder* R package and aligned using MAFFT. Sequence logos were calculated using the Berkeley WebLogo server for the aligned TIR sequences from *Z. passerinii* and *Z. tritici* separately (33). For phylogenetic reconstruction, all detected copies of *Styx* across the five *Zymoseptoria* species were aligned jointly with either the *Styx* consensus sequence or the coding sequence of the integrase locus. The resulting alignments were trimmed using the *remove_reference_gaps_in_alignment.pl* script (https://github.com/lakras/bio-helper-scripts/blob/main/aligned-fasta) to the consensus sequence or the integrase gene. Phylogenetic trees were then inferred with FastTree v2.1.11_1 software using 1000 bootstraps and the generalized time-reversible model (-nt -boot 1000 -seed 1253 -gtr options) (34). Variable positions in the sequence alignments were identified in the trimmed alignments using the *snp-sites* v2.5.1 software (35). Resulting vcf files were simplified to table formats using the *VcfSimplify.py* script (https://github.com/everestial/VCF-Simplify). Biallelic single-nucleotide polymorphisms (SNPs) with C or G as reference allele and T or A as alternative allele, respectively, were considered RIP-like mutations. Tabular data were analyzed in R version 4.1.2 and visualized using the *ggplot2* v3.3.5 package (36).

Identification of recombination blocks in the pedigree

To recapitulate recombination events across the pedigree, we assigned chromosomal blocks in each of the 10 progeny genomes to the parent of origin (1A5 or 1E4) using pairwise whole-genome alignments with *nucmer* (MUMmer version 4.0.0rc1) (37). For the alignment step, we used a minimum length of a single exact match of 100 bp, a maximum gap between two matches in a cluster of 10 bp, a minimum cluster of matches of 1000 and an alignment extension in poorly scoring regions of 200 bp (options -l 100 -g 10 -c 1000 -b 200). We filtered the resulting alignments for a minimum identity of 99% and a minimum length of 50 000 bp using *delta-filter* (options -i 99 -l 50000). The filtered alignment coordinates were then converted into a tab-delimited format with the *show-coords* tool (-THrd options). To assign parental origins of progeny chromosomal blocks, we concatenated filtered match coordinates for both parental genomes. We assigned the parent of origin for a chromosomal block based on the presence of matches satisfying the >50 000 bp alignment length and ≥99% identity criteria. To assign the parental origin of individual *Styx* copies in progeny genomes, we matched the chromosomal coordinates of *Styx* copies with the chromosomal block information on the parent of origin using *bedtools intersect* version v2.30.0 (38). Shifts in homology from one parent to the other along progeny chromosomes were considered as the most likely recombination breakpoints. To reduce false positives, alignment blocks where both parents showed >99.8 percentage identity and one of the two parental alignment overlapped by >50% of their length were left out. Aligned blocks were only retained if the length exceeded 100 kb. Finally, predicted recombination breakpoints separated by <20 kb were merged into a single recombination event given that closely spaced recombination breakpoints are unlikely due to crossover interference (39).

Styx expression analysis

We investigated *Styx* expression in the *Zymoseptoria* genus using RNA sequencing (RNA-seq) data generated in axenic growth conditions. Datasets from 17 *Z. tritici* isolates were downloaded from the National Center for Biotechnology Information (NCBI) BioProject PRJNA559981. In addition, we used Sequence Read Archive (SRA) RNA-seq experiment accessions SRX4341756, SRX4341752 and SRX4341751 for isolate 1E4 and SRX4341748, SRX4341749 and SRX4341750 for isolate 1A5. For *Z. pseudotritici*, *Z. ardabiliae*, *Z. brevis* and *Z. passerinii*, RNA-seq datasets were retrieved from the NCBI BioProjects PRJNA277173, PRJNA277174, PRJNA277175 and PRJNA639021, respectively. The raw reads were trimmed using Trimmomatic v0.39 and mapped to the genome matching the source of the data using STAR v2.7.10a/2.7.9a while allowing for multiple mapped reads (*-outFilterMultiNmax* 100 -*winAnchorMultimapNmax* 200 parameters) (40,41). Gene and transposon family expression were assessed using TEtranscripts v2.2.1 (*-mode* multi) (42). For downstream analyses, read counts were normalized to counts per million of reads in R using the *calcNormFactors* function from the *edgeR* package v3.34.1 (method = 'TMM') (43).

Chromatin immunoprecipitation sequencing

We performed chromatin immunoprecipitation sequencing (ChIP-seq) analyses of the histone modifications H3K27me3 and H3K27me2 of *Z. tritici* isolates and two sister species (*Z. brevis* and *Z. pseudotritici*) cultured in carbon-limited medium (i.e. minimal medium) to match RNA-seq assay conditions (NCBI accession PRJNA1036405). Isolates were grown in Vogel's Medium N (minimal medium) until hyphae formation for 8 days at 18°C. Chromatin immunoprecipitation was performed according to a previously established protocol (44). The ChIP-seq library was prepared for sequencing and analyzed using NovaSeq 6000 in paired-end mode with a read length of 150 bp. ChIP-seq raw reads were trimmed using Trimmomatic v.0.32 with parameters ILLUMINACLIP:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36 (41). For each ChIP-seq dataset, reads were mapped to the corresponding genome assembly with bowtie2 v.2.4 and the –very-sensitive-local parameter (45). Duplicated reads were removed with the GATK Picard MarkDuplicates function v.4.2.4.1 (46). ChIP-seq peak calling was performed with Homer v. 4.11 with parameters makeTagDirectory -keepOne -mapq 10 and findPeaks -style histone -region -size 800 -minDist 1600 -C 0 (47). Quality as fraction of reads in peak was assessed with the parse2wigdir+ function from DROMPAplus (48). For both investigated methylation marks, we computed per gene coverage using the *bed_coverage* function from the *valr* R package v0.7.0.

Estimation of *Styx* copy numbers in genome resequencing datasets

To assess the geographic and temporal variation in *Styx* copy numbers in *Z. tritici*, we used a 923-genome panel of Illumina sequenced genomes (49). Accession numbers for the NCBI

SRA are available from (49). The collection covers 42 countries and all continents where the pathogen has been recorded. We trimmed raw Illumina reads to remove adapters and retain only high-quality bases (LEADING:15 TRAILING:15 SLIDINGWINDOW:5:15 MINLEN:50) with Trimmomatic v.0.39 (41). We used the *Styx* consensus sequence (29) as a reference sequence to map reads (as single reads ignoring read pair information) with the option –very-sensitive-local using bowtie2 v.2.4.1 (45). Read alignments per analyzed genome were used to derive the number of reads aligning to the *Styx* consensus with the *idxstats* option of *samtools* v.1.10 (50).

Results

Evolutionary origins and reactivation across speciation events

The *Styx* transposon is one of the two most active transposons in the fungal wheat pathogen *Z. tritici* shown to recently create new copies and segregate significant copy number variation within the species (49,51). *Styx* is restricted to the *Zymoseptoria* genus (27). To recapitulate the evolutionary history of the element, we searched 20 reference-quality genomes defining the global pangenome of *Z. tritici* and 4 genomes of additional species of the genus (28,29). *Zymoseptoria tritici* diverged from its closest sister species *Z. pseudotritici* ∼11 000 years ago (52). *Zymoseptoria tritici* originated in the Middle East and subsequently colonized North Africa and Europe with the latest spread occurring in the Americas and Oceania following the establishment of wheat cultivation in the past centuries (49,53). Individual *Zymoseptoria* sp. genomes carry between 0 and 34 *Styx* copies for a total of 200 discovered copies (Figure 1A and Supplementary Tables S1 and S2). The highest number of copies was found in an Argentinian *Z. tritici* isolate (Arg00; 34 copies) followed by the sister species *Z. pseudotritici* and *Z. passerinii* (25 and 23 copies, respectively). Isolates sampled near the center of origin of *Z. tritici* carry no copies of the element (IR01_26b, IR01_48b, ISY92, KE94, TN09, YEQ92). High copy numbers of *Styx* in sister species and the absence of *Styx* in center-of-origin populations suggest that *Styx* was nearly eliminated from *Z. tritici* early in speciation. *Styx* carries highly conserved TIRs with 97–100% identity overall. Single TIRs (i.e. TIRs without a matching partner) are present in all genomes of the genus, including in the *Z. tritici* center-of-origin populations (Figure 1A). Alignment of TIR sequences delineates two groups, with one group being exclusive to *Z. passerinii* and one group shared among the other members of the genus. The two TIR groups diverge at a 10-bp motif in the center of the TIR, suggesting that the two TIR groups define independent transposition activity of *Styx* (Figure 1B).

The majority (∼52%) of *Styx* copies are 7927 bp in size corresponding to the consensus sequence (Figure 1C). The *Styx* carries four open reading frames, one encoding the putative transposase, two genes with putative DNA binding motifs (GATA and CCHC domains) and a gene with a domain of unknown function (DUF3577) (Figure 1D). Sequences of atypical length are strongly differentiated and likely constitute degenerated copies. The *Z. pseudotritici* genome carries 20 copies ranging from 8232 to 8240 bp in length sharing a unique 285 bp insertion flanked by a tandem repeat of 32 bp suggesting the emergence of a *Styx* subfamily (Figure 1D). Four *Z. ardabiliae* copies carry a similar insertion consistent with a *Styx* subfamily diversification within the genus. The longest *Styx* variant is shared among two species and may constitute a derivative of the ancestral *Styx* present at the origin of the genus. The expansion outside of the *Z. tritici* center of origin was driven by a shorter variant. We performed an ancestral state reconstruction of long versus short *Styx* sequence variants to assess the evolutionary history of *Styx* variants (Figure 1E). The ancestral *Styx* element was most likely shorter lacking the characteristic 285 bp sequence insertion shared by *Z. pseudotritici* and *Z. ardabiliae*. Sequences of the transposase transcript (as defined by the *Z. tritici* consensus) cluster at the species level except for a copy in the genome of the *Z. tritici* strain IPO323, which may represent an ancestral polymorphism shared with *Z. ardabiliae* (Figure 1E). The highest degree of conservation of the transposase locus is found in *Z. tritici* among isolates from Europe, the Americas and Australia with <10 mutations overall. All four open reading frames of the *Styx* are transcribed both on the host and in culture condition in *Z. tritici* (27). Yet, we found the *Styx* element to be largely silenced in culture condition in the sister species (Figure 1F). In contrast with the low expression in these sister species, the high sequence similarity of *Styx* copies within *Z. tritici* is coinciding with high levels of transcription (Figure 1F). Consistent with the observed silencing, no sister species *Styx* carries an intact integrase coding sequence.

TEs can be silenced by the host genome through chromatin reorganization, often mediated by the methylation of the lysine at position 27 of the histone 3 (54–56). To address the putative impact of *Styx* insertions on the chromatin landscape, we performed ChIP-seq for two histone 3 methylation marks, H3K27me3 and H3K4me2, and on 17 *Z. tritici* isolates with chromosome-level genome assemblies. *Styx* copies show evidence for both H3K4me2 and H3K27me3 histone methylation (Figure 2A–E and Supplementary Table S3). We analyzed histone methylation patterns of genes adjacent to a polymorphic insertion locus segregating *Styx* presence/absence variation among the 17 *Z. tritici* isolates. Histone methylation marks among single-copy orthologous genes next to *Styx* insertions were largely conserved independent of the presence of *Styx* (Figure 2B and C). Consistent with these findings, gene expression was also largely stable across orthologous genes and independent of the presence of *Styx* (Figure 2D). We investigated three outliers for changes in gene expression associated with a *Styx* insertion. The distance between the coding sequence and *Styx* was 53, 254 and 261 bp, respectively, for genes encoding a leucine carboxyl methyltransferase, a GNAT acetyltransferase and a sporulation protein, respectively. The changes in gene expression following *Styx* insertion were >7-fold (Figure 2D and Supplementary Table S3). However, we found that the difference in gene expression did not correlate with changes in H3K27me3 or H3K4me2 histone methylation marks. We found the most significant changes in H3K27me3 marks in coding sequences following a *Styx* insertion in two genes encoding a metallohydrolase/oxidoreductase and a short protein of unknown function, respectively (Supplementary Table S3). For plant pathogens such as *Z. tritici*, effectors perform essential functions during host infection. We identified 15 genes encoding predicted effectors within 25 kb of a *Styx*

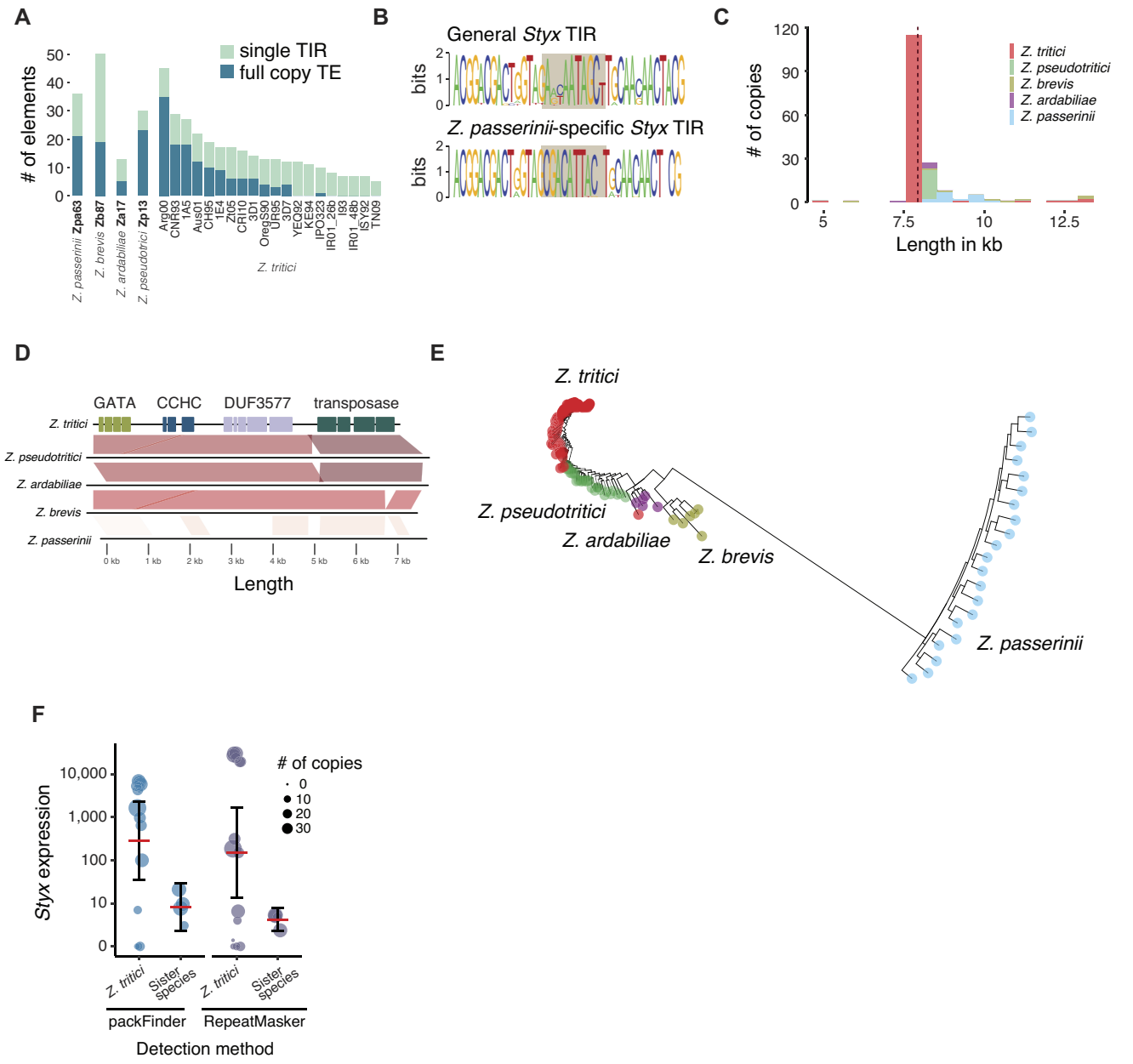


Figure 1. Structure, evolution and expression dynamics of the *Styx* TE in the *Zymoseptoria* genus of plant pathogens. (**A**) Number of *Styx* copies identified in 24 complete genomes of the *Zymoseptoria* genus. Full elements and single TIRs are shown separately. (**B**) Sequence logo showing the conservation of the TIR sequence identified in *Z. tritici* and *Z. passerinii,* respectively. (**C**) Length variation of *Styx* copies across the four *Zymoseptoria* species restricting for copies shorter than 13 kb. (**D**) Synteny of the five representative *Styx* sequences from *Z. tritici, Z. pseudotritici, Z. ardabiliae, Z. brevis* and *Z. passerinii* showing syntenic sequences and the insertion present in *Z. pseudotritici*. Darker shades indicate higher sequence identity. (**E**) Phylogenetic tree based on the integrase coding sequence of the different *Styx* copies. The tree was rooted using the single degenerate copy present in the *Z. tritici* reference isolate IPO323. (**F**) Expression in counts per million of the *Styx* element as annotated by packFinder and RepeatMasker in *Z. tritici* isolates and the four other sister species. Dot size shows the element copy number in the respective isolate.

insertion and 13 of these genes showed variation in gene expression or histone methylation marks associated with the *Styx* insertion (Supplementary Table S3).

In sister species, most *Styx* copies carry no facultative repressive or euchromatic histone methylation marks in contrast to *Z. tritici* (Figure 3A and Supplementary Table S4). This is consistent with the largely silenced state of *Styx* in sister species genomes and indicates that *Styx* max be covered by obligate repressive histone methylation marks. In *Z. passerinii*, *Styx* copies are highly divergent consistent with the accumulation of random mutations (Figure 3B). In contrast, *Z. tritici* copies are highly similar and most divergence was caused by mutations typically generated by RIP genomic defenses (i.e. C→T and G→A transitions; Wilcoxon rank sum test *P*-value <0.05). Around 38% of *Z. tritici Styx* copies retained a pairwise identity ≥99%, consistent with ongoing mobilization of the element and weak effect of RIP (Figure 3C). In *Z. ardabiliae*, *Z. passerinii* and *Z. pseudotritici*, copies share a pairwise identity between 90% and 99%. In *Z. brevis*, 18% of the *Styx* copies share ≥90% identity and only ∼4% of the copies share ≥99% identity. In conjunction, the degree of sequence conservation, copy number increase and transcriptional activity strongly suggest that *Styx* was recently reactivated in *Z. tritici*.

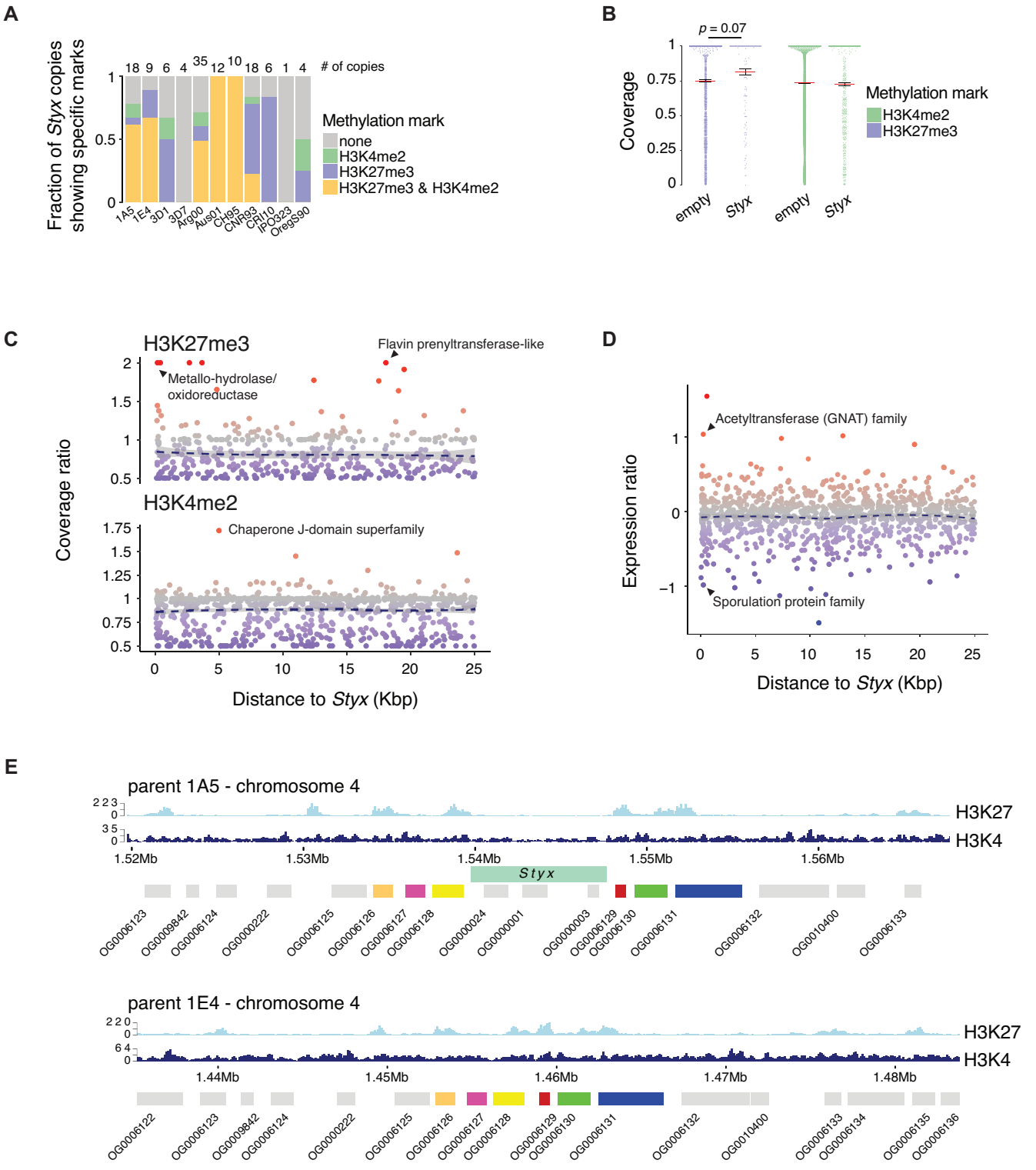


Figure 2. Histone methylation marks near *Styx* loci in *Z. tritici*. (**A**) Fraction of the total number of *Styx* elements showing H3K4me2, H3K27me3, both or no methylation marks. Numbers on top indicate the total number of *Styx* copies present in the isolate. (**B**) Single-copy gene coverage 25 kb up- and downstream of *Styx* insertion loci for H3K4me2 and H3K27me3 methylation marks in *Z. tritici* isolates with or without *Styx* present. The *P*-value of the post-hoc Tukey honestly significant difference (HSD) test for the analysis of variance is shown for H3K27me3 marks. (**C**) Single-copy genes 25 kb up- and downstream of *Styx* insertion loci. H3K4me2 and H3K27me3 coverage ratios, expressed as the coverage in isolates with the *Styx* insertion divided by the coverage in isolates without the insertion, are shown as a function of the distance to the *Styx* insertion. (**D**) Single-copy genes 25 kb up- and downstream of *Styx* insertion loci. The expression ratio was calculated as the $\log_{10}$ of the expression in isolates with the *Styx* insertion divided by the expression in isolates without the insertion. (**E**) Case study of a syntenic locus showing *Styx* presence (in isolate 1A5) and absence (in isolate 1E4). RNA-seq read and histone methylation mark (H3K4me2 and H3K27me3) coverages are shown. Read density is expressed as the number of reads overlapping 100-bp windows. The lower boxes represent orthologous genes identified in both genomes. The top rectangle represents the *Styx* element present in the 1A5 isolate.

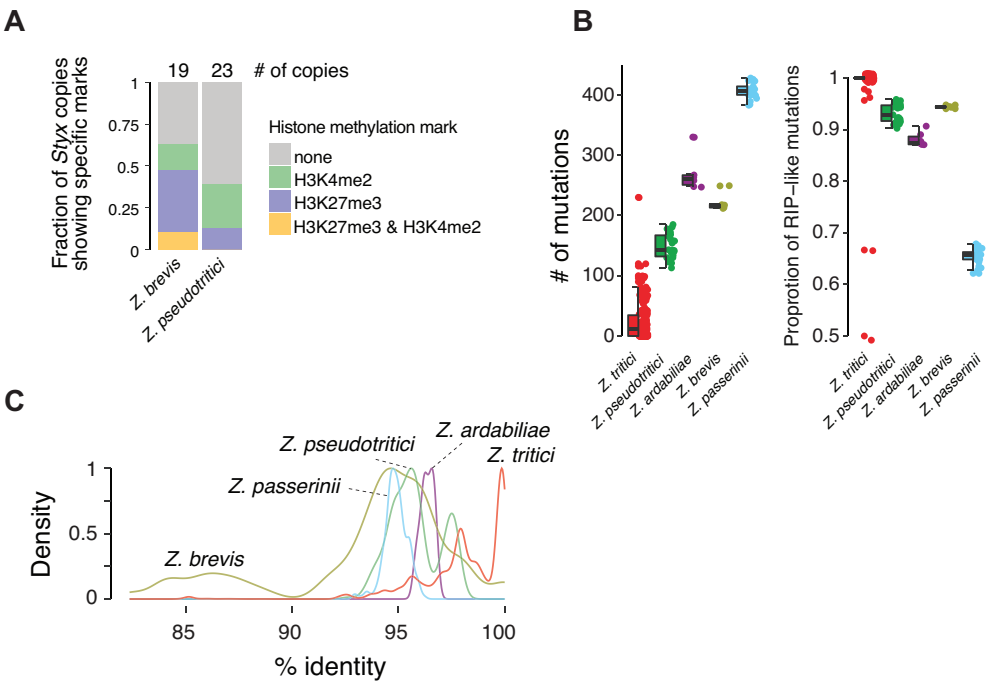


Figure 3. *Styx* copies and signatures of decay among *Zymoseptoria* sister species. (**A**) Fraction of the total number of *Styx* elements showing H3K4me2, H3K27me3, both or no methylation marks in four species of the *Zymoseptoria* genus. Numbers on top indicate the total number of *Styx* copies present in the genome. (**B**) Number of mutations detected in the integrase coding sequence for all *Styx* copies in the five *Zymoseptoria* species. C→T and G→A transitions are considered RIP-like mutations. (**C**) Pairwise identity based on reciprocal blast search of *Styx* copies. The analysis was performed individually for each of the five *Zymoseptoria* species.

Recent *Styx* expansion in Southern Hemisphere populations

To recapitulate the temporal and spatial dynamics of the *Styx* expansion in *Z. tritici*, we analyzed a 923-genome panel of isolates collected from wheat fields across the world. To account for limitations of short-read sequenced genomes, we mapped individual sequencing reads to TE consensus sequences to estimate copy numbers (Supplementary Figure S1). We validated the accuracy of the copy number estimation by analyzing a set of 10 isolates with both short-read data and completely assembled genomes available (Figure 4A). We find substantial *Styx* copy number variation within the species and across geographic regions (Figure 4B). Genomes sequenced from center-of-origin populations in the Middle East and North Africa are largely devoid of *Styx* copies. European and North American genomes show an average of 8–9 *Styx* copies per genome, while the Oceanian and South American populations carry on average 17 and 20 copies per genome, respectively (Figure 4B). The highest copy numbers globally were found in genomes from Argentina with a maximum of 56 copies (average of 31 copies). In contrast to strong geographic effects worldwide on *Styx* copy numbers, copy numbers were largely stable across the sampling period 1999–2016 in Europe with overall moderate levels of *Styx* (Figure 4C).

Newly created copies during the *Styx* expansion may have started to accumulate mutations. Based on the 923-genome panel, we identified a total of 1707 single-nucleotide variants affecting 21% of the 7927-bp *Styx* consensus sequence. Consistent with the high copy numbers and recent expansion, *Styx* copies in Oceanian and South American populations were highly similar (Figure 4D). Center-of-origin populations tend to carry more sequence variants consistent with the presence of older *Styx* copies. Isolates from Africa carry an average of 447 variants per copy, which is consistently higher than in other regions (Tukey's HSD test, $P < 0.01$; Figure 5A). Mutations in *Styx* affected primarily noncoding regions (~43%; Figure 5B). An additional 30.5% and 4.5% of the mutations were missense and stop gain variants, respectively. C→T and G→A transitions dominate the mutation spectrum consistent with RIP genomic defenses driving the divergence of *Styx* copies (Figure 5C). *Styx* copies in Oceanian genomes have a higher proportion of frameshift and missense variants despite overall lower differentiation among copies (Tukey's HSD test, $P < 0.001$; Figure 5D). The broad genomic survey of the species reveals a strong association of rapid *Styx* copy number expansion following founder events at the origin of Oceanian and South American populations, as well as a concurrent degradation of genomic defenses against *Styx*.

Styx mobilization and chromosome-scale rearrangements upon meiosis

Highly active TEs such as the *Styx* element may show observable rates of transposition even over single generations. Investigating the youngest insertions of *Styx* provides a more complete spectrum of insertion sites as selection may not have eliminated yet all deleterious new insertions. We analyzed evidence for newly inserted copies of *Styx* in a four-generation pedigree initiated by two isolates collected from a European wheat field (1A5 and 1E4) (57). The pedigree comprised a total of 10 progenies and revealed a large rearrangement of chromosome 17 most likely triggered by copies of *Styx* (Figure 6A) (30). The parents 1A5 and 1E4 carry 18 and 9 *Styx* copies, respectively, and progenies carry 9–16 copies (Figure 6B). We analyzed recombination tracks using whole-genome

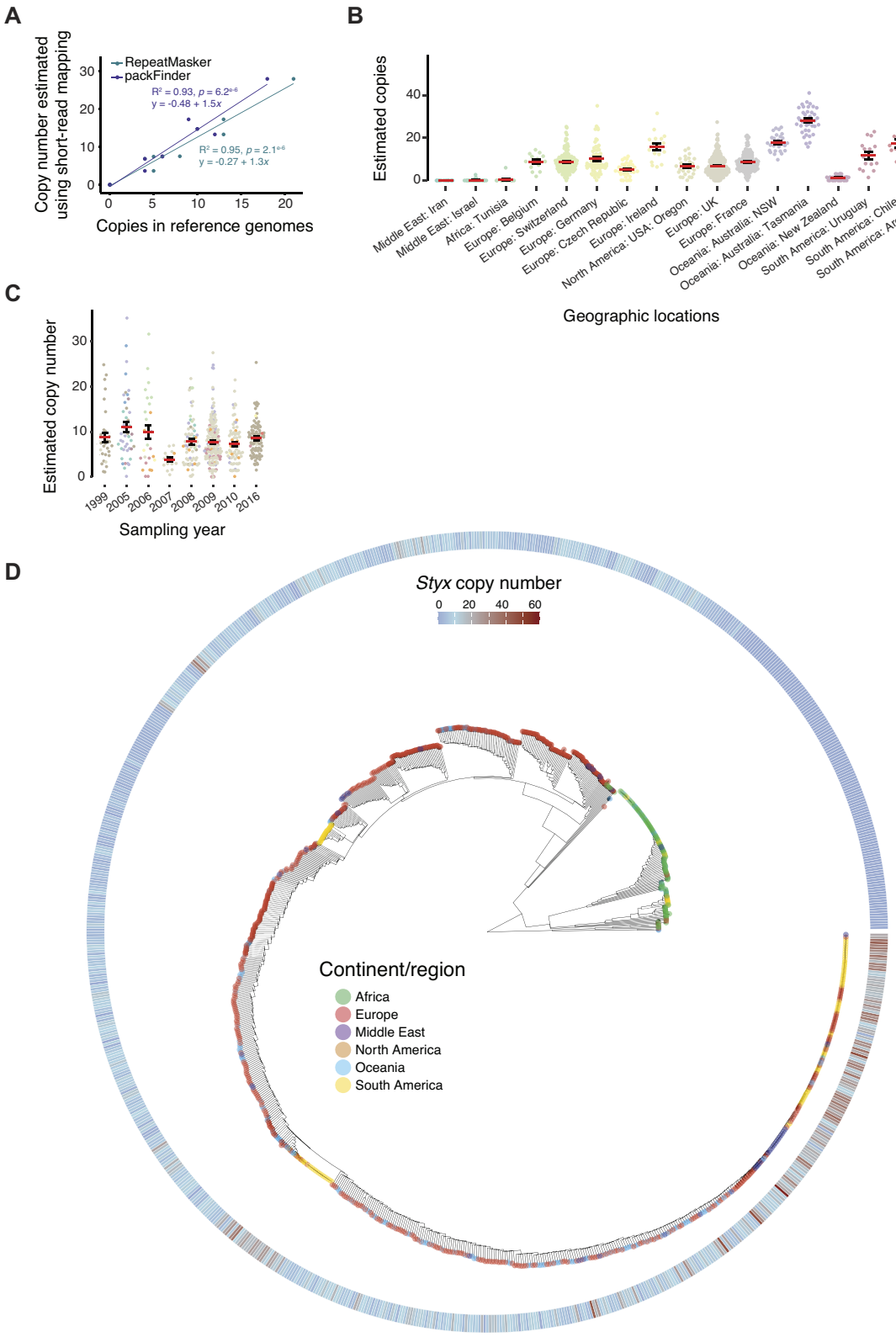


Figure 4. Reactivation of the *Styx* element in the wheat pathogen *Z. tritici* tracked by a 900-genome panel. (**A**) Correlation between the number of annotated *Styx* copies from RepeatMasker- or packFinder-based annotation of chromosome-level assemblies and the estimated copy number from short-read mapping on the consensus sequence of the *Styx* element. (**B**) Estimated *Styx* copy number in *Z. tritici* isolates sampled across the species geographic range (17 geographic locations with >10 isolates for a total of 856 isolates). (**C**) Estimated *Styx* copy number in isolates sampled in Europe between the years 1999 and 2016. (**D**) Phylogenetic tree based on SNPs identified against the *Styx* consensus sequence in 923 worldwide *Z. tritici* isolates. The heatmap on the outside of the tree shows the estimated *Styx* copy number per sequenced genome.

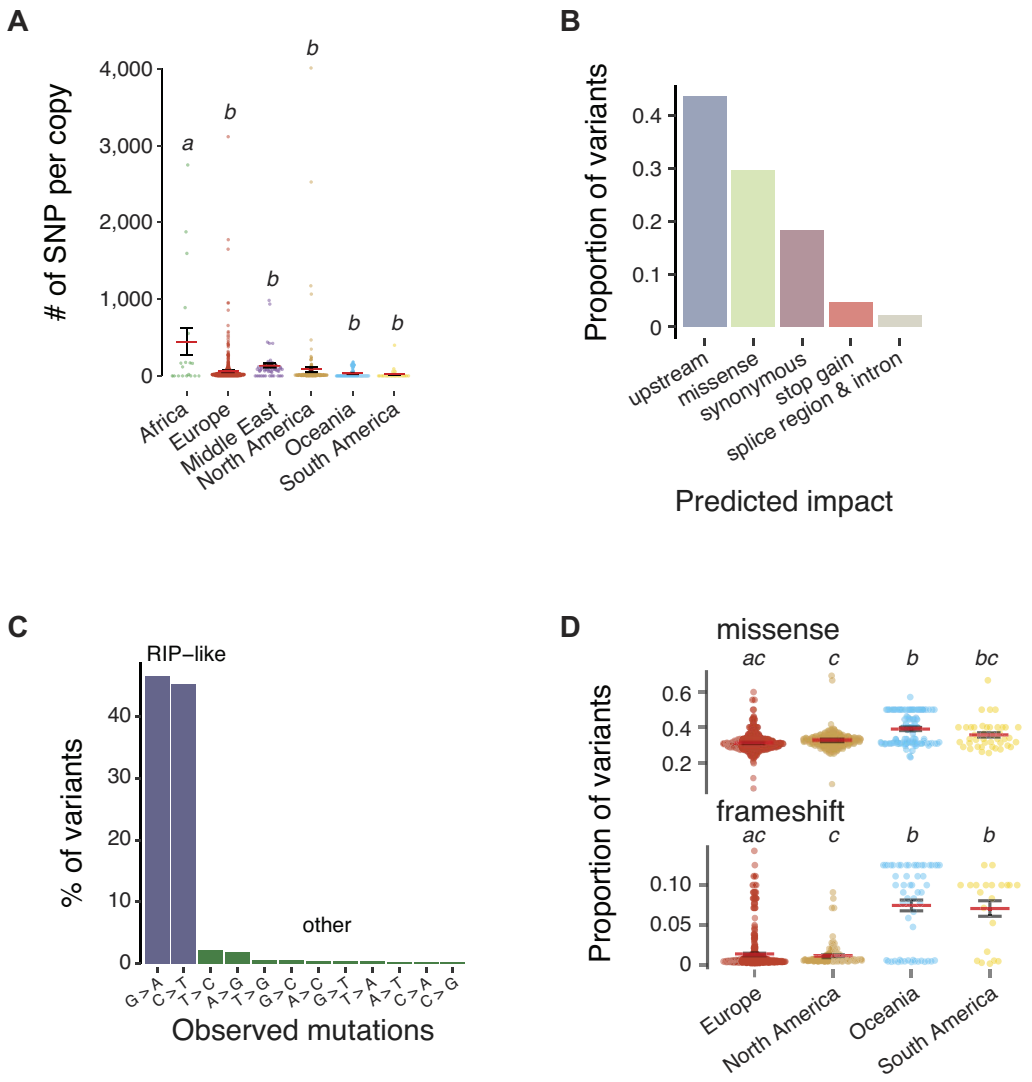


Figure 5. Mutation accumulation of the *Styx* element in *Z. tritici* driven by genomic defenses. (**A**) Average number of variants detected in reads from each of the 923 worldwide *Z. tritici* isolates mapped against the *Styx* consensus sequence and normalized by the estimated *Styx* copy number. (**B**) Predicted effect of variants detected in the read alignment among all isolates. (**C**) Proportion of biallelic variants classified according to the type of transition or transversion. C→T and G→A transitions are considered RIP-like mutations. (**D**) Proportion of variants annotated as missense and frameshift. Letters indicate significant differences (Tukey's HSD test, $P < 0.05$).

alignments of each progeny genome against the two parental genomes to differentiate transposition creating a new *Styx* copy from recombination introducing existing *Styx* copies into a new background. Using this recombination map established for each progeny, we tracked vertical inheritance of each *Styx* copy in the pedigree (Figure 6C). We found that both parental genomes contributed equally to copy numbers in the progeny with an average of 3.8 and 4 copies originating from the 1A5 and 1E4 parents, respectively. A further 44 *Styx* copies (∼36%) from all progenies were found in chromosomal locations incompatible with vertical inheritance from either of the two parental backgrounds and, hence, constitute new insertions (Figure 6D).

We searched for ancestral sequences at the origin of each new *Styx* insertion, including the 6-bp target site duplication sites. We discarded loci where parental genomes segregated presence–absence of *Styx* copies. Each progeny genome carries between 1 and 10 new *Styx* insertions for a total of 32 distinct insertion events (Figure 6E). New insertions cover 7927 ± 1 bp of sequence matching the full-length *Styx* element and locate to 8 out of 13 core chromosomes, as well as 5 out of 8 accessory chromosomes. Approximately half of the *Styx* insertions were vertically transmitted in the pedigree. Consistently with this, half of the insertions identified at the third and fourth meiotic rounds were vertically inherited (i.e. M3 and M4; Figure 6F). New insertions occurred in regions with higher TE density compared to the genome-wide average (Figure 6G; Wilcoxon rank sum test *P*-value <1e−3). New insertions tend to persist in the genomes as the number of vertically inherited copies increased further down the pedigree (i.e. meiotic round 3 to 4), although overall *Styx* copy numbers remained stable consistent with recombination reducing *Styx* copies in progeny genomes (Figure 6F).

We more closely analyzed the association between recombination breakpoints and *Styx* copies. We identified a total of 138 recombination breakpoints in the 10 progeny genomes with an average of ∼14 breakpoints per progeny (Figure 7A and Supplementary Table S5). We assessed co-occurrence in

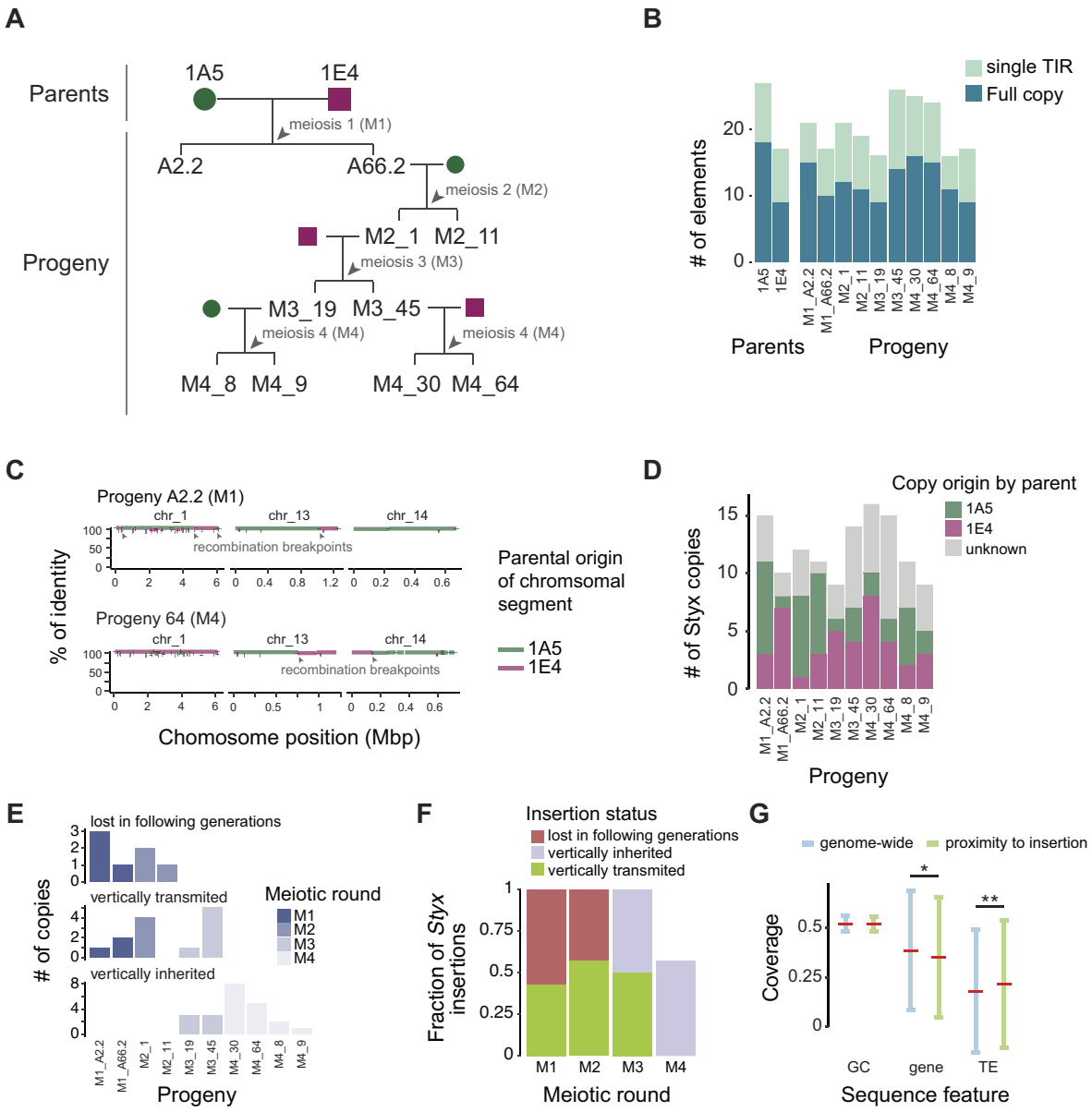


Figure 6. Four-generation pedigree analyses reveal active transposition of the *Styx* element. (**A**) Schematic of the analyzed *Z. tritici* pedigree with parent–offspring relations. (**B**) Number of *Styx* copies identified in the pedigree. Full elements and single TIRs are shown separately. (**C**) Illustration of the chromosomal block synteny analyses of the progeny isolates M1_A2.2 and M4_64 showing evidence for recombination events on core chromosomes 1 and 13 and accessory chromosome 14. (**D**) Parental origin of the *Styx* copies in different progeny isolates. (**E**) Status of the *Styx* copies identified in each meiotic round along the progeny (M1 to M4). (**F**) Fraction of the *Styx* copies identified at each meiotic round (M1 to M4) that were lost, vertically inherited or vertically transmitted along the progeny. (**G**) Isochore, gene and transposon coverage in proximity to *Styx* insertions (windows of 50 kb centered on *Styx* copies) compared to a genome-wide set of 50-kb windows. Asterisks indicate the result of a pairwise Wilcoxon rank sum test (*$P < 0.05$ and **$P < 0.01$ after Holm correction).

50-kb windows and found that *Styx* copies were over 2-fold enriched near recombination breakpoints (Fisher's exact test odds ratio = 2.4, *P*-value = 0.03). Genomic features such as isochores can co-vary with recombination rates (58,59). In the pedigree, recombination breakpoints occurred predominantly in gene-rich and TE-poor regions sharing similarities with the main genomic niche of *Styx* (Tukey's HSD test, $P < 1e{-}16$; Figure 7B). The most striking association of *Styx* copies with recombination breakpoints is a large chromosomal translocation involving core chromosomes 6 and 12 (Figure 7C). The two *Styx* copies inherited from the 1A5 parental genome were near the synteny breakpoints at the origin of the fused core chromosomes in progeny M4_8 (Figure 7C).

The recency of the new *Styx* insertions enables to track the genealogical history of copies across pedigree generations. For this, we performed full-length multiple sequence alignments of all *Styx* copies detected among parental isolates and progenies (Figure 7D). The 32 new *Styx* insertions separate into two distinct clades with one composed of 21 sequences being nearly identical to 11 copies of *Styx* identified in the parental genome 1A5. A second clade is composed of 11 insertions nearly identical to three *Styx* found in the parental genome

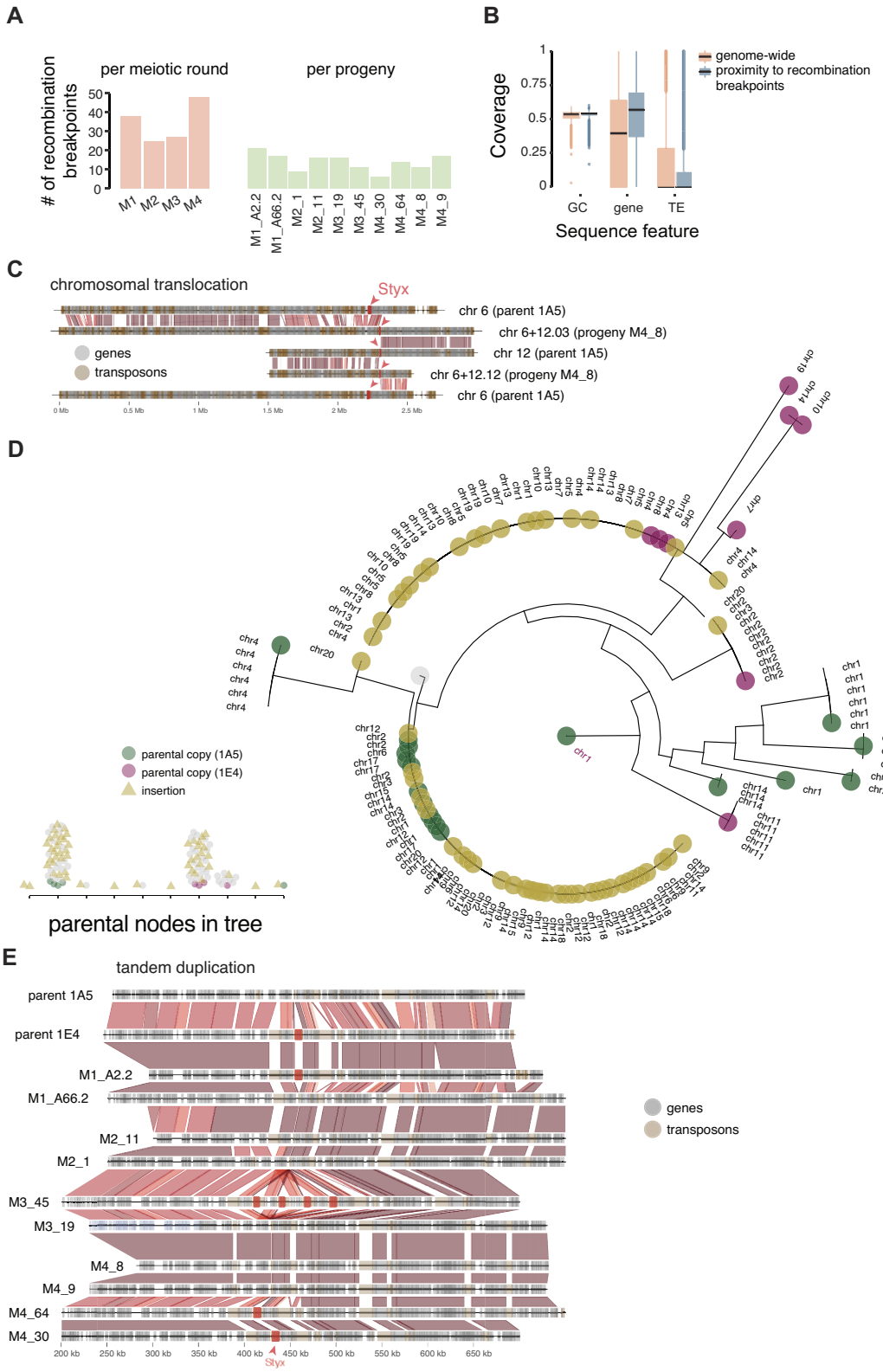


Figure 7. Chromosomal translocations triggered by new *Styx* insertions and insertion preference near recombination breakpoints. (**A**) Number of recombination breakpoints identified in the progeny genomes. Numbers are shown per meiotic round (M1 to M4) and separately for each progeny isolate. (**B**) Isochore, gene and TE coverage at recombination breakpoints compared to genome-wide 50-kb windows. (**C**) Synteny plot illustrating the chromosomal translocation that occurred between chromosomes 6 and 12 in the progeny isolate M4_8. *Styx* copies are highlighted with red arrows. (**D**) Phylogenetic tree based on full *Styx* sequences detected in parental and progeny genomes. Vertically inherited *Styx* copies are differentiated by colors from new insertions. (**E**) Synteny plot illustrating the *Styx* tandem duplication that occurred on chromosome 8 in the progeny isolate M3_45. *Styx* copies, TEs and genes are highlighted by coloured boxes.

1E4 (Figure 7D). We retraced the origin of a new *Styx* insertion on chromosome 1 found in progeny M2_1 (position 3 640 021 bp with target site duplication ATCGAG) to a 1A5 parental copy on chromosome 2 (position 52 220 bp with target site duplication CGTGAT). We also retraced a second insertion on chromosome 3 found in the progeny M4_64 (position 1 432 987 bp with target site duplication TGTGAT) to a 1E4 parental copy on chromosome 2 (position 3 217 078 bp with target site duplication CATAAG; Figure 7D). Importantly, the parental copy on chromosome 2 was retained in the M4_64 progeny, suggesting that *Styx* likely mobilized through a non-replicative mechanism and that the host DNA repair machinery likely promoted this copy number increase. Interestingly, the 1E4 parental copy encodes a non-functional integrase coding sequence indicating that even degenerate copies of the *Styx* can be mobilized by an integrase protein encoded by intact copies. Two *Styx* insertions each on chromosomes 14 and 15 were recurrent insertion events within 100 kb distance from each other (Supplementary Figure S2). Overall, 44% (8 out of 18) of the new insertions on the same chromosome occurred within 150 kb from the parental copy. We observed a similar transposition pattern of *Styx* copies across *Zymoseptoria* species. Overall ∼51% of *Styx* copies are clustered on the same chromosome and separated by <150 kb from each other. Physical proximity of new *Styx* insertions suggests that chromosomal conformation (i.e. contact) may be relevant to the transposition of the element.

We identified a total of 25 solo-TIRs among progenies, which is comparable to the abundance of solo-TIRs across *Zymoseptoria* species (Figure 6B). However, solo-TIRs are readily created *de novo* as identified in the progeny isolate M3_45 with three new solo-TIRs compared to parental genomes. Analyzing synteny at the corresponding locus on chromosome 8 showed that the three solo-TIRs originated from a triplication of a *Styx* element in the parent 1E4 (Figure 7E). The parental *Styx* duplicated and inserted three times near the original location creating two to three additional copies of six genes near the original *Styx* element. The six genes include a duplicated secreted endonuclease gene, a triplicated gene encoding a secreted protein of unknown function, a twice duplicated gene encoding a membrane-bound protein, a duplicated gene encoding a predicted protein kinase and two genes encoding for proteins of unknown function (duplicated and triplicated, respectively). Across *Zymoseptoria* species, we find ∼19% of all *Styx* copies within 50 kb of two or more gene duplicates (23 of 122 total copies). Similarly, we found tandem-inverted *Styx* copies co-localized on the same chromosome arm. Hence, the *Styx* transposition may be driven by proximal insertions and topological preferences (Supplementary Figure S3).

Given the ability of the *Styx* element to mobilize over single rounds of meiosis and trigger sequence rearrangements, we investigated whether genomic defenses counteracted *Styx* activity. The average number of *Styx* copies remained largely stable in the pedigree despite evidence for new insertions (Figure 6B–E). Hence, either strong purifying selection eliminated progeny with higher *Styx* copy numbers or *Styx* copies were readily excised. We find evidence for excision at a *Styx* locus of the 1A5 parental chromosome 10, which is missing *Styx* in two progeny isolates. The excision event is further supported by the retention of parental sequences surrounding the element (Supplementary Figure S4). Synteny analyses of the excision locus revealed three paralogous genes on either side of the parental *Styx*. The duplicated genes in the flanking region likely served as homology anchors for the excision of the *Styx* copy (Supplementary Figure S4). Taken together, *Styx* mobilization is likely an agent triggering chromosomal rearrangements over single generations and impacts adjacent chromosomal regions by possibly favoring gene duplications.

Styx activity escapes genomic defenses by the RIP machinery

RIP is a mechanism that inactivates TEs in fungal genomes. Mutational signatures of active RIP were found across the *Zymoseptoria* genus (60). Here, we analyzed *Styx* copies in the progeny to quantify the occurrence of *de novo* RIP-like mutations. We focused on the 31 *Styx* copies that persisted for one to four rounds of meiosis and identified mutations in new copies using the parental copy of the element as a reference. Over 93% of the *Styx* copies remained identical throughout the pedigree (29/31). The two *Styx* copies with point mutations have 5 + 12 and 9 + 10 RIP-like mutations, respectively (i.e. C→T and G→A transitions). None of the new *Styx* copies in the pedigree showed signatures of RIP. To address whether RIP is broadly ineffective against TEs, we investigated the accumulation of *de novo* mutations in four additional TEs with varying copy numbers and expression profiles (RLC_Deimos, DXX_Birute, DTA_Vera and DTT_Tapputi). Identical to the procedure used to recover *Styx* copies, we analyzed target site duplication sequences to recover orthologous copies among pedigree genomes. The DTA_Vera and DTT_Tapputi elements have 12 and 14 copies in the pedigree, respectively, although copies share only low sequence identity and were likely not targeted by RIP (0.55 and 0.43 maximal pairwise sequence identity, respectively). For the elements DXX_Birute and RLC_Deimos, *de novo* mutation accumulation was only detected in the high-copy RLC_*Deimos* (74 copies in the 1A5 parent) despite copies of the element sharing between 0.4 and 0.9 sequence identity. The 21 detected mutations were restricted to 8 out of 142 copies and were only found in progeny genomes of the first round of meiosis (i.e. M1_A66.2 and M1.A2.2). Given the older sequencing technology used to generate PacBio long reads (RSII versus Sequel), we cannot rule out that non-polished sequencing errors persisted in the assembly as artifacts. Altogether, we conclude that genomic defenses in *Z. tritici*, including RIP, are unable to efficiently target *Styx* or other TE copies. The rapid expansion of the *Styx* transposon in the pedigree is consistent with the expansion dynamics observed at the species level with major consequences for chromosomal integrity and the faithful transmission of genetic information.

Discussion

A recent origin of *Styx* within the *Zymoseptoria* genus

TEs are major drivers of genome evolution over deep timescales. Here, we unravel the recent evolutionary history of the *Styx* transposon and how this recapitulates major phases of TE family dynamics. The *Styx* recently differentiated into subelements with distinct TIRs and sequence length concomitant with speciation of five closely related *Zymoseptoria* species. Consistent with a recent origin, two of the four *Styx* coding sequences are exclusively found in *Zymoseptoria* species (27). The encoded DDD/E transposase suggests a common origin with members of the IS3EU DNA transposon

superfamily (61). TE families typically differentiate rapidly with few documented examples of ongoing activity such as the high-copy miniature non-autonomous *mPing* and the two autonomous *Ping* and *Pong* elements in rice retaining signatures of a common ancestor (19). Given the shared coding sequences with IS3EU family members, *Styx* likely evolved through gene co-option in the ancestor of the five *Zymoseptoria* species. The IS3EU transposase family shares similarities with bacterial insertion sequences from the IS3 family (61,62). Furthermore, horizontal transfer of DDD/E TE families is widespread and found in multiple kingdoms (63,64). However, no *Styx*-like elements were found in species outside of the *Zymoseptoria* genus.

Independent waves of *Styx* reactivation

Sequence identity patterns of *Styx* copies in the *Zymoseptoria* genus are consistent with at least two recent independent reactivation events in the ancestors of *Z. pseudotritici* and *Z. tritici* (Figure 8). In *Z. tritici*, high-quality genomes analyzed across the global distribution range carry highly heterogeneous copy numbers suggesting very recent transposition activity (51). In concordance, we find that the element is highly expressed in many *Z. tritici* isolates and silenced in other species of the genus indicating that copies were repressed following the transposition bursts in these species (Figure 8). Host control over TEs can be mediated by mutations affecting transposon mobility (18,19,65). Here, we investigated proximal mechanisms associated with *Styx* mobilization and repression. The TIRs of the *Z. passerinii Styx* subelement are mutated compared to the dominant TIR sequence found in the genus likely affecting *Styx* mobility. Similarly, the recent *mPing* burst in rice was likely facilitated by a point mutation adjacent to the *mPing* TIRs (19). How transposons reactivate over evolutionary timescales remains largely unknown. However, external stressors such as temperature can promote transposon activity (11,16). In *Z. tritici*, nutrient limitation and stress imposed by plant immune defense favor derepression in a TE-specific manner (14). It remains unknown whether resistant wheat cultivars and increased fungicide applications are constituting sufficiently strong factors to cause lasting TE derepression in *Z. tritici*. Fungicide exposure under laboratory conditions can lead to transposon activity in other pathogens (66).

The genome-wide transposon content in *Z. brevis and Z. passerinii* is higher than that in *Z. tritici* consistent with more young elements and recent activity (60). However, the same two species have stronger signatures of RIP-like mutations across their TE repertoire suggesting that higher TE activity is counterbalanced by TE defenses. In agreement with RIP being active in the genus, a large fraction of the mutations identified across *Styx* copies were most likely generated by TE defenses. Low activity of RIP in *Z. tritici* is most likely explained by the near-complete loss of a DNA methyltransferase known to promote cytosine methylation and RIP in *Neurospora crassa* (67–69). Taken together, the loss of an active DNA methyltransferase and the high transcriptional levels of *Styx* in *Z. tritici* suggest that the element likely escapes genomic defenses beyond RIP. Such defenses would likely be associated with transcriptional interference such as quelling and methylation induced premeiotically. Additional defense mechanisms such as meiotic silencing by unpaired DNA may be active against *Styx*; however, such a mechanism is unlikely to be active in the *Zymoseptoria* genus given the absence of a *sad-2* homolog (70). We show therefore a parallel between *Styx* copy number and the presence of *DIM2* in the genus that suggests that RIP might contribute to *Styx* control.

H3K9me3 and H3K27me3 histone modifications are broadly conserved mechanisms of transposon control in eukaryotes (54,71,72). In *Z. tritici*, >80% of TEs are targeted by H3K9me3 modifications and are likely constitutively silenced (73). However, approximately a third of *Z. tritici* TEs are associated with H3K27me3 modifications, which mediate chromosomal integrity in the species (73). Here, we show that *Styx* copies are targeted by both H3K4me2 and H3K27me3 modifications linking *Styx* loci to the typically most dynamic genomic regions in fungal plant pathogens (73–75).

A population-level perspective on TE reactivation

Styx copy number estimates across >900 *Z. tritici* genomes vary by an order of magnitude across the global distribution range. The element is nearly absent from the center-of-origin populations in the Middle East and North Africa. *Styx* copies culminate in populations from Oceania and South America suggesting a recent reactivation in these regions. In agreement, we find that *Styx* copies from Oceania and South America accumulated very few mutations and copies retained high sequence identity. The transposon content of genomes is expected to result from the equilibrium established between transposon activity, host repressive mechanisms and purifying selection (1). However, bursts in transposon activity and population bottlenecks following founder events are impacting transposon dynamics as well (76). The evolutionary history of *Z. tritici* is tightly linked to wheat domestication in the Fertile Crescent followed by stepwise introduction events to new continents (49). Populations outside of the Middle East and Europe have both gained in transposon content and overall reduced RIP-like mutations (49). Colonization events over the past centuries in Australia induced population bottlenecks leading to reduced genetic diversity and likely less efficient selection. Such relaxation in selection may have weakened genomic defenses against transposons. Relaxation of genomic defenses likely underpins also incipient genome size expansion within the species (23). *Styx* follows the same trend of higher copies numbers in regions with weakened genomic defenses (i.e. RIP-like mutations). The proximal trigger of *Styx* reactivation outside of the *Z. tritici* center of origin may well have been the loss of *DIM2* underpinning RIP activity.

Styx escapes genomic defenses and destabilizes chromosomal sequences

How transposons activate and reintegrate into chromosomal sequences remains poorly understood. A particular challenge arises from the action of purifying selection removing an unknown subset of all insertions. In addition, population bottlenecks and founder events can arbitrarily eliminate insertions from populations. Here, we used a pedigree to track the *Styx* element over generations using complete genome sequences. *Styx* is indeed mobile creating dozens of new copies while maintaining parental copies suggesting that transposition occurs upon DNA replication. In bacteria, transposons of the IS200 and IS605 families have been described to mobilize through a 'peel-and-paste, cut-and-copy' mechanism. Targeted double-strand breaks follow single-strand transposon excision events and allow faithful preservation of the donor

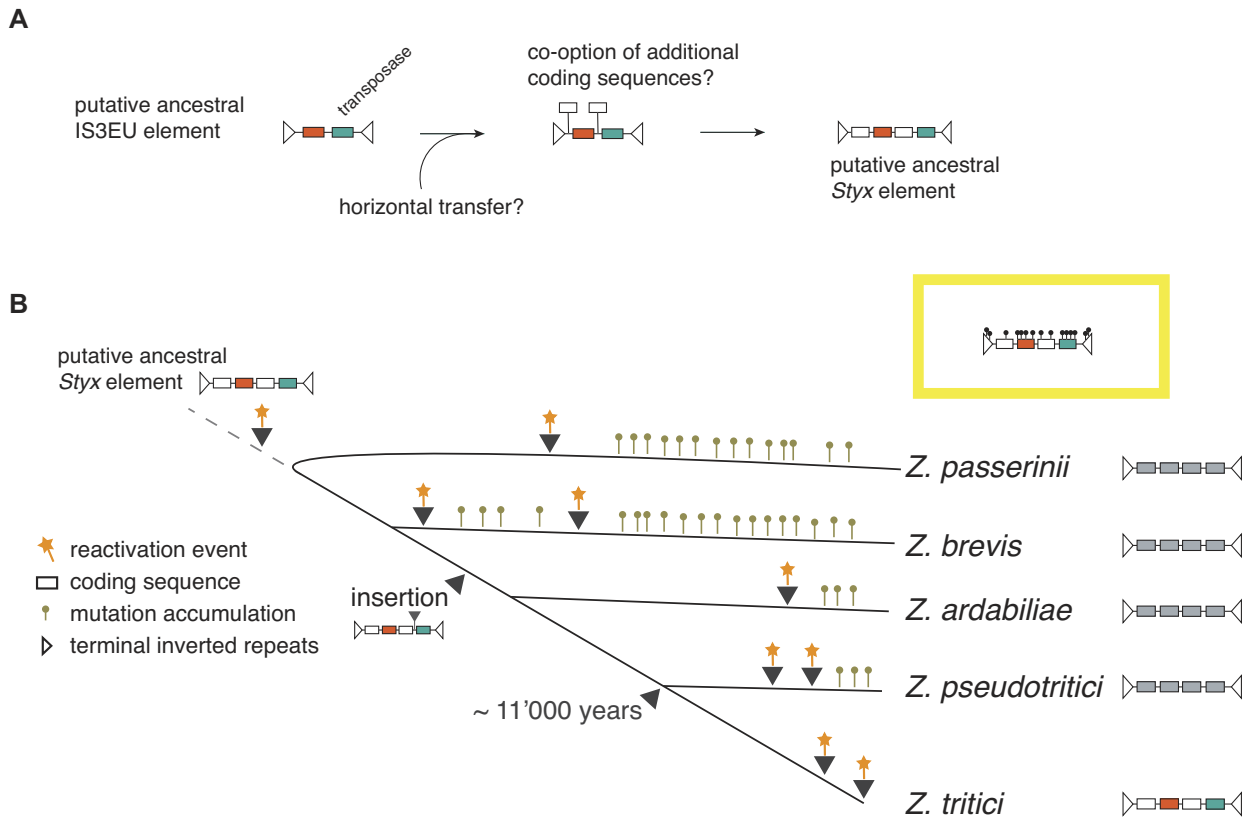


Figure 8. Proposed model illustrating the sequence of evolutionary events at the origin of the multiple *Styx* expansions in the *Zymoseptoria* genus. (**A**) *Styx* likely originated from an ancestral class II IS3EU element encoding two proteins, including a DDD/E transposase required for mobilization. After a putative horizontal transfer in an ancestor to the *Zymoseptoria* genus, the element co-opted two additional coding sequences to form the *Styx* element. (**B**) The newly formed *Styx* element possibly activated first prior to speciation in the genus. Following speciation, *Styx* experienced substantial sequence divergence. Based on species-level copy identity, *Styx* likely reactivated independently two times in *Z. brevis*, *Z. pseudotritici* and *Z. tritici*. In *Z. tritici*, the mobilization likely occurred following the spread of the pathogen outside of the center of origin in the Middle East.

copy through homologous recombination (77–79). Such indirect replicative transposition is consistent with the substantial gains in copy number observed across the species range. By tracking new copies to their progeniture sequence in parental genomes, we found that both copies with functional and degenerate transposase coding sequences are mobile. Degenerate copies likely benefited from transposase expression of intact copies. Mobility retention of degenerate *Styx* may favor even further sequence reduction to reduce into miniature inverted-repeat TEs as shown for the *mPing* elements (7,19).

The new *Styx* insertions in the *Z. tritici* progeny are found preferentially closer to genes. However, *Styx* shows no detectable insertion site sequence preference suggesting that the heterogeneous distribution in the genome across the species most likely arises from preferences for open chromatin but may also be influenced by random factors such as genetic drift. In the pedigree, *Styx* loci display precise excisions mediated by recombination of flanking sets of paralogous genes. Furthermore, recombination breakpoints are preferred *Styx* insertion sites in the pedigree. Associations of TEs and recombination breakpoints were previously observed for *Drosophila P*-elements inducing recombination in males and promoting large deletions and duplications of flanking regions (80). TE-mediated recombination such as that driven by the RAG recombinase system evolved from transposon domestication (81). Reshuffling of vertebrate immune loci is mediated by the RAG recombinase system with suppressed transposition activity (81).

Some of the most consequential effects of unchecked TE activity in genomes are large sequence rearrangements. CACTA elements, for instance, are associated with chromosomal rearrangements that led to the formation of the R-r complex that controls the production of anthocyanin in maize (82). LINE-1 insertions have been associated with large deletions and the amplification of oncogenes in cancer cells (83). The *Styx* element is associated with multiple chromosomal rearrangements over only four generations and several dozen new insertions overall. An investigation of an unassembled chromosome 17 variant recovered from the same progeny pedigree suggested that *Styx* is associated with additional rearrangements (84). Three out of six analyzed chromosomal breakpoints associated with a destabilized chromosome 17 overlap with *Styx* copies (84). Hence, the observed chromosomal rearrangements were likely promoted by non-allelic homologous recombination at *Styx* copies and have the potential to initiate chromosomal sequence meltdowns in at least one accessory chromosome. Deleterious effects of *Styx* are likely compounded by the tendency to generate paralogous gene copies. This is consistent with double-strand breaks being involved in the *Styx* replicative transposition. Double-strand breaks would have favored the observed tandem duplications of the element and concurrent gene duplications in flanking regions. Paralogs in proximity to *Styx* insertion sites are particularly noteworthy given the very low rate of gene duplicate retention compared to other fungi likely as a consequence of active RIP (29,85). Taken together, *Styx* plays complex roles in the

Zymoseptoria genus from influencing asexual reproduction, contributing to gene duplications, chromosomal rearrangements and fusions. The lack of control over *Styx* transposition in recently established populations highlights the tenuous and transitory control host genomes exert over selfish elements.

Data availability

The data analyzed in the frame of this study are available from ENA and NCBI repositories as indicated in Supplementary Table S1. Sequencing reads resulting from the ChIP-seq experiments are available on NCBI under the project PRJNA1036405. Figure source datasets are available from Zenodo (https://doi.org/10.5281/zenodo.10184730). Additional short-read genome datasets are listed in Supplementary Data S1 of (49).

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We thank group members for fruitful discussions.

Author contributions: T.B. and D.C. conceived the study; T.B. performed analyses; S.M.T. performed experiments; A.F. and S.M.T. provided datasets; D.C. provided funding and supervised the work; T.B. and D.C. wrote the manuscript with input from S.M.T. and A.F.

Funding

Swiss National Science Foundation [173265 and 201149 to D.C.] Funding for open access charge: Swiss National Science Foundation.

Conflict of interest statement

None declared.

References

1. Wells,J.N. and Feschotte,C. (2020) A field guide to eukaryotic transposable elements. *Annu. Rev. Genet.*, **54**, 539–561.
2. Argueso,J.L., Westmoreland,J., Mieczkowski,P.A., Gawel,M., Petes,T.D. and Resnick,M.A. (2008) Double-strand breaks associated with repetitive DNA can reshape the genome. *Proc. Natl Acad. Sci. U.S.A.*, **105**, 11845–11850.
3. Catoni,M., Jonesman,T., Cerruti,E. and Paszkowski,J. (2019) Mobilization of Pack-CACTA transposons in *Arabidopsis* suggests the mechanism of gene shuffling. *Nucleic Acids Res.*, **47**, 1311–1320.
4. Uchiyama,T., Hiura,S., Ebinuma,I., Senda,M., Mikami,T., Martin,C. and Kishima,Y. (2013) A pair of transposons coordinately suppresses gene expression, independent of pathways mediated by siRNA in *Antirrhinum*. *New Phytol.*, **197**, 431–440.
5. Dubin,M.J., Mittelsten Scheid,O. and Becker,C. (2018) Transposons: a blessing curse. *Curr. Opin. Plant Biol.*, **42**, 23–29.
6. Batzer,M.A. and Deininger,P.L. (2002) Alu repeats and human genomic diversity. *Nat. Rev. Genet.*, **3**, 370–379.
7. Chen,J., Lu,L., Robb,S.M.C., Collin,M., Okumoto,Y., Stajich,J.E. and Wessler,S.R. (2020) Genomic diversity generated by a transposable element burst in a rice recombinant inbred population. *Proc. Natl Acad. Sci. U.S.A.*, **117**, 26288–26297.
8. Slotkin,R.K. and Martienssen,R. (2007) Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.*, **8**, 272–285.
9. Yang,G., Zhang,F., Hancock,C.N. and Wessler,S.R. (2007) Transposition of the rice miniature inverted repeat transposable element *mPing* in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. U.S.A.*, **104**, 10962–10967.
10. Ito,H., Gaubert,H., Bucher,E., Mirouze,M., Vaillant,I. and Paszkowski,J. (2011) An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature*, **472**, 115–120.
11. Ito,H., Yoshida,T., Tsukahara,S. and Kawabe,A. (2013) Evolution of the ONSEN retrotransposon family activated upon heat stress in Brassicaceae. *Gene*, **518**, 256–261.
12. Naito,K., Zhang,F., Tsukiyama,T., Saito,H., Hancock,C.N., Richardson,A.O., Okumoto,Y., Tanisaka,T. and Wessler,S.R. (2009) Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature*, **461**, 1130–1134.
13. Makarevitch,I., Waters,A.J., West,P.T., Stitzer,M., Hirsch,C.N., Ross-Ibarra,J. and Springer,N.M. (2015) Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLoS Genet.*, **11**, e1004915.
14. Fouché,S., Badet,T., Oggenfuss,U., Plissonneau,C., Francisco,C.S. and Croll,D. (2020) Stress-driven transposable element de-repression dynamics and virulence evolution in a fungal pathogen. *Mol. Biol. Evol.*, **37**, 221–239.
15. Le,T.N., Schumann,U., Smith,N.A., Tiwari,S., Khang Au,P.C., Zhu,Q.H., Taylor,J.M., Kazan,K., Llewellyn,D.J., Zhang,R., *et al.* (2014) DNA demethylases target promoter transposable elements to positively regulate stress responsive genes in *Arabidopsis*. *Genome Biol.*, **15**, 458.
16. Roquis,D., Robertson,M., Yu,L., Thieme,M., Julkowska,M. and Bucher,E. (2021) Genomic impact of stress-induced transposable element mobility in *Arabidopsis*. *Nucleic Acids Res.*, **49**, 10431–10447.
17. Oggenfuss,U. and Croll,D. (2022) Recent transposable element bursts triggered by insertions near genes in a fungal pathogen. bioRxiv doi: https://doi.org/10.1101/2022.07.13.499862, 13 July 2022, preprint: not peer reviewed.
18. Scheuermann,B., Diem,T., Ivics,Z. and Andrade-Navarro,M.A. (2019) Evolution-guided evaluation of the inverted terminal repeats of the synthetic transposon *Sleeping Beauty*. *Sci. Rep.*, **9**, 1171.
19. Chen,J., Lu,L., Benjamin,J., Diaz,S., Hancock,C.N., Stajich,J.E. and Wessler,S.R. (2019) Tracking the origin of two genetic components associated with transposable element bursts in domesticated rice. *Nat. Commun.*, **10**, 641.
20. Selker,E.U. (2002) Repeat-induced gene silencing in fungi. *Adv. Genet.*, **46**, 439–450.
21. Goodwin,S.B., M'Barek,S.B., Dhillon,B., Wittenberg,A.H.J., Crane,C.F., Hane,J.K., Foster,A.J., van der Lee,T.A.J., Grimwood,J., Aerts,A., *et al.* (2011) Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. *PLoS Genet.*, **7**, e1002070.
22. Fones,H. and Gurr,S. (2015) The impact of *Septoria tritici* Blotch disease on wheat: an EU perspective. *Fungal Genet. Biol.*, **79**, 3–7.
23. Oggenfuss,U., Badet,T., Wicker,T., Hartmann,F.E., Singh,N.K., Abraham,L., Karisto,P., Vonlanthen,T., Mundt,C., McDonald,B.A., *et al.* (2021) A population-level invasion by transposable elements triggers genome expansion in a fungal pathogen. *eLife*, **10**, e69249.
24. Krishnan,P., Meile,L., Plissonneau,C., Ma,X., Hartmann,F.E., Croll,D., McDonald,B.A. and Sánchez-Vallet,A. (2018) Transposable element insertions shape gene regulation and melanin production in a fungal pathogen of wheat. *BMC Biol.*, **16**, 78.
25. Steinhauer,D., Salat,M., Frey,R., Mosbach,A., Luksch,T., Balmer,D., Hansen,R., Widdison,S., Logan,G., Dietrich,R.A., *et al.*

(2019) A dispensable paralog of succinate dehydrogenase subunit C mediates standing resistance towards a subclass of SDHI fungicides in *Zymoseptoria tritici*. *PLoS Pathog.*, **15**, 616904.
26. Singh,N.K., Badet,T., Abraham,L. and Croll,D. (2021) Rapid sequence evolution driven by transposable elements at a virulence locus in a fungal wheat pathogen. *BMC Genomics*, **22**, 393.
27. Wang,C., Milgate,A.W., Solomon,P.S. and McDonald,M.C. (2021) The identification of a transposon affecting the asexual reproduction of the wheat pathogen *Zymoseptoria tritici*. *Mol. Plant Pathol.*, **22**, 800–816.
28. Feurtey,A., Lorrain,C., Croll,D., Eschenbrenner,C., Freitag,M., Habig,M., Haueisen,J., Möller,M., Schotanus,K. and Stukenbrock,E.H. (2020) Genome compartmentalization predates species divergence in the plant pathogen genus *Zymoseptoria*. *BMC Genomics*, **21**, 588.
29. Badet,T., Oggenfuss,U., Abraham,L., McDonald,B.A. and Croll,D. (2020) A 19-isolate reference-quality global pangenome for the fungal wheat pathogen *Zymoseptoria tritici*. *BMC Biol.*, **18**, 12.
30. Badet,T., Croll,D., Fouché,S., Hartmann,F.E., Zala,M. and Croll,D. (2021) Machine-learning predicts genomic determinants of meiosis-driven structural variation in a eukaryotic pathogen. *Nat. Commun.*, **12**, 3551.
31. Gisby,J.S. and Catoni,M. (2022) The widespread nature of Pack-TYPE transposons reveals their importance for plant genome evolution. *PLoS Genet.*, **18**, e1010078.
32. Katoh,K. and Standley,D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.*, **30**, 772–780.
33. Crooks,G.E., Hon,G., Chandonia,J.M. and Brenner,S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.
34. Price,M.N., Dehal,P.S. and Arkin,A.P. (2009) FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.*, **26**, 1641–1650.
35. Page,A.J., Taylor,B., Delaney,A.J., Soares,J., Seemann,T., Keane,J.A. and Harris,S.R. (2016) SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb. Genomics*, **2**, e000056.
36. Valero-Mora,P.M. (2010) ggplot2: elegant graphics for data analysis. *J. Stat. Softw.*, **35**, 260.
37. Marçais,G., Delcher,A.L., Phillippy,A.M., Coston,R., Salzberg,S.L. and Zimin,A. (2018) MUMmer4: a fast and versatile genome alignment system. *PLoS Comput. Biol.*, **14**, e1005944.
38. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
39. Croll,D., Lendenmann,M.H., Stewart,E. and McDonald,B.A. (2015) The impact of recombination hotspots on genome evolution of a fungal plant pathogen. *Genetics*, **201**, 1213–1228.
40. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
41. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
42. Jin,Y., Tam,O.H., Paniagua,E. and Hammell,M. (2015) TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics*, **31**, 3593–3599.
43. Robinson,M.D., McCarthy,D.J. and Smyth,G.K. (2009) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.
44. Soyer,J.L., Möller,M., Schotanus,K., Connolly,L.R., Galazka,J.M., Freitag,M. and Stukenbrock,E.H. (2015) Chromatin analyses of *Zymoseptoria tritici*: methods for chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq). *Fungal Genet. Biol.*, **79**, 63–70.
45. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
46. Babadi,M., Lee,S.K., Smirnov,A., Lichtenstein,L., Gauthier,L.D., Howrigan,D.P. and Poterba,T. (2018) Abstract 2287: Precise common and rare germline CNV calling with GATK. *Cancer Res.*, **78**, 2287.
47. Heinz,S., Benner,C., Spann,N., Bertolino,E., Lin,Y.C., Laslo,P., Cheng,J.X., Murre,C., Singh,H. and Glass,C.K. (2010) Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.
48. Nakato,R. and Sakata,T. (2021) Methods for ChIP-seq analysis: a practical workflow and advanced applications. *Methods*, **187**, 44–53.
49. Feurtey,A., Lorrain,C., McDonald,M.C., Milgate,A., Solomon,P.S., Warren,R., Puccetti,G., Scalliet,G., Torriani,S.F.F., Gout,L., *et al.* (2023) A thousand-genome panel retraces the global spread and adaptation of a major fungal crop pathogen. *Nat. Commun.*, **14**, 1059.
50. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
51. Oggenfuss,U. and Croll,D. (2023) Recent transposable element bursts are associated with the proximity to genes in a fungal plant pathogen. *PLoS Pathog.*, **19**, e1011130.
52. Stukenbrock,E.H., Banke,S., Javan-Nikkhah,M. and McDonald,B.A. (2007) Origin and domestication of the fungal wheat pathogen *Mycosphaerella graminicola* via sympatric speciation. *Mol. Biol. Evol.*, **24**, 398–411.
53. Zhan,J., Pettway,R.E. and McDonald,B.A. (2003) The global genetic structure of the wheat pathogen *Mycosphaerella graminicola* is characterized by high nuclear diversity, low mitochondrial diversity, regular recombination, and gene flow. *Fungal Genet. Biol.*, **38**, 286–297.
54. Déléris,A., Berger,F. and Duharcourt,S. (2021) Role of Polycomb in the control of transposable elements. *Trends Genet.*, **37**, 882–889.
55. Cui,X., Jin,P., Cui,X., Gu,L., Lu,Z., Xue,Y., Wei,L., Qi,J., Song,X., Luo,M., *et al.* (2013) Control of transposon activity by a histone H3K4 demethylase in rice. *Proc. Natl Acad. Sci. U.S.A.*, **110**, 1953–1958.
56. Lindehell,H., Schwartz,Y.B. and Larsson,J. (2023) Methylation of lysine 36 on histone H3 is required to control transposon activities in somatic cells. *Life Sci. Alliance*, **6**, e202201832.
57. Lendenmann,M.H., Croll,D. and McDonald,B.A. (2015) QTL mapping of fungicide sensitivity reveals novel genes and pleiotropy with melanization in the pathogen *Zymoseptoria tritici*. *Fungal Genet. Biol.*, **80**, 53–67.
58. Tortereau,F., Servin,B., Frantz,L., Megens,H.J., Milan,D., Rohrer,G., Wiedmann,R., Beever,J., Archibald,A.L., Schook,L.B., *et al.* (2012) A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content. *BMC Genomics*, **13**, 586.
59. Montoya-Burgos,J.I., Boursot,P. and Galtier,N. (2003) Recombination explains isochores in mammalian genomes. *Trends Genet.*, **19**, 128–130.
60. Lorrain,C., Feurtey,A., Ller,M.M., Haueisen,J. and Stukenbrock,E. (2021) Dynamics of transposable elements in recently diverged fungal pathogens: lineage-specific transposable element content and efficiency of genome defenses. *G3 (Bethesda)*, **11**, jkab068.
61. Kojima,K.K. (2019) Structural and sequence diversity of eukaryotic transposable elements. *Genes Genet. Syst.*, **94**, 233–252.
62. Siguier,P., Gourbeyre,E. and Chandler,M. (2014) Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol. Rev.*, **38**, 865–891.
63. Baril,T. and Hayward,A. (2022) Migrators within migrators: exploring transposable element dynamics in the monarch butterfly, *Danaus plexippus*. *Mobile DNA*, **13**, 5.
64. Dupeyron,M., Baril,T. and Hayward,A. (2021) Broadscale evolutionary analysis of eukaryotic DDE transposons. bioRxiv

doi: https://doi.org/10.1101/2021.09.26.461848, 26 September 2021, preprint: not peer reviewed.
65. Skipper,K.A., Andersen,P.R., Sharma,N. and Mikkelsen,J.G. (2013) DNA transposon-based gene vehicles—scenes from an evolutionary drive. *J. Biomed. Sci.*, **20**, 92.
66. Chen,F., Everhart,S.E., Bryson,P.K., Luo,C., Song,X., Liu,X. and Schnabel,G. (2015) Fungicide-induced transposon movement in *Monilinia fructicola*. *Fungal Genet. Biol.*, **85**, 38–44.
67. Dhillon,B., Cavaletto,J.R., Wood,K.V. and Goodwin,S.B. (2010) Accidental amplification and inactivation of a methyltransferase gene eliminates cytosine methylation in *Mycosphaerella graminicola*. *Genetics*, **186**, 67–77.
68. Freitag,M., Williams,R.L., Kothe,G.O. and Selker,E.U. (2002) A cytosine methyltransferase homologue is essential for repeat-induced point mutation in *Neurospora crassa*. *Proc. Natl Acad. Sci. U.S.A.*, **99**, 8802–8807.
69. Moller,M., Habig,M., Lorrain,C., Feurtey,A., Haueisen,J., Fagundes,W.C., Alizadeh,A., Freitag,M. and Stukenbrock,E.H. (2021) Recent loss of the Dim2 DNA methyltransferase decreases mutation rate in repeats and changes evolutionary trajectory in a fungal pathogen. *PLoS Genet.*, **17**, e1009448.
70. Shiu,P.K.T., Zickler,D., Raju,N.B., Ruprich-Robert,G. and Metzenberg,R.L. (2006) SAD-2 is required for meiotic silencing by unpaired DNA and perinuclear localization of SAD-1 RNA-directed RNA polymerase. *Proc. Natl Acad. Sci. U.S.A.*, **103**, 2243–2248.
71. Janssen,A., Colmenares,S.U. and Karpen,G.H. (2018) Heterochromatin: guardian of the genome. *Annu. Rev. Cell Dev. Biol.*, **34**, 265–288.
72. Kabi,M. and Filion,G.J. (2021) Heterochromatin: did H3K9 methylation evolve to tame transposons?*Genome Biol.*, **22**, 325.
73. Möller,M., Schotanus,K., Soyer,J.L., Haueisen,J., Happ,K., Stralucke,M., Happel,P., Smith,K.M., Connolly,L.R., Freitag,M., *et al.* (2019) Destabilization of chromosome structure by histone H3 lysine 27 methylation. *PLoS Genet.*, **15**, e1008093.
74. Cook,D.E., Kramer,H.M., Torres,D.E., Seidl,M.F. and Thomma,B.P.H.J. (2020) A unique chromatin profile defines adaptive genomic regions in a fungal plant pathogen. *eLife*, **9**, e62208.
75. Xie,J., Tang,W., Lu,G., Hong,Y., Zhong,Z., Wang,Z. and Zheng,H. (2023) Histone H3K27me3 methylation regulates the expression of secreted proteins distributed at fast-evolving regions through transcriptional repression of transposable elements. *J. Integr. Agric.*, **22**, 3059–3068.
76. Charlesworth,B. and Charlesworth,D. (1983) The population dynamics of transposable elements. *Genet. Res.*, **42**, 1–27.
77. Meers,C., Le,H., Pesari,S.R., Hoffmann,F.T., Walker,M.W.G., Gezelle,J. and Sternberg,S.H. (2023) Transposon-encoded nucleases use guide RNAs to selfishly bias their inheritance. bioRxiv doi: https://doi.org/10.1101/2023.03.14.532601, 29 March 2023, preprint: not peer reviewed.
78. Karvelis,T., Druteika,G., Bigelyte,G., Budre,K., Zedaveinyte,R., Silanskas,A., Kazlauskas,D., Venclovas,Č. and Siksnys,V. (2021) Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. *Nature*, **599**, 692–696.
79. He,S., Corneloup,A., Guynet,C., Lavatine,L., Caumont-Sarcos,A., Siguier,P., Marty,B., Dyda,F., Chandler,M. and Ton Hoang,B. (2015) The IS200/IS605 family and "peel and paste" single-strand transposition mechanism. *Microbiol. Spectr.*, **3**, https://doi.org/10.1128/microbiolspec.MDNA3-0039-2014.
80. Preston,C.R., Sved,J.A. and Engels,W.R. (1996) Flanking duplications and deletions associated with P-induced male recombination in *Drosophila*. *Genetics*, **144**, 1623–1638.
81. Zhang,Y., Cheng,T.C., Huang,G., Lu,Q., Surleac,M.D., Mandell,J.D., Pontarotti,P., Petrescu,A.J., Xu,A., Xiong,Y., *et al.* (2019) Transposon molecular domestication and the evolution of the RAG recombinase. *Nature*, **569**, 79–84.
82. Walker,E.L., Robbins,T.P., Bureau,T.E., Kermicle,J. and Dellaporta,S.L. (1995) Transposon-mediated chromosomal rearrangements and gene duplications in the formation of the maize R-r complex. *EMBO J.*, **14**, 2350–2363.
83. Rodriguez-Martin,B., Alvarez,E.G., Baez-Ortega,A., Zamora,J., Supek,F., Demeulemeester,J., Santamarina,M., Ju,Y.S., Temes,J., Garcia-Souto,D., *et al.* (2020) Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. *Nat. Genet.*, **52**, 306–319.
84. Fouché,S., Oggenfuss,U., McDonald,B.A. and Croll,D. (2023) Recurrent chromosome destabilization through repeat-mediated rearrangements in a fungal pathogen. bioRxiv doi: https://doi.org/10.1101/2023.07.14.549097, 15 July 2023, preprint: not peer reviewed.
85. Duplessis,S., Cuomo,C.A., Lin,Y.C., Aerts,A., Tisserant,E., Veneault-Fourrey,C., Joly,D.L., Hacquard,S., Amselem,J., Cantarel,B.L., *et al.* (2011) Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc. Natl Acad. Sci. U.S.A.*, **108**, 9166–9171.

Received: April 9, 2023. **Revised:** November 30, 2023. **Editorial Decision:** December 7, 2023. **Accepted:** December 11, 2023