Historic transposon mobilisation waves create distinct pools of adaptive variants in a major crop pathogen

Tobias Baril[1,*], Guido Puccetti[1], Daniel Croll[1,*]

[1] Laboratory of Evolutionary Genetics, Institute of Biology, University of Neuchâtel, Rue Emile-Argand 11, 2000 Neuchâtel, Switzerland

[*]Correspondence: tobias.baril@unine.ch, daniel.croll@unine.ch

ORCIDs: TB (0000-0002-5936-7531), GP (0000-0002-7747-4306), DC(0000-0002-2072-380X).

Data Availability

The data underlying this article are available in the article and in its online supplementary material on Zenodo (DOI: 10.5281/zenodo.15125975). The manually-curated TE library is available from Zenodo (DOI: 10.5281/zenodo.8379980), and has been submitted to the Dfam consortium (Dfam Release 3.9).

Conflict of Interest

The authors declare no conflicts of interest.

Abstract

Transposable elements (TEs) can drive the evolution of host-pathogen interactions and gains in antimicrobial resistance. However, how adaptive TEs arise in populations and historical contingencies affect TE dynamics remains unknown. Fungal pathogens in agriculture provide unique frameworks to address such questions due to the availability of spatially explicit sampling and well-characterized niche conditions. We characterised TE evolutionary dynamics using an extensive intraspecies sampling of 1,953 genomes across the global distribution range of the major fungal wheat pathogen *Zymoseptoria tritici*. Employing a pangenomic approach, we characterise genomic diversity and benchmark methods to robustly infer TE insertion polymorphism, before systematically assessing TEs as a source of adaptive variation. We annotated ~3.2 million TE loci among genomes, finding substantial variation in TE content within and among populations. TE activity surged during the pathogen's expansion from its centre of origin in the Middle East, with unique TE activity profiles arising in derived populations. TE-mediated adaptation emerged from distinct waves of TE mobilization. The highest rates of TE activity were observed over timescales as short as 25 years. 45 TE loci showing local adaptation signatures within 1kb of 49 host genes were identified, with adaptive TE insertions likely related to adaptation to antifungals and the plant host environment. This work highlights the power of vast genomic datasets to unravel intraspecies TE invasion histories and pinpoint factors likely driving recent adaptation. This argues for a shift in focus to incorporate deep population-level TE activity surveys in our pursuit to uncover the molecular drivers of adaptive evolution.

Introduction

Exposure to new environments can shift fitness optima and challenge naïve species to adapt. The ability to rapidly respond to novel stressors can provide a significant evolutionary advantage. Genomic variation arising through mutation can modify an organism's ability to survive and reproduce in changing environments, and much focus is given to the role of single-nucleotide polymorphisms (SNPs) in the emergence of adaptive variation, whilst transposable elements (TEs) are often ignored [1]. However, there is growing evidence highlighting the importance of TE-derived genomic variation underpinning traits such as xenobiotic resistance[2–4], stress responsiveness[5] and environmental adaptation[6]. Despite this, our understanding of the genome-wide role of TEs and the extent to which TEs are a general source of adaptive genomic variation remains largely unexplored.

TEs are DNA sequences capable of moving from one genomic location to another[7], and can represent significant portions of the genomes of eukaryotes, comprising over two-thirds of the human genome[8] and over 90% of the maize genome[9]. However, other lineages have been more successful in suppressing deleterious TE activity, with TEs contributing only 0.13% of total genome size in *Trichoplax* (currently the lowest content among Animalia)[10] and rarely over 5% of total genome size in yeasts[11,12]. Autonomous TEs encode domains for their transposition activity, rendering them candidate sources of genomic variation for rapid host adaptation by moving '*plug-and-play*' units around host genomes. TEs are also known to be activated under stress[13–15], coupling exposure to novel environments with transposition activity to give rise to adaptive variation.

The effects of transposition events vary depending on TE insertion sites within host genomes. Through their duplication and mobilisation, TEs generate diverse mutations including the generation and modification of gene regulatory networks, chromosomal rearrangements, exon shuffling, and donation of coding sequence through domestication and co-option[16–21]. TEs have facilitated rapid adaptive evolution through single transposition events across the tree of life. For example, TE insertions have led to the evolution of V(D)J recombination for adaptive immunity in jawed vertebrates[22], the transition from a commensal to pathogenic

lifestyle in *Escherichia coli*[23], cold-stress adaptation in blood oranges[24], xenobiotic resistance in insects[2,3,25,26] and fungi[4,27,28], industrial melanism in the peppered moth[29], and effector genes essential for host infection in oomycetes and fungi[15,27]. TE activity is also likely a significant source of genomic variation at the population level, with intra-species TE activity associated with adaptation to cold[6] and urban environments[30], as well as modifying inter-individual immune responses[31]. Beneficial TE insertions are expected to experience strong positive selection and rapid fixation in populations, whilst the majority of TE insertions are expected to range from nearly neutral to deleterious. TEs exerting highly deleterious effects are efficiently removed via purifying selection[32], whilst the efficiency of selection is reduced against TEs with less deleterious effects, such as non-autonomous and fragmented TE insertions[33,34].

Bursts of TE activity in response to environmental change are hypothesised to be important drivers of micro- and macro-evolutionary events, whilst generating and maintaining genomic diversity[35,36]. Recent work highlights the power of TEs to invade and subsequently propagate in host genomes to generate significant genomic diversity through repeated waves of TE activity over timescales as short as 50 years[36,37]. Crucially, evidence mechanistically linking such TE waves to adaptive evolution is currently lacking, and previous studies have been unable to systematically assess this on a global scale due to challenges associated with incomplete sampling and analytical limitations. Nonetheless, it is hypothesised that new alleles generated by waves of TE activity may become adaptive in response to environmental stress. Identifying adaptive TEs remains challenging, in part due to a lack of available large-scale population datasets to systematically characterise TE-host evolutionary dynamics. Our understanding of TE-host dynamics has also been limited by a focus on a few model systems. However, genome dynamics following TE invasions vary significantly among eukaryotic lineages. For example, transition-transversion bias was generalised across metazoans, but counter examples exist[38]. Further, population-level studies often identify TEs with adaptive signatures but fail to link these signatures to the underlying historic context of TE-host dynamics. This leads to questions concerning the mode and tempo of TE-mediated adaption. Whilst TEs are regarded as sources of variation for rapid adaptation, it remains challenging to date

TE activity and characterise the timescales during which TEs provide an adaptive benefit.

Here, we focus on one of the best characterized fungal species in terms of TE content in the genome. The ascomycete *Zymoseptoria tritici* is a major globally-distributed fungal pathogen of wheat causing significant yield losses[39]. The centre of origin is located in the Middle East, where its emergence accompanied the domestication of wheat[40]. *Z. tritici* has shown rapid adaptive responses across all major wheat-producing areas to strong selective pressures, including host resistance and fungicide use[41]. Extensive standing genetic variation is observed within and among distinct global populations of *Z. tritici*[42–46], which provides the capacity to respond rapidly to novel stressors[5]. Specific examples of TE-mediated adaptation exist, including multidrug resistance[4], epigenetic regulation of key genes involved in host plant infection[15], and melanin production as a response to stressful environments[5]. In parallel to the global expansion of the pathogen, TE content of the genome has likely expanded as a consequence of less efficient genomic defences[42,45]. The well-documented co-evolutionary history between *Z. tritici* and domesticated wheat presents a unique opportunity to explore TE-host evolutionary dynamics within defined timescales.

Here, we aim to determine the tempo and adaptive roles of intraspecies TE invasions using one of the largest intraspecies genome sequencing datasets for eukaryotes. We manually curated 331 distinct TE families and subsequently identified TE insertion and absence polymorphisms in 1,953 genomes sampled across the global range of *Z. tritici*[42,46]. Using this extensive global panel, we characterise population-specific waves of TE activity throughout the pathogen's evolutionary history. We assess TE locus frequencies among populations and identify candidate TEs with signatures of local adaptation integrating data on fungicide resistance and bioclimatic variables. Overall, we find repeated instances of TEs exhibiting signatures of positive selection in a population-specific manner and associated with distinct TE activation waves in the species history.

Results

Pangenome-Informed TE discovery and benchmarking

To maximise capture of intraspecies TE diversity in our analysis, we adopted a pangenomic approach to TE curation. Earl Grey[47] was used to *de novo* annotate the reference-genome quality pangenome comprising of 19 genomes representing the global diversity of *Z. tritici*, and four closely related sister species (*Z. pseudotritici, Z. passerinii, Z. ardabiliae,* and *Z. brevis*). Following manual curation, the library contains 331 TE consensus sequences for 92 DNA transposons, 22 long-interspersed nuclear elements (LINEs), 65 long terminal repeat retrotransposons (LTRs), 31 miniature inverted terminal repeat elements (MITEs), 11 rolling circles (also known as *helitrons*), 1 short interspersed nuclear element (SINE), 1 terminal-repeat retrotransposon in miniature (TRIM), and 105 unclassified elements[48].

Detecting TE insertions in short-read whole genome sequencing data is inherently challenging due to the difficulty of assembling repetitive DNA sequences from short reads. To tackle this problem, we performed extensive benchmarking based on 11 methodologies implemented in the McClintock2 meta-pipeline[49]. Performance was highly variable, consistent with previous findings[49]. Across all benchmarks, the highest-performing methodologies were TEFLoN[50] and PoPoolationTE[51] for reference and non-reference TE insertions, and TEMP2[52] for non-reference insertions. We adopted an approach that combines the best-performing reference and non-reference TE detection tools to generate a comprehensive non-redundant TE annotation set for each isolate. We use TEFLoN as the base annotation set, before supplementing this with unique TE annotations from PoPoolationTE and TEMP2, using the latter specifically for non-reference TE detection. We repeated the above benchmarking analyses with this improved methodology and could confirm an increase in TE detection performance (Supplementary File 1). By combining TEFLoN, PoPoolationTE, and TEMP2, we maximise TE detection for each isolate whilst minimising false positive rates. A consequence of our approach is that some TEs might be missed, particularly non-reference TEs, as well as TEs in short-read datasets of lower sequencing coverage depth (Discussed in Supplementary File 1).

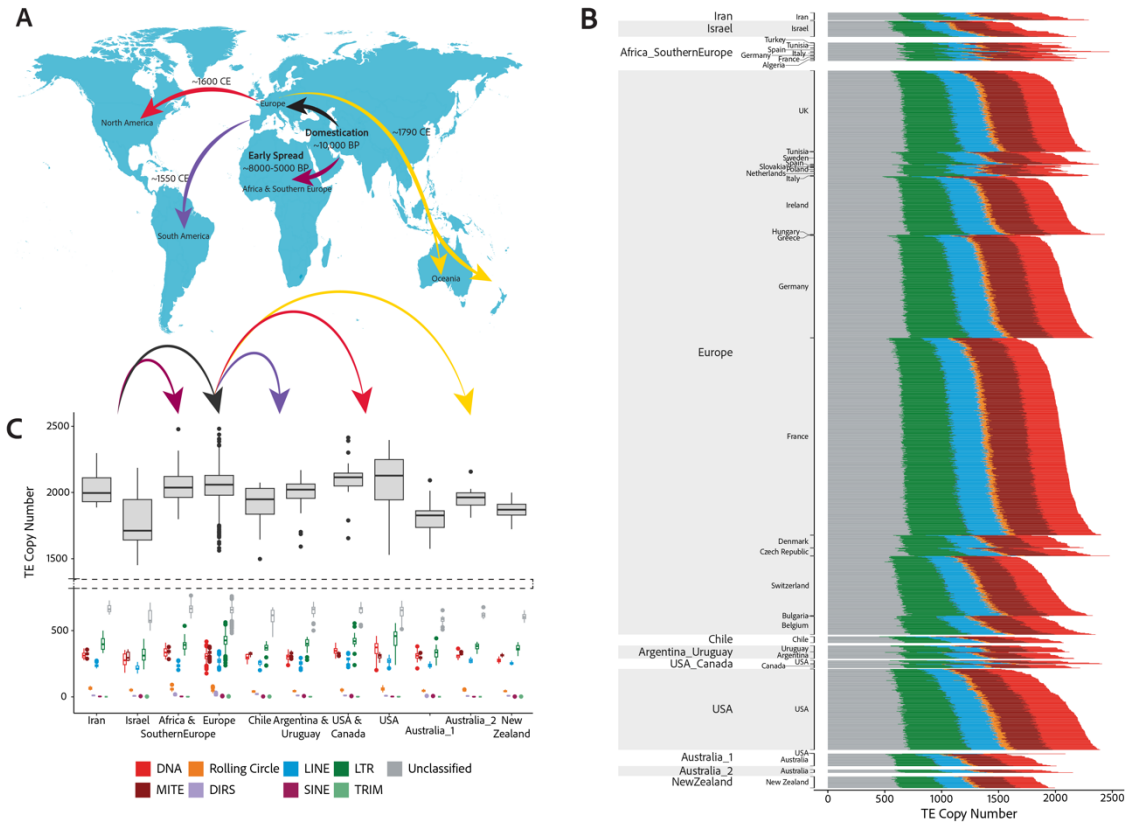
A Highly Dynamic Transposable Element Landscape Following Global Expansion

We analysed 1,953 isolates of *Z. tritici* sampled across the globe. Isolates were assigned to 12 genetic clusters based on genome-wide polymorphism, consistent with previous analyses[42,46] (Figure 1A). Isolates that could not be clearly assigned to a single cluster (n = 384) were not included in analyses requiring explicit population assignment retaining a total of 1,569 isolates. From most ancient to most recent in terms of colonisation history, populations are named based on the origin of member isolates: *i.e.* Iran, Israel, Africa & Southern Europe, Europe, USA & Canada, USA, Chile, Argentina & Uruguay, Australia_1, Australia_2, and New Zealand.

Among all isolates, 3,201,910 TE loci were detected (μ = 2,041; min = 1,451, isolate ST92ISR_Ar_16h from Israel; max = 2,481, isolate 08STCZ016 from Europe; Supplementary File 2). There are significant differences in TE abundance among different populations (One-way analysis of variance: $F_{10,1558} = 28.25$, $p < 0.01$), illustrating the dynamic nature of TE landscapes among populations despite the separation by only short evolutionary timescales, with the Iranian and Israeli populations arising ~10,000 years ago, and the most recent Oceanian populations (Australia_1, Australia_2, and New Zealand) arising ~1790 AD. TE copy number has increased significantly as populations spread beyond the Middle East and colonised Africa, Europe, and the Americas, suggesting an increase in TE activity associated with colonisation of new environments (Figure 1C; Supplementary Table S1). However, populations in Australia and New Zealand are not following this pattern, and show a significant reduction in TE abundance compared to the Iran, Africa & Southern Europe, Europe, USA & Canada, USA, and Argentina & Uruguay clusters (Supplementary Table S1). This is consistent with increased TE activity as *Z. tritici* spread out of the Middle East ~10,000 years ago, into Africa (~8,000-5,000 years ago), Europe, and the Americas (~1550-1600 AD), followed by a reduction in TE activity, and TE degradation or removal as the pathogen colonised Oceania (~1790 AD) (Figure 1).

205 In addition to significant variation among populations, inter-individual differences
206 in TE abundance are also considerable in scale, with standard deviations in TE copy
207 number ranging from 59 (New Zealand) to 203 (Israel) (Figure 1C). Consistent with
208 the reduced TE abundance observed in Oceanian populations, we also observe a
209 decrease in TE abundance variation among Oceanian isolates. This decrease could
210 be attributed to the reduced level of genetic diversity observed in these
211 populations as a result of the colonization bottleneck. However, the TE decrease
212 observed in Oceania may also be explained by a dampening of TE activity following
213 counterselection against deleterious insertions.

214



215

216 **Figure 1. Global dynamics of TE variation among populations of *Z. tritici*. A.** A
217 schematic illustrating the global spread of wheat across the globe. **B.** TE copy
218 number for each isolate split by high-level TE classifications, as indicated in the key.
219 Isolates are grouped by country of sampling and genetic cluster membership,
220 ordered from most ancient to most recent using a colonisation timeline based on
221 global wheat domestication. **C.** Population-wide mean TE copy numbers. Top

222 boxes represent total TE copy numbers. Bottom boxes show TE copy numbers split
223 by high-level TE classifications, as indicated in the key.

224

225 In all populations, LTR retrotransposons are the most abundant TEs (global μ = 384,
226 min = 319 (Israel), max = 438 (USA)). Most populations share a similar profile in terms
227 of TE abundance, characterised by high numbers of LTR retrotransposons (global
228 μ = 384), followed by MITEs (global μ = 321), DNA TEs (global μ = 313), LINEs (global
229 μ = 254), Rolling Circles (global μ = 50), DIRS elements (global μ = 10), SINEs (global
230 μ = 2), and TRIMs (global μ = 1). However, in North America, DNA TEs have expanded
231 to overtake MITEs as the second-most abundant TE classification (USA: DNA μ =
232 363, MITE μ = 306; USA & Canada: DNA μ = 347, MITE μ = 328; Australia_1: DNA μ =
233 307, MITE μ = 296) (Figure 1, Supplementary Table S2).

234 Beyond changes in TE activity during global expansion, we also observe large intra-
235 population variation in TE copy number. High levels of intra-population variation
236 are hallmarks of rapid evolution on very short timescales. Within the USA
237 population, we find individuals falling into two groups characterised by TE
238 abundances either below or above 2,000 copies. These subpopulations highlight
239 rapid increases in TE abundance within the USA population, where 96% of the
240 isolates sampled in the 1990's contain <2,000 TEs and 83% of the isolates sampled
241 in the 2010's contain >2,000 TEs. When comparing isolates sampled in the 1990's to
242 those sampled in the 2010's, the mean TE abundance has increased from 1,840 to
243 2,227 TEs per isolate, driven mostly by LTR retrotransposon activity, consistent with
244 previous findings (Oggenfuss et al. 2021), and DNA TE activity (LTR
245 retrotransposons: $\mu_{1990's}$ = 372, $\mu_{2010's}$ = 485; DNA TEs: $\mu_{1990's}$ = 315, $\mu_{2010's}$ = 398)(Figure
246 1B).

247 Given that TE activity is overall detrimental to host genome integrity, most TE
248 families are expected to be repressed and maintained at low copy numbers.
249 Consistent with this, 261 (79%) TE families in *Z. tritici* have a maximum population
250 mean copy number < 10. Conversely, we also observe large increases in the
251 abundance of specific TE families as *Z. tritici* expanded from the centre of origin in
252 the Middle East. 18 TE families show mean TE copy number variation of > 10 among
253 populations (Figure 2B). The spread of *Z. tritici* into North America ~1550 AD has led

to increases in TE activity, where LTR retrotransposons and rolling circle TEs increased in copy number to their highest levels in North American populations. More recently, DNA TEs have expanded in the Oceanian populations representing the last colonized continent. DIRS elements are found at very low numbers across all populations except for the Chilean population where these have expanded 3-fold in copy number (Figure 2B). In contrast, LINEs show a stepwise increase tracking the colonisation timeline of populations, suggesting low levels of ongoing activity rather than TE burst dynamics as observed for other TEs.

Beyond population-specific differences in total TE abundance, we observe substantial levels of population-specific TE activity. There are significant differences among genomes in their affinity for highly abundant TE families (*i.e.*, defined as having ≥ 20 copies in at least 1 isolate; Pearson's $\chi^2_{118889} = 132{,}250$; $p < 0.01$; Figure 2A). Clustering genomes based on their affinity for highly abundant TE families largely recapitulates the species' population history, which is indicative of population-specific TE-host dynamics as an important source of genomic differentiation (Figure 2A). The most abundant TE family in the centre of origin populations is a LINE (*family_1507*; Iran μ = 64; Israel μ = 61). This family largely ceased activity during the global expansion of *Z. tritici*, persisting in populations at a mean abundance of between 61 (Israel and USA) and 66 (Africa & Southern Europe). In contrast, the *Copia* LTR retrotransposon, *family_71*, shows significant activity in the USA, Europe, Australia_1, and Australia_2 populations (USA μ = 81, Australia_1 μ = 65, Europe μ = 61, Australia_2 μ = 67) (Figure 2B, Figure 3B). This *Copia* LTR retrotransposon is the most active TE in *Z. tritici* but reached peak activity only after persisting at low copy numbers in early-colonising populations (Iran μ = 17; Israel μ = 10, Africa & Southern Europe μ = 7; Figure 2B, Figure 3B). The pattern observed for this *Copia* element is consistent with the TE burst model, whereby the exposure to abiotic stress in a novel environment during range expansion results in massive bursts of specific TE families[35]. Such large increases in TE copy number have the potential to severely impact host genome integrity, as demonstrated for the *Styx* element (*family_11*)[20], even whilst proliferating to a lesser extent than *family_71*. *Styx* is absent in the Iranian population and found at its highest mean copy number in Australia, expanding throughout the global expansion (Supplementary Table S3).

Figure 2. TE activity gains and losses during global pathogen expansion. **A.** Affinity of individual isolates for the most abundant TE families (≥ 20 copies in at least one isolate) using Pearson's χ^2 residuals. Negative affinities (i.e. repulsions) are illustrated with blue shades, whilst positive affinities (i.e. attractions) are illustrated with red shades, as indicated in the key. TE family classifications are illustrated on the top edge, along with TE family numbers corresponding to the species TE library[48]. Isolate population membership is illustrated on the left edge. Boxes highlight patterns of TE expansion or contraction as labelled. **B.** TE burst profiles showing mean population-wide TE abundance for families with an abundance variation ≥ 10 among populations. Populations are ordered from most ancient to most recent based on the global spread of wheat. F*amily_71*, which shows a large expansion in the USA population, is highlighted by a box in panels A and B.

TE contributions to a highly dynamic genome

The genome of *Z. tritici* is shaped by ongoing TE activity contributing to high levels of genomic diversity. To quantify the exact contributions to genetic variability, we assessed polymorphism created by the presence or absence of TEs across loci. 1,477

insertion sites are found in at least one genome in every population, constituting a pool of persistent TE variants, whilst 214,312 sites are unique to genomes from a single population (Figure 3A). Population-specific waves of TE activity produced 28,740 population-specific polymorphic TE loci arising in populations established in the last ~500 years with the USA contributing 21,120 loci alone (Supplementary Table S3).

The majority of TE loci show high TE insertion frequencies within populations (defined as the TE being at frequencies ranging 10-95%, also termed “common” insertions). The mean number of loci at such high TE frequencies is 1,025 (49.85% of total TE loci; Europe) to 1,370 loci (65.40% of total TE loci; USA). TE mean frequencies at the loci range between 54.64% (Iran) and 68.62% (Australia_1) (Supplementary Table S4). The number of loci carrying rare TEs per isolate is much lower and ranges between 100 (4.95% of total TE loci; Iran) and 407 (19.81% of total TE loci; Europe). Loci with fixed TEs were ranging between 232 (13.15% of total TE loci; Israel) and 893 (45.63% of total TE loci; Australia_2). Large variation around the mean TE frequencies may underpin evolutionary potential for the species. TE activity has occurred over short evolutionary timescales, with the majority of the TE activation waves occurring in the last ~500 years. The recency of TE activation waves raises the question as to why such activity is tolerated by the genome and whether such activity could have produced adaptive benefits.

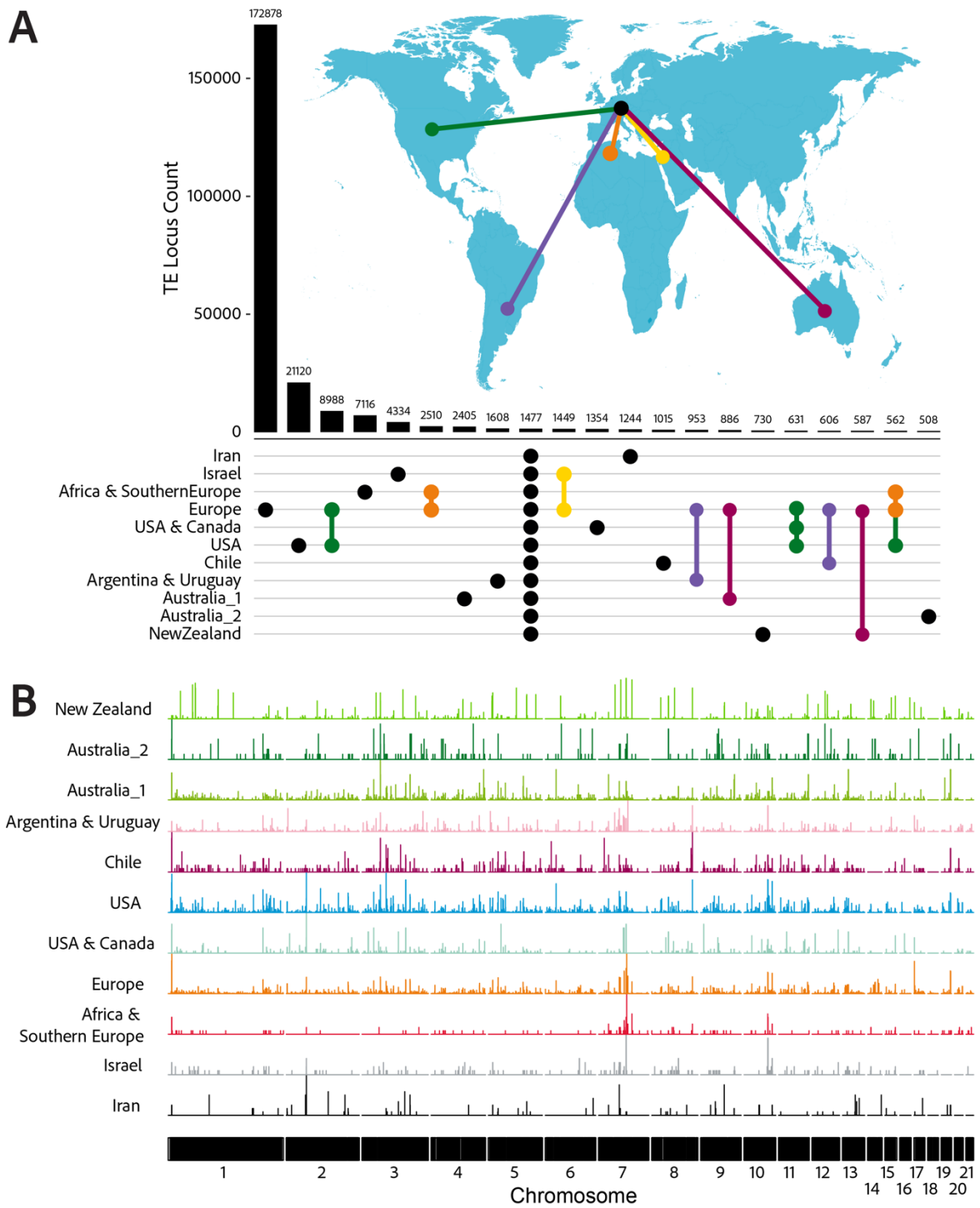


Figure 3. TE activity leads to population differentiation across the globe. A. TE loci identified in each population of *Z. tritici,* grouped by populations containing each TE insertion. Bars are in descending order by locus count. Dots indicate populations where at least one isolate contains the TE locus. Black lines connect populations that share TE loci under their respective quantification bars. Coloured connections indicate shared loci among populations indicated on the map. **B.** Expansion of the highly active LTR, *family_71* and occupancy across genomic loci. Peaks indicate TE insertion frequency across chromosomes. Populations are ordered following global pathogen spread.

Candidate Adaptive TEs Facilitating the Global Pathogen Spread

TEs rising to high frequencies in specific populations are prime candidates underpinning recent adaptation, driven by positive selection in extant populations due to a positive effect on host fitness[53]. We identified the strongest candidate TEs with signatures of local adaptation by detecting TE loci with outlier differentiation among populations using BayPass[54]. We retained 45 consistently identified TE loci for further inspection (Figure 4). Nine loci were discarded as the genomic context could not be confidently assessed. The 36 remaining adaptive TE loci were proximal to 49 host genes within a 1,000bp window. We identified conserved functions for 20 genes[55,56] (Figure 4; Supplementary Table S5; Supplementary Table S6). The most abundant GO terms included Zinc ion binding (GO:0008270; n=5); ATP binding (GO:0005524; n=4); Mitochondrion (GO:0005739; n=3); Membrane (GO:0016020; n=3) and ATP hydrolysis activity (GO:0016997; n=3). Adaptive TE loci are found in the promoter and 5' UTR regions of 30 distinct host genes, where insertions have the potential to modify host gene expression (Figure 4; Supplementary Table S6). Six host genes contain adaptive TEs in their coding sequence, whilst four genes contain candidate TEs in their 3' UTR and 12 genes carry TEs in their downstream region (Supplementary Table S6).

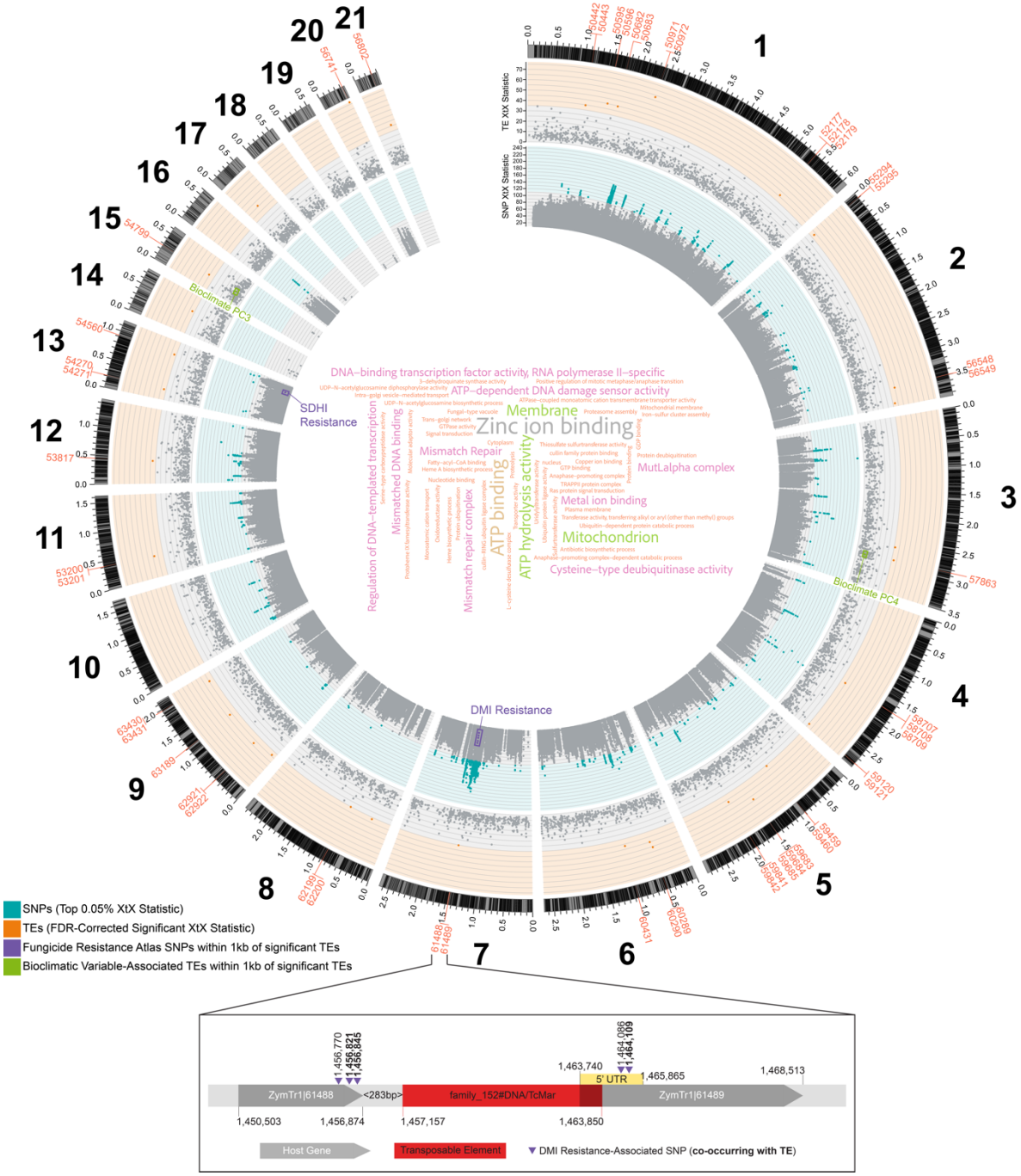


Figure 4. Adaptive TE loci detected by outlier search across the global pathogen distribution range. TE and SNP annotations mapping to chromosomes of the *Z. tritici* reference genome IPO323. Local adaptation candidates with significant XtX values are found in their respective shaded regions, with SNPs on the inside track, and TEs on the outside track. SNPs implicated in fungicide resistance[46] and TEs associated with bioclimatic variables within 1,000bp of candidate TEs are labelled, as indicated in the key. Black bars on the outermost track show IPO323 gene annotations, with grey illustrating intergenic regions. Numbers indicate genomic coordinates in megabases (Mb). Genes within 1,000bp of candidate TEs are labelled outside the gene track. The wordcloud in the centre of the circos plot illustrates the GO terms assigned to genes nearby candidate TEs, with size proportional to the number of genes annotated with each GO term. Inset gene region schematic illustrates a single candidate TE and the host genome landscape surrounding

362 this insertion. Host genes and fungicide-resistance SNPs are labelled, as indicated in the
363 key. SNPs in bold are found to colocalise with the candidate TE insertion in the Argentina
364 and Uruguay population.

365

366 To assess the adaptive role of TEs in the pathogen's historic expansion, we assigned
367 adaptive TE loci to individual historic activation waves. Furthermore, we integrated
368 know-how from a large antifungal resistance mapping study on *Z. tritici*, as well as
369 climatic data. Two adaptive TEs are found within 1,000bp of 6 SNPs associated with
370 fungicide resistance (Figure 4). The LTR retrotransposon (*family_1437*) on
371 chromosome 13 is at common frequencies in Africa & Southern Europe, Europe,
372 and New Zealand, and co-localises with a SNP associated with resistance to the
373 succinate dehydrogenase inhibitor (SDHI) pydiflumetofen. This TE is also located
374 downstream of the gene ZymTr1|54560 of unknown function. F*amily_1437* showed
375 peak activity in the Iran and Israel populations, showing that the TE was inserted
376 well before the application of fungicides in agriculture (Figure 5). This supports the
377 role of TEs as a source of standing genetic variation that can provide a host benefit
378 much later in the species' evolution when novel challenges arise.

379 The DNA element (*family_152*) is found on chromosome 7 at common frequency in
380 Argentina & Uruguay, and co-localises with 5 SNPs associated with resistance to
381 the demethylase inhibitor (DMI) mefentrifluconazole. This TE is located
382 downstream of gene ZymTr1|61488 (Proteasome assembly, GO:0043248) and
383 upstream of gene ZymTr1|61489 (DNA-binding transcription factor activity
384 GO:0000981 among other encoded functions). In contrast to *family_1437*,
385 *family_152* shows a peak activity wave in the most recent Oceanian populations,
386 and a smaller more ancient activity wave in the North American populations
387 (Figure 2B, Figure 5). This is consistent with a narrow time span between the
388 emergence of the adaptive TE and strong selection pressures by DMIs potentially
389 favouring the TE (Figure 4).

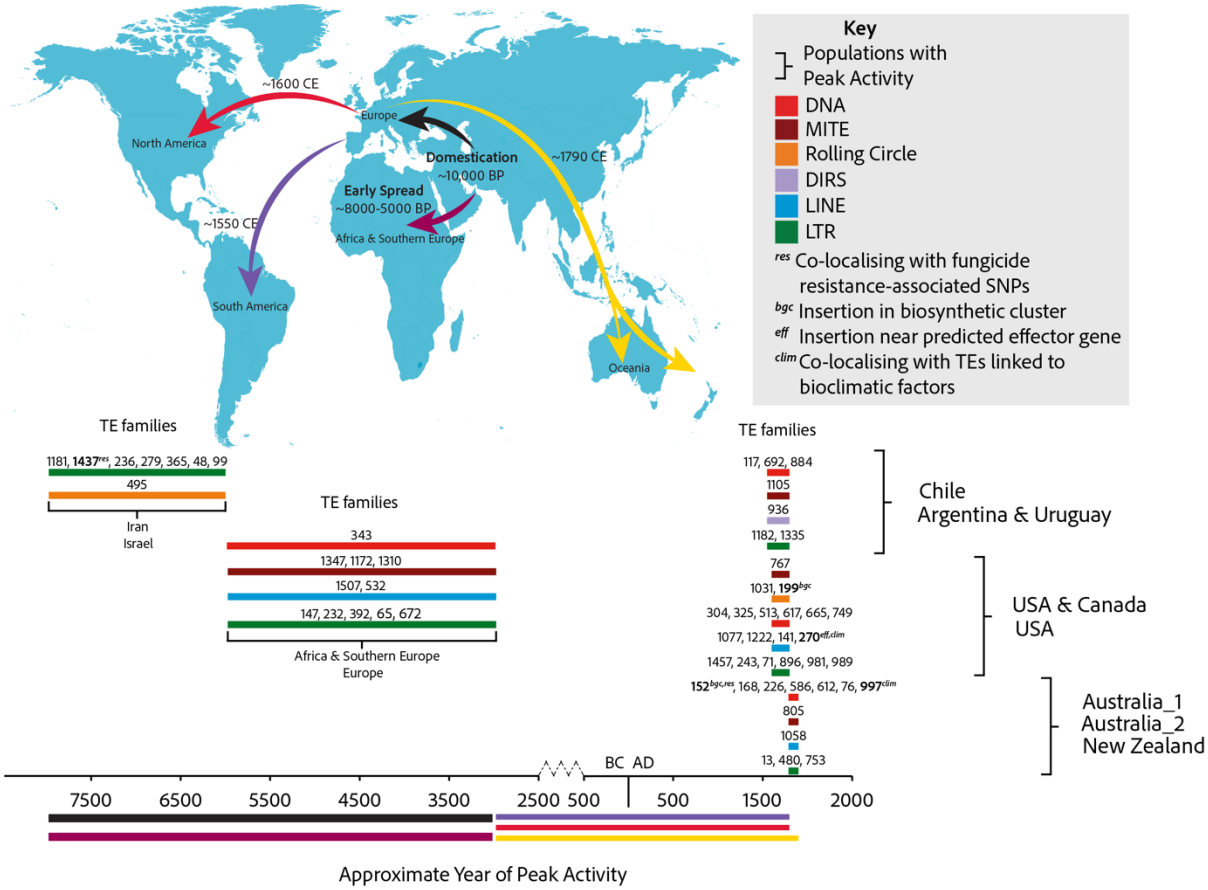


Figure 5. Adaptive TEs emerging through historic TE activation waves. The map shows the global colonisation timeline of *Z. tritici* following wheat domestication. Bars width reflects the rough window of colonisation for the populations with peak activity as shown in the key. Colours under the timeline correspond to those shown in the map arrows. TE families are shown by their family number corresponding to their names in the curated library. TE family peak activity is estimated based on the populations showing the highest mean copy number for TE families with a population mean TE copy number range > 5. Candidate TE families associated with host functions are shown in bold, with host functions and TE classifications indicated in the key.

Next, we determined whether the resistance-associated SNP allele correlated with the inserted TE. Only 16 isolates, all from the European population, have the resistance allele co-occurring with the presence of *family_1437* on chromosome 13 (Figure 6A), whilst 146 isolates possess the resistance SNP allele without the TE, and 168 isolates possess the TE with the susceptible SNP allele. The majority of isolates, 1,238, contain the susceptible SNP allele and lack the TE insertion. A more nuanced pattern is observed at the locus on chromosome 7, where the TE insertion is only present in isolates from Argentina & Uruguay, and all isolates with this TE (TE locus frequency = 50%, present in 14 isolates) have the resistance allele at the same three

410 of five loci associated with resistance to mefentrifluconazole (positions: 1,456,821;
411 1,456,845; 1,464,109). 3 isolates (11%) possess resistance alleles at these 3 positions in
412 the absence of the LTR insertion, whilst the remaining 11 isolates (39%) in Argentina
413 & Uruguay do not contain the LTR insertion or any of the resistance-associated
414 SNPs (Figure 6). The common frequency of this LTR insertion and a co-occurrence
415 with 3 SNPs associated with fungicide resistance at this locus provides evidence to
416 support a putative association between the candidate TE and local adaptation
417 through modifying a fungicide resistance phenotype in the Argentina & Uruguay
418 population, however functional characterisation will be required to confirm a
419 mechanistic role for the TE insertion.

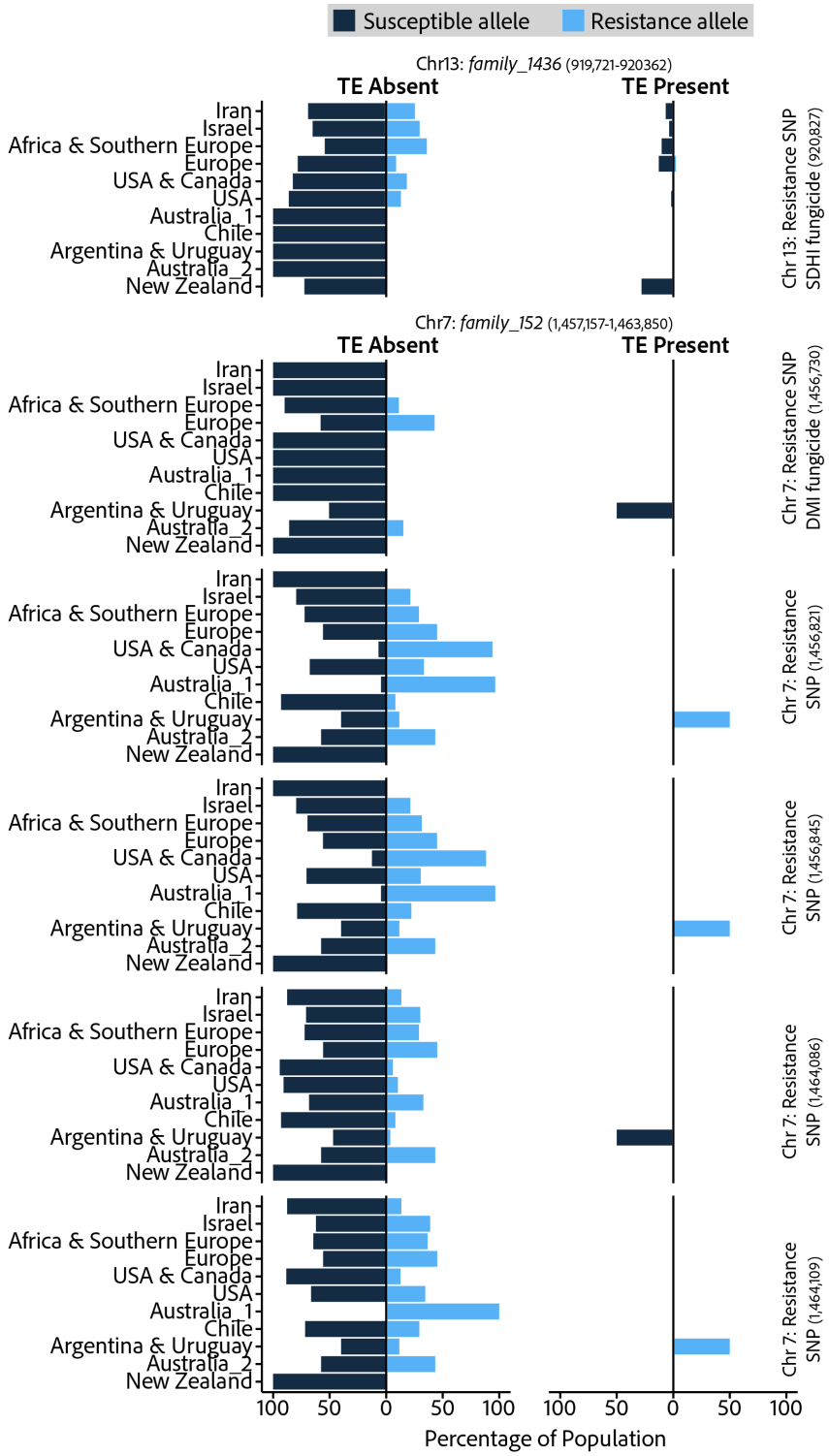


Figure 6. Co-occurrence of fungicide resistance-associated loci and proximal adaptive TE insertions. Populations are ordered using the colonisation timeline of wheat (Figure 1). Each row is a distinct fungicide resistance-associated loci (SNPs), with bars indicating the presence of susceptible or resistance alleles, as indicated in the key. Left-hand plots show the proportion of isolates within a single population with susceptible or resistance SNPs in the absence of the proximal TE. Right-hand plots show the proportion of isolates within a single population with susceptible or resistance SNPs in the presence of the proximal TE.

448 Throughout the global expansion of *Z. tritici*, populations have been exposed to
449 heterogeneous selection driven by climate including changes in temperature,
450 precipitation, and seasonality. Recently, 21 copy number variants (CNVs) were
451 found to be significantly associated with climatic factors at 14 independent loci[57].
452 We performed genome-wide environmental association mapping on TE loci and
453 10 principle components representing variation among 76 bioclimatic variables.
454 We identified 2 TE loci significantly associated with variations in climate
455 (Supplementary Table S6) including the same LINE located 1,516bp upstream of a
456 putative effector on chromosome 3 (*family_270*) is also located 873bp downstream
457 of a *PIF-Harbinger* DNA TE (*family_997*) that is significantly associated with
458 bioclimate PC4, with the top positive loadings associated with precipitation in the
459 warmest and driest quarters, precipitation in the driest month, annual
460 precipitation, and the minimum monthly climate moisture index (Supplementary
461 Table S6; Supplementary Table S7). On chromosome 15, an LTR element,
462 *family_492*, is nested at a locus containing an unclassified TE (*family_1037*), which
463 is significantly associated with bioclimate PC3, with the top positive loadings
464 associated with precipitation in the wettest and coldest quarters, maximum
465 monthly climate moisture index, annual precipitation, and net primary
466 productivity (Supplementary Table S6; Supplementary Table S7). However, in both
467 cases, there is no co-occurrence of bioclimate-associated TEs with local adaptation
468 TE candidates.

469 The closest polymorphic TE locus to an effector (Zt_3_904) is a LINE (*family_270*),
470 found 1,516bp upstream of the transcription start site at common frequency in the
471 Chile population (Supplementary Table S6). Zt_3_904 is a putative suppressor of
472 flg22-induced reactive oxygen species (ROS) burst[58], with potential importance in
473 successful host colonisation to overcome plant defences and maintain a redox
474 environment conducive to cell viability[59]. 3 candidate TEs are associated with host
475 genes within 3 separate biosynthetic gene clusters (BGCs) (Supplementary Table
476 S6). A rolling circle element on chromosome 1 is associated with genes involved in
477 terpene synthase synthesis, whilst a DNA TE insertion on chromosome 11 is
478 associated with an NRPS-like BGC, and a MITE on chromosome 12 is associated
479 with a Non-alpha poly-amino acids (NAPAA) BGC. In all three cases the genes have
480 likely auxiliary BGC functions rather than being the core gene.

Discussion

Despite growing evidence for TE polymorphism facilitating host adaptation, the TE dynamics leading up to the emergence of adaptive variation remain poorly understood. The extensive genomic resources for *Z. tritici*, combined with knowledge on the pathogen colonisation timeline over the last ~10,000 years, provided a unique opportunity to explore population-level TE-host evolutionary dynamics in the largest global sampling dataset to date. This framework enabled us to characterise population-specific waves of TE activity associated with the pathogen's global spread and facilitated the systematic characterisation of the dynamics underlying the emergence of TE-mediated adaptive variation. Curated TE databases are largely deficient in fungal sequences. As such, little is known about the diversity of TEs across the fungal kingdom despite the clinical and agricultural relevance of fungal pathogens and their TE-mediated adaptation.

We characterised TE activity across the colonisation timeline for *Z. tritici*, and found extensive population-specific activity, with TE derepression coinciding with expansion into novel environments. Importantly, different colonisation stages are characterised by the proliferation of distinct TE families, and at different intensities, demonstrating the highly localised nature of TE activity among distinct populations. For example, the early spread of *Z. tritici* into Africa and Europe led to a wave of peak activity in 10 distinct TE families, whilst expansion into North America led to a wave of peak activity in 19 different TE families, including an intense increase in activity of the LTR retrotransposon *family_71*, which quadrupled in copy number during this colonisation stage. Following increased TE family activity, a species may see the emergence of host benefits mediated by beneficial TE insertions. However, these two stages of TE impacts on the host genome remain poorly connected in nearly all systems studied to date. Among TE insertions with likely beneficial effects, we found the corresponding peak TE family activity to be predominantly in the last 500 years (8/11 TE families). This timeframe matches the pathogen's expansion into North and South America, and Oceania. TE insertions

likely impacted host functions such as fungicide resistance, BGCs, adaptation to the climate and effector genes suggesting a short time lag between the TE activity burst and adaptive significance. Conversely, the peak activity of the LTR retrotransposon, *family_1427*, dates back to the origin of the species in the centre of origin but the association with fungicide resistance has only been created in recent decades given the recency of such treatments in agriculture. This suggests that even older TE activation episodes can act as reservoirs for genomic variation of current adaptation processes.

Combining our findings, we propose that the expansion of *Z. tritici* into novel environments has resulted in TE family derepression in an environment-dependent manner. This localised TE activity, in combination with limited gene flow and the emergence of bottlenecks in founder populations, has resulted in significant levels of genomic differentiation among *Z. tritici* populations at the TE level. This variation can then provide a host benefit through (i) contributing to extant levels of standing genetic variation facilitating adaptation in derived populations, or (ii) provide 'pre-adaptation' genetic variation to facilitate rapid adaptation thousands of generations later. Supporting this, we observe massive increases in TE activity over timescales as short as 25 years. To our knowledge, increases in TE copy number of this magnitude (mean TE copy number increase = 387) over such a short timescale have not previously been observed in natural populations. Assuming a generation time of one year, in the North American population we observe a mean transposition rate of 15.48 successful new insertions per generation, whilst ~0.0040-0.40 SNPs per generation are expected to arise at a neutral mutation rate of 10^{-8}-10^{-10} substitutions per base pair per generation[60–62]. Therefore, the potential of TEs to generate new variants is magnitudes higher than SNPs. Furthermore, TE insertion variants may encode regulatory or functional domains expanding the repertoire of functional consequences for the genome. Previous work has focused on the role of SNPs in the rapid adaptation of *Z. tritici* [46,63–66], and the extent of TE activity observed argues for a shift in focus to incorporate TEs into future investigations to determine the genetic basis of rapid adaptation in pathogens such as *Z. tritici*.

544 Connecting adaptive TE insertions to the underlying TE activity in the genome
545 reveals important features of TE activation lifecycles. Most TE families in *Z. tritici* are
546 retained at low copy numbers, with a few families dominating the TE landscape.
547 The LTR retrotransposon, *family_71*, shows a unique expansion of a magnitude not
548 observed for any other families. This proliferation is consistent with the
549 derepression of this TE family following the expansion of *Z. tritici* out of the Middle
550 East, providing support for the role of stress in the induction of transposition
551 activity[6,13,15]. Most research on TE dynamics in genomes only considers a subset of
552 actual TE activity, as only TEs that successfully insert into host genomes are
553 identified. Consequently, the host-TE interactions resulting in the success of this
554 LTR family whilst others persist at low levels remain elusive, and we require an
555 improved understanding of TE success rates to characterise the dynamics that
556 result in either TE families being driven to extinction or TE families expanding in
557 host genomes and contributing to host evolution.

558 In a gene-dense genome, the likelihood of a successful TE insertion is reduced and
559 detrimental effects on host genes are more likely. Hence, purifying selection on TEs
560 should be more intense in gene-dense genomes. Furthermore, fixation of TEs in
561 gene regions where they can exert regulatory effects on host genes is expected to
562 provide a benefit to the host and be positively selected. Fixation resulting from
563 inefficient purifying selection or genetic drift should also be less likely given the
564 large effective population sizes of the pathogen[33,34]. We identified 45 TEs that have
565 risen to high population frequencies (≥ 10% and < 95%) with significant signatures
566 of local adaptation among populations. Because TEs can act as "*plug-and-play*"
567 genomic units, they can provide host benefits by facilitating complex local
568 adaptation. Such strong selection pressure to fix adaptive TE insertions might have
569 been conferred by xenobiotic resistance [67], novel environments such as extreme
570 cold and human-impacted landscapes [6,30], as well as host-pathogen interactions
571 [15,68]. Consistent with this, we identified 2 candidate TE insertions associated with
572 SNPs conferring resistance to fungicides, with a co-occurrence between a DNA TE
573 and SNPs conferring resistance to mefentrifluconazole in the Argentina & Uruguay
574 population. TE-mediated xenobiotic resistance mechanisms have been previously
575 reported and functionally characterised in several insect lineages[69–73], and
576 resistance to DMIs has been functionally demonstrated in *Penicillium digitatum*[74],
577 which was conferred by MITE-altered expression of PdCYP51B. In *Z. tritici*,

comprehensive antifungal resistance mapping efforts revealed that resistance loci were overrepresented in repetitive regions of the genome suggesting a broad involvement of TE-driven adaptive variants[46,63]. Given the strong selection pressure imposed on plant pathogen populations by fungicides, and the ability of TEs to exert regulatory effects on nearby host genes, TE insertion variants should be the primary focus for the detection of emerging resistance in the field.

The advances in population-level sequencing of TE variants and connecting such polymorphism to adaptive trait variation raise major questions how much of the total adaptive variation in species gene pools has been generated by TE activity alone. Quantifying these contributions requires characterisation of TE dynamics beyond traditional model systems and species groups. The approach taken here illustrates the widespread and repeated contributions of TEs to the generation of genomic novelty within a single species in the understudied fungal kingdom. We shed light on the power of TEs to generate extensive genomic variation within and among populations over extremely short timescales, measured in decades rather than millions of years. We find that exposure to novel environments, and the establishment of new populations, underpins distinct waves of TE activity generating adaptive genetic variation and geographic differentiation. These waves of TE activity result in large numbers of new insertions even in the suppressive landscape of a gene-dense genome. The dynamism of TE activity may overcome limitations in species gene pools and drive responses to changing environments more quickly than previously expected. Our findings support tight integration of TE variant monitoring into investigations of rapid adaptation, particularly in systems where TEs might be derepressed. The activation of TEs in response to novel environments also sheds light on the potential impact that climate change could have on host genomes, particularly related to the movement and expansion of pathogens relevant for food security and human health. However, routine implementation of long-read sequencing in population studies will be required to fully resolve TE dynamics. Beyond this, future research needs to broaden sampling of TE diversity beyond traditionally studied species groups.

Materials and Methods

Generation of a Manually-Curated TE Library for *Zymoseptoria tritici*

19 reference-quality genome assemblies representing the global diversity of *Z. tritici* and 4 reference-quality genome assemblies for the sister species (*Z. pseudotritici, Z. passerinii, Z. ardabiliae,* and *Z. brevis*)[75–78] were annotated with Earl Grey (v3.0; https://github.com/TobyBaril/EarlGrey)[47] to *de novo* detect and generate putative TE consensus sequences. TE consensus sequences were first clustered using CD-Hit-Est (v4.8.1)[79,80] to reduce redundancy in the initial library by grouping sequences with 90% similarity across at least 80% of the total length of the longer sequence to prevent collapsing of chimeric consensus sequences. Consensus sequences were manually curated following the protocols described by[81]. Genomic copies of each TE with 1,000bp of flanking sequence at each end were obtained from all 23 genome assemblies using a "BLAST, Extract, Align, Trim" process [82]. The 25 longest hits, along with up to 75 randomly selected hits, were obtained for each TE query sequence. Multiple sequence alignments were generated using MAFFT --auto (v7.505)[83] before T-COFFEE (v13.45.0.4846264)[84] was used to remove columns composed of ≥80% gaps. Resultant sequence alignments were manually curated to define boundaries of single-copy DNA sequence (i.e. visible TE boundaries), identify TE subfamilies, and to remove regions of low conservation. New majority-rule consensus sequences were constructed with EMBOSS (v6.6.0) cons[85] after which TE consensi were visually inspected using TE-Aid (https://github.com/clemgoub/TE-Aid/) to identify diagnostic features characteristic of the main TE classifications[86]. The presence of TIRs was recorded, along with their sequence, along with homology to curated TEs in Dfam (v3.7)[87,88], which was attained with nhmmscan (HMMER v3.3.2)[89]. These sources of information were combined to manually classify each TE consensus sequence using the convention 'ZymTri_2023_family_n#classification/family' for compatibility with RepeatMasker. Consensus sequences classified with low confidence are labelled with both '?' and '_LowConf' appended to the name. Following *de novo* TE extension, curation, and new consensus generation, a second redundancy filter for the final TE library was run in CD-Hit-Est to collapse TEs to

640 family-level using the '80-80-80 rule' (≥80% identity over ≥80% sequence length
641 over ≥80bp)[81,86]. CD-Hit-Est alignments were manually inspected to select the
642 representative sequence in each cluster by choosing the most intact sequence
643 with the highest classification confidence level. Chimeric sequences erroneously
644 clustered with their individual counterparts were manually separated to retain
645 sequences for the chimeric TE and their individual elements.

646

647 In total, 331 TE consensus sequences are found in the *Z. tritici* library, of which 199
648 are confidently classified, 27 are classified with low confidence, and 105 are putative
649 TEs labelled as '*Unclassified*'. The manually-curated TE library can be freely
650 accessed at https://doi.org/10.5281/zenodo.8379981[48,90].

651

652 Sourcing of Short-Read Genomes for the global panel

653 We obtained short-read genomes provided from [42] and [46]. Samples cover the global
654 distribution of *Z. tritici*, from the centre of origin in the Middle East to the most
655 recently derived populations in Oceania. Samples were taken over a span of
656 decades, with the earliest samples isolated in 1981, and the latest in 2019. Genome
657 information is available in supplementary table S8.

658

659 SNP Calling

660 We generated a pangenome graph from the reference-quality *Z. tritici* genome
661 assemblies using REVEAL (REcursive Exact-matching ALigner)
662 (https://github.com/jasperlinthorst/reveal), starting with the IPO323 assembly and
663 iteratively aligning subsequent *Z. tritici* genome assemblies. The finished draft
664 genome was made to conform to the layout of the initial backbone sequence of
665 IPO323 (finish –order=chains). Unmapped fragments of at least 10kb were
666 appended to the pangenome fasta with headers informative of their genomic
667 locations in their respective source genomes.

668

669 Variant calling was performed following the protocols and using scripts described
670 in [42]. Briefly, sequence variants were called using short-read mapping to the
671 pangenome assembly representing the global diversity of *Z. tritici*. Reads were
672 filtered and trimmed using Trimmomatic (v0.40)[91] to remove adapter sequences,
673 trimming leading and trailing bases with a quality lower than 15, and removing
674 sequences shorter than 50bp. Trimmed reads were mapped to the pangenome
675 assembly using bowtie2 (v2.5.2)[92,93]. GATK4 (v4.4.0.0)[94] was used to call short
676 variants with the commands HaplotypeCaller, Combine GVCFs, and
677 GenotypeGVCFs, setting ploidy to 1 and the maximum number of alternative alleles
678 to 2.

679

680 Quality filtering was performed over several steps, starting with a standard set of
681 hard filters using the GATK quality metrics with thresholds set using visual
682 inspection of the metrics across the called variants. The per-site filters used were:
683 FS>10, MQ<20, QD<20, ReadPosRankSum, MQRankSum, and BaseQRankSum
684 between -2 and 2. Genotypes with depth <3 were also removed.

685

686 Genotyping errors can occur due to misalignment or the presence of repeated
687 sequences in the genome. In the case of *Z. tritici*, erroneous variant calls can arise
688 where near-equal numbers of reads support the reference and alternative allele,
689 which should be recognised as errors in a haploid organism. To resolve these errors,
690 genotypes with <90% of reads supporting the called allele were discarded.
691 Following this, included samples were filtered with vcftools (v0.1.16)[95] –missing-indv
692 and –depth options to remove samples with >20% missing data and a mean depth
693 of coverage <6 on core chromosomes of the IPO323 backbone of the pangenome
694 assembly. Clonal, or near-clonal, samples were removed by creating a network of
695 isolates with an identity-by-state value >0.99 as calculated by plink (v1.90b6.21)[96].
696 Subgraphs representing groups of clones were extracted in R (v4.3.3)[97] using the
697 tidygraph and ggraph packages[98,99] in the RStudio IDE[100,101]. For each network, the
698 isolate with the lowest amount of missing data was retained, whilst all other

699 isolates were discarded. Following these filtering steps, the final isolate count was
700 1,953 and the total variant count (indels and SNPS) was 11,066,828.

701

702 Population Structure Analyses

703 To estimate the global population structure of *Z. tritici*, the total SNP set was
704 filtered to sample 1 SNP/kb, with no missing data, and a minor allele frequency of
705 0.05, resulting in a total of 10,704 SNPs. A clustering analysis was run using the snmf
706 clustering method in the LEA R package[102]. Clustering was run for K values (number
707 of clusters) ranging from 1 to 15 with 10 repetitions per K. The best K was calculated
708 using the entropy method in snmf to determine the best-fit model, in addition to
709 manual evaluation based on smallest cluster size and the number of isolates
710 assigned to a cluster with a coefficient of >0.75. For downstream analysis, isolates
711 were assigned to discrete populations at the coefficient threshold at K=13. Two
712 clusters predominantly assigned to isolates of European origin were collapsed into
713 a single Europe cluster, in line with previous studies which assign European isolates
714 to a single dominant cluster[42]. This resulted in a total of 12 distinct genetic clusters,
715 including one labelled as "Hybrid", containing isolates that could not be assigned
716 at the coefficient threshold.

717

718 Benchmarking Short-Read Transposable Element Annotation
719 Tools

720 Several tools exist for annotating TEs from short-read sequencing to facilitate
721 population-level studies. 12 tools are included in the McClintock2 meta-pipeline
722 (https://github.com/bergmanlab/mcclintock)[49], namely: ngs_te_mapper[103],
723 ngs_te_mapper2[104], PoPoolationTE[105], PoPoolationTE2[106], RelocaTE[107], RelocaTE2[108],
724 RetroSeq[109], TEBreak[110], TEFLoN[50], TE-locate[111], TEMP[112], and TEMP2[52]. We were
725 unable to successfully run relocaTE2 and therefore we evaluated the performance
726 of 11 methodologies to determine which should be used to detect TE insertions in
727 our genome panel. A combination of tools is potentially required to detect both
728 reference and non-reference TE insertions from short-read datasets. In this context,

729 reference insertions are those present in the reference genome (i.e. the assembly
730 which short-reads are mapped to), whilst non-reference insertions are those that
731 are present in the short reads, but missing in the reference genome. To determine
732 the best tool, or combination of tools, to annotate TEs across 1,953 genomes, we
733 employed several benchmarking analyses for method selection. We consistently
734 ran all methods in McClintock2 with default parameters.

735

736 To assess tool annotation performance (analysis 1), we used chromosome 1 of the
737 CRI10 *Z. tritici* reference genome[75] as the "reference" assembly, with TEs annotated
738 by RepeatMasker (-s -norna) (v4.1.4)[113]) using our *de novo* TE library. 100x depth short
739 reads were generated using ART (version MountRainier-2016-06-05)[114] (-ss HS25 --
740 noALN -l 100 --paired --mflen 375 --sdev 110 --fcov 100) and a modified version of
741 chromosome 1 of CRI10 with 36 extra random TE sequences inserted at known
742 positions. The aim of analysis 1 is for each tool to successfully annotate all
743 "*reference*" TEs (i.e. those present in the unmodified chromosome 1 CRI10
744 reference) along with 36 "*non-reference*" TEs (i.e those present in the short reads
745 from modified CRI10, but not in the "reference" assembly to which reads are being
746 mapped).

747

748 To assess the ability of tools to avoid false positive annotations (analysis 2), we used
749 the modified chromosome 1 of CRI10 (i.e. with the extra 36 random TE sequences
750 added) as the "reference" assembly, with TEs annotated by RepeatMasker using
751 our *de novo* TE library. 100x depth short reads were generated using ART (-ss HS25
752 --noALN -l 100 --paired --mflen 375 --sdev 110 --fcov 100) and the unmodified version
753 of chromosome 1 of CRI10. The aim of analysis 2 is for each tool to successfully
754 annotate only "*reference*" TEs, but missing the 36 artificially added TE insertions,
755 which are not present in the short-read dataset. Here, any TEs labelled as non-
756 reference are false-positives, or mislabelled if they are indeed reference insertions.

757

758 To assess the consistency between TE polymorphism calls from short-read and
759 long-read data, (analyses 3 and 4), we used the IPO323 genome assembly as the

760 “reference” assembly, with TEs annotated by RepeatMasker using our *de novo* TE
761 library. For analysis 3, McClintock2 methods were used to map short reads from the
762 CRI10 isolate[42] to annotate TEs. For analysis 4, pbsv
763 (v2.9.0)(https://github.com/PacificBiosciences/pbsv) was used to call structural
764 variation (SV) using long reads from the same CRI10 isolate[75]. The aim of analyses 3
765 and 4 is to use both long and short reads from the same isolate (i.e. genetically
766 identical read sets) to determine whether SV calls using long-read data support TE
767 absence (i.e. deletions in comparison to the reference) and non-reference TE
768 presence (i.e. TEs present in CRI10 but absent in IPO323).

769

770 For analyses 1 and 2, reference TE loci were intersected with the TE annotations
771 GFF3 file produced by each method in McClintock2 using intersectBed in BedTools
772 (v2.31.0)[115]. True positives are defined as shared hits between the reference and
773 method GFF3 files. False negatives are defined as annotations in the reference
774 GFF3 missing in the method GFF3. False positives are defined as annotations
775 present in the method GFF3 but missing in the reference GFF3. Method
776 annotations were labelled based on the TE classification and insertion type (i.e
777 reference or non-reference TE) to evaluate overall method performance using R.
778 Pearon’s Chi-Squared Test for Independence was performed in R (chisq.test) to test
779 for significant differences among TE annotation methods.

780

781 For analyses 3 and 4, TE loci annotations generated by McClintock methods in
782 analysis 3 were intersected with the SV calls generated by pbsv in analysis 4 using
783 intersectBed in BedTools. Shared and unique loci were analysed in R and labels
784 assigned to evaluate the level of evidence supporting a call from the short-read
785 analyses. An annotation using a McClintock2 method was labelled as “supported”
786 by long-read data if the McClintock2 annotation is a non-reference insertion, and
787 the SV call is an insertion. Additionally, a reference annotation using McClintock2
788 methods was labelled as “supported” if there was no SV call, as reference insertions
789 should not generate SV calls, as they are present in both the reference genome (i.e.
790 IPO323) and the long-read dataset (CRI10 reads). McClintock2 method calls were
791 labelled as “dubious” if: (i) the McClintock2 call is a non-reference insertion, and the

792 SV call is “imprecise”; (ii) the McClintock2 call is a reference insertion, and the SV
793 call is “imprecise”, “deletion”, or “insertion”, as TE insertions are likely to have
794 degraded to different extents between CRI10 and IPO323. McClintock2 method
795 calls were labelled as “unsupported” if: (i) the McClintock2 call is a non-reference
796 insertion, and the SV call is “deletion”, “translocation”, “inversion”, or “duplication”;
797 (ii) the McClintock2 call is a reference insertion, and the SV call is “inversion”,
798 “translocation”, or “duplication”; (iii) the McClintock2 method call is a non-reference
799 insertion, and it is missing from the SV calls. Pearson’s Chi-Squared Test for
800 Independence was performed in R (chisq.test) to test for significant differences
801 among TE annotation methods.

802

803 Following benchmarking, we decided to combine three methods: TEFLoN[50],
804 PoPoolationTE[51], and TEMP2[52], with the latter used only to annotate non-reference
805 insertions. The non-redundant set of annotations from this combination of
806 methods was benchmarked using analyses 1 to 4 to assess the improvement
807 gained by combining methods. For all downstream analyses, this combination of
808 methods is used and a non-redundant set of TE annotations is generated using
809 intersectBed in BedTools and custom scripts to add PoPoolationTE and TEMP2
810 non-reference annotations to TEFLoN annotations. Full benchmarking results are
811 provided in supplementary file 1.

812

813 Before annotating the *Z. tritici* global panel, we wanted to understand the potential
814 impact of variation in sequencing technology and read depth on our ability to
815 detect TEs. We took the isolate with the most reads in each collection and
816 downsampled these to read thresholds calculated based on the isolate from the
817 collection with the lowest read count and shortest reads (Hartmann_FstQst_2015).
818 All test isolate read sets were downsampled to the following read count thresholds
819 using seqtk sample (v1.4-r122)(https://github.com/lh3/seqtk) with randomly
820 specified seeds per paired-end read set: (i) 9,860,000; (ii) 8,410,000; (iii) 7,480,000;
821 (iv) 6,540,000; (v) 5,610,000; (vi) 4,670,000; (vii) 3,740,000; (viii) 2,800,000; (ix) 1,
822 870,000; (x) 935,000. Models were fitted between read count and TE locus count in
823 R and the minimum threshold for each dataset was determined as the minimum

824 number of reads before the change in gradient was ≥0.1. Each isolate dataset was
825 labelled as "accepted" if the read count was above this threshold, "borderline" if the
826 read count was within 10% of the threshold, and "rejected" if the read count was
827 >10% below the threshold. In total, 418 isolates were accepted, 104 were borderline,
828 and 1,431 were rejected. Given the large number of rejected isolates, we retained all
829 isolates in the global panel for downstream analyses. Consequently, we likely
830 underestimate the TE content in rejected isolates, and therefore will only detect
831 the strongest differences in TE locus frequency among populations, and miss more
832 subtle and potentially more recent patterns.

833

834 Annotating Transposable Element Insertions in the global panel

835 Following benchmarking, all 1,953 short-read genomes were analysed with
836 McClintock2 methods TEFLoN, PoPoolationTE, and TEMP2 (non-reference only) to
837 annotate TEs in each isolate. Here, each set of sequencing reads was mapped to
838 the pangenome assembly generated for this study (see methods: SNP Calling).
839 McClintock2 was also provided with the *de novo* TE library for *Z. tritici* in FASTA
840 format, a GFF3 file containing the TE annotations for the pangenome assembly,
841 and a taxonomy file providing mapping between each unique TE in the GFF3 file
842 and its respective classification. For each run, the resultant TE annotations were
843 combined to make a non-redundant set comprising positions identified by
844 TEFLoN, PoPoolationTE, and TEMP2, using bedtools to add unique positions from
845 PoPoolationTE and TEMP2 to the TEFLoN annotations.

846

847 By default, McClintock filters annotations for quality. For a TE annotation to be
848 called, all three methods require all TE annotations to be supported by ≥10% of the
849 total reads, whilst TEFLoN also requires a minimum of 3 reads supporting insertion
850 presence. TEFLoN also requires non-reference hits to be "*supported by both 5' and*
851 *3' breakpoints*", PoPoolationTE requires insertions to be supported by "*reads on*
852 *both the forward and reverse strand*", and TEMP2 requires insertions to be
853 supported by "*reads mapping to both ends of the insertion*". We visualised the
854 frequency of reads supporting each TE annotation using R and ggplot2 (v3.5.0)[116] to

855 determine if further quality filtering was needed. Given that *Z. tritici* is haploid, we
856 set a further quality threshold to remove any positions with a read frequency ≤0.5,
857 equivalent to a TE annotation having a 50% chance of being false. For each isolate,
858 ~90% of total TE annotations were supported by ≥90% of mapped reads supporting
859 the presence of a TE insertion.

860

861 TE locus information was parsed in R to label every locus from each isolate. Each
862 locus was given an individual ID, and loci were defined as shared among isolates if
863 annotations were for the same TE family on the same contig, at overlapping
864 coordinates and/or within 100bp of each other. Each locus was labelled with
865 diagnostic information: (i) a unique numeric locus identifier; (ii) as "reference" if it
866 was present in the pangenome assembly, or "non-reference" if it was absent; (iii)
867 the contig where the annotation is found; (iv) the TE family identifier from the *de*
868 *novo* TE library; (v) Either "rare" if the TE is present in < 10% of the isolates in the
869 respective population, "common" if the TE is present in ≥ 10% and < 95% of the
870 isolates in the respective population, and "fixed" if the TE is present in > 95% of the
871 isolates in the respective population, following the recommendations of [53].

872

873 To determine the importance of TEs in driving population differentiation, a TE locus
874 presence/absence matrix was generated in R. TE loci were filtered to remove
875 positions present in <10 isolates, reducing the locus count to 17,129. A principal
876 component analysis (PCA) was performed using prcomp in R and results were
877 visualised with ggplot2. All subsequent downstream analyses were performed
878 using R, in the Rstudio IDE, and the tidyverse (v2.0.0), data.table (v1.14.10), and
879 magrittr (v2.0.3) packages. Plots were generated with ggplot2[116] and KaryoploteR[117].

880

881 Detection of Adaptive Transposable Element Insertions

882 Pangenome-wide scans for adaptive differentiation were performed using the
883 core model in BayPass (v2.4)[54], which accounts for the correlation structure among
884 allele frequencies and facilitates the detection of putative TEs subject to adaptive

885 variation based on the XtX statistic[118]. As described by [54], the core model has several
886 advantages over other methods of detecting adaptive variation as it explicitly
887 accounts for the demographic history of populations (covariance matrix Ω), whilst
888 making no assumptions about the underlying demography.

889

890 Only isolates that were confidently assigned to a genetic cluster were included,
891 resulting in a dataset containing 1,569 individuals. TE loci were filtered to maximise
892 our ability to detect putatively adaptive loci with high confidence: (i) TE loci unique
893 to a single isolate were removed; (ii) TE loci were removed if they were not found in
894 at least 10% of the isolates of at least one population (*i.e.* TE loci rising to common
895 frequency in at least one population were retained); (iii) TE loci were removed if
896 they were not polymorphic in at least one population. This reduced our analysis to
897 11,073 TE loci. The core model was provided with the relatedness matrix used to
898 assign isolates to their respective genetic clusters (see methods: Determining the
899 population structure of *Zymoseptoria tritici*), to ensure consistency in the
900 estimations of population structure. The genotyping input file was generated from
901 the TE presence/absence matrix used above in R. Briefly, the genotyping file
902 contained TE presence and absence counts (in lieu of reference and alternative
903 allele counts if SNPs were used) per site and per population. TE presence counts
904 were used as the alternative count, whilst TE absence counts were used as the
905 reference count[119].

906

907 The BayPass core model was run with 40 pilot runs of 1,000 iterations, followed by
908 burn-in of 1,000,000 iterations and 10,000,000 model runs sampled with a thinning
909 interval of 1,000 to result in a total value count of 10,000. The model was run three
910 times with different seeds to verify convergence among separate model estimates.
911 Following model runs, TEs with very low allele frequencies (MAF <0.01) based on
912 the mean of the posterior distribution of the frequency of the reference allele
913 across populations for each site were filtered, as recommended in the BayPass
914 manual (https://forgemia.inra.fr/mathieu.gautier/baypass_public/)[54]. To control for
915 multiple testing, the p-value distributions were manually inspected. As the p-value
916 distributions were uniform, a false discovery rate correction was applied using the

917 Benjamini-Hochberg method in R. We focused on XtX statistic scores below an
918 FDR-corrected significance threshold of 0.05. Loci were then filtered further so that
919 only TEs on the IPO323 chromosomal-level backbone of the pangenome were
920 considered, as these contained genomic sufficient context along with gene
921 annotations to facilitate downstream analyses.

922

923 Pangenome-wide scans were also performed using the same parameters and
924 three long model runs using SNP loci to assess the co-localisation among
925 differentiated TE and SNP loci. Similarly, only isolates assigned to genetic clusters
926 were included (1,569 individuals). Loci with >10% missing data were removed,
927 resulting in 4,189,176 SNPs. The genotyping input file was generated from the SNP-
928 calling VCF in R, resulting in a file containing SNP allele calls per site and per
929 population. Consistent with the TE analysis, SNPs with very low allele frequencies
930 (MAF <0.01) based on the mean of the posterior distribution of the frequency of the
931 reference allele across populations for each site were filtered. P-value distributions
932 were manually inspected and a q-value transformation was applied to control for
933 multiple testing in R using the qvalue package(v2.36.0)[120]. SNPs were defined as
934 significant if their XtX statistics were in the top 0.05% of values, and their associated
935 q-values were below 0.05.

936

937 Integration of bioclimatic, fungicide resistance and gene 938 function datasets

939 To gain insight into the potential mechanistic impacts of candidate TE loci, we
940 obtained: (i) Gene annotations for the reference assembly IPO323[121]; (ii) SNP loci
941 significantly associated with fungicide resistance from a *Z. tritici* fungicide
942 resistance atlas[46]; (iii) Data for 76 bioclimatic variables from the Climatologies at
943 high resolution for the earth's land surface areas dataset (CHELSA v2.1)
944 (https://chelsa-climate.org/bioclim/)[122,123]; (iv) Effector gene amino acid sequences
945 from [58] (v) Biosynthetic cluster annotations for the reference assembly IPO323[57].

946

947 Genes within 1,000bp of candidate TE loci were obtained using BedTools window.
948 Candidate gene amino acid sequences were annotated with InterProScan (v5.67-
949 99.0)[55,56] to assign predicted gene functions, GO terms, and MetaCyc and Reactome
950 pathway memberships.

951

952 SNPs with significant associations to fungicide resistance within 1,000bp of
953 candidate TE loci were obtained with BedTools window. Information on the specific
954 fungicides to which resistance is conferred and SNPeff predicted impacts were
955 obtained from the fungicide resistance atlas.

956

957 For each isolate, geographical coordinates were obtained from attached metadata.
958 In some cases, assigned coordinates were inferred based on the most precise
959 sampling data available, as previously described[42]. The sampling coordinates for
960 each isolate were used to approximate the environmental conditions from all 76
961 bioclimatic variables, which are averaged over the period 1981-2010[122,123]. We
962 generated a pseudo-VCF containing all TE calls for each isolate, where 0
963 represented TE absence and 1 represented TE presence (Supplementary File 3). We
964 identified genotype-to-environment associations among TE loci and bioclimatic
965 variables using vcf2gwas (https://github.com/frankvogt/vcf2gwas)[124], specifying
966 the use of univariate linear mixed models (-lmm). Given the high correlation
967 among distinct bioclimatic variables (Supplementary Figure S1), we also performed
968 PCA-based dimensionality reduction on the bioclimatic variables within vcf2gwas
969 (--PCA 10), which were then used as the phenotypes for the genotype-to-
970 environment association. Following this, TEs with significant genotype-to-
971 environment associations were filtered to retain those within 1,000bp of candidate
972 TE loci.

973

974 Predicted effector gene amino acid sequences were obtained[58] and corresponding
975 loci in the IPO323 reference assembly used in this study were identified using
976 tblastn[82] followed by manual curation (Supplementary Table S9). A bed file was

977 generated and BedTools window was used to identify putative effector genes
978 within 1,000bp of candidate TE loci.

979

980 Biosynthetic gene clusters (BGCs) are arrays of genes involved in the production of
981 diverse secondary metabolites[125], and they often exhibit large presence/absence
982 variation among individuals and populations[57]. We obtained BGC annotations for
983 the IPO323 reference assembly from [57], and cross-referenced the gene members
984 with genes located with 1,000bp of candidate TEs to assign predicted BGC
985 functions to genes at these loci.

986

987 Plots to illustrate the association among candidate TEs and aforementioned data
988 sources were generated using circos (v0.69-8)[126].

989

990

991 # Acknowledgements

992 DC was supported by the Swiss National Science Foundation (grant 201149) by an
993 Innosuisse grant (32532.1 IP-LS).

994

References

1. Merenciano, M. & González, J. The Interplay Between Developmental Stage and Environment Underlies the Adaptive Effect of a Natural Transposable Element Insertion. *Mol. Biol. Evol.* **40**, (2023).

2. Schmidt, J. M. *et al.* Copy number variation and transposable elements feature in recent, ongoing adaptation at the Cyp6g1 locus. *PLoS Genet.* **6**, e1000998 (2010).

3. Puinean, A. M. *et al.* Amplification of a cytochrome P450 gene is associated with resistance to neonicotinoid insecticides in the aphid Myzus persicae. *PLoS Genet.* **6**, 1–11 (2010).

4. Omrane, S. *et al.* Plasticity of the MFS1 Promoter Leads to Multidrug Resistance in the Wheat Pathogen Zymoseptoria tritici. *mSphere* **2**, (2017).

5. Krishnan, P. *et al.* Transposable element insertions shape gene regulation and melanin production in a fungal pathogen of wheat. *BMC Biol.* **16**, 78 (2018).

6. Tahami, M. S. *et al.* Transposable elements in a cold-tolerant fly species, Drosophila montana: a link to adaptation to the harsh cold environments. *bioRxiv* 2024.04.17.589934 (2024) doi:10.1101/2024.04.17.589934.

7. McClintock, B. “Controlling Elements and the Gene.” *Cold Spring Harb. Symp. Quant. Biol.* **21**, 197–216 (1956).

8. de Koning, A. P. J., Gu, W., Castoe, T. A., Batzer, M. A. & Pollock, D. D. Repetitive elements may comprise over Two-Thirds of the human genome. *PLoS Genet.* **7**, (2011).

9. Vicient, C. M. Transcriptional activity of transposable elements in maize. *BMC Genomics* **11**, 601 (2010).

10. Canapa, A., Barucca, M., Biscotti, M. A., Forconi, M. & Olmo, E. Transposons, Genome Size, and Evolutionary Insights in Animals. *Cytogenet. Genome Res.* **147**, 217–239 (2015).

11. Bleykasten-Grosshans, C. & Neuvéglise, C. Transposable elements in yeasts. *C. R. Biol.* **334**, 679–686 (2011).

12. Wöstemeyer, J. & Kreibich, A. Repetitive DNA elements in fungi (Mycota): impact on genomic architecture and evolution. *Curr. Genet.* **41**, 189–198 (2002).

13. Chénais, B., Caruso, A., Hiard, S. & Casse, N. The impact of transposable elements on eukaryotic genomes: From genome size increase to genetic adaptation to stressful environments. *Gene* **509**, 7–15 (2012).

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

15. Fouché, S. *et al.* Stress-Driven Transposable Element De-repression Dynamics and Virulence Evolution in a Fungal Pathogen. *Mol. Biol. Evol.* **37**, 221–239 (2020).

16. Sundaram, V. *et al.* Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res.* **24**, 1963–1976 (2014).

17. Bourque, G. *et al.* Ten things you should know about transposable elements. *Genome Biol.* **19**, 199 (2018).

18. Cosby, R. L., Chang, N.-C. & Feschotte, C. Host–transposon interactions: conflict, cooperation, and cooption. *Genes Dev.* **33**, 1098–1116 (2019).

19. Wells, J. N. & Feschotte, C. A Field Guide to Eukaryotic Transposable Elements. *Annu. Rev. Genet.* **54**, 539–561 (2020).

20. Badet, T., Tralamazza, S. M., Feurtey, A. & Croll, D. Recent reactivation of a pathogenicity-associated transposable element is associated with major chromosomal rearrangements in a fungal wheat pathogen. *Nucleic Acids Res.* **52**, 1226–1242 (2024).

21. Schrader, L. & Schmitz, J. The impact of transposable elements in adaptive evolution. *Mol. Ecol.* (2018) doi:10.1111/mec.14794.

22. Kapitonov, V. V. & Koonin, E. V. Evolution of the RAG1-RAG2 locus: Both proteins came from the same transposon. *Biol. Direct* **10**, 1–8 (2015).

23. Proença, J. T., Barral, D. C. & Gordo, I. Commensal-to-pathogen transition: One-single transposon insertion results in two pathoadaptive traits in Escherichia coli -macrophage interaction. *Sci. Rep.* **7**, 4504 (2017).

24. Butelli, E. *et al.* Retrotransposons Control Fruit-Specific, Cold-Dependent Accumulation of Anthocyanins in Blood Oranges. *Plant Cell* **24**, 1242–1255 (2012).

25. Daborn, P. J. *et al.* A single p450 allele associated with insecticide resistance in Drosophila. *Science* **297**, 2253–2256 (2002).

26. de Little, S. C., Edwards, O., van Rooyen, A. R., Weeks, A. & Umina, P. A. Discovery of metabolic resistance to neonicotinoids in green peach aphids (Myzus persicae) in Australia. *Pest Manag. Sci.* **73**, 1611–1617 (2017).

27. Möller, M. & Stukenbrock, E. H. Evolution and genome architecture in fungal plant pathogens. *Nat. Rev. Microbiol.* **15**, 771 (2017).

28. Lorrain, C., Oggenfuss, U., Croll, D., Duplessis, S. & Stukenbrock, E. Transposable elements in fungi: Coevolution with the host genome shapes, genome architecture, plasticity and adaptation. *Encyclopedia of* (2021) doi:10.1016/B978-0-12-819990-9.00042-1.

29. Hof, A. E. V. t. *et al.* The industrial melanism mutation in British peppered moths is a transposable element. *Nature* **534**, 102–105 (2016).

30. Vargas-Chavez, C. *et al.* Transposable element variants and their potential adaptive impact in urban populations of the malaria vector Anopheles coluzzii. *Genome Res.* **32**, 189–202 (2022).

31. Ullastres, A., Merenciano, M. & González, J. Regulatory regions in natural transposable element insertions drive interindividual differences in response to immune challenges in Drosophila. *Genome Biol.* **22**, 265 (2021).

32. Rouzic, A. L., Boutin, T. S. & Capy, P. Long-term evolution of transposable elements. *Proceedings of the National Academy of Sciences* **104**, 19375–19380 (2007).

33. Charlesworth, B. The Population Genetics of Transposable Elements. *Populaion Genetics and Molecular Evolution* 219–232 (1985).

34. Charlesworth, B., Sniegowski, P. & Stephan, W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371**, 215–220 (1994).

35. Belyayev, A. Bursts of transposable elements as an evolutionary driving force. *J. Evol. Biol.* **27**, 2573–2584 (2014).

36. Scarpa, A. *et al.* Double trouble: two retrotransposons triggered a cascade of invasions in Drosophila species within the last 50 years. *Nat. Commun.* **16**, 516 (2025).

37. Pianezza, R., Scarpa, A., Narayanan, P., Signor, S. & Kofler, R. Spoink, a LTR retrotransposon, invaded D. melanogaster populations in the 1990s. *PLoS Genet.* **20**, e1011201 (2024).

38. Keller, I., Bensasson, D. & Nichols, R. A. Transition-transversion bias is not universal: a counter example from grasshopper pseudogenes. *PLoS Genet.* **3**, e22 (2007).

39. Fones, H. & Gurr, S. The impact of Septoria tritici Blotch disease on wheat: An EU perspective. *Fungal Genet. Biol.* **79**, 3–7 (2015).

40. Stukenbrock, E. H., Banke, S., Javan-Nikkhah, M. & McDonald, B. A. Origin and domestication of the fungal wheat pathogen Mycosphaerella graminicola via sympatric speciation. *Mol. Biol. Evol.* **24**, 398–411 (2007).

41. Petit-Houdenot, Y., Lebrun, M.-H. & Scalliet, G. Understanding plant-pathogen interactions in Septoria tritici blotch infection of cereals. in *Achieving durable disease resistance in cereals* 263–302 (Burleigh Dodds Science Publishing, London, 2021).

42. Feurtey, A. *et al.* A thousand-genome panel retraces the global spread and adaptation of a major fungal crop pathogen. *Nat. Commun.* **14**, 1059 (2023).

43. Singh, N. K., Karisto, P. & Croll, D. Population-level deep sequencing reveals the interplay of clonal and sexual reproduction in the fungal wheat pathogen Zymoseptoria tritici. *Microb. Genom.* **7**, (2021).

44. Abraham, L. N., Oggenfuss, U. & Croll, D. Population-level transposable element expression dynamics influence trait evolution in a fungal crop pathogen. *MBio* **15**, e0284023 (2024).

45. Oggenfuss, U. *et al.* A population-level invasion by transposable elements triggers genome expansion in a fungal pathogen. *Elife* **10**, (2021).

46. Puccetti, G. *et al.* A large European diversity panel reveals complex azole fungicide resistance gains of a major wheat pathogen. *bioRxiv* 2025.03.19.644200 (2025) doi:10.1101/2025.03.19.644200.

47. Baril, T., Galbraith, J. & Hayward, A. Earl Grey: A Fully Automated User-Friendly Transposable Element Annotation and Analysis Pipeline. *Mol. Biol. Evol.* **41**, (2024).

48. Baril, T. & Croll, D. A pangenome-guided manually curated library of transposable elements for Zymoseptoria tritici. *BMC Res. Notes* **16**, (2023).

49. Chen, J., Basting, P. J., Han, S., Garfinkel, D. J. & Bergman, C. M. Reproducible evaluation of transposable element detectors with McClintock 2 guides accurate inference of Ty insertion patterns in yeast. *Mob. DNA* **14**, 8 (2023).

50. Adrion, J. R., Song, M. J., Schrider, D. R., Hahn, M. W. & Schaack, S. Genome-Wide Estimates of Transposable Element Insertion and Deletion Rates in Drosophila Melanogaster. *Genome Biol. Evol.* **9**, 1329–1340 (2017).

51. Kofler, R. *et al.* PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS One* **6**, e15925 (2011).

52. Yu, T. *et al.* A benchmark and an algorithm for detecting germline transposon insertions and measuring de novo transposon insertion frequencies. *Nucleic Acids Res.* **49**, e44 (2021).

53. Rech, G. E. *et al.* Population-scale long-read sequencing uncovers transposable elements associated with gene expression variation and adaptive signatures in Drosophila. *Nat. Commun.* **13**, 1948 (2022).

1135 54. Gautier, M. Genome-Wide Scan for Adaptive Divergence and Association with
1136 Population-Specific Covariates. *Genetics* **201**, 1555–1579 (2015).

1137 55. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification.
1138 *Bioinformatics* **30**, 1236–1240 (2014).

1139 56. Blum, M. *et al.* The InterPro protein families and domains database: 20 years
1140 on. *Nucleic Acids Res.* **49**, D344–D354 (2021).

1141 57. Tralamazza, S. M., Gluck-Thaler, E., Feurtey, A. & Croll, D. Copy number variation
1142 introduced by a massive mobile element facilitates global thermal adaptation
1143 in a fungal wheat pathogen. *Nat. Commun.* **15**, 5728 (2024).

1144 58. Thynne, E. *et al.* An array of*Zymoseptoria tritici*effectors suppress plant
1145 immune responses. *bioRxiv* (2024) doi:10.1101/2024.03.12.584321.

1146 59. Singh, Y., Nair, A. M. & Verma, P. K. Surviving the odds: From perception to
1147 survival of fungal phytopathogens under host-generated oxidative burst. *Plant*
1148 *Communications* **2**, 100142 (2021).

1149 60. Ene, I. V. *et al.* Global analysis of mutations driving microevolution of a
1150 heterozygous diploid fungal pathogen. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E8688–
1151 E8697 (2018).

1152 61. Zhu, Y. O., Siegal, M. L., Hall, D. W. & Petrov, D. A. Precise estimates of mutation
1153 rate and spectrum in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E2310-8 (2014).

1154 62. Edwards, H. M. & Rhodes, J. Accounting for the Biological Complexity of
1155 Pathogenic Fungi in Phylogenetic Dating. *J Fungi (Basel)* **7**, (2021).

1156 63. Dutta, A., McDonald, B. A. & Croll, D. Combined reference-free and multi-
1157 reference based GWAS uncover cryptic variation underlying rapid adaptation
1158 in a fungal plant pathogen. *PLoS Pathog.* **19**, e1011801 (2023).

1159 64. Zhong, Z. *et al.* A small secreted protein in Zymoseptoria tritici is responsible
1160 for avirulence on wheat cultivars carrying the Stb6 resistance gene. *New*
1161 *Phytol.* **214**, 619–631 (2017).

1162 65. Hartmann, F. E., Sánchez-Vallet, A., McDonald, B. A. & Croll, D. A fungal wheat
1163 pathogen evolved host specialization by extensive chromosomal
1164 rearrangements. *ISME J.* **11**, 1189–1204 (2017).

1165 66. Amezrou, R. *et al.* A secreted protease-like protein in Zymoseptoria tritici is
1166 responsible for avirulence on Stb9 resistance gene in wheat. *PLoS Pathog.* **19**,
1167 e1011376 (2023).

1168 67. Baril, T., Pym, A., Bass, C. & Hayward, A. Transposon accumulation at xenobiotic
1169 gene family loci in aphids. *Genome Res.* **33**, 1718–1733 (2023).

68. Fouché, S., Oggenfuss, U., Chanclud, E. & Croll, D. A devil's bargain with transposable elements in plant pathogens. *Trends Genet.* **38**, 222–230 (2022).

69. Salces-Ortiz, J., Vargas-Chavez, C., Guio, L., Rech, G. E. & González, J. Transposable elements contribute to the genomic response to insecticides in Drosophila melanogaster. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **375**, 20190341 (2020).

70. Gilbert, C., Peccoud, J. & Cordaux, R. Transposable Elements and the Evolution of Insects. *Annu. Rev. Entomol.* **66**, 355–372 (2021).

71. Li, X., Schuler, M. A. & Berenbaum, M. R. Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. *Annu. Rev. Entomol.* **52**, 231–253 (2007).

72. Itokawa, K., Komagata, O., Kasai, S., Masada, M. & Tomita, T. Cis-acting mutation and duplication: History of molecular evolution in a P450 haplotype responsible for insecticide resistance in Culex quinquefasciatus. *Insect Biochem. Mol. Biol.* **41**, 503–512 (2011).

73. Panini, M. *et al.* Transposon-mediated insertional mutagenesis unmasks recessive insecticide resistance in the aphid Myzus persicae. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).

74. Sun, X. *et al.* PdMLE1, a specific and active transposon acts as a promoter and confers Penicillium digitatum with DMI resistance. *Environ. Microbiol. Rep.* **5**, 135–142 (2013).

75. Badet, T., Oggenfuss, U., Abraham, L., McDonald, B. A. & Croll, D. A 19-isolate reference-quality global pangenome for the fungal wheat pathogen Zymoseptoria tritici. *BMC Biol.* **18**, 12 (2020).

76. Goodwin, S. B. *et al.* Finished genome of the fungal wheat pathogen Mycosphaerella graminicola reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. *PLoS Genet.* **7**, e1002070 (2011).

77. Plissonneau, C., Hartmann, F. E. & Croll, D. Pangenome analyses of the wheat pathogen Zymoseptoria tritici reveal the structural basis of a highly plastic eukaryotic genome. *BMC Biol.* **16**, 5 (2018).

78. Feurtey, A. *et al.* Genome compartmentalization predates species divergence in the plant pathogen genus Zymoseptoria. *BMC Genomics* **21**, 588 (2020).

79. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).

80. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).

81. Goubert, C. *et al.* A beginner's guide to manual curation of transposable elements. *Mob. DNA* 1–98 (2022).

82. Camacho, C. *et al.* BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 1–9 (2009).

83. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

84. Notredame, C., Higgins, D. G. & Heringa, J. T-coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217 (2000).

85. Rice, P., Longden, L. & Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277 (2000).

86. Wicker, T. *et al.* A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**, 973–982 (2007).

87. Hubley, R. *et al.* The Dfam database of repetitive DNA families. *Nucleic Acids Res.* **44**, D81–D89 (2016).

88. Storer, J., Hubley, R., Rosen, J., Wheeler, T. J. & Smit, A. F. The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mob. DNA* **12**, 2 (2021).

89. Wheeler, T. J. & Eddy, S. R. nhmmer: DNA homology search with profile HMMs. *Bioinformatics* **29**, 2487–2489 (2013).

90. Baril, T. & Croll, D. Zymoseptoria tritici pangenome-guided TE consensus library. Zenodo https://doi.org/10.5281/ZENODO.8379981 (2023).

91. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

92. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

93. Langmead, B., Wilks, C., Antonescu, V. & Charles, R. Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* **35**, 421–432 (2019).

94. Van der Auwera, G. A. & O'Connor, B. D. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. ("O'Reilly Media, Inc.," 2020).

95. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

96. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

97. R Core Team. R: A Language and Environment for Statistical Computing. Preprint at https://www.R-project.org (2023).

98. Pedersen, T. L. tidygraph: A tidy API for graph manipulation. *R package version* **1**, (2020).

99. Pedersen, T. L. ggraph: An implementation of grammar of graphics for graphs and networks. *R package version* **2**, 1 (2020).

100. Rstudio Team. RStudio: Integrated Development Environment for R. *RStudio, PBC. Boston, MA.* **700**, 879 (2020).

101. Racine, J. S. RSTUDIO: A PLATFORM-INDEPENDENT IDE FOR R AND SWEAVE. *J. Appl. Econometrics* **27**, 167–172 (2013).

102. Frichot, E. & François, O. LEA: An R package for landscape and ecological association studies. *Methods Ecol. Evol.* (2015) doi:10.1111/2041-210X.12382.

103. Linheiro, R. S. & Bergman, C. M. Whole genome resequencing reveals natural target site preferences of transposable elements in Drosophila melanogaster. *PLoS One* **7**, e30008 (2012).

104. Han, S. *et al.* Transposable element profiles reveal cell line identity and loss of heterozygosity in Drosophila cell culture. *Genetics* **219**, (2021).

105. Kofler, R., Betancourt, A. J. & Schlötterer, C. Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in Drosophila melanogaster. *PLoS Genet.* **8**, e1002487 (2012).

106. Kofler, R., Gómez-Sánchez, D. & Schlötterer, C. PoPoolationTE2: Comparative Population Genomics of Transposable Elements Using Pool-Seq. *Mol. Biol. Evol.* **33**, 2759–2764 (2016).

107. Robb, S. M. C. *et al.* The use of RelocaTE and unassembled short reads to produce high-resolution snapshots of transposable element generated diversity in rice. *G3* **3**, 949–957 (2013).

108. Chen, J., Wrightsman, T. R., Wessler, S. R. & Stajich, J. E. RelocaTE2: a high resolution transposable element insertion site mapping tool for population resequencing. *PeerJ* **5**, e2942 (2017).

109. Keane, T. M., Wong, K. & Adams, D. J. RetroSeq: transposable element discovery from next-generation sequencing data. *Bioinformatics* **29**, 389–390 (2013).

110. Schauer, S. N. *et al.* L1 retrotransposition is a common feature of mammalian hepatocarcinogenesis. *Genome Res.* **28**, 639–653 (2018).

111. Platzer, A., Nizhynska, V. & Long, Q. TE-Locate: A Tool to Locate and Group Transposable Element Occurrences Using Paired-End Next-Generation Sequencing Data. *Biology* **1**, 395–410 (2012).

112. Zhuang, J., Wang, J., Theurkauf, W. & Weng, Z. TEMP: a computational method for analyzing transposable element polymorphism in populations. *Nucleic Acids Res.* **42**, 6826–6838 (2014).

113. Smit, A. F. A., Hubley, R. R. & Green, P. R. RepeatMasker Open-4.0. *http://repeatmasker.org* Preprint at (2013).

114. Huang, W., Li, L., Myers, J. R. & Marth, G. T. ART: a next-generation sequencing read simulator. *Bioinformatics* **28**, 593–594 (2012).

115. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

116. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*. (Springer, 2016).

117. Gel, B. & Serra, E. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* **33**, 3088–3090 (2017).

118. Günther, T. & Coop, G. Robust identification of local adaptation from allele frequencies. *Genetics* **195**, 205–220 (2013).

119. Bogaerts-Márquez, M., Guirao-Rico, S., Gautier, M. & González, J. Temperature, rainfall and wind variables underlie environmental adaptation in natural populations of Drosophila melanogaster. *Mol. Ecol.* **30**, 938–954 (2021).

120. Dabney, A., Storey, J. D. & Warnes, G. Q-value estimation for false discovery rate control. *Medicine* **344**, 48 (2004).

121. Lapalu, N. *et al.* Improved gene annotation of the fungal wheat pathogen Zymoseptoria tritici based on combined Iso-Seq and RNA-Seq evidence. *bioRxiv* (2023) doi:10.1101/2023.04.26.537486.

122. Karger, D. N. *et al.* Climatologies at high resolution for the earth's land surface areas. *Sci Data* **4**, 170122 (2017).

123. Karger, D. N. *et al.* Data from: Climatologies at high resolution for the earth's land surface areas. *Climatologies at high resolution for the earth's land surface areas* https://doi.org/10.5061/dryad.kd1d4 (2018).

124. Vogt, F., Shirsekar, G. & Weigel, D. vcf2gwas: Python API for comprehensive GWAS analysis using GEMMA. *Bioinformatics* **38**, 839–840 (2022).

125. Keller, N. P. Translating biosynthetic gene clusters into fungal armor and weaponry. *Nat. Chem. Biol.* **11**, 671–677 (2015).

126. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).