Dynamics of transposable elements in recently diverged fungal pathogens: lineage-specific transposable element content and efficiency of genome defences

Cécile Lorrain[*, †, ‡1], Alice Feurtey[*, †1,2], Mareike Möller[*, † 3], Janine Haueisen[*, †] and Eva Stukenbrock[*, †]

* Environmental Genomics, Max Planck Institute for Evolutionary Biology, 24306 Plön, Germany

† Environmental Genomics, Christian-Albrechts University of Kiel, 24118 Kiel, Germany

‡ Université de Lorraine/INRAE, UMR 1136 Interactions Arbres/Microorganismes, INRAE Centre Grand Est – Nancy, Champenoux 54280, France

[1] Present address: Pathogen Evolutionary Ecology, ETH Zurich, 8092 Zürich, Switzerland

[2] Present address: Laboratory of Evolutionary Genetics, Institute of Biology, University of Neuchâtel, 2000 Neuchâtel, Switzerland.

[3] Present address: Department of Biochemistry and Biophysics, Oregon State University, Corvallis, OR, United States of America

Running title: Variation in TE content shapes fungal genomes

Corresponding authors:

Cécile Lorrain

Pathogen Evolutionary Ecology, ETH Zurich, Universität strasse 2, 8092 Zürich, Switzerland

clorrain@ethz.ch

Eva Stukenbrock

Max Planck for Evolutionary Biology, August-Thienemann-Straße 2, 24306 Plön, Germany & Christian-Albrechts University of Kiel, Christian-Albrechts-Platz 4, 24118 Kiel, Germany

estukenbrock@bot.uni-kiel.de

Phone: +49 431880 6368

Keywords: Transposable elements, effectors, genome architecture, *Zymoseptoria tritici*, repeat-induced point mutations, genome plasticity

Abstract

Transposable elements (TEs) impact genome plasticity, architecture and evolution in fungal plant pathogens. The wide range of TE content observed in fungal genomes reflects diverse efficacy of host-genome defence mechanisms that can counter-balance TE expansion and spread. Closely related species can harbour drastically different TE repertoires, suggesting variation in the efficacy of genome defences. The evolution of fungal effectors, which are crucial determinants of pathogenicity, has been linked to the activity of TEs in pathogen genomes. Here we describe how TEs have shaped genome evolution of the fungal wheat pathogen *Zymoseptoria tritici* and four closely related species. We compared *de novo* TE annotations and Repeat-Induced Point mutation signatures in thirteen genomes from the *Zymoseptoria* species-complex. Then, we assessed the relative insertion ages of TEs using a comparative genomics approach. Finally, we explored the impact of TE insertions on genome architecture and plasticity. The thirteen genomes of *Zymoseptoria* species reflect different TE dynamics with a majority of recent insertions. TEs associate with distinct genome compartments in all *Zymoseptoria* species, including chromosomal rearrangements, genes showing presence/absence variation and effectors. European *Z. tritici* isolates have reduced signatures of Repeat-Induced Point mutations compared to Iranian isolates and closely related

57 species. Our study supports the hypothesis that ongoing but moderate TE mobility in

58 *Zymoseptoria* species shapes pathogen genome evolution.

Introduction

Transposable elements (TEs), DNA elements that can replicate through transposition (i.e. independently of the host genome replication machinery) are ubiquitous in Eukaryotic genomes. The TE content in fungal plant-pathogen genomes covers a wide range: from less than 1% to more than 90% of the genome in *Fusarium graminearum* and *Blumeria graminis,* respectively (Cuomo *et al.* 2007; Frantzeskakis *et al.* 2018). TEs are categorized into two classes, retrotransposons (class I) and DNA transposons (class II), based on their mechanism of transposition (Wicker *et al.* 2007). Retrotransposons replicate using an RNA intermediate to insert at a new position and DNA transposons replicate either by a mechanism of direct excision from double-stranded DNA (subclass I) or using single-strand excision followed by a rolling-circle mechanism (subclass II) (Wicker *et al.* 2007). TE classes are divided into orders that contain various numbers of superfamilies and families, which are categorized by coding sequence structure (Wicker *et al.* 2007). TEs can be autonomous (e.g. LTRs and TIRs) or non-autonomous (e.g. SINEs and MITEs). The latter relies on the replication machinery of autonomous TEs to transpose (Wicker *et al.* 2007). TE activity (i.e. transposition) is known to have an overall negative impact on host fitness (Horváth *et al.* 2017). As a result, TEs engage in a co-evolutionary arms race dynamic with the host genome (Biémont 2010; Castanera *et al.* 2016).

Fungal genomes have evolved diverse genome defence mechanisms to regulate TE expansions. In addition to histone modifications and DNA methylation (Deniz *et al.* 2019), a fungal-specific mechanism called Repeat-Induced Point mutation (RIP) specifically mutates duplicated sequences such as TEs (Gladyshev 2017). RIP induces a dinucleotide bias in duplicated sequences by mutating G:C into A:T; this bias can be measured in fungal genomes, and quantified as a RIP signature (Gladyshev and Kleckner 2014, 2017). In “RIPed” genomes, RIP-induced mutation can result in Large RIP Affected Regions (LRARs) that are large genomic regions consecutively affected by RIP (van Wyk *et al.* 2019). Genomes of several fungal pathogens have relatively high TE contents, while simultaneously exhibiting signatures of RIP (Gao *et al.* 2011; Gioti *et al.* 2013; Grandaubert *et al.* 2014; Dhillon *et al.* 2014; Fokkens *et al.* 2018). It remains unclear how TEs can maintain stable proportions in their host genomes with defence mechanisms such as RIP. It is worth to note that the maintenance of RIP mechanisms is costly for the host and RIP can be lost (Galagan and Selker 2004). In the genus *Neurospora*, closely related species have variation in RIP signatures, and *Neurospora* species with reduced RIP signatures exhibit TE expansions (Gioti *et al.* 2013). This dynamic between TE expansions and host defense mechanisms is summarized as the burst and decay model of TE evolution (Arkhipova 2018). This model assumes that TEs are active

(i.e., they burst in transposition or expansion) until they are inactivated by host defence mechanisms such as RIP (i.e., they decay). As a consequence of the RIP-induced overaccumulation of mutations in TEs, dating specific TE families' invasions is challenging (Grandaubert *et al.* 2014).

TE insertions and how the host respond to them, shape genome architecture. TEs can increase genomic plasticity by promoting chromosomal rearrangements and compartmentalization, duplicating or deleting genes, and altering gene expression. Often, the TE content is linked to compartmentalization of the genome with TE-enriched genomic compartments or regions associating with specific epigenetic signatures and changes in CG content (Duplessis *et al.* 2011; Bertazzoni *et al.* 2018; Frantzeskakis *et al.* 2018; Chen *et al.* 2018; Stam *et al.* 2018). Genome compartmentalization into TE-rich and TE-poor regions can sometimes be observed at the chromosomal level. Some fungal species harbor dispensable or accessory chromosomes, which contain a higher proportion of TEs than the core chromosomes as observed in the wheat pathogen *Zymoseptoria tritici* (Croll and McDonald 2012). Despite being considered as "junk DNA" until recently, TEs have been shown in recent years to participate in adaptation to environmental changes of their hosts. For instance, in plant-associated fungi, TEs are de-repressed during stressful conditions such as the early stage of plant infection (Fouché *et al.* 2020). TEs are also

119 physically associated with pathogenicity-related genes (i.e., effectors) of various
120 fungal pathogen species, suggesting a role in effector gene diversification (Rouxel *et*
121 *al.* 2011; Grandaubert *et al.* 2014; Soyer *et al.* 2014; Dong *et al.* 2015; Fokkens *et al.*
122 2018; Fouché *et al.* 2018). In this way TEs can have mediated mutational changes
123 with benefits to the host organism. TEs can even be co-opted, or domesticated by
124 the host, and evolved to have a new function in the host. For instance, the effector
125 AvrK1 in *Blumeria graminis* f.sp. *hordei* is directly derived from a LINE
126 retrotransposon (Amselem *et al.* 2015b). In *Z. tritici*, LTR retrotransposon insertion
127 upstream of a multidrug efflux transporter has conferred fungicide resistance
128 (Omrane *et al.* 2017b). Multiple studies demonstrate the importance of TEs for fungal
129 pathogenicity, yet little is known about the extent to which TEs represent key players
130 in the global evolution of the genomes of fungal pathogens.
131
132 The wheat pathogen *Z. tritici* has emerged as a model to study fungal genome and
133 TE evolution. Closely related species of *Z. tritici* have been isolated from wild
134 grasses and barley, and include *Zymoseptoria passerinii*, *Zymoseptoria ardabiliae*,
135 *Zymoseptoria brevis* and *Zymoseptoria pseudotritici* (Stukenbrock *et al.* 2007, 2012).
136 The genomes of the five species comprise conserved core chromosomes and
137 variable accessory chromosomes which show low gene density, low transcriptional
138 activity, and enrichment of the heterochromatin-associated histone mark H3K27me3

(Goodwin *et al.* 2011; Kellner *et al.* 2014; Schotanus *et al.* 2015; Feurtey *et al.* 2019). TEs have been investigated in the economically important wheat pathogen *Z. tritici*, describing for example the occurrence of RIP in this species through the quantification of C to T transition mutations for the reference isolate IPO323 (Dhillon *et al.* 2014). Several associations have been made between TEs and *Z. tritici* virulence and fungicide resistance. *Z. tritici* genes encoding effectors and transporters involved in fungicide resistance physically associate with TEs (Omrane *et al.* 2017b; Hartmann *et al.* 2018; Oggenfuss *et al.* 2020). Recent studies demonstrated that TEs in *Z. tritici* are de-repressed during early stages of wheat infection (Fouché *et al.* 2020). Interestingly, the TE content varies between closely related *Zymoseptoria* species(Grandaubert *et al.* 2015). TEs in *Z. tritici* accumulate in recently founded populations outside of the center of origin due to relaxed purifying selection (Fouché *et al.* 2020; Oggenfuss *et al.* 2020; Badet *et al.* 2020). Despite these recent studies, little is yet known about how TEs and genome defence mechanisms have co-evolved and shaped host genomes over larger time scales among the *Zymoseptoria* genus.

In this study, we explore TE dynamics in thirteen *Zymoseptoria* genomes, including nine *Z. tritici* genomes and one genome from each of the four closely related species, *Z. passerinii*, *Z. ardabiliae*, *Z. brevis* and *Z. pseudotritici*. We specifically

addressed the following questions: i) How do TE distributions and insertion ages impact genome architecture and plasticity? ii) Does TE content correlate with gene presence/absence variation among genomes? iii) What is the extent of variation in RIP among genomes? To answer these questions, we annotated the TE content of each of the thirteen genomes *de novo* and analysed TE landscapes within and among genomes. We found evidence for variation in the efficiency of RIP among *Z. tritici* genomes and among the different species genomes.

Material and Methods

Genomic and gene expression data

All *Zymoseptoria* spp. isolates used in this study come from publicly available genome assemblies. We used recently assembled genomes of nine *Z. tritici* isolates sampled from Europe and Iran. Zt05 was isolated from wheat in Denmark, Zt10 and Zt289 from wheat in Iran and Zt469 from *Aegilops* sp. in Iran (Grandaubert *et al.* 2015; Feurtey *et al.* 2019; Möller *et al.* 2020). We also included the genomes of five *Z. tritici* isolates for which recent Pacbio assemblies were published (four from Switzerland Zt1A5; Zt1E4; Zt3D1; Zt3D7 (Plissonneau *et al.* 2018). Finally, we used one genome from each of four closely related species *Z. ardabiliae* Za17, *Z. brevis* Zb87; *Z. passerinii* Zpa63 and *Z. pseudotritici* Zp13 (Feurtey *et al.* 2019). Gene expression during wheat infection from (Haueisen *et al.* 2019), updated expression

profiles on the last versions of genome assemblies and new gene predictions of the three *Z. tritici* isolates IPO323, Zt05 and Zt10, were used as described in (Feurtey *et al.* 2019). In summary, RNAseq was performed during wheat infection time-course using strand-specific RNA-libraries from Illumina HISeq2500 sequencing, with 100pb single-end reads. A total of 89.5 to 147.5 million reads per sample was obtained (Haueisen *et al.* 2019). To simplify the expression data, we combined all time-points from wheat infection and calculated expression levels in Transcripts Per Million (TPM) as described in (Feurtey *et al.* 2019).

Annotation of repeated elements and relative age of insertion

We used the REPET pipeline (https://urgi.versailles.inra.fr/Tools/REPET; (Quesneville *et al.* 2005; Flutre *et al.* 2011) to annotate the repeat regions of *Z. ardabiliae* Za17, *Z. brevis* Zb87, *Z. pseudotritici* Zp13, *Z. passerinii* Zpa63 and the nine *Z. tritici* isolates as described in (Feurtey *et al.* 2019). Briefly, we identified repeats in each genome by building TE consensus sequences, as a proxy for TE ancestral sequence. Each consensus is derived from a multiple alignment of TEs in clusters. We then mined each *Zymoseptoria* spp. genome using the constructed TE consensus library to recover TE copies belonging the same consensus. TE annotation metrics are summarized in Table S1.

We assessed relative ages of TE insertions in *Zymoseptoria* spp. using the REPET similarity-based approach to measure the distributions of TE clusters' sequence

198 identities. Based on the burst and decay evolution of TEs (Roessler *et al.* 2018), we
199 analysed sequence divergence of individual element clusters to assess the relative
200 age of TE insertions in each genome (i.e. TE spread in host genome). For this, we
201 need to assume that RIP-induced mutations occurrence is constant over time for
202 each genome. Based on this assumption, the extent of sequence similarity is
203 proportional to the divergence time of copies. It is thereby possible to compare
204 relative insertion ages of TE insertions within genomes (Figure S1). We assessed
205 sequence similarity within each TE cluster by comparing each TE copy to the
206 consensus sequence.

207 **Analysis of transposon genomic environments**

208 To further investigate intra-specific variation in TE content, we conducted a detailed
209 comparison of the core chromosomes of the nine *Z. tritici* isolates. By focusing only
210 on the core chromosomes, we avoid an overestimation of TE insertions variation due
211 to presence/absence polymorphisms of the accessory chromosomes. Transposon
212 densities along *Z. tritici* isolates' chromosomes were measured using 100 kbp
213 windows using bedtools "makewindows". We fixed window coordinates based on the
214 reference genome of IPO323 using orthologous genes. The closest genes
215 neighbouring each window's borders were extracted from the reference IPO323
216 genome using the bedtools "closest" function (Quinlan and Hall 2010). Orthologs of
217 these neighbouring genes were then extracted after identification using the program

Proteinortho (Lechner *et al.* 2011). The number of transposons per window was calculated using bedtools "intersect" function (Quinlan and Hall 2010). The results were visualized using the software Circos (Krzywinski *et al.* 2009).

To assess associations between the vicinity of TEs and genes potentially involved in plant infection, we used the previous functional annotations of predicted effectors and orthologous genes of Za17, Zpa63, Zb87 and Zp13, and the three *Z. tritici* isolates IPO323, Zt05 and Zt10 (Feurtey *et al.* 2019). We annotated the genes with presence/absence variation (PAV) among the thirteen *Zymospetoria* species-complex genomes. For this we used PoFF, an extension of the software Proteinortho which integrates data on conserved synteny to detect orthologous relationships (Lechner *et al.* 2011). We differentiated genes as follows: 1) showing PAV among all thirteen genomes as *PAV genes*, 2) genes present in all thirteen *Zymoseptoria* genomes as *Core genes* and 3) genes present on all nine *Z. tritici* isolates as *Core Z. tritici genes*. We also included predicted genes with TE-like domains (e.g. transposase) in the TE repertoires to avoid considering TEs as PAV genes. To statistically assess associations between the vicinity of transposons and gene categories, we used the R package regioneR (Gel *et al.* 2015). We used the function "meanDistance" to test whether specific gene categories are closer to transposons than a random distribution. We performed the permutation test with the

237 "randomizeRegion" function and 1000 permutations. Randomizations were
238 performed per chromosome.
239 We quantified TE impacts on intra-specific small-scale chromosomal rearrangements
240 in the nine isolates of *Z. tritici*. For this, we used the software SyRi (Goel *et al.* 2019)
241 to identify synteny breaks of a minimum of 500bp. We assessed the distances from
242 these synteny breaks to the closest transposons using bedtools "closest" function
243 (Quinlan and Hall 2010).

244 **Repeat-Induced Point mutation (RIP) analysis**

245 Repeat-Induced Point (RIP) mutation indices were calculated, and Large RIP-
246 affected genomic regions were determined, using the RIPper software
247 (https://github.com/TheRIPper-Fungi/TheRIPper/ (van Wyk *et al.* 2019). Regions of
248 more than 4000 bp that are consecutively affected by RIP are considered to be
249 "large RIP affected genomic regions". For genome-wide RIP index assessments, we
250 used default parameters of parameters of 1000bp windows with a 500bp step size.
251 RIP Composite index values were calculated as follows: (TpA/ ApT) – (CpA + TpG/
252 ApC + GpT). A region is affected by RIP when the Composite index is > 0 (van Wyk
253 *et al.* 2019). To calculate the RIP Composite index of each transposon copy, we
254 used 50bp non-overlapping windows using homemade script.

255 **Data Availability**

All data produced and analysed in this study including transposable elements annotations and consensus libraries are available at {ZENODO DOI TO BE ADDED WHEN ACCEPTED}.

Results

Transposable elements content varies in the genomes of the *Zymoseptoria* species-complex

Overall, the TE proportions of *Z. passerinii*, *Z. ardabiliae*, *Z. brevis* and *Z. pseudotritici* genomes are higher than the TE proportions of *Z. tritici* isolates (Table S1; Figure 1). Outside of *Z. tritici*, TE content ranges from 6.9 Mb in the *Z. ardabiliae* Za17 genome (18.2% of the genome) to 12.9 Mb in the *Z. passerinii* Zpa63 genome (31.4% of the genome; Figure 1B; Table S1). Among *Z. tritici* isolates, TE content ranges from 5.70 Mb in the Iranian isolate Zt289 genome (14.5% of the genome) to 8.21 Mb in the Zt05 genome (19.8% of the genome) (Figure 1B; Table S1). Most of the TE coverage consists of LTR-retrotransposons among all genomes of *Zymoseptoria* species. LTRs elements represent from 3.2 Mb to 7.4 Mb of the genomes of *Z. tritici* (Zt289) and *Z. passerinii,* respectively. Among these LTRs, we only found the *Copia* and *Gypsy* elements, except in the Iranian *Z. tritici* isolate Zt469, in which we also identified elements belonging to the unique *Bel-Pao* family (representing 0.85 Mb; Table S1). LINE elements are completely absent from the

genome of *Z. pseudotritici* and comprise only 0.29% (0.1 Mb) of the genome of *Z. ardabiliae* and 0.5% (0.2 Mb) of the genome of *Z. passerinii*. In contrast, *Z. brevis* Zb87 appears as an outlier: its genome contains 1.2 Mb (2.9% of the genome) of LINE elements. Among *Z. tritici* isolates, LINE content ranges from 0.29% (in Zt469) to 3.83% (in Zt05). Taken together, these results suggest recent invasion and variable expansions of LINEs in *Z. tritici* and *Z. brevis* (Figure 1; Table S1).

Many recent TE insertions postdate diversification of the *Zymoseptoria* genus

One third of the TEs in the different *Zymoseptoria* genomes are relatively recent insertions. Sequence similarities between TEs and their cognate consensus sequences in the genomes of *Zymoseptoria* spp. range from 65.4% to 100% of sequence identity (Figure 1C). Based on the approach used in Maumus & Quesneville (2014), we defined TE categories based on thresholds as follows: 1) copies with less than 85% sequence identity to the consensus comprise *old* insertions and 2) copies with 85% to 95% sequence identity are *intermediate* insertions and 3) copies with more than 95% identity with the consensus sequence represent *recent* insertions. Based on these criteria, we show that *old* insertions represent 33% (*Z. passerinii* - 5.6 Mb) to 46.6% (*Z. pseudotritici* - 5.6 Mb) of the total TE content, while *recent* insertions represent 29.3% (*Z. tritici* Zt3D1 - 3.4Mb) to 40.5% (*Z. brevis* - 6.2Mb) of transposons (Figure 1C; Figure S1). The majority of *recent* TE insertions are retrotransposons while the majority of *old* insertions are

295 retrotransposons and DNA transposons (Figure S2). Taken together these results
296 suggest that at least a fraction of TEs have been recently active within genomes.
297 Variation in TE insertions among *Z. tritici* core chromosomes indicates recent TE
298 transposition activity, mostly driven by LTRs. The proportion of TEs per core
299 chromosome or contig varies from 5.7% (chr11 of Zt469) to 33.3% (chr9 of Zt05)
300 (Figure 2). In addition, isolates with very close TE content per chromosome show
301 very different TE distribution patterns (Figure 2A; Figure S3A). For instance, Zt1A5
302 and Zt1E4 exhibit the same number of TE insertions on chromosome 9 (163 TEs),
303 however the distribution of these copies is different between the two genomes
304 (Figure 2A). We counted up to three times more TE insertions per chromosome
305 among the nine *Z. tritici* isolates (e.g. 68 vs 198 on chromosome 8) (Figure 2B).
306 Variation is mostly driven by LTR retrotransposon content (Figure S3B). Taken
307 together these results indicate that each *Z. tritici* genome reflects independent TE
308 insertion events and the recent transposition activity of few specific TE orders (e.g.
309 LTRs).

310 **Past and present TE activity has shaped the genomes of *Zymoseptoria* species**

311 TE-rich accessory chromosomes represent an ancestral trait of the genome
312 architecture among the *Zymoseptoria* species-complex. In agreement with a
313 previous study (Grandaubert *et al.* 2015), transposons in this study are enriched on
314 accessory chromosomes (Figure 3A). Rearrangements in the genomes also co-

315 occur in the vicinity of TEs, indicating the TEs' potential role in causing structural
316 variation among *Zymoseptoria* genomes. We identified a large interspecies
317 chromosomal inversion on contig_65 of *Z. brevis* Zb87 and contig_76 of *Z.*
318 *pseudotritici* Zp13 compared to the chromosome 2 of *Z. tritici* IPO323 (Feurtey *et al.*
319 2019). This large rearranging region actually consists of several inversions between
320 clusters of TEs (Figure S4). We identified three TE clusters in Zb87 (46 TEs
321 representing 0.12 Mb) and Zp13 (30 TEs representing 0.28 Mb) that surround the
322 inverted loci and are absent in *Z. tritici* IPO323 (Figure S4). We further scrutinized
323 the link between TEs and rearrangements by pairwise comparisons of the nine *Z.*
324 *tritici* isolates to the reference IPO323 (Table S3). Of all inverted regions identified
325 per isolate, 27% (Zt469) to 72% (Zt1A5) are flanked (separated by less than 10bp)
326 by transposons both upstream and downstream (Table S3). In total, we counted from
327 15 (Zt10) to 52 (Zt1E4) intra-specific inversions comprising more than 500 bp in
328 pairwise comparisons to the reference isolate IPO323 (Table S3). The cumulative
329 size of these inversions ranges from ~136.7 kb to ~845.2 kb in Zt1E4 and Zt469,
330 respectively. From 10 (Zt469) to 128 (Zt05) genes are located in inverted genomic
331 regions (Table S3). One inverted region of 170kb is found on chromosome 13 in the
332 reference genome but absent from the other isolates. Taken together these results
333 indicate that TEs have shaped genome rearrangements and genome
334 compartmentalization in *Zymoseptoria* species.

Presence/absence variation and effector genes associate with transposable elements

Presence/absence variable (PAV) genes are enriched in the vicinity of TEs while core genes are located further apart than expected by chance. To assess the potential impact of TEs on the high gene presence/absence variation described in *Z. tritici* (Plissonneau *et al.* 2018; Badet *et al.* 2019), we explored the genes close to TEs. We found that PAV genes are significantly closer to TEs in the genomes of *Z. ardabiliae*, *Z. brevis*, and the three *Z. tritici* isolates IPO323, Zt05 and Zt10 (permutation test of 1000 iterations, p-value< 0.05) (Figure 3B; Table S4). On the contrary, the distance between core genes and TEs is significantly higher than expected from a random distribution (permutation test of 1000 iterations p-value< 0.05). The exceptions are in *Z. pseudotritici*, and *Z. passerinii* in which observed distances between PAV genes and core genes to TEs were not significantly different from a random distribution (Table S4).

To assess if particular TE families are associated with genes, we tested the distribution of the major TE families from class I (i.e. *Copia*, *Gypsy* and *LINE* elements) and class II (*TIR*, *Helitron* and *MITE* elements) (Table S5). In all *Zymoseptoria* genomes *TIR* and *MITE* elements are located significantly closer to PAV genes than expected from a random distribution (applying a permutation test of 1000 iteration p-value< 0.05) (Table S5). Also *Copia* elements are significantly closer

355 to PAV genes than expected from a random distribution in *Z. tritici* (IPO323 and
356 Zt10) and *Z. ardabiliae*. In contrast, class I *Gypsy* elements are more distant to PAV
357 genes than expected from a random distribution in all *Zymoseptoria* genomes even
358 though they are the most abundant TEs (permutation test of 1000 iteration p-value<
359 0.05) (Table S5). In conclusion, different TE families are found closer to PAV genes
360 among the genomes of the *Zymoseptoria* species-complex.
361 Effector genes are enriched in the vicinity of TEs in the genomes of the
362 *Zymoseptoria* species-complex. We used the previously annotated effector genes of
363 *Z. ardabiliae*, *Z. brevis*, *Z. pseudotritici*, *Z. passerinii* and the three *Z. tritici* isolates
364 IPO323, Zt05 and Zt10 (Feurtey *et al.* 2019) and tested whether they are enriched in
365 the vicinity of TEs. We show that these effector genes are significantly closer to TEs
366 than expected if they were randomly distributed in *Zymoseptoria* genomes
367 (permutation test of 1000 iteration p-value< 0.05) (Figure 3B; Table S4). Effector
368 genes associate with TEs in each genome, but not with the same elements. Effector
369 genes of *Z. tritici* IPO323 associate with *TIR* elements while in Zt10 effector genes
370 associate with *Helitron* and *Copia* elements but not *TIR* elements (Table S5). As an
371 example, the effector gene Zt09_13_00269 shows a presence/absence
372 polymorphism among *Zymoseptoria tritici* isolates (Figure 3C). This effector is absent
373 from the *Aegilops*-infecting isolate Zt469 which could indicate a potential link with
374 host specificity. Zt09_13_00269 is surrounded by different TEs both up- and

375 downstream. Interestingly, this effector candidate is expressed during wheat
376 infection in the reference isolate IPO323 (56.9 transcripts per million - TPM), while
377 not expressed in Zt05 and Zt10 during wheat infection (1.23 TPM and 4.13 TPM
378 respectively) (Table S6). A specific MITE transposon localizes 881 bp upstream of
379 Zt09_13_00269 in IPO323 and in Zt289 but not in the other isolates (Figure 3C). In
380 conclusion, effectors genes are enriched in the vicinity of TEs suggesting that TEs
381 could play a role in the evolution of pathogenicity-related genes in the *Zymoseptoria*
382 species-complex.

383 **Signatures of Repeat-Induced Point mutations are reduced on TEs of**
384 **European *Z. tritici* isolates compared to other *Zymoseptoria* genomes**

385 RIP efficacy recently decreased in European *Z. tritici* isolates. We here evaluated the
386 RIP signatures in the *Zymoseptoria* genomes to estimate to what extent RIP affects
387 different species of the genus. To this end, we scanned each genome and calculated
388 RIP indices (see Methods). In total, RIP signatures affect between 17.4% (in Zt3D7)
389 and 34.5% (in Zpa63) of the total genomic TE content (Table 1). These RIP
390 proportions correspond to the TE content of each genome indicating RIP has been -
391 at least recently - active in the genomes of *Zymoseptoria* species. Therefore, as the
392 TE content in the genomes of *Zymoseptoria* spp. are recent insertions mostly and
393 RIP is found also on recent copies, it indicates that the RIP has been recently active
394 in the genomes. RIP efficacy is not equivalent for all types of TEs (van Wyk *et al.*

2019). The small-sized and non-autonomous TEs such as MITE DNA-transposons are less affected by RIP (Figure S7). Overall, in *Zymoseptoria*, the vast majority of TEs carries RIP signatures (86-99% of TEs are RIPed), based on the estimation of RIP composite indices per 50 bp windows of TEs (Figure 4). The large number of RIPed TEs suggest a RIP-induced decay of TE copies in the genomes of *Zymoseptoria* species. However, 8.1% to 13.5% of TEs in the European *Z. tritici* isolates do not have any RIP signature, while this percentage is only of 1 to 6.8% for TEs of the Iranian *Z. tritici* isolates (Figure 4; Figure S6). This reduction of RIP can be linked to the variation in TE repertoires among the genomes of *Z. tritici* isolates. We observed that the proportion of MITE elements (i.e. elements that are less affected by RIP) do not correlate with this reduction of RIP in copies. The Iranian *Z. tritici* isolates comprise from 29 to 121 MITE copies while European isolates comprise from 68 to 156 MITE copies (Table S1). The reduction of RIP signatures in European *Z. tritici* isolates indicates a relaxation of the RIP efficacy in those isolates, potentially due to a different composition in TE repertoires.

In addition, the average size of Large RIP-affected regions (LRARs) can be used as a proxy for RIP mechanism efficiency because it reflects to which extent a highly repeated region is affected by RIP. In the genomes of *Zymoseptoria* species, LRARs have an average size between 10.3 kb (in Zt1E4) and 25.4 kb (in Za17) and comprise large AT-rich regions (Table 1). LRAR average sizes in the *Z. tritici*

415 European isolates are reduced compared to the other members of the *Zymoseptoria*
416 species-complex (Table 1). It is worth noting that *Z. tritici* genomes with lower LRARs
417 average size do not have significantly lower TE content compared to the Iranian
418 isolates. This indicate that the average size of LRARs is a good indicator for RIP
419 efficiency. Besides TE sequences, only a small number of genes exhibited RIP
420 signatures, including 43 genes in *Z. pseudotritici* and 92 genes in *Z. tritici* Zt05
421 (Table S7). Taken together, we conclude that RIP is highly efficient to inactivate TEs
422 in the genomes of *Zymoseptoria* species but show evidence of relaxation in the
423 genomes of the European isolates of *Z. tritici*.

424 **Discussion**

425 Growing evidence demonstrates that transposable elements represent key players
426 for the evolution and adaptation of fungal plant pathogens (Möller and Stukenbrock
427 2017). We investigated how past and present transposition events have shaped
428 genome evolution of a major wheat pathogen and its wild-grass infecting sister-
429 species. Within the genus *Zymoseptoria,* TEs associate with effector genes and PAV
430 genes, as previously described for isolates of *Z. tritici* (Plissonneau *et al.* 2018). This
431 suggests that TEs are either directly or indirectly involved in evolution of effectors
432 and PAV genes in *Zymoseptoria* species (Faino *et al.* 2016). TEs may therefore
433 represent a major driver of *Zymoseptoria* genome evolution.

434

435 Our detailed transposon analysis confirmed a conserved genome
436 compartmentalization of TEs in accessory regions among species of the
437 *Zymoseptoria* genus. TE accumulation is often associated with genome size
438 expansions (Raffaele and Kamoun 2012). However, genome sizes of *Zymoseptoria*
439 spp. are comparatively stable; even a duplication of the transposon content does not
440 cause genome size variation larger than 2 Mb (Feurtey *et al.* 2019). This can be
441 explained by the high content of accessory chromosomes showing
442 presence/absence variation: indeed, these chromosomes are small but have a high
443 content of TEs (Grandaubert *et al.* 2015). This was previously shown for *Z. tritici*, and
444 here we show that genome compartmentalization involving dynamic TE content
445 represents a more ancestral trait of the genomes of these fungi (Goodwin *et al.*
446 2011). Purifying selection pressure is higher on the gene-dense core chromosomes
447 of *Z. tritici* compared to the transposon-dense accessory chromosomes
448 (Grandaubert *et al.* 2019). It is possible that relaxed selection acting on accessory
449 genome compartments allows TEs to accumulate in these regions.

450

451 TEs are involved either directly or indirectly in the genome plasticity of fungi. In this
452 study, we explored the impact of TEs on genome architecture and gene
453 presence/absence variation between and within species of *Zymoseptoria*. TEs
454 associate with regions exhibiting inter- and intra-species chromosomal

455 rearrangements. Furthermore, we found that genes affected by presence/absence
456 variation between and within species are located close to TEs. Such association of
457 gene presence/absence with TEs has also been observed in other plant pathogenic
458 fungi such as the rice blast *M. oryzae* (Yoshida *et al.* 2016; Bao *et al.* 2017).
459
460 TEs associate with effector genes in *Zymoseptoria* species, supporting the
461 importance of TE-driven gene evolution for pathogenicity. In the wheat pathogen
462 *Pyrenophora tritici-repentis*, the pathogenicity-related protein ToxA, is found in the
463 two other wheat-associated pathogens *Parastagonospora nodorum* and *Bipolaris*
464 *sorokiniana* (McDonald *et al.* 2019). Horizontal transfer of the ToxA encoding gene
465 was demonstrated between these wheat-associated species and transfer was
466 mediated by a DNA transposon from the hAT family (McDonald *et al.* 2019). TE
467 insertions that facilitate variation of pathogenicity-related traits have been reported in
468 several other fungal pathogens, including *L. maculans*, *F. oxysporum*, *M. oryzae* and
469 *Verticillium dahliae* (Rouxel *et al.* 2011; Amselem *et al.* 2015a; Faino *et al.* 2016;
470 Fokkens *et al.* 2018).
471
472 The majority of TEs in the genomes of *Zymoseptoria* species represents young
473 insertions. TE insertions in these closely related species are species- or strain-
474 specific and one third comprise copies of highly conserved sequences. Extensive

475 copy number variation along core chromosomes in the nine *Z. tritici* isolates
476 strengthens our conclusion that the majority of TEs has spread recently. A study of
477 TE dynamics in worldwide *Z. tritici* populations demonstrated expansions of TE
478 families in isolates of recently founded populations compared to isolates collected
479 close to the center of origin (Oggenfuss *et al.* 2020). More investigations are needed
480 to understand how TE expansions in *Z. tritici* closely related species occurred,
481 particularly for the higher repeated genomes of *Z. passerinii* and *Z. brevis*. It would
482 be interesting to explore if TEs also accumulate more in genomes of *Z. passerinii*
483 and *Z. brevis* isolates outside of the center of origins or if populations show different
484 TE accumulation patterns. Understanding how TEs have accumulated stronger in
485 these genomes requires comparing more *Z. passerinii* and *Z. brevis* isolates.
486 The high level of young TE insertions in *Zymoseptoria* species contrasts with the
487 high level of RIP mutations found in TE copies. The classical model of TE burst and
488 decay dynamics states that TE proliferation is counterbalanced by genome defence
489 mechanisms with more or less efficacy, which eventually leads to TE elimination
490 (Arkhipova 2018). Here, we observe high level of RIP signatures even on recent TE
491 copies suggesting that most TEs are likely inactive. Based on this, we speculate that
492 the majority of TEs in *Zymoseptoria* were affected by RIP shortly after their insertion.
493 It is worth noting that inactive TEs can still spread in genomes to a certain extent.

494 Indeed, recombination between homologous regions can lead to duplications
495 (Bourque *et al.* 2018).
496
497 The highly efficient RIP system may impede evolutionary innovation among
498 *Zymoseptoria.* RIP prevents gene duplications in addition to TE duplication, and
499 gene duplications are considered a major driver of genome evolution (Galagan and
500 Selker 2004). Although RIP can be 'leaky' and induce mutation on single-copy
501 genes, which has been demonstrated as efficient mechanism for effector
502 diversification in some fungal species (Rouxel *et al.* 2011). We found however very
503 few genes harbouring RIP signatures among the genomes of the *Zymoseptoria*
504 species, which could indicate that such mechanism of effector diversification via RIP
505 scarcely occur in *Zymoseptoria* species. The details of how RIP mechanism occurs
506 in *Zymoseptoria* spp. remain largely unknown. However, one key player in RIP is the
507 DNA methyltransferase DIM-2. RIP reduction in European *Z. tritici* isolates correlates
508 with a deficiency in DNA methylation and absence of DIM-2 proteins in European *Z.*
509 *tritici* isolates (Dhillon *et al.* 2010; Möller *et al.* 2020). Host genome defences against
510 TEs represent a potential fitness cost notably with regard to gene duplication for
511 rapid adaptation (Galagan and Selker 2004). For example, there are almost no gene
512 duplications in the genome of the model species *N. crassa* because of extremely
513 high RIP efficacy (Galagan and Selker 2004). In *Z. tritici*, a recent study

514 demonstrated the TE-induced expansion of MSF transporters (Oggenfuss *et al.*
515 2020). The latter are mostly involved in fungicide resistance in *Z. tritici* (Omrane *et al.*
516 2017a).
517
518 Moderate transposon activity has the potential to be advantageous for rapid
519 adaptation of fungal plant-pathogens. We propose that the relaxation of transposition
520 repression by RIP in *Z. tritici* isolates outside of the center of origin could represent
521 an advantage for rapid adaptation. For example, the increase of TE insertions in *Z.*
522 *tritici* could be advantageous during migration to new environments following wheat
523 deployment across the world (Oggenfuss *et al.* 2020). In the cereal pathogen *M.*
524 *oryzae*, DIM-2 is functional in isolates infecting wheat, rice and common millet, but
525 isolates infecting foxtail millet carry a non-functional DIM-2 variant. There may be a
526 link between DNA methylation of transposons and infection success of *M. oryzae* on
527 specific host plant species (Ikeda *et al.* 2013). Based on the findings of our study,
528 relaxation in RIP efficacy among *Z. tritici* isolates might potentially be advantageous
529 for rapid adaptation to new wheat cultivars.
530
531 **Acknowledgements**
532 CL was funded by the Institut national de la recherche agronomique (INRA) in the
533 framework of a “Contrat Jeune Scientifique” and by the Labex ARBRE (Lab of

534 Excellence ARBRE). Genome research in the group of EHS is funded by the Max
535 Planck Society and CIFAR. AF is funded by the DFG Priority Program SPP1819. CL
536 performed data and results analyses. EHS and CL contributed to the design and
537 implementation of the research. All authors contributed to the writing of the
538 manuscript. All authors read and approved the manuscript.

539

540

References

Amselem J., M. Vigouroux, S. Oberhaensli, J. K. M. Brown, L. V Bindschedler, *et al.*, 2015a Evolution of the EKA family of powdery mildew avirulence-effector genes from the ORF 1 of a LINE retrotransposon. BMC Genomics 16: 917. https://doi.org/10.1186/s12864-015-2185-x

Amselem J., M.-H. Lebrun, and H. Quesneville, 2015b Whole genome comparative analysis of transposable elements provides new insight into mechanisms of their inactivation in fungal genomes. BMC Genomics 16: 141. https://doi.org/10.1186/s12864-015-1347-1

Arkhipova I. R., 2018 Neutral Theory, Transposable Elements, and Eukaryotic Genome Evolution. Mol. Biol. Evol. 35: 1332–1337. https://doi.org/10.1093/molbev/msy083

Badet T., U. Oggenfuss, L. Abraham, B. A. McDonald, and D. Croll, 2019 A 19-isolate reference-quality global pangenome for the fungal wheat pathogen Zymoseptoria tritici. bioRxiv 803098. https://doi.org/10.1101/803098

Badet T., U. Oggenfuss, L. Abraham, B. A. McDonald, and D. Croll, 2020 A 19-isolate reference-quality global pangenome for the fungal wheat pathogen Zymoseptoria tritici. BMC Biol. 18: 12. https://doi.org/10.1186/s12915-020-0744-3

Bao J., M. Chen, Z. Zhong, W. Tang, L. Lin, *et al.*, 2017 PacBio Sequencing Reveals Transposable Elements as a Key Contributor to Genomic Plasticity and Virulence Variation in Magnaporthe oryzae. Mol. Plant 10: 1465–1468.

Bertazzoni S., A. H. Williams, D. A. Jones, R. A. Syme, K.-C. Tan, *et al.*, 2018 Accessories Make the Outfit: Accessory Chromosomes and Other Dispensable DNA Regions in Plant-Pathogenic Fungi. Mol. Plant-Microbe Interact. 31: 779–788. https://doi.org/10.1094/MPMI-06-17-0135-FI

Biémont C., 2010 A brief history of the status of transposable elements: from junk DNA to major players in evolution. Genetics 186: 1085–1093. https://doi.org/10.1534/genetics.110.124180

Bourque G., K. H. Burns, M. Gehring, V. Gorbunova, A. Seluanov, *et al.*, 2018 Ten things you should know about transposable elements. Genome Biol. 19: 199. https://doi.org/10.1186/s13059-018-1577-z

Castanera R., L. López-Varas, A. Borgognone, K. LaButti, A. Lapidus, *et al.*, 2016 Transposable Elements versus the Fungal Genome: Impact on Whole-Genome Architecture and Transcriptional Profiles. PLoS Genet. 12. https://doi.org/10.1371/journal.pgen.1006108

Chen E. C. H., E. Morin, D. Beaudet, J. Noel, G. Yildirir, *et al.*, 2018 High intraspecific genome diversity in the model arbuscular mycorrhizal symbiont Rhizophagus irregularis. New Phytol. 220: 1161–1171. https://doi.org/10.1111/nph.14989

Croll D., and B. A. McDonald, 2012 The Accessory Genome as a Cradle for Adaptive Evolution in Pathogens. PLOS Pathog. 8: e1002608. https://doi.org/10.1371/journal.ppat.1002608

Cuomo C. A., U. Güldener, J. R. Xu, F. Trail, B. G. Turgeon, *et al.*, 2007 The Fusarium graminearum genome reveals a link between localized polymorphism and pathogen specialization. Science (80-.). 317: 1400–1402. https://doi.org/10.1126/science.1143708

Deniz Ö., J. M. Frost, and M. R. Branco, 2019 Regulation of transposable elements by DNA modifications. Nat. Rev. Genet. 20: 417–431.

590 Dhillon B., J. R. Cavaletto, K. V Wood, and S. B. Goodwin, 2010 Accidental
591 Amplification and Inactivation of a Methyltransferase Gene Eliminates Cytosine
592 Methylation in Mycosphaerella graminicola. Genetics 186: 67 LP – 77.
593 https://doi.org/10.1534/genetics.110.117408
594 Dhillon B., N. Gill, R. C. Hamelin, and S. B. Goodwin, 2014 The landscape of
595 transposable elements in the finished genome of the fungal wheat pathogen
596 Mycosphaerella graminicola. BMC Genomics 15: 1132.
597 https://doi.org/10.1186/1471-2164-15-1132
598 Dong S., S. Raffaele, and S. Kamoun, 2015 The two-speed genomes of filamentous
599 pathogens: waltz with plants. Curr. Opin. Genet. Dev. 35: 57–65.
600 https://doi.org/10.1016/j.gde.2015.09.001
601 Duplessis S., C. A. Cuomo, Y. C. Lin, A. Aerts, E. Tisserant, *et al.*, 2011 Obligate
602 biotrophy features unraveled by the genomic analysis of rust fungi. Proc. Natl.
603 Acad. Sci. U. S. A. 108: 9166–9171. https://doi.org/10.1073/pnas.1019315108
604 Faino L., M. F. Seidl, X. Shi-Kunne, M. Pauper, G. C. M. van den Berg, *et al.*, 2016
605 Transposons passively and actively contribute to evolution of the two-speed
606 genome of a fungal pathogen. Genome Res. 26: 1091–1100.
607 https://doi.org/10.1101/gr.204974.116
608 Feurtey A., C. Lorrain, D. Croll, C. Eschenbrenner, M. Freitag, *et al.*, 2019 Genome
609 compartmentalization predates species divergence in the plant pathogen genus
610 Zymoseptoria bioRxiv 864561.
611 https://doi.org/10.1101/864561
612 Flutre T., E. Duprat, C. Feuillet, and H. Quesneville, 2011 Considering Transposable
613 Element Diversification in De Novo Annotation Approaches, (Y. Xu, Ed.). PLoS
614 One 6: e16526. https://doi.org/10.1371/journal.pone.0016526
615 Fokkens L., S. Shahi, L. R. Connolly, R. Stam, S. M. Schmidt, *et al.*, 2018 The multi-
616 speed genome of Fusarium oxysporum reveals association of histone
617 modifications with sequence divergence and footprints of past horizontal
618 chromosome transfer events. bioRxiv 465070. https://doi.org/10.1101/465070
619 Fouché S., C. Plissonneau, and D. Croll, 2018 The birth and death of effectors in
620 rapidly evolving filamentous pathogen genomes. Curr. Opin. Microbiol. 46: 34–
621 42.
622 Fouché S., T. Badet, U. Oggenfuss, C. Plissonneau, C. S. Francisco, *et al.*, 2020
623 Stress-Driven Transposable Element De-repression Dynamics and Virulence
624 Evolution in a Fungal Pathogen. Mol. Biol. Evol. 37: 221–239.
625 https://doi.org/10.1093/molbev/msz216
626 Frantzeskakis L., B. Kracher, S. Kusch, M. Yoshikawa-Maekawa, S. Bauer, *et al.*,
627 2018 Signatures of host specialization and a recent transposable element burst
628 in the dynamic one-speed genome of the fungal barley powdery mildew
629 pathogen. BMC Genomics 19. https://doi.org/10.1186/s12864-018-4750-6
630 Galagan J. E., and E. U. Selker, 2004 RIP: the evolutionary cost of genome defense.
631 Trends Genet. 20: 417–423. https://doi.org/10.1016/j.tig.2004.07.007
632 Gao Q., K. Jin, S.-H. Ying, Y. Zhang, G. Xiao, *et al.*, 2011 Genome Sequencing and
633 Comparative Transcriptomics of the Model Entomopathogenic Fungi
634 Metarhizium anisopliae and M. acridum, (M. Achtman, Ed.). PLoS Genet. 7:
635 e1001264. https://doi.org/10.1371/journal.pgen.1001264
636 Gel B., A. Díez-Villanueva, E. Serra, M. Buschbeck, M. A. Peinado, *et al.*, 2015
637 regioneR: an R/Bioconductor package for the association analysis of genomic
638 regions based on permutation tests. Bioinformatics btv562.
639 https://doi.org/10.1093/bioinformatics/btv562

Gioti A., J. E. Stajich, and H. Johannesson, 2013 Neurospora and the dead-end hypothesis: genomic consequences of selfing in the model genus. Evolution 67: 3600–3616. https://doi.org/10.1111/evo.12206

Gladyshev E., and N. Kleckner, 2014 Direct recognition of homology between double helices of DNA in Neurospora crassa. Nat. Commun. 5: 3509. https://doi.org/10.1038/ncomms4509

Gladyshev E., and N. Kleckner, 2017 DNA sequence homology induces cytosine-to-thymine mutation by a heterochromatin-related pathway in Neurospora. Nat. Genet. 49: 887–894. https://doi.org/10.1038/ng.3857

Gladyshev E., 2017 Repeat-Induced Point Mutation and Other Genome Defense Mechanisms in Fungi. Microbiol. Spectr. 5. https://doi.org/10.1128/microbiolspec.FUNK-0042-2017

Goel M., H. Sun, W.-B. Jiao, and K. Schneeberger, 2019 SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. Genome Biol. 20: 277. https://doi.org/10.1186/s13059-019-1911-0

Goodwin S. B., S. Ben M'Barek, B. Dhillon, A. H. J. Wittenberg, C. F. Crane, *et al.*, 2011 Finished Genome of the Fungal Wheat Pathogen Mycosphaerella graminicola Reveals Dispensome Structure, Chromosome Plasticity, and Stealth Pathogenesis. PLoS Genet. 7: e1002070. https://doi.org/10.1371/journal.pgen.1002070

Grandaubert J., R. G. Lowe, J. L. Soyer, C. L. Schoch, A. P. Van de Wouw, *et al.*, 2014 Transposable element-assisted evolution and adaptation to host plant within the Leptosphaeria maculans-Leptosphaeria biglobosa species complex of fungal pathogens. BMC Genomics 15: 891. https://doi.org/10.1186/1471-2164-15-891

Grandaubert J., A. Bhattacharyya, and E. H. Stukenbrock, 2015 RNA-seq-Based Gene Annotation and Comparative Genomics of Four Fungal Grass Pathogens in the Genus Zymoseptoria Identify Novel Orphan Genes and Species-Specific Invasions of Transposable Elements. G3 Genes, Genomes, Genet. 5: 1323–33. https://doi.org/10.1534/g3.115.017731

Grandaubert J., J. Y. Dutheil, and E. H. Stukenbrock, 2019 The genomic determinants of adaptive evolution in a fungal pathogen. Evol. Lett. 3: 299–312. https://doi.org/10.1002/evl3.117

Hartmann F. E., B. A. McDonald, and D. Croll, 2018 Genome-wide evidence for divergent selection between populations of a major agricultural pathogen. Mol. Ecol. 27: 2725–2741. https://doi.org/10.1111/mec.14711

Haubold B., F. Klötzl, and P. Pfaffelhuber, 2015 andi: Fast and accurate estimation of evolutionary distances between closely related genomes. Bioinformatics 31: 1169–1175. https://doi.org/10.1093/bioinformatics/btu815

Haueisen J., M. Möller, C. J. Eschenbrenner, J. Grandaubert, H. Seybold, *et al.*, 2018 Highly flexible infection programs in a specialized wheat pathogen. Ecol. Evol. https://doi.org/10.1002/ece3.4724

Haueisen J., M. Möller, C. J. Eschenbrenner, J. Grandaubert, H. Seybold, *et al.*, 2019 Highly flexible infection programs in a specialized wheat pathogen. Ecol. Evol. 9: 275–294. https://doi.org/10.1002/ece3.4724

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

Ikeda K. I., B. Van Vu, N. Kadotani, M. Tanaka, T. Murata, *et al.*, 2013 Is the fungus Magnaporthe losing DNA methylation? Genetics 195: 845–855.

https://doi.org/10.1534/genetics.113.155978
Kellner R., A. Bhattacharyya, S. Poppe, T. Y. Hsu, R. B. Brem, *et al.*, 2014 Expression profiling of the wheat pathogen Zymoseptoria tritici reveals genomic patterns of transcription and host-specific regulatory programs. Genome Biol. Evol. 6: 1353–1365. https://doi.org/10.1093/gbe/evu101
Krzywinski M., J. Schein, I. Birol, J. Connors, R. Gascoyne, *et al.*, 2009 Circos: an information aesthetic for comparative genomics. Genome Res. 19: 1639–1645. https://doi.org/10.1101/gr.092759.109
Lechner M., S. Findeiß, L. Steiner, M. Marz, P. F. Stadler, *et al.*, 2011 Proteinortho: Detection of (Co-)orthologs in large-scale analysis. BMC Bioinformatics 12: 124. https://doi.org/10.1186/1471-2105-12-124
McDonald M. C., A. P. Taranto, E. Hill, B. Schwessinger, Z. Liu, *et al.*, 2019 Transposon-Mediated Horizontal Transfer of the Host-Specific Virulence Protein ToxA between Three Fungal Wheat Pathogens. MBio 10: e01515-19. https://doi.org/10.1128/mBio.01515-19
Möller M., and E. H. Stukenbrock, 2017 Evolution and genome architecture in fungal plant pathogens. Nat. Rev. Microbiol. 15: 756–771. https://doi.org/10.1038/nrmicro.2017.76
Möller M., M. Habig, C. Lorrain, A. Feurtey, J. Haueisen, *et al.*, 2020 Recent loss of the Dim2 cytosine DNA methyltransferase impacts mutation rate and evolution in a fungal plant pathogen. bioRxiv 2020.03.27.012203. https://doi.org/10.1101/2020.03.27.012203
Oggenfuss U., T. Badet, T. Wicker, F. E. Hartmann, N. K. Singh, *et al.*, 2020 A population-level invasion by transposable elements in a fungal pathogen. bioRxiv. https://doi.org/10.1101/2020.02.11.944652
Omrane S., C. Audéon, A. Ignace, C. Duplaix, L. Aouini, *et al.*, 2017a Plasticity of the MFS1 promoter leads to multi drug resistance in the wheat pathogen Zymoseptoria tritici. mSphere 1–42.
Omrane S., C. Audéon, A. Ignace, C. Duplaix, L. Aouini, *et al.*, 2017b Plasticity of the MFS1 Promoter Leads to Multidrug Resistance in the Wheat Pathogen Zymoseptoria tritici. mSphere 2: e00393-17. https://doi.org/10.1128/mSphere.00393-17
Plissonneau C., F. E. Hartmann, and D. Croll, 2018 Pangenome analyses of the wheat pathogen Zymoseptoria tritici reveal the structural basis of a highly plastic eukaryotic genome. BMC Biol. 16: 5. https://doi.org/10.1186/s12915-017-0457-4
Quesneville H., C. M. Bergman, O. Andrieu, D. Autard, D. Nouaud, *et al.*, 2005 Combined Evidence Annotation of Transposable Elements in Genome Sequences. PLoS Comput. Biol. 1: e22. https://doi.org/10.1371/journal.pcbi.0010022
Quinlan A. R., and I. M. Hall, 2010 BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841–842. https://doi.org/10.1093/bioinformatics/btq033
Raffaele S., and S. Kamoun, 2012 Genome evolution in filamentous plant pathogens: why bigger can be better. Nat. Rev. Microbiol. 10: 417–430. https://doi.org/10.1038/nrmicro2790
Roessler K., A. Bousios, E. Meca, and B. S. Gaut, 2018 Modeling Interactions between Transposable Elements and the Plant Epigenetic Response: A Surprising Reliance on Element Retention. Genome Biol. Evol. 10: 803–815. https://doi.org/10.1093/gbe/evy043
Rouxel T., J. Grandaubert, J. K. Hane, C. Hoede, A. P. van de Wouw, *et al.*, 2011

740 Effector diversification within compartments of the Leptosphaeria maculans
741 genome affected by Repeat-Induced Point mutations. Nat. Commun. 2: 202.
742 https://doi.org/10.1038/ncomms1189
743 Schotanus K., J. L. Soyer, L. R. Connolly, J. Grandaubert, P. Happel, *et al.*, 2015
744 Histone modifications rather than the novel regional centromeres of
745 *Zymoseptoria tritici* distinguish core and accessory chromosomes. Epigenetics
746 Chromatin 8: 41. https://doi.org/10.1186/s13072-015-0033-5
747 Soyer J. L., M. El Ghalid, N. Glaser, B. Ollivier, J. Linglin, *et al.*, 2014 Epigenetic
748 Control of Effector Gene Expression in the Plant Pathogenic Fungus
749 Leptosphaeria maculans. PLOS Genet. 10: e1004227.
750 Stam R., M. Münsterkötter, S. D. Pophaly, L. Fokkens, H. Sghyer, *et al.*, 2018 A New
751 Reference Genome Shows the One-Speed Genome Structure of the Barley
752 Pathogen Ramularia collo-cygni. Genome Biol. Evol. 10: 3243–3249.
753 https://doi.org/10.1093/gbe/evy240
754 Stukenbrock E. H., S. Banke, M. Javan-Nikkhah, and B. A. McDonald, 2007 Origin
755 and domestication of the fungal wheat pathogen Mycosphaerella graminicola
756 via sympatric speciation. Mol. Biol. Evol. 24: 398–411.
757 https://doi.org/10.1093/molbev/msl169
758 Stukenbrock E. H., W. Quaedvlieg, M. Javan-Nikhah, M. Zala, P. W. Crous, *et al.*,
759 2012 *Zymoseptoria ardabiliae* and *Z. pseudotritici*, two progenitor species of the
760 septoria tritici leaf blotch fungus *Z. tritici* (synonym: *Mycosphaerella*
761 *graminicola*). Mycologia 104: 1397–407. https://doi.org/10.3852/11-374
762 Vaghefi N., J. R. Kikkert, M. D. Bolton, L. E. Hanson, G. A. Secor, *et al.*, 2017 De
763 novo genome assembly of Cercospora beticola for microsatellite marker
764 development and validation. Fungal Ecol. 26: 125–134.
765 https://doi.org/10.1016/J.FUNECO.2017.01.006
766 Wicker T., F. Sabot, A. Hua-Van, J. L. Bennetzen, P. Capy, *et al.*, 2007 A unified
767 classification system for eukaryotic transposable elements. Nat. Rev. Genet. 8:
768 973–982. https://doi.org/10.1038/nrg2165
769 Wyk S. van, C. H. Harrison, B. D. Wingfield, L. De Vos, N. A. van der Merwe, *et al.*,
770 2019 The RIPper, a web-based tool for genome-wide quantification of Repeat-
771 Induced Point (RIP) mutations, (I. Khan, Ed.). PeerJ 7: e7447.
772 https://doi.org/10.7717/peerj.7447
773 Yoshida K., D. G. O. Saunders, C. Mitsuoka, S. Natsume, S. Kosugi, *et al.*, 2016
774 Host specialization of the blast fungus Magnaporthe oryzae is associated with
775 dynamic gain and loss of genes linked to transposable elements. BMC
776 Genomics 17: 370. https://doi.org/10.1186/s12864-016-2690-6
777
778

Figures legends

Figure 1: Transposable element content and identity variation in the *Zymoseptoria* genus

A) Phylogenetic tree constructed by the software Andi (Haubold *et al.* 2015), using whole genome assemblies. The tree was rooted with *Cercospora beticola* (Vaghefi *et al.* 2017) as outgroup (data not shown). B) Bars represent TE content (%) per genome estimated after REPET (Flutre *et al.* 2011) annotation. Colours represent TE order coverage with retrotransposons (LTR, LINE and other class I orders in warm colors) and DNA transposons (TIR, MITE and other class II orders in cold colors). C) Sequence identity distribution between TE copies to their cognate consensus. Each dot represents the median sequence identity of TE cluster. Boxplots are coloured in regards to the species and isolate geographical origin.

Figure 2: **Transposable element content variation along chromosomes of nine *Z. tritici* isolates.**

A) Circos plot of the TE content along chromosomes of nine *Z. tritici* isolates. The first track represents the karyotype of the reference isolate IPO323; the second track shows a heatmap of TE density per 100kb windows in IPO323. Track three to ten represent the density of TE copies per 100kb windows in other *Z. tritici* isolates.

799 Chromosomal coordinates refer to the closest orthologous gene projected on the
800 IPO323 genome. Darker colors correspond to a higher TE density. B) TE content
801 variation of the 13 core chromosomes of *Z. tritici*. The boxplots represent the
802 distribution of TE coverage percentage per chromosome or contig among the nine *Z.*
803 *tritici* isolates.
804
805 **Figure 3**: **Transposable element insertions shape genome**
806 **compartmentalization and plasticity**
807 A) Bars represents mean percentage of TE content per genome compartment of
808 *Zymoseptoria* species of core (dark grey) and non-core (light grey)
809 chromosomes/contigs. Boxplots describe the mean TE coverage percentage per
810 genome compartment among all thirteen genomes of *Zymoseptoria* species. P-value
811 was estimated using a Kruskal-Wallis test. B) Distribution of distance between
812 effectors (blue), species- (red) and genus-specific (dark orange) genes and core
813 genes (light orange) to the closest TE per chromosome/contig. Mean distances and
814 permutation test results per gene category are summarised in Table S3. C) Example
815 of a presence/absence polymorphism of the candidate effector Zt09_13_00269
816 identified in (Haueisen *et al.* 2018) in the vicinity of TE-rich region. TEs are shown in
817 orange, genes in grey and the effector of interest is shown in blue boxes. Connecting
818 lines in grey represent the orthologous genes in each genome.

Figure 4: Repeat Induced Point (RIP) mutations in transposable elements of *Zymoseptoria* spp. genomes

Histograms of Composite RIP index (CRI) frequencies of TEs estimated using a 50bp sliding windows approach as follows: CRI = (TpA/ ApT) – (CpA + TpG/ ApC + GpT) for A) *Z*. *passerinii*, Z*. ardabiliae*, *Z. brevis* and *Z. pseudotritici*, and B) Iranian *Z. tritici* isolates and C) European *Z. tritici* isolates. Vertical dash lines exhibit the threshold (0) above which CRI values indicate a RIP signature.

Figure S1: Burst and decay TE evolution. Insertion of a TE copy (brown box) at a locus of the host genome, followed by further insertions during time and an accumulation of mutations (dark red triangles) and structural modifications such as partial deletions (dash lines) and/or insertions of other TEs (gold box). The older the insertion the more variants it accumulates. Therefore, TE families with old insertions are less similar to their consensus sequence while younger insertions are highly similar to the consensus. Consensus sequence represents the ancestral sequence of each annotated TE cluster (Flutre *et al.* 2011).

Figure S2: Relative abundance of TE divergent (left panel) and conserved (right panel) copies per genome. The divergent copies have less than 85% of sequence

identity to their cognate consensus while conserved copies have more than 95% of sequence identity to the consensus. Retrotransposons are represented in warm colors (LTR in light orange, LINE in orange and other retrotransposons orders in red) and DNA transposons are represented in cold colors (TIR in dark blue, MITE in blue and other DNA transposons in light blue). Uncategorized TEs are represented in grey.

Figure S3: Intraspecific variation in TE coverage per chromosome/contig

A) Total TE coverage per core chromosome/contig in the genomes of *Z. tritici* isolates. B) TE orders coverage contributions per core chromosome/contig of nine *Z.tritici* isolates.

Figure S4: Chromosomal inversion identified in *Z. tritici* IPO323 compared to *Z. brevis* Zb87 and *Z. pseudotritici* Zp13. TEs are represented in orange, genes in grey and effectors genes in blue. Links indicate orthologous genes. Arrows represent effector genes loci.

Figure S5: Comparison of Repeat-induced point mutation (RIP) composite index variation per sample group. Distribution comparison of RIP index (CRI) frequencies of TEs estimated using a 50bp sliding windows approach as follows: CRI

859 = (TpA/ ApT) – (CpA + TpG/ ApC + GpT) for Sister species: *Z. passerinii*, Z.
860 *ardabiliae*, *Z. brevis* and *Z. pseudotritici* (light orange), Iranian *Z. tritici* isolates (light
861 blue) and C) European *Z. tritici* isolates (dark blue). P-values were estimated using
862 Kruskal-Wallis test.
863
864 **Figure S6: Repeat-induced point mutation (RIP) composite index per TE copy**
865 **of all *Zymoseptoria* species.** Mean RIP composite index (CRI) frequencies of TEs
866 estimated using a 50bp sliding windows approach as follows: CRI = (TpA/ ApT) –
867 (CpA + TpG/ ApC + GpT) for each TE copy per order from Class I and Class II.
868
869 **Table S1: TE content in thirteen genomes of *Zymoseptoria* genus metrics.** TE
870 classification was performed based on Wicker's classification system (Wicker *et al.*
871 2007). The two classes (Class I and Class II) of TE are subdivided into subclasses,
872 orders and superfamilies as follows: Long terminal repeats (LTR) elements,
873 *Dictyostelium* intermediate repeat sequence (DIRS), Penelope-like elements (PLEs),
874 Long INterspersed Elements (LINEs) and Short INterspersed Elements (SINEs).
875 Terminal inverted repeat (TIR), Crypton, Helitron, Maverick and Miniature inverted-
876 repeat transposable element (MITEs).
877
878 **Table S2: Metrics output from REPET annotation of TEs.**

879

880 **Table S3: Chromosomal inversions between *Z. tritici* isolates compared to the**

881 **reference IPO323.**

882

883 **Table S4: Permutations tests based on mean distance of TEs to genes.**

884

885 **TableS5: Permutations tests based on mean distance of TE orders to genes.**

886

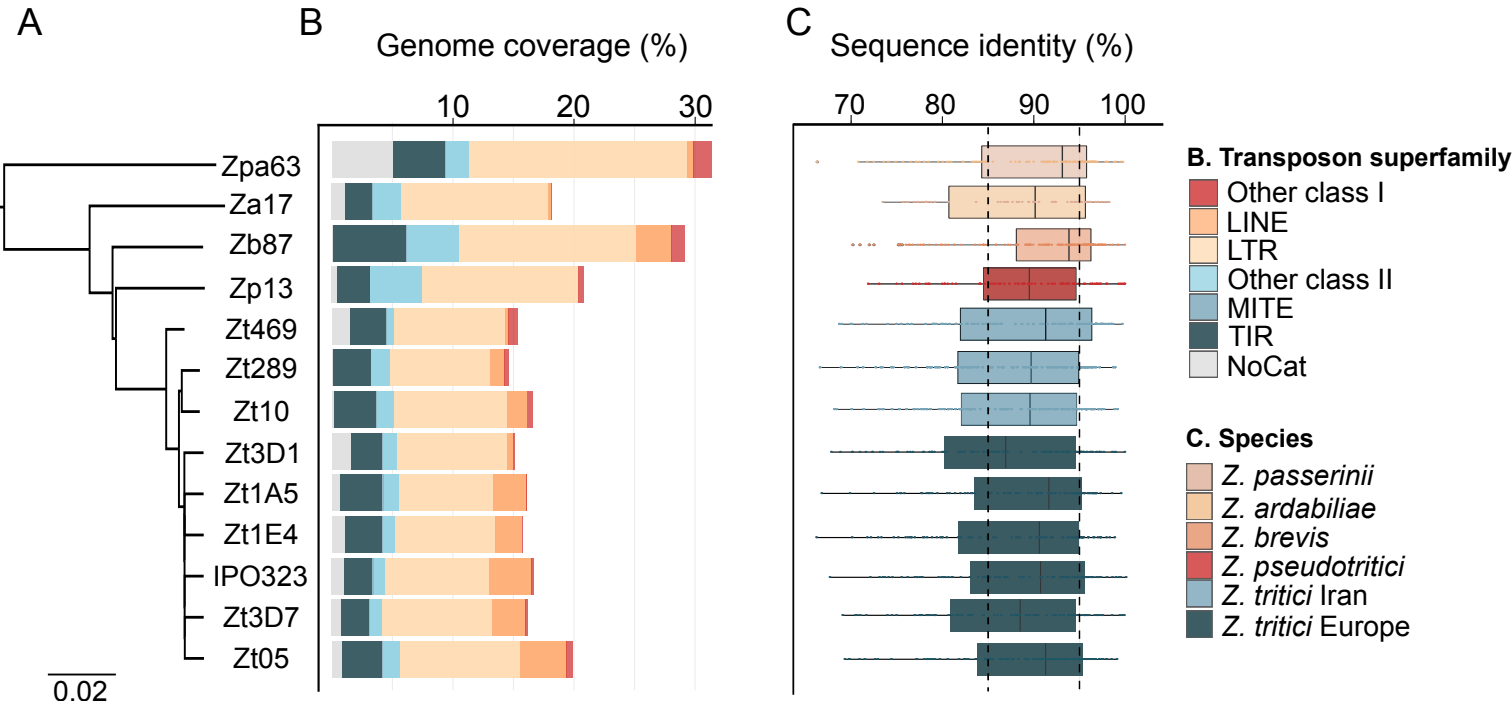
887 **TableS6: Effectors candidates and their neighbouring TEs**

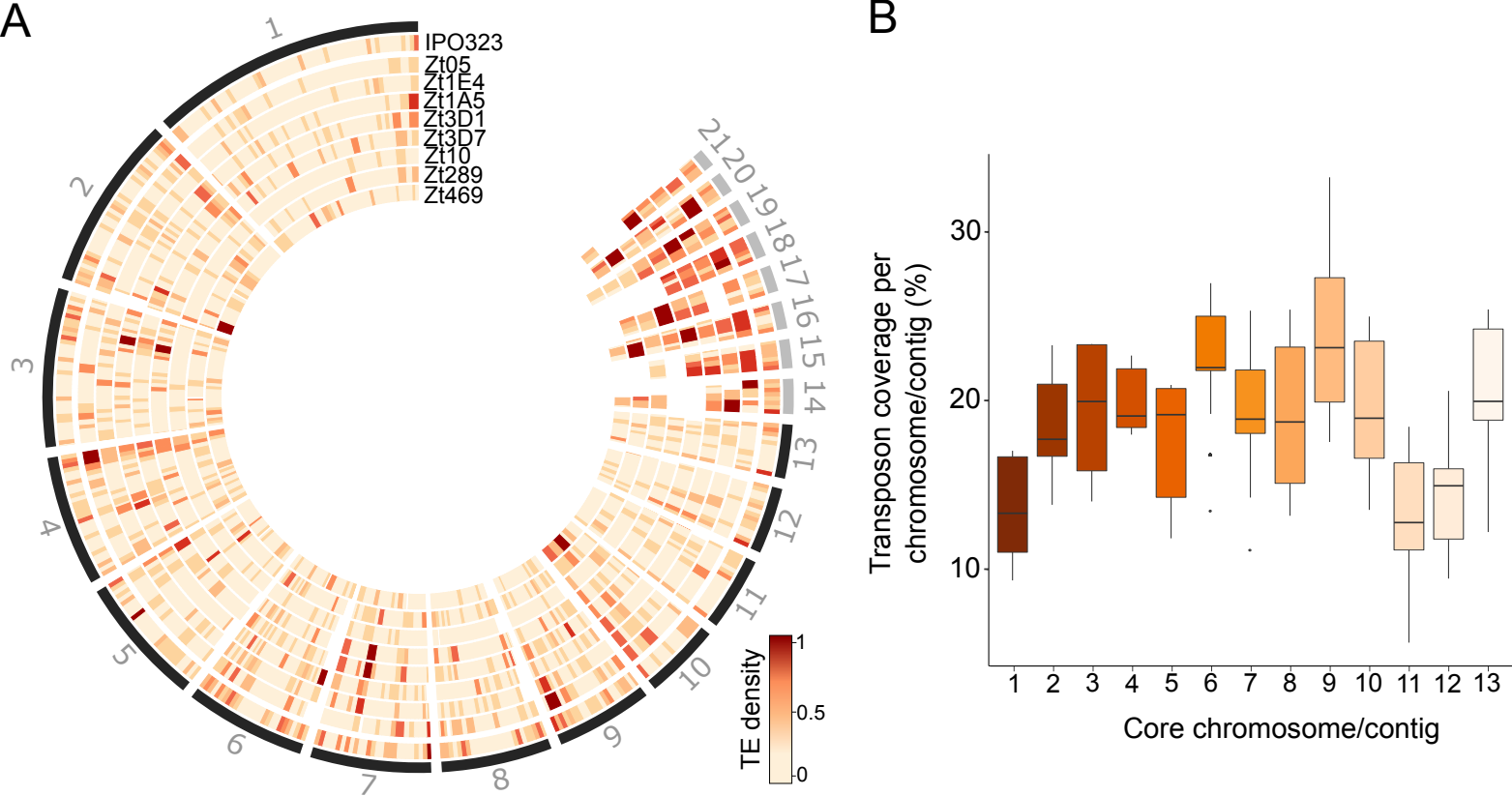
888

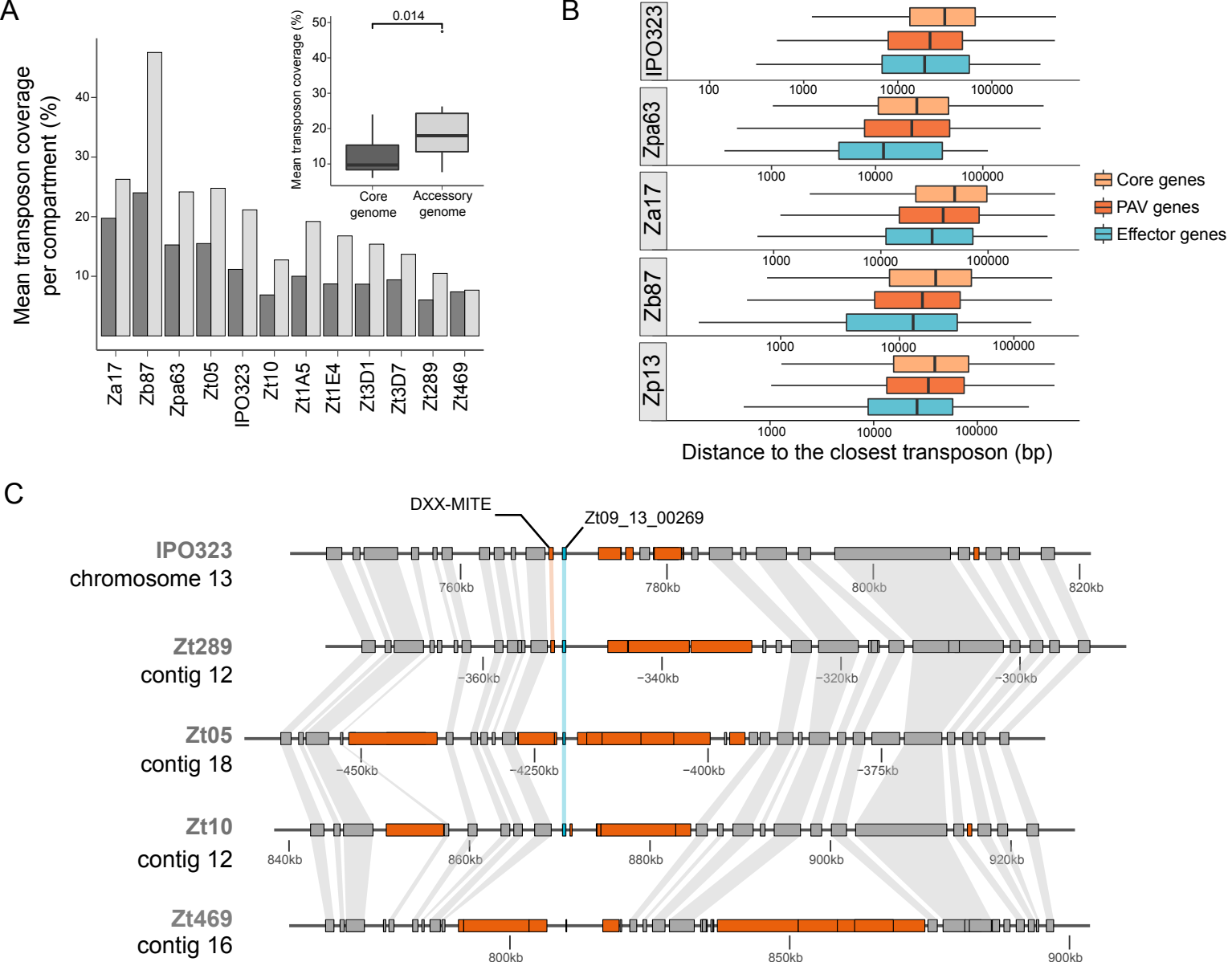
889 **Table S7: Genes affected by Repeat-induced point mutations signatures**

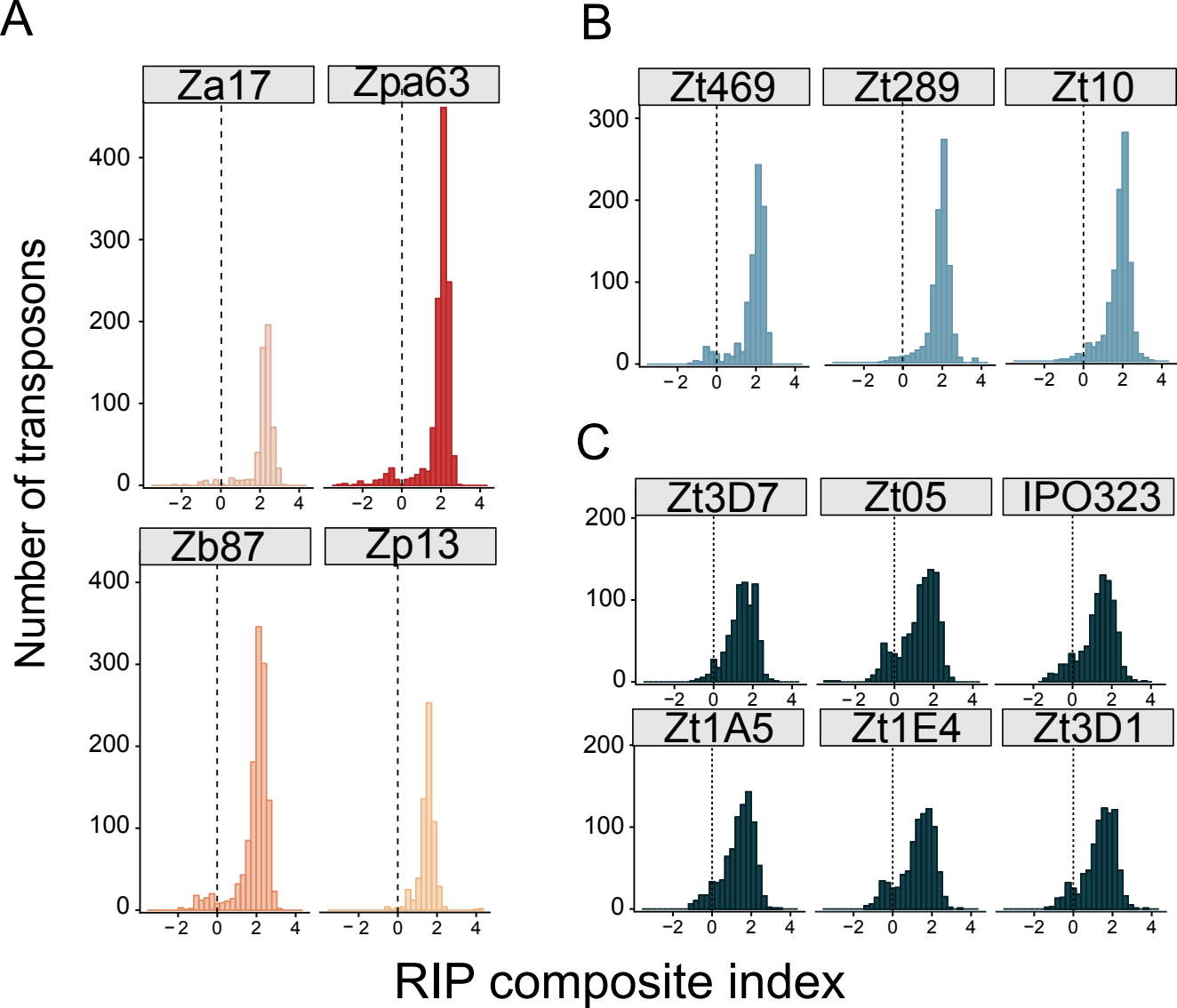
890

891


A
B
Genome coverage (%)
10
20
30
C
Sequence identity (%)
70
80
90
100
Zpa63
Za17
Zb87
Zp13
Zt469
Zt289
Zt10
Zt3D1
Zt1A5
Zt1E4
IPO323
Zt3D7
Zt05
0.02
B. Transposon superfamily
Other class I
LINE
LTR
Other class II
MITE
TIR
NoCat
C. Species
Z. passerinii
Z. ardabiliae
Z. brevis
Z. pseudotritici
Z. tritici Iran
Z. tritici Europe


A
IPO323
Zt05
Zt1E4
Zt1A5
Zt3D1
Zt3D7
Zt10
Zt289
Zt469
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
TE density
1
0.5
0
B
Transposon coverage per chromosome/contig (%)
30
20
10
1
2
3
4
5
6
7
8
9
10
11
12
13
Core chromosome/contig


A
Mean transposon coverage per compartment (%)
Za17
Zb87
Zpa63
Zt05
IPO323
Zt10
Zt1A5
Zt1E4
Zt3D1
Zt3D7
Zt289
Zt469
Mean transposon coverage (%)
0.014
Core genome
Accessory genome
B
IPO323
Zpa63
Za17
Zb87
Zp13
Core genes
PAV genes
Effector genes
Distance to the closest transposon (bp)
C
DXX-MITE
Zt09_13_00269
IPO323
chromosome 13
Zt289
contig 12
Zt05
contig 18
Zt10
contig 12
Zt469
contig 16


A
B
C
Za17
Zpa63
Zb87
Zp13
Zt469
Zt289
Zt10
Zt3D7
Zt05
IPO323
Zt1A5
Zt1E4
Zt3D1
Number of transposons
RIP composite index

1 **Table 1: Repeat-induced point mutation signatures in the genomes of *Zymoseptoria* species**

2

	Z. ardabiliae (Za17)	*Z. brevis* (Zb87)	*Z. pseudotritici* (Zp13)	Z. passerinii (Zpa63)	*Z. tritici* (Zt469)	*Z. tritici* (Zt289)	*Z. tritici* (Zt10)	*Z. tritici* (IPO323)	*Z. tritici* (Zt05)	*Z. tritici* (Zt1E4)	*Z. tritici* (Zt3D1)	*Z. tritici* (Zt1A5)	*Z. tritici* (Zt3D7)
Genome Size (Mbp)	38.1	41.6	40.3	41.4	42.9	39.0	39.2	39.7	41.2	38.7	40.6	39.7	37.9
GC content of entire genome (%)	51.42	49.97	51.6	49.71	48.52	51.53	51.77	52.13	51.94	52.26	52.01	52.2	52.22

Total estimated Genome-wide RIP* (%)	22.85	30.76	25.84	34.52	17.44	18.55	22.05	19.78	21.77	16.65	19.55	17.41	17.4
Average size of LRAR (kbp)	25.4	19.1	24.1	19.8	18.2	15.9	16.8	13.0	13.5	10.4	11.7	11.1	11.7
Average GC Content of	41.05	40.66	44.33	40.8	41.81	42.74	42.84	43.48	43.64	43.57	43.64	43.59	43.39

LRAR ** **(%)**													
Sum Of all LRAR (Mbp)	7.3	11.9	9.3	12.2	7.0	6.3	8.1	7.2	8.2	5.6	7.1	6.2	5.9

*RIP: Repeat-induced point mutation; ** LRAR: Large RIP Affected Regions (van Wyk *et al.* 2019)