

Population-level transposable element expression dynamics influence trait evolution in a fungal crop pathogen

Leen Nanchira Abraham,¹ Ursula Oggenfuss,¹ Daniel Croll¹

AUTHOR AFFILIATION See affiliation list on p. 17.

ABSTRACT The rapid adaptive evolution of microbes is driven by strong selection pressure acting on genetic variation. How adaptive genetic variation is generated within species and how such variation influences phenotypic trait expression is often not well understood though. We focused on the recent activity of transposable elements (TEs) using deep population genomics and transcriptomics analyses of a fungal plant pathogen with a highly active content of TEs in the genome. *Zymoseptoria tritici* causes one of the most damaging diseases on wheat, with recent adaptation to the host and environment being facilitated by TE-associated mutations. We obtained genomic and RNA-sequencing data from 146 isolates collected from a single wheat field. We established a genome-wide map of TE insertion polymorphisms in the population by analyzing recent TE insertions among individuals. We quantified the locus-specific transcription of individual TE copies and found considerable population variation at individual TE loci in the population. About 20% of all TE copies show transcription in the genome suggesting that genomic defenses such as repressive epigenetic marks and repeat-induced polymorphisms are at least partially ineffective at preventing the proliferation of TEs in the genome. A quarter of recent TE insertions are associated with expression variation of neighboring genes providing broad potential to influence trait expression. We indeed found that TE insertions are likely responsible for variation in virulence on the host and potentially diverse components of secondary metabolite production. Our large-scale transcriptomics study emphasizes how TE-derived polymorphisms segregate even in individual microbial populations and can broadly underpin trait variation in pathogens.

IMPORTANCE Pathogens can rapidly adapt to new hosts, antimicrobials, or changes in the environment. Adaptation arises often from mutations in the genome; however, how such variation is generated remains poorly understood. We investigated the most dynamic regions of the genome of *Zymoseptoria tritici*, a major fungal pathogen of wheat. We focused on the transcription of transposable elements. A large proportion of the transposable elements not only show signatures of potential activity but are also variable within a single population of the pathogen. We find that this variation in activity is likely influencing many important traits of the pathogen. Hence, our work provides insights into how a microbial species can adapt over the shortest time periods based on the activity of transposable elements.

KEYWORDS transposable elements, gene expression, rapid adaptation, plant pathogen, fungi *Zymoseptoria tritici*

Rapid adaptive evolution enables microbial species to cope with challenging environmental conditions including climate change. Most evidence for rapid adaptation comes from experimental studies applying artificial selection pressures (1). How adaptive genetic variation is generated or maintained in natural populations

Editor Alexander Idnurm, University of Melbourne, Melbourne, Australia

Address correspondence to Daniel Croll, daniel.croll@unine.ch.

The authors declare no conflict of interest.

Received 20 October 2023

Accepted 22 January 2024

Published 13 February 2024

Copyright © 2024 Abraham et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

remains poorly explored. Transposable elements (TEs) were recently recognized as key drivers of adaptive genetic variation within species and even single populations (2–4). TEs are genomic sequences that can be mobilized, and high activity in some species was linked to the rapid gain of adaptive variation (5–8). Based on the mode of proliferation, TEs are categorized into retrotransposons, which copy *via* an RNA intermediate, and DNA transposons which can excise and integrate into a different locus. TE-encoded transposase proteins mediate the excision and integration of TEs into the genome (9). TEs create genetic variation in populations through their mobilization. New TE copies can affect gene functions, for example, by insertion into *cis*-regulatory elements or serve as alternative promoters that render a gene more responsive (10, 11). Similarly, TE exaptation into coding sequences can produce novel regulatory sequences and lead to intronization or exonization events with beneficial or detrimental effects for the host (12, 13). Also, TEs can be an important source of epigenetic modifications regulating gene expression (14). Thus, TEs actively inserting into new genomic loci can produce a multitude of effects at the level of gene functions and expression of phenotypic traits (15–17).

TE insertion polymorphisms (TIPs) within species are characterized by individual genotypes differing in the presence or absence of a specific TE at a specific locus. Such insertion polymorphism has well-documented effects in plants such as *Arabidopsis thaliana* (18) and *Oryza sativa* (19). However, TIPs are also widespread in fungal genomes of plant pathogens including *Zymoseptoria tritici* and *Parastagonospora nodorum* (3, 20). The effects of TE insertions segregating within species largely depend on the site of insertion. Due to epigenetic silencing, most TE copies in eukaryotic genomes are not transcriptionally active (21). TE-rich genomic regions carry repressive epigenetic marks, such as histone modifications and DNA methylation. These modifications prevent the transcriptional machinery from accessing TE sequences, thereby silencing TE copies. Beyond epigenetic silencing of TE-rich regions, genomes can encode specific machinery to defend against active transposition. This includes RNA interference, which can bind and cleave transcribed TE sequences (22, 23). In some fungal genomes, defenses expand beyond interference and include repeat-induced point mutations (RIP) as a highly targeted mechanism to counteract TE activity. RIP introduces C→T mutations in repetitive DNA during meiosis and is thought to largely prevent the retention of duplicated sequences in the genome (24).

Assessing the landscape of TEs in the genome helps distinguish between silenced and deactivated TEs from copies with the potential to proliferate. Genomes of fungal pathogens including the plant, animal, and the broad host range pathogens carry large TE-rich regions that also encode virulence-related genes (25–28). The TE-rich regions influence genome plasticity and effector expression in the rice blast pathogen *Magnaporthe oryzae* (29–31). Effector genes in the pathogen *Leptosphaeria maculans* on *Brassica* undergo rapid evolutionary and epigenetic changes near TEs (32). The wheat pathogen *Pyrenophora tritici-repentis* shows higher TE content in pathogenic isolates compared to nonpathogenic isolates, with evidence of TE-mediated effector diversification and movement of virulence factors (33–35). The economically important wheat pathogen *Z. tritici* carries one of the most dynamic fungal genomes with a TE content varying from 16% to 24% among isolates (36). Some TEs share a similar epigenetic niche in the genome as effector genes. Through this co-localization, some TEs and effector genes experience concurrent de-repression during plant infection (37). Silenced TEs near an effector are linked to reduced expression and higher damage on a specific wheat cultivar (37–39). TE insertions upstream of the gene encoding the transcription factor Zmr1 regulate melanin production, a pigment required for pathogen survival during stressful conditions (40). The insertion of a retrotransposon into the promoter of a gene encoding a transporter increases fungicide efflux and contributes to multi-drug resistance (41). Similarly, TE insertions are linked to the adaptive deletion of a gene encoding an effector, which is likely recognized by the plant host (42). The activity of TEs is generally high in the pathogen and varies with geography (3, 36, 43, 44). This is partially explained by

the recent loss of genomic defenses against TEs with DNA methylation deactivated at the origin of the species and RIP likely losing efficacy during the colonization of new continents (43, 45, 46).

In this study, we interrogated large transcriptomic, genomic, and phenotypic data sets of a *Z. tritici* population to identify associations between transcriptionally active TEs and adaptive trait variation. We further analyzed variation in TE transcription among individuals and quantified the strength of genomic defenses against TE loci and repressive epigenetic marks. Finally, to capture TE-driven phenotypic variation in the pathogen, we associated metabolite production variation and pathogenicity-related traits of the fungal population with TE insertion polymorphisms.

RESULTS

A diverse pool of active TEs in the genome of *Z. tritici*

The TEs in the *Z. tritici* genome were previously identified by screening copies of consensus sequences in a panel of 19 reference-quality genomes to maximize the discovery of low-copy TEs (36). Consensus sequences were defined for TE copies with an identity >80% and >80% length covered. TEs were classified based on protein functions encoded by open reading frames. Short non-autonomous TEs were classified separately (36). We assessed the strength of selective constraints against TEs by analyzing the distribution of TE families relative to gene elements (Table S1 at <https://doi.org/10.5281/zenodo.10020096>). Based on the telomere-to-telomere assembled reference genome IPO323, we found that ~10% ($n = 398$) of TEs in the genome were located within 1 kb upstream of coding regions likely overlapping with regulatory elements (Fig. 1A). Similarly, exons (6%), 3' untranslated regions (UTRs, 6%), and downstream regions (<1 kb, 10%) also carried high proportions of TEs near genes. Introns and 5' UTRs showed low TE counts (1% and 2%, respectively; Fig. 1A). Across gene elements, only few DNA transposons inserted into exons, while retrotransposons made up the highest proportion (Fig. 1B). To investigate the distribution of major TE families across gene elements, we focused on the most abundant TE families (≥ 25 copies in the genome) including five DNA transposons, three retrotransposons, and two unassigned TE families (Fig. 1C and D). We found that 56% of the copies of the DNA transposon *Harbinger* (DTH) *Donna* were integrated into introns. The tight association of *Donna* with some intron sequences explains previous findings of identical intron sequences in unrelated genes (47). Similarly, more than half of the copies of the DTT_Lise (Tc1-*Mariner* superfamily) and a *Ty1/Copia* element (RLC *Deimos*) were integrated into upstream regions of genes (Fig. 1C). TE families *Gliese*, the MITEs *Troll*, and *Goblin* showed the highest insertion proportions in 3' UTRs. We next analyzed associations of TE insertions and gene expression levels under nutrient-limited conditions. TEs inserted most frequently close to the 10% lowest expressed genes and the insertion events were dominated by retrotransposons (Fig. 1E). *Ty3/mdg-4* retrotransposons and unassigned TEs show the strongest skew toward integration close to lowest expressed genes (Fig. 1F; Fig. S1). On the contrary, DNA transposons are less depleted near strongly expressed genes. Overall, the TE distribution in the genome is strongly correlated with the transcriptional landscape of genes.

TE insertion dynamics in a field population

TEs in the species show recent activity and contribute to genome-wide polymorphism (44). We examined TE-generated polymorphism based on TIPs for a deeply sampled single field population of genetically diverse isolates (Table S2 at <https://doi.org/10.5281/zenodo.10020096>). We used *ngs_te_mapper* to scan 139 whole-genome sequencing data sets individually aligned to the reference genome. TEs absent in the reference genome were rare in the population ("non-reference TEs"; on average carried by 5% of all isolates) in contrast to TEs with a copy present in the reference genome ("reference TEs"; Fig. 2A). This is consistent with the expectation that TEs present in the reference genome are more likely to be at a higher frequency among other isolates as well. DTX is the most

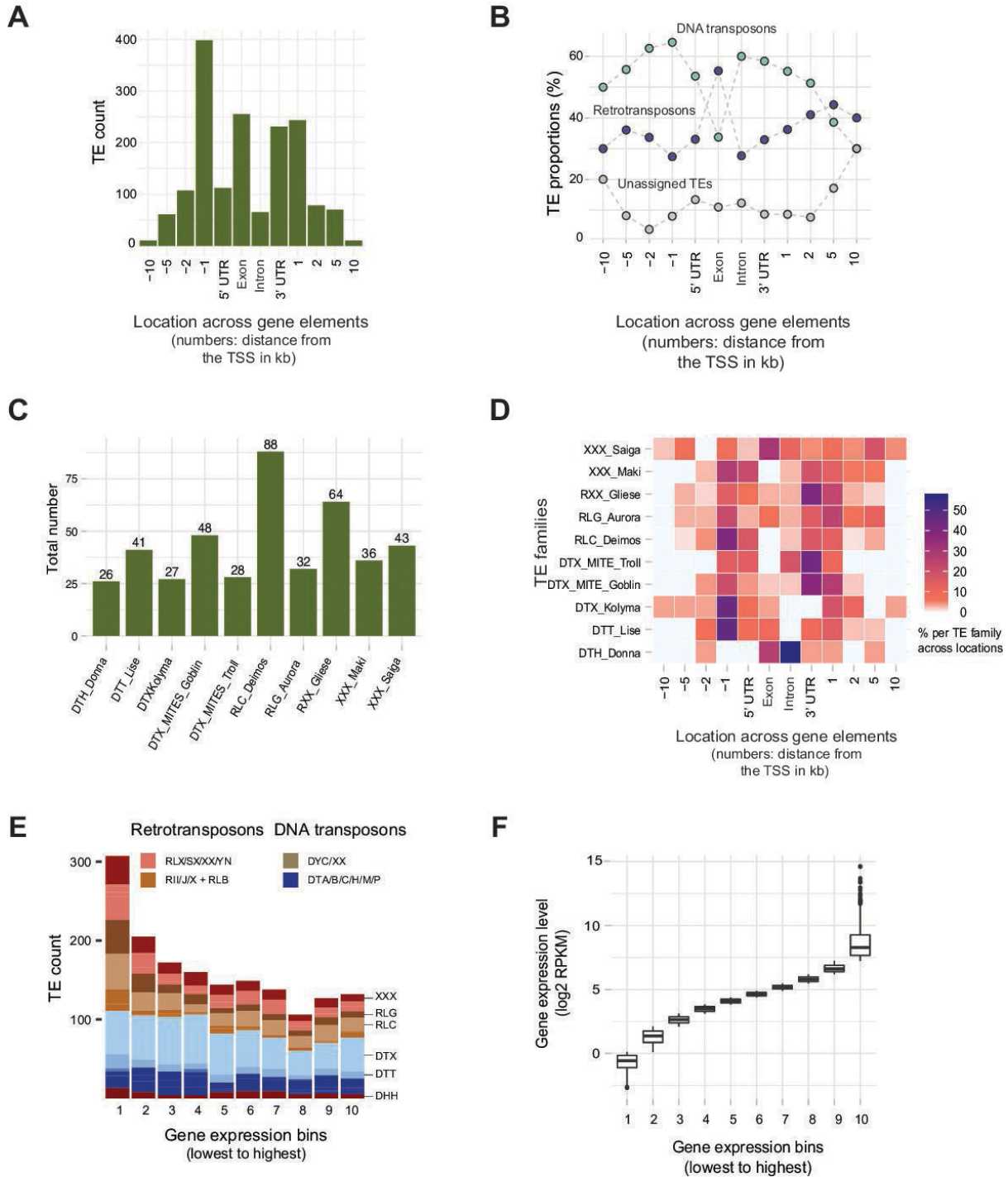


FIG 1 Distribution of transposable elements (TEs) in gene elements in the reference genome of *Zymoseptoria tritici*. (A) Genomic localization of TEs ($n = 1,640$) in gene elements and 10 kb windows upstream and downstream of the transcription start site (TSS). (B) Distribution of DNA and retrotransposons across gene elements and 10 kb windows upstream and downstream of the gene (887 DNA transposons, 597 retrotransposons, 156 unclassified elements). The dotted lines represent the distribution of TE classes. Exons and introns are grouped into a single category per gene. (C) A total number of major TE families (with ≥ 25 copies in the genome) inserted into gene elements. (D) Distribution of the 10 most abundant TE families in gene elements and 10 kb windows upstream and downstream of TSS. Percentages refer to the portion of element copies located in particular gene elements. (E) Count of TEs localized close to genes categorized according to 10 bins of gene expression. The weakest expressed genes co-localized with the highest number of TEs. (F) Variation in gene expression within each of the 10 gene expression bins.

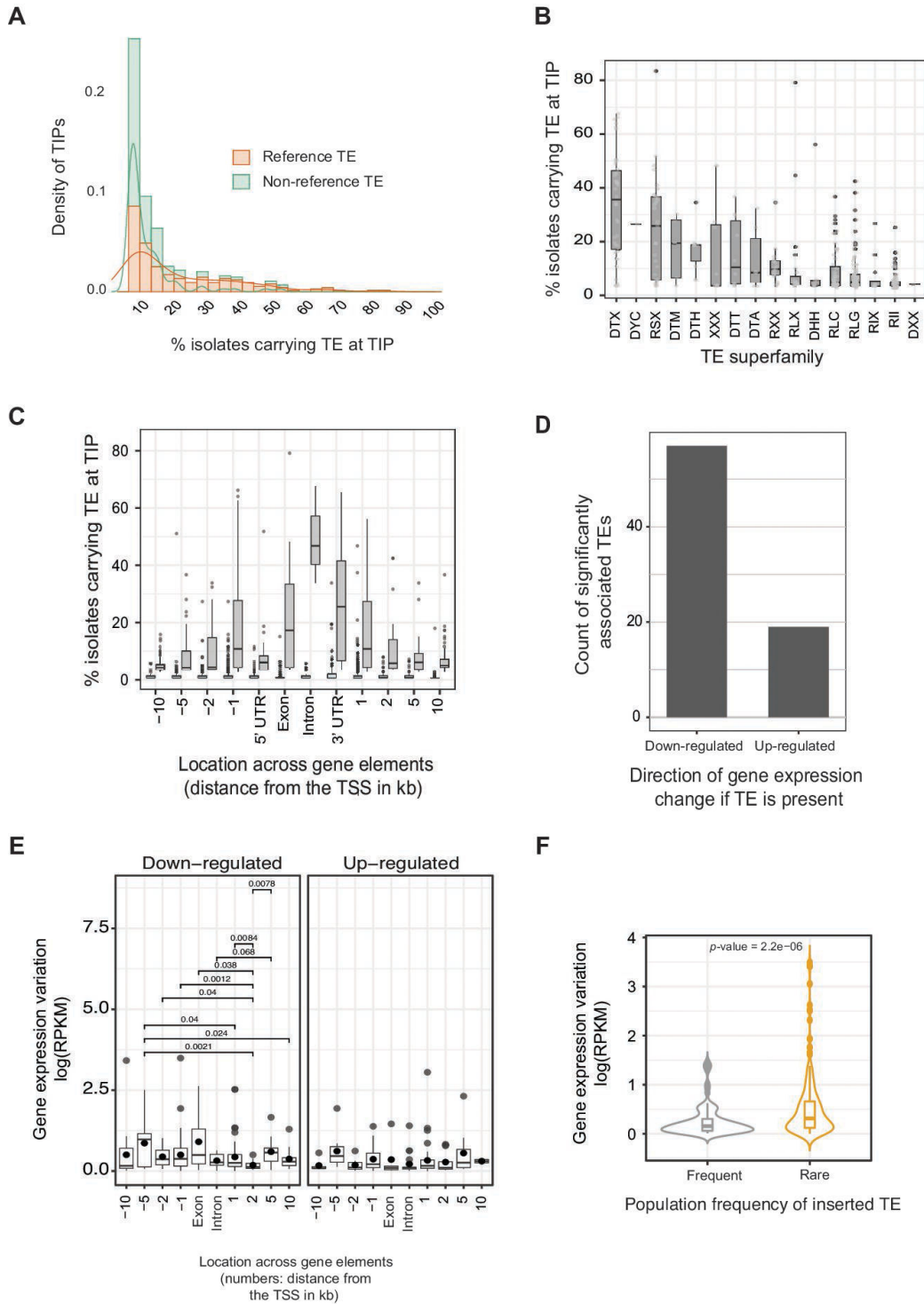


FIG 2 TE insertion polymorphisms (TIPs) in a single field population and association with gene expression variation. (A) Frequency of TEs present in the reference genome (“reference TEs”) or only detected in other isolates (“non-reference TEs”). A total of 3,662 non-reference TEs and 446 reference TEs were analyzed. (B) The frequency of the inserted TE at TIP loci is expressed as the percentage of the isolates carrying the TE. TIPs were binned by TE superfamilies for visualization. (C) The frequency of the inserted TE at TIP loci is expressed as the percentage of the isolates carrying the TE binned by location across gene elements (± 10 kb window). (D) Count of significantly associated TIPs with neighboring gene expression variation. The classification is based on the up- or downregulation of the neighboring gene if the TE is present. (E) Association of inserted TEs at TIPs with the down- or upregulation of the neighboring gene shown for TIPs located in different gene elements. Significant (Continued on next page)

FIG 2 (Continued)

associations are shown with the respective *P*-value. (F) Gene expression variation of genes with a neighboring TIP according to the frequency of the inserted TE. TIPs with rare TE presence (<20 isolates; <15%) are significantly associated with genes showing higher gene expression variation. The statistical test is robust to the exclusion of outliers (values > 1.5).

frequent TE superfamily showing TIPs in the population followed by SINE (RSX) and mutator elements (DTM; Fig. 2B). TIPs inserted into introns had the highest population frequency compared to TEs inserted in other gene elements (Fig. 2C). For example, 46% of isolates contained intron insertions of the TIR DTX elements. Furthermore, 79% of isolates from the population carried LTR elements (RLX) in exons (Fig. S2). Given the high numbers of TIPs near coding sequences, we analyzed associations of TIPs with gene expression (Table S3 at <https://doi.org/10.5281/zenodo.10020096>). We analyzed genes with TIPs showing a minor allele frequency >5% and being located within 10 kb of the gene. We compared transcript levels between isolates with and without TE insertions for a total of 354 TIPs. Approximately 21.4% ($n = 76$) of the TIPs in the population showed a significant association with the expression of a neighboring gene (*P*-value < 0.05). Overall, the majority of TE insertions near a gene tend to reduce the gene transcription levels (Fig. 2D). Interestingly, the distance at which the TIP is located compared to the gene with associated gene expression variation seems not to be an important factor determining the extent of gene expression variation (Fig. 2E). Next, we analyzed the population frequency of the inserted TEs and gene expression variation among isolates. We found that rare TE insertions at the population level were significantly associated with higher expression variation of neighboring genes (Fig. 2F) and the association is robust to the removal of outliers with log (RPKM) >1.5. In summary, the genomic landscape of polymorphic TE insertions reflects gene expression variation within the pathogen population.

Locus-specific transcription of TEs

Transcription of TEs can lead to transposition and copy number increases depending on the type of TE. To identify the pool of transcribed TEs from each TE family, we analyzed poly-A tail enriched RNA-seq data. We generated locus-specific expression estimates based on uniquely mapping RNA-sequencing reads to identify variation in transcriptionally active copies of the same TE family (Table S4 at <https://doi.org/10.5281/zenodo.10020096>). Reads mapping to multiple loci of the same TE family were distributed proportionally across loci according to the algorithm to account for relative differences in transcriptional activity among TE families. We found that 43% of all TE copies per genome show evidence of transcription and substantial variation among individual TE families and superfamilies (Fig. 3A; Fig. S3). We observed the highest transcription levels for LTRs including RLB_BEL1, RLX_LARD, and RLG *Ty3/mdg-4* (Fig. 3B). Although the number of TE copies significantly differs among chromosomes, we found no significant variation in transcription (Fig. S4). Sequence similarity of recently transposed TE copies can be a challenge for the unique mapping of sequencing reads to individual loci. Hence, locus-specific analyses are likely biased against the youngest TEs (identified by the presence of near-identical copies) in the genome. To assess this risk, we reconstructed the phylogenetic tree of an MITE family using sequences of individual TE copies in the genome. We assessed transcript abundance for each locus in association with the terminal branch length of the individual copy. In general, short branch lengths indicate rapid expansion of copies. We found no meaningful association between the sequence similarity of copies of a TE family and transcript abundance suggesting that expression quantification introduces no major bias against young TEs as illustrated for a MITE family (Fig. 3C) and an RLX_Lard_Gridr (Fig. S5). Overall, transcription of MITE copies near genes tends to be positively correlated with the expression of genes, in particular at distances of 0–2 kb from the TSS (Fig. 3D) suggesting shared epigenetic effects.

The genomic context and the identity of individual TEs are potentially significant factors explaining genome-wide TE transcription and insertion activity. First, we found

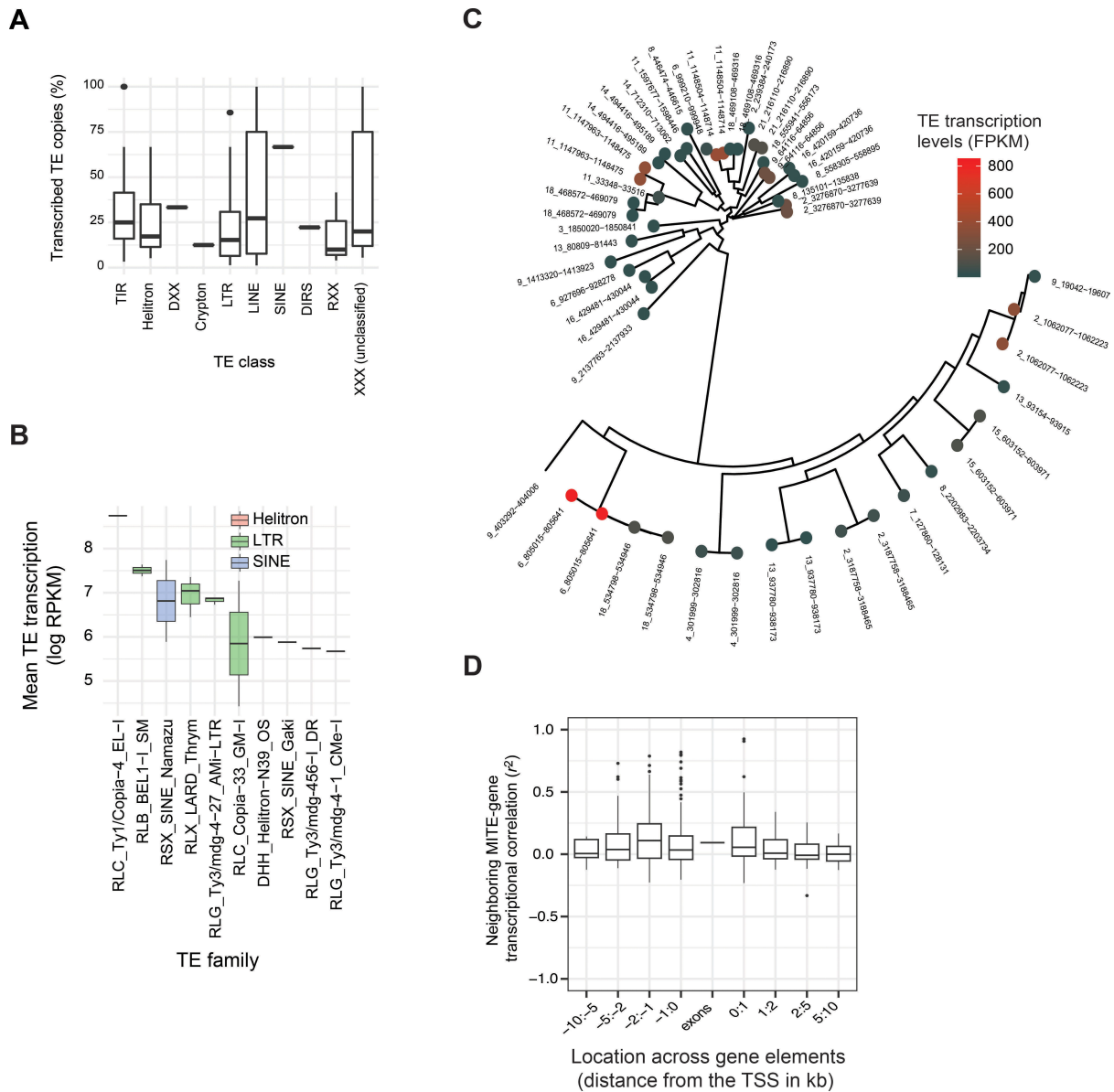


FIG 3 Population-level variation of locus-specific transcription of TEs. (A) Percentage of transcribed TE copies per TE family grouped by TE superfamily. (B) Transcription levels of individual copies of the ten most transcribed TE families in the genome. (C) Neighbor-joining tree of all copies of the MITE *Undine* family discovered in the IPO323 reference genome. Colors indicate transcription levels for each copy. (D) Correlation of transcription between MITE copies and neighboring genes. The correlation values are summarized based on the distance of the MITE from the TSS.

that the highest number of TEs were located within 1 kb upstream of the gene transcription start sites (Fig. 4A). We also found that TE copies further away from the TSS tend to show higher transcriptional variation among isolates compared to TE copies < 1 kb from genes but at comparably transcription levels (Fig. 4A). Epigenetic factors such as histone methylation marks including H3K4me2, H3K9me2, H3K9me3, and H3K27me3 can have important effects on gene and TE transcription (48, 49). In addition, repeat-induced point (RIP) mutations, and GC content of TE loci and their association with transcription can be important indicators of the activity of TEs (24, 50). We found a negative correlation between TE transcription levels and the phylogenetic distance among TE copies of the same family (i.e., branch length) as well as counts of RIP mutations. This is consistent with the action of RIP as a genomic defense against TEs. Consistent with expectations, euchromatic histone methylation marks such as H3K4m3 were positively correlated

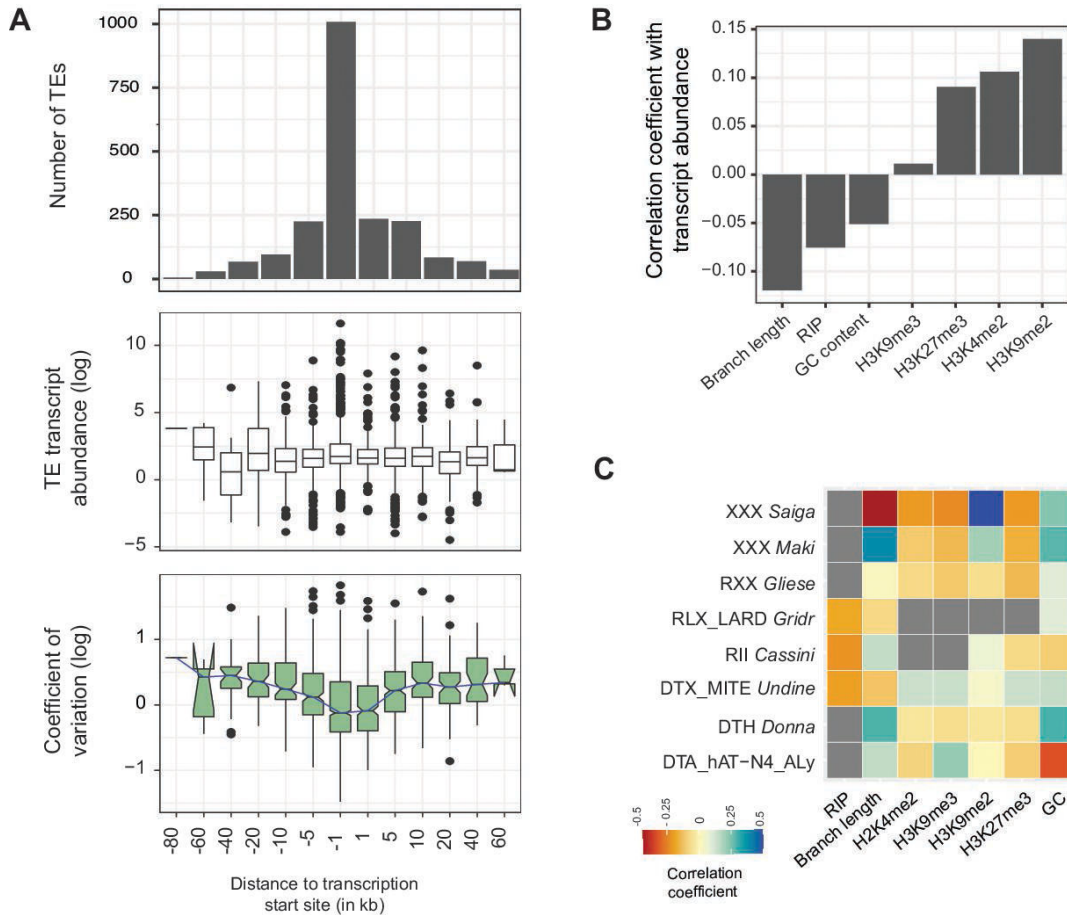


FIG 4 Transcriptional landscape of TEs across the genome. (A) Number of TEs, transcript abundance, and coefficient of variation of TEs localized upstream and downstream of the TSS of a gene (number of TEs = 2083). (B) Correlation of transcript abundance detected at individual TE loci and branch length of the TE copy compared to other TE copies of the same family, the number of RIP mutations, GC content, and different histone methylation marks (assessed for $n = 455$ TE loci). (C) Analyses of the eight most transcribed TEs (>25 transcribed copies each). Correlation matrix of transcription of individual TE loci and genomic features as shown in (B).

with the transcription of TEs. Surprisingly, also repressive histone methylation marks (H3K27m3 and H3K9m2) were positively correlated with TE transcription (Fig. 4B).

Analysis of correlates among genomic features and TE transcription at the level of individual TE families revealed a more complex landscape of associations compared to the global level of all TEs combined (Fig. 4C). The non-autonomous TE family DTX MITE *Undine* showed a negative correlation between transcription and the phylogenetic distance of the TE copy (using terminal branch lengths) and a slightly positive correlation with both repressive and euchromatic histone methylation mark (Fig. 4C). The uncharacterized TE *Saiga* showed both a strong negative correlation with branch length of the copy and a strong positive with GC content and repressive histone methylation marks. We further investigated interactions of the genomic environment with TE transcriptional activity. We found a positive relationship between the proportion of transcribed copies of a TE in a genome with the percentage of isolates carrying expressed TE copies. For instance, among DNA transposons, the MITE *Undine* had more than 50% of the copies expressed in the reference genome and more than 75% of the *Undine* copies were transcribed among isolates. For the DTA *hAT* element, less than 30% of the copies were expressed in the genome as well as among isolates (Fig. S6; Table S5 at <https://doi.org/10.5281/zenodo.10020096>).

Enrichment of repeat-induced polymorphisms in regulatory regions of TEs

We used genome-wide association mapping to systematically identify single nucleotide polymorphisms (SNPs) associated with variation in TE transcription among isolates. We associated SNPs within a 5 kb distance both upstream and downstream of the TE with the transcriptional variation of the TE in the population. We found at least one significantly associated SNP in proximity to 22 TE loci with nearly all significant SNPs located within 2 kb of the TE (Fig. 5A). Regression slopes for transcriptional variation association mapping are positive if the alternative allele is associated with higher transcription. A negative slope stems from higher transcription associated with the reference genome allele. We found no meaningful effect of reference vs alternative SNP allele or distance of the SNP to the TE (Fig. 5B). Next, we assessed whether there was an enrichment for RIP-like mutations among TE transcription-associated polymorphisms. In *Z. tritici*, RIP induces C→T transitions at CpA sites (45). We analyzed all SNPs in the genome associated either with TE transcription or gene transcription (51). Then, we assessed whether the polymorphism is likely RIP associated. We found that for TE transcription-associated SNPs, the odds ratio of the SNP being caused by RIP was 2.70 (CI 0.52–13.98; Fig. 5C). In contrast for gene transcription-associated SNPs, the odds ratio of the SNP being caused by RIP was only 1.03 (CI 0.93–1.14). Given the overlapping confidence intervals, it remains

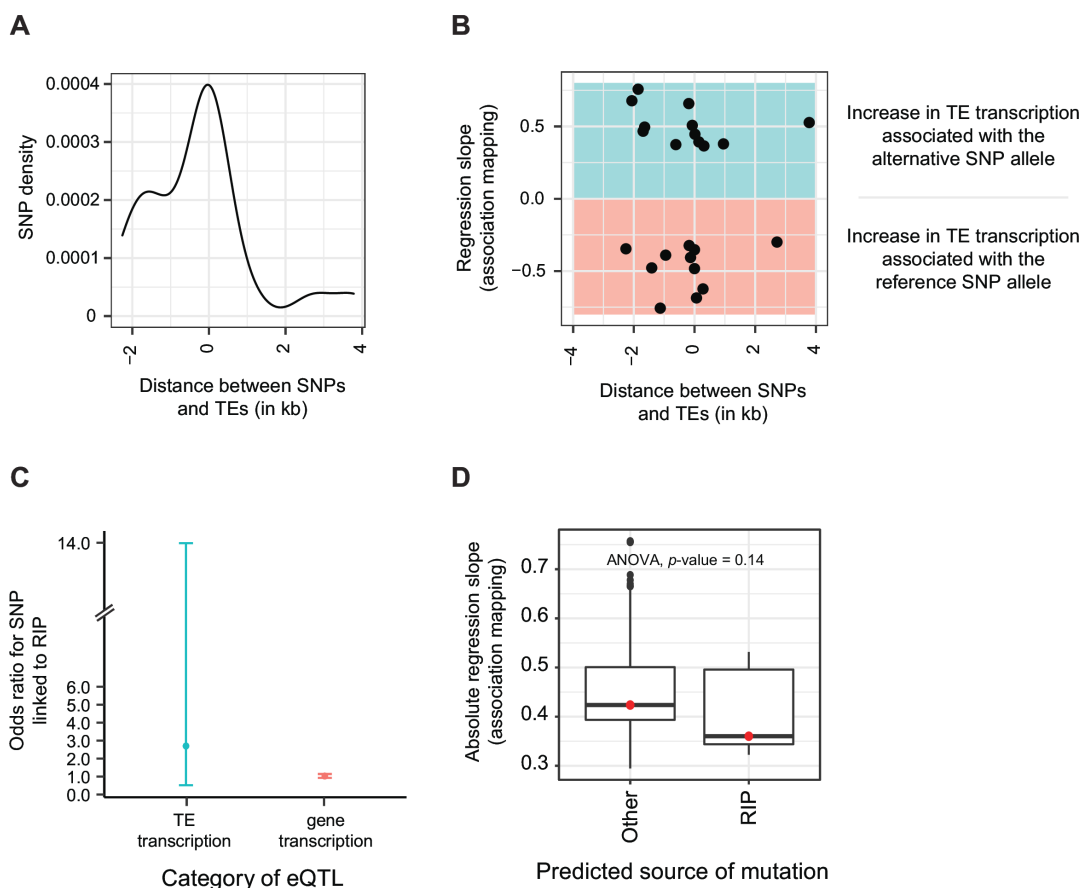


FIG 5 Regulatory variants associated with TE transcription. (A) Distribution of SNPs associated with TE transcription variation among isolates. The majority of the significant SNPs are located within a 2 kb distance upstream and downstream of TE loci ($n = 22$). (B) Regression slope (i.e., effect size) of significantly associated SNPs with TE transcription variation (for $n = 22$ associated SNPs). The colors represent negative and positive regression slopes highlighting either a positive expression association with the reference or an alternative allele at the SNP. (C) Enrichment of RIP-like mutations in SNPs associated with TE transcription variation compared to SNPs associated with gene expression variation. An odds ratio >1 indicates an enrichment of RIP-like mutations and 95% confidence intervals are shown. A total of 197 SNPs were analyzed for an association with TE transcription variation and 283,867 SNPs were analyzed for an association with gene expression variation. (D) Absolute regression slope (i.e., effect size) of RIP-like vs other SNPs associated with TE transcription variation (for $n = 22$ associated SNPs).

unclear whether RIP plays a more pronounced role in generating SNPs governing TE transcription variation rather than gene transcription. Finally, we analyzed whether the strength of the TE transcriptional association expressed by the regression slope differs depending on whether the SNP was likely caused by RIP or not. We found no significant differences in regression slopes depending on the source of the mutation (Fig. 5D).

Phenotypic variation associated with recent TE insertions in a population

Insertions caused by TEs can have consequences for the expression of phenotypic traits. However, a range of additional genetic variants are likely to impact trait variation as well. To assess the likelihood of TE loci influencing trait expression independently of other polymorphisms, we analyzed patterns of linkage disequilibrium (LD) between TIPs and nearby SNPs. We found that most TIPs were on average in low LD with SNPs within a 600 bp window compared to LD between pairs of SNPs in the same window (Fig. 6A; Table S6 at <https://doi.org/10.5281/zenodo.10020096>). However, SNPs and TIPs tended to be at a longer distance within the window compared to SNP-SNP distances. This phenomenon is likely to explain the low LD observed between SNPs and TIPs (Fig. S7). To identify associations of TIPs with phenotypic trait variation, we analyzed a series of trait data sets generated for the same population including pathogenicity trait measurements and variation in metabolite production (Table S7 at <https://doi.org/10.5281/zenodo.10020096>). Wheat leaf lesion area produced by individual isolates showed an association significant at the 10% false discovery rate (FDR) threshold with a TIP caused by a 6 kb *Ty1/Copia* element (RLC *Deimos*) located on chromosome 9 (Fig. 6B). The associated TIP is within ~5 kb of a gene encoding a cell wall-degrading enzyme (i.e., glycosyl-hydrolase family 47) secreted during host infection. We found no significant difference in the expression of the gene encoding the cell wall-degrading enzyme between isolates carrying the *Ty1/Copia* element or not (*t*-test *P*-value = 0.44; Fig. 6C).

Given the large scope for TE insertions associated with gene transcription variation, we screened the population for additional phenotypic readouts. We analyzed a data set of metabolite production profiles of each isolate assessed under culture conditions (Table S8 at <https://doi.org/10.5281/zenodo.10020096>) (52). We found that 65% of all TIPs were significantly associated with variation in intensity in at least one metabolite (Table S9 at <https://doi.org/10.5281/zenodo.10020096>). Out of the significantly associated TIPs, 13 were significantly associated with at least 20 different metabolites (Fig. S8). We found that TIPs located within ~1 kb upstream of the TSS of genes have a stronger association with variation in metabolite production than TIPs further away from the TSS (Fig. 6D and E). The enrichment of significant TIPs within coding regions and near TSS is consistent with the observations from a metabolome GWAS (52) performed on the same study population using genome-wide SNPs. Overall, 26% of all significant associations with individual metabolite production variation using TIP-GWAS were not previously identified using SNPs only (Fig. 6F).

Based on the metabolite-TIP GWAS, we identified a significant association for a DTX MITE insertion with the metabolite profile *m/z* intensity 245.1866. The associated MITE insertion is in the intergenic region of a polyketide synthase (PKS) secondary metabolite gene cluster located on chromosome 10 (position 427,490–477,261 bp) associated with higher metabolite production (Fig. 7A). The insertion occurred upstream of the gene encoding a NmrA-like family protein (Zt09_10_00163). NmrA is a negative transcriptional regulator involved in nitrogen metabolite repression in pathogenic fungi (53). We found that the MITE insertion was significantly associated with lower expression of the NmrA-like family protein (Fig. 7B and C).

DISCUSSION

We performed a comprehensive analysis of transcriptional variation generated by individual TEs in a pathogen population and consequences for phenotypic trait evolution. Recent TE insertions generated presence-absence variation among isolates (i.e.,

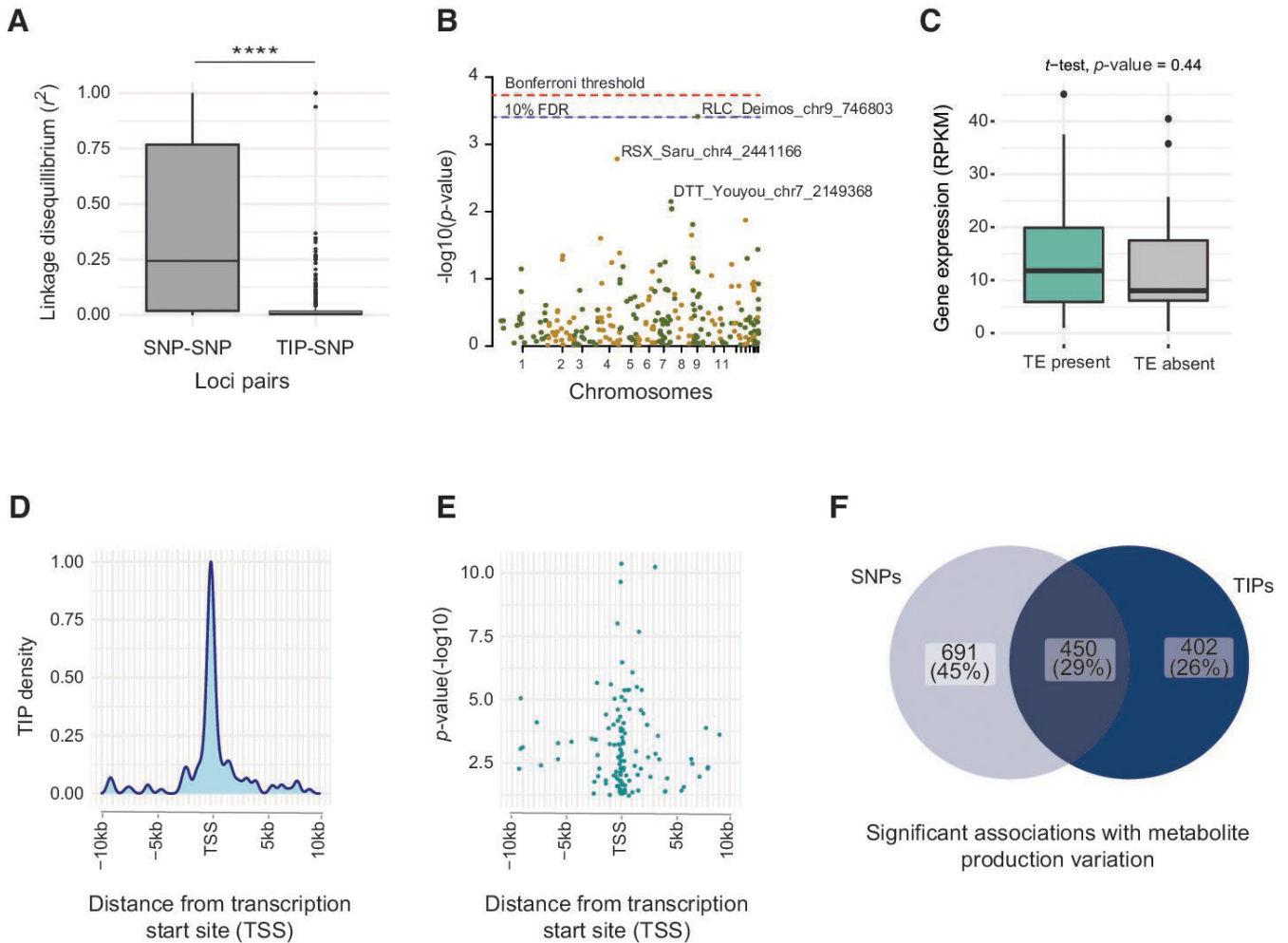


FIG 6 Phenotypic-genotype associations at TIP loci. (A) Linkage disequilibrium (LD) between TIPs and SNPs in 5 kb windows as well LD of SNP-SNP pairs in the same windows (number of SNPs in analyzed TIP windows = 12,832). (B) Manhattan plot for TIP association mapping with virulence (measured by the percent leaf area covered by lesion). Colors refer to TIPs located on distinct chromosomes. (C) Gene expression variation of a gene encoding a cell wall-degrading enzyme according to the presence or not of a TE at the TIP (t -test, P -value = 0.44). (D) Distribution of TIPs ($n = 1213$) is significantly associated with variation in metabolome peak intensities near transcription start sites (TSS) of genes. (E) P -value distribution of TIPs associated with metabolome peak intensities near the TSS of genes. The most significantly associated TIPs are located closest to the TSS. (F) Venn diagram representing the overlap of significant association from TIP metabolome GWAS and SNP-based metabolome GWAS.

TIPs) with selection likely playing an important role in filtering new insertions. Beyond the single field population, the species contains multiple, highly active TEs with diverse effects on trait expression, epigenetic variation, and even genome size (37, 44, 54). We identified considerable intraspecific variation in the transcription of individual TE copies emphasizing the importance of the genomic background for controlling TE transcription. Finally, TE insertions likely influence the gene expression of neighboring genes and the expression of traits including metabolites, which can mediate organismal interactions.

We found approximately a fifth of all TE copies in the genome with transcriptional activity. This indicates significant potential for TE mobility depending on the mode of proliferation. Assessing locus-specific transcription is challenging though given the repetitive nature of TEs. Our approach was based on short RNA-seq reads uniquely mapped to specific TE loci in the genome (55). However, not all TE-derived RNA-seq reads can be assigned using this restrictive approach. All non-uniquely mapped TE-derived reads are distributed proportionally among TE loci of the same family. This latter step enables accurate quantification of TE transcription among TE families at the detriment of

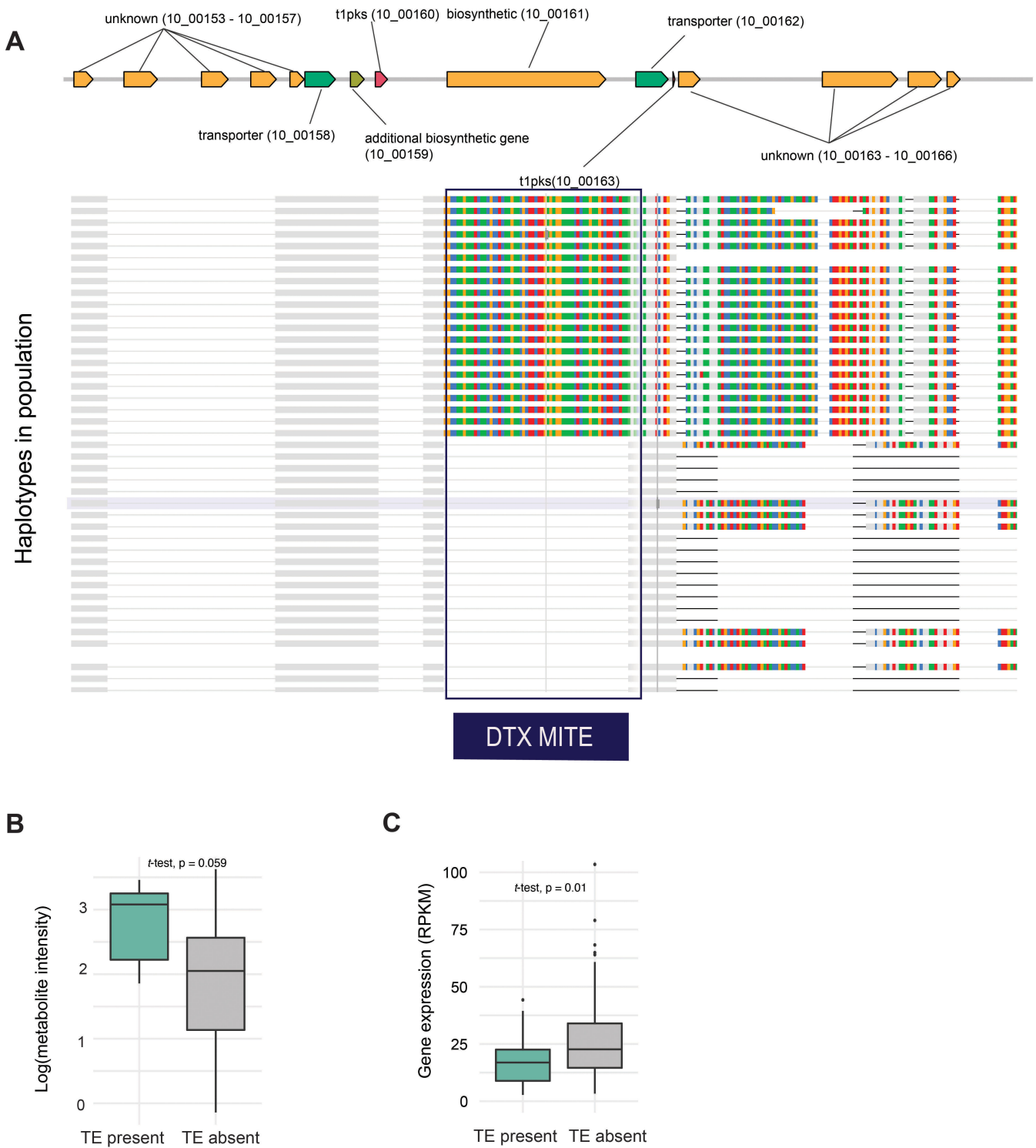


FIG 7 Genomic context of the MITE insertion in a secondary metabolite gene cluster. (A) The top drawing shows the genomic organization of the PKS1 secondary metabolite gene cluster. Genes are labeled with functional annotation and genomic location. The multiple sequence alignment highlights the conserved MITE insertion shared among isolates (i.e., haplotypes). (B) Association of metabolite production intensity with the MITE presence/absence variation. (C) Association of the *NmrA*-like gene expression variation associated with the MITE presence/absence variation. Number of isolates with TE present = 17 and isolates with TE absent = 129.

accurate TE quantification among TE loci of the same TE family. Long-read transcriptome sequence (e.g., Iso-Seq) largely solves the challenge of unique read mapping; however,

such approaches come at the expense of coverage and, hence, quantification accuracy. The compact genome of *Z. tritici* (~40 Mb) and comparatively low TE content (16%–24%) facilitates accurate read mapping compared to large animal genomes for which the approach was initially developed (55).

A high number of TE insertions are not fixed within the population suggesting ongoing activity creating new copies. However, some segregating TE insertions may be old, and estimating the age of an insertion is difficult without access to allelic variants of an inserted TE and surrounding sequences. If purifying selection is strong though, TE insertions should remain at low frequency and tend to be young as shown by a global TE analysis of *Z. tritici* populations (3). Consistent with strong selection, TE insertions were more likely to occur in non-coding regions than exon or intron sequences carrying a high likelihood of deleterious effects. We also found that the identity of TEs is an important factor explaining their distribution across gene elements with, for example, some TEs nearly exclusively associated with intronic insertions. Some introns in *Z. tritici* were identified as non-autonomous TEs because unrelated genes across the genome harbor nearly identical intron sequences (47, 56). Overall, the species harbors dozens of polymorphisms for the presence-absence of entire intron sequences consistent with the recent mobility of intronic sequences (47). Retrotransposons were overrepresented in exonic sequences compared to DNA transposons. This may be explained at least partially by gene models including open reading frames of TEs or chimeric genes. TE frequencies at TIPs tend to be low except the intronic TEs, which are most likely domesticated with no or only minor deleterious effects. Overall, the TE landscape of a single population likely largely reflects insertion constraints and counter-selection against TE insertions.

Active transcription of a TE can generate new TE insertions in the genome depending on the mechanism of propagation, as well as impact the structure and function of the genome (15, 57). Selection acting against deleterious TE insertions is complemented by specific genomic defense mechanisms that evolved to counteract TE activity in fungal genomes. These mechanisms include RIP mutations acting against duplicated sequences and epigenetic silencing reducing the transcriptional activity of TEs and accessibility of DNA (14, 21, 24). We found mostly low to no transcription of TE copies affected by RIP suggesting that the defense mechanism effectively silences TEs similar to *N. crassa* (50). The most abundant TE family (i.e., the MITE *Undine*) showed the highest transcriptional variation for individual loci in the population. The same TE is also highly upregulated during plant infection and has undergone a massive recent expansion (37, 44). Hence, variation in MITE transcription near coding sequences likely reflects epigenetic variation. As MITEs are non-autonomous elements relying on transposition functions provided by full-length TEs elsewhere in the genome, transcription of MITEs is most likely uninformative about their potential to spread in the genome. Recapitulating MITE mobilization would require analyses at the population level to match MITE sequence characteristics (i.e., terminal sequences or target site duplications) to full-length TEs potentially serving as vehicles for MITE transposition. The weak effect of genomic defenses against MITEs is also likely related to the short length (~53 bp), which renders RIP-based defenses ineffective (20, 37). The heterogeneous effects of RIP against different classes of TEs are compounded by the relaxation of RIP genomic defenses within the species following the expansion from the center of origin in the Middle East (43, 45, 58).

Transcription and distribution in the genome were highly uneven among TE families consistent with differences in recent activity and insertion preferences matching evidence from other kingdoms (59). The *Mutator* elements of maize typically integrate into gene-rich regions (60). In *S. cerevisiae*, *Ty1/Copia* integrates preferentially upstream of genes transcribed by the RNA polymerase III (Pol III) (61, 62). *Ty1* is also enriched in heterochromatic regions triggered by the recognition of heterochromatin during integration and then perpetuates the heterochromatic mark by triggering epigenetic modifications at new insertion loci (63). The *P* elements in *Drosophila* preferentially integrate at replication origins in the genome to support transposition (64). We also find that TE insertions are underrepresented near highly expressed genes similar to

the observation in *D. nasuta* (65). Those observations may however be the result of a TE survivor bias given the typically slow evolutionary rates of highly expressed genes given the strong purifying selection (65, 66). Under some conditions, the spread of heterochromatin from TE insertion sites can induce epigenetic silencing of neighboring genes (14). Consistent with this possibility, we found that TE insertions are significantly associated with expression variation of neighboring genes. TE insertion tends to be associated with lower gene transcription levels consistent with previous analyses of gene expression analyses over the course of plant infections (37). We found surprising positive associations between repressive heterochromatin and increases in TE transcription. Without access to histone methylation data at the population level, interpreting potential mechanistic links between the two factors is challenging. It is important to note also that histone methylation data were collected in less stressful environmental conditions than the TE expression quantification. Based on environmental cues and stress induction, TEs show distinct de-repression patterns throughout an infection with correlated responses of genes in proximity (67). Our findings expand our understanding of the coordinated expression of TEs and adjacent genes by showing that these patterns can be driven by polymorphic TE insertions.

Crucial for the adaptive evolution of microbial species, TE insertion activity in genomes can drive phenotypic trait evolution through their impact on gene structure and regulation. We showed that variation in pathogenicity-related traits and metabolite production is most likely underpinned by recent TE insertion activity in functionally important regions of the genome. Our TIP-GWAS showed that two-thirds of all TIPs showed a significant association with at least one metabolite intensity profile. The specific biological roles of the different metabolites are poorly understood with few exceptions (52, 68). However, individual metabolites can play major roles in species interactions with competitors or hosts (40, 69). Similarly, TE insertions in crop plants are associated with a wide range of agronomic traits and secondary metabolites such as in tomatoes (4, 70). Metabolite production-associated TIPs in *Z. tritici* showed enrichment in coding regions and transcription start sites consistent with these regions having a higher potential for functional consequences. Given the large number of TE insertions associated across many genes for metabolite production provides a vast potential for rapid evolution of the species from standing variation in single field populations. TE-driven adaptation within species and populations is likely to proceed at a more rapid pace given the potential for stronger phenotypic variation among genotypes at TIPs. Consistent with individual findings across plants, animals, and fungal species, active TEs in the genome can underpin the most recent adaptive evolution of large effect size. The power of TEs to drive evolutionary change stems from their potential to affect the expression of multiple nearby genes, modulate their response to external stress but also to rearrange large sequence segments including the deletion and duplication of gene sets. Recent discoveries of large TEs with the ability to conjugate dozens of genes and hundreds of kilobases of sequences (71) exemplify the driving force transposable elements play underpinning the rapid evolution of microbial species.

MATERIALS AND METHODS

Fungal isolates and sequencing

We analyzed *Z. tritici* isolates collected from a wheat field in Eschikon, Switzerland. Wheat cultivars were planted in a randomized block design and infections occurred naturally from local inoculum or adjacent fields (72). To isolate strains, individual cirri from pycnidia were identified on infected leaves and plated on a yeast sucrose broth (YSB) solid media plate with kanamycin 50 $\mu\text{g mL}^{-1}$ and incubated at 18 °C. After a week, a single colony from each plate was inoculated in a yeast-sucrose broth (YSB) and incubated on a shaking incubator at 18°C for 8 days at 140–180 rpm (72). Total genomic DNA was extracted from 139 *Z. tritici* isolates YSB cultures using the QIAGEN DNAeasy Plant Mini Kit. Illumina libraries were prepared using the TruSeq Nano DNA

Library Prep kit. Sequencing was performed in 100 bp paired-end mode on a HiSeq 4000 at the iGE3 sequencing platform (Geneva, Switzerland). Raw reads are available on the NCBI Short Read Archive under the BioProject PRJNA596434. For RNA sequencing, the same isolates (146 isolates were cultured in a Vogel's Medium N (Minimal) modified as ammonium nitrate replaced with potassium nitrate and ammonium phosphate (73) without sucrose and agarose to induce hyphal growth. Total RNA was isolated from the filtered mycelium after 10–15 days using the NucleoSpin RNA Plant and Fungi kit. The RNA concentrations and integrity were checked using a Qubit 2.0 Fluorometer and an Agilent 4200 TapeStation system, respectively. Only high-quality RNA (RIN >8) was used to prepare TruSeq-stranded mRNA libraries with a 150 bp insert size and including a poly-A enrichment step. Sequencing was performed on an Illumina HiSeq 4000 in 100 bp single-end mode. Raw reads for RNAseq are available on the NCBI Short Read Archive under the BioProject PRJNA650267 (52; Table S10 at <https://doi.org/10.5281/zenodo.10020096>).

Identification of TE insertion polymorphism in *Z. tritici*

Whole-genome Illumina sequencing reads were quality checked using FastQC [version 0.11.5 (74)] and trimmed with Trimmomatic version 0.36 (75) to remove adapter sequences and low-quality reads with parameters ILLUMINA-CLIP: TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDING WINDOW:4:15 MINLEN:36. To detect TE insertions, we used the R-based tool *ngs_te_mapper* version 79ef861f1d52cdd08eb2d51f145223fad0b2363c (76) integrated into the McClintock pipeline version 20cb912497394fabddcdaa175402adacf5130bd1 (77). Filtering and validation of the identified insertion polymorphisms followed previously established protocols for the same species (3). To confirm the presence of predicted non-reference TEs, we extracted the reads mapped near the predicted insertion site and assessed whether the target site duplication represented a break in the alignment as expected. Using spliced junction reads, we analyzed whether a gap was suggested in the region of the position indicating the absence of a reference genome TE copy in a particular isolate.

Locus-specific TE and gene expression analyses

Locus-specific expression profiles for TEs were generated using mapped RNA-seq data with the tool SQuIRE (v0.9.9.9a-beta) (55). TE annotation of the reference genome IPO323 was retrieved from (36). The SQuIRE “Map” mode was used to align RNA-seq data and SQuIRE “Count” with the `-EM` parameter to perform the estimation-maximization algorithm to quantify TE expression considering both uniquely mapped and multi-mapped reads typically found in repetitive sequences such as TEs. TE-derived reads normalized by fragments per kilobase of exon per million (FPKM) were filtered to keep the reads originating from the annotated strand direction in the IPO323 reference genome. Locus-specific expression analyses can be influenced by high sequence similarity among copies in recently expanded TE families. To assess the impacts of high sequence similarity among copies, we have generated a phylogenetic tree of the MITE family *Undine* using individual locus sequences. Branches were annotated with corresponding transcript abundance. We used neighbor-joining tree estimation (`njs`) in the R (78) package *ape* (version 5.5) (79) and the transcript abundance was plotted using the R package *ggtree* (version 3.0.4) (80). In addition, gene expression profiles were analyzed using QTLtools (version 1.1) (81). RNA-seq reads mapped to the IPO323 reference genome were analyzed based on transcript-trained, high-quality gene models (82). Read counts were summarized with QTLtools `--quan` mode. Only reads with a minimum Phred mapping quality >10 were kept for further analyses. Reads were normalized using the `--rpkm` (reads per kilobase of transcript per million reads mapped) from QTLtools.

Association mapping analyses for expression variation

SNPs used for the association mapping were retrieved from a variant calling procedure performed with raw sequencing reads checked by FastQC (version 0.11.5) (83) and trimmed with Trimmomatic (75) to remove adapter sequences and low-quality reads with parameters ILLUMINACLIP: TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDING WINDOW:4:15 MINLEN:36. Trimmed sequences were aligned to the *Z. tritici* IPO323 reference genome (see above) using bowtie2 version 2.3.4.3 (84) with the option `--very-sensitive-local`. Variant calling was performed using Haplotypecaller integrated into the Genome Analysis Toolkit (GATK) v. 4.0.11.0 (85). We retained SNPs with QUAL >1000, AN >20, QD >5.0, MQ >20.0, ReadPosRankSum_lower = 2.0, ReadPosRankSum_upper = 2.0, MQRankSum_lower = 2.0, MQRankSum_upper = 2.0, BaseQRankSum_lower = 2.0, BaseQRankSum_upper = 2. Variants passing quality filtration were further filtered to remove multiallelic SNPs using the *bcftools* (version 1.9) `--norm` option. Variants were filtered to keep only variants genotyped in at least 90% of the individuals and common variants $\geq 5\%$ using the *VCFtools* `--max-missing` and *bcftools* `-q 0.05` minor option (86, 87). Expression quantitative trait loci (eQTL) for variation in locus-specific TE expression were searched using QTLtools (version 1.1) (81) in the *cis* `--permutation` mode with 1,000 permutations and 5 kb *cis* windows surrounding each TE locus. The permutation *P*-values were false discovery rate (FDR) corrected to identify the top eQTL (5% threshold). We also analyzed gene expression variation associated with TIPs. For this, only genes within 5 kb of each TIP were tested for associations with gene expression levels. Fold change in gene expression was calculated from the ratio of mean gene expression between isolates with TE insertion and isolates without TE insertion. A Wilcoxon test was performed to assess the significance of the association.

Phenotype-genotype association mapping for TIPs

We analyzed genome-wide TIPs for associations with phenotypic trait variation (i.e., TIP-GWAS). To assess variation in virulence phenotype variation, we performed an infection assay on the Swiss winter wheat cultivar Claro grown in a growth chamber (72). We used diluted (2×10^5 spores/mL in 15 mL of sterile water containing 0.1% TWEEN2) 8-day-old YSB-grown *Z. tritici* spore suspension to infect the 3-week-old wheat plant. After spray inoculation, the plants were kept at 100% humidity for 21 days. Leaf lesions were assessed using ImageJ. The ratio of the total lesion area and total leaf area was calculated to obtain the percent leaf area covered by lesions (PLACL) (88).

The metabolome composition of the pathogen population was assayed previously using untargeted metabolite profiling based on UPLC-HRMS (52). *Z. tritici* isolates grown in YSB for 8 days were filtered through cheesecloth to remove hyphae and washed in milli-Q water to remove media traces. The spores were suspended and lyophilized to extract metabolites in 1 mL of HPLC-grade methanol. The extract was centrifuged at 15,000 rpm for 5 min to pellet down debris and this last step was repeated until a clear supernatant was recovered. Untargeted metabolite profiling was carried out by UHPLC-HRMS using an Acquity UPLC coupled to a Synapt G2 QTOF mass spectrometer (Waters, Inc.). Formic acid (0.05%) in water as mobile phase A and formic acid (0.05%) in acetonitrile as mobile phase B with a gradient of 0–100% B in 10 min, holding at 100% B for 2.0 min, re-equilibration at 0% B for 3.0 min was used. Samples were analyzed using mass spectrometric parameters of 50–1,200 Da, 0.2 s scan time, 120°C source temperature, 2.5 kV capillary voltage, 25V cone voltage, 900 L/h desolvation gas flow, and 400°C, 20 L/h cone gas flow, and 4 eV collision energy (low energy acquisition function) or 15–50 eV collision energy (high energy acquisition function). Recordings were made using Masslynx XS v.4.1 (Waters Inc.). Detecting markers with Markerlynx XS was performed with the following parameters: initial and final retention times 1.5 and 10 min, mass range 85–1200 Da, mass window 0.02 Da, retention time window 0.08 min, intensity threshold 500 counts, automatic peak width calculations, de-isotoping applied.

We used untransformed relative abundance values for each peak from the metabolome analysis for metabolome variation association mapping.

TIPs with a minor allele frequency of >5% (for either the presence or absence of the TE at the locus) were used for association mapping with phenotypic trait variation based on mixed linear models and performed likelihood ratio tests ($-2 \ln$) with the GEMMA version 0.98.3 (89). The standardized relatedness matrix calculated from individual genotypes was used to correct for uneven relatedness in association mapping. Association P -values were considered significant using Bonferroni (90) multiple comparison corrections. The Bonferroni threshold was calculated by dividing the nominal threshold of $\alpha = 0.05$ by the total number of TIPs used in our GWAS ($n = 192$). In addition, we considered a 5% false discovery rate (FDR) threshold using the $p.adjust$ function in the R package *stat* (91). Linkage disequilibrium (r^2) between TIPs and SNPs in a 5 kb upstream and downstream distance from the TIP locus were calculated using the “ $-hap-r2$ ” in VCFtools v. 0.1.15 (86). Linkage disequilibrium analyses between TIP alleles (presence/absence) and calculated the r^2 for all the SNPs in the window to the TIP locus polymorphism. We also calculated r^2 for all pairwise SNP combinations in the same windows.

ACKNOWLEDGMENTS

RNA-seq data generated for this manuscript were obtained in collaboration with the Genetic Diversity Centre (GDC), ETH Zurich, and the iGE3 platform of the University of Geneva.

L.N.A. and D.C. conceived the study. L.N.A. generated the RNA-seq data. U.O. generated the T.E. polymorphism data. L.N.A. performed data analyses. L.N.A. wrote the manuscript with D.C.

AUTHOR AFFILIATION

¹Laboratory of Evolutionary Genetics, Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland

PRESENT ADDRESS

Leen Nanchira Abraham, Institute of Plant Sciences, University of Cologne, Cologne, Germany

Ursula Oggenfuss, Department of Microbiology and Immunology, University of Minnesota Medical School, Minneapolis, Minnesota, USA

AUTHOR ORCID*s*

Leen Nanchira Abraham  <http://orcid.org/0000-0002-9122-6456>

Ursula Oggenfuss  <http://orcid.org/0000-0001-9291-9185>

Daniel Croll  <http://orcid.org/0000-0002-2072-380X>

AUTHOR CONTRIBUTIONS

Leen Nanchira Abraham, Conceptualization, Formal analysis, Visualization, Writing – original draft, Writing – review and editing | Ursula Oggenfuss, Conceptualization, Data curation, Writing – review and editing | Daniel Croll, Conceptualization, Funding acquisition, Project administration, Supervision, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

All RNA-seq data sets are available from the NCBI Sequence Read Archive BioProject PRJNA596434. Supplementary tables provide all TE insertion, expression, metabolome, and association mapping data and are accessible on Zenodo (<https://doi.org/10.5281/10.5281/10.5281>)

zenodo.10020096). [data set] Leen Nanchira Abraham, Croll Daniel; 2022; PRJNA650267 : Population-level transcriptome of *Zymoseptoria tritici*; NCBI Sequence Read Archive.

ADDITIONAL FILES

The following material is available [online](#).

Supplemental Material

Supplemental Figures (mBio02840-23-s0001.pdf). Figures S1–S7.

REFERENCES

- Barrett RDH, Schluter D. 2008. Adaptation from standing genetic variation. *Trends Ecol Evol* 23:38–44. <https://doi.org/10.1016/j.tree.2007.09.008>
- De Kort H, Legrand S, Honnay O, Buckley J. 2022. Transposable elements maintain genome-wide heterozygosity in inbred populations. *Nat Commun* 13:7022. <https://doi.org/10.1038/s41467-022-34795-4>
- Oggenfuss U, Badet T, Wicker T, Hartmann FE, Singh NK, Abraham L, Karisto P, Vonlanthen T, Mundt C, McDonald BA, Croll D. 2021. A population-level invasion by transposable elements triggers genome expansion in a fungal pathogen. *Elife* 10:e69249. <https://doi.org/10.7554/eLife.69249>
- Cai X, Lin R, Liang J, King GJ, Wu J, Wang X. 2022. Transposable element insertion: a hidden major source of domesticated phenotypic variation in *Brassica rapa*. *Plant Biotechnol J* 20:1298–1310. <https://doi.org/10.1111/pbi.13807>
- Britten RJ. 2010. Transposable element insertions have strongly affected human evolution. *Proc Natl Acad Sci U S A* 107:19945–19948. <https://doi.org/10.1073/pnas.1014330107>
- Hosaka A, Kakutani T. 2018. Transposable elements, genome evolution and transgenerational epigenetic variation. *Curr Opin Genet Dev* 49:43–48. <https://doi.org/10.1016/j.gde.2018.02.012>
- Wei KHC, Mai D, Chatla K, Bachtrog D. 2022. Dynamics and impacts of transposable element proliferation in the *Drosophila nasuta* species group radiation. *Mol Biol Evol* 39:msac080. <https://doi.org/10.1093/molbev/msac080>
- Chuong EB, Elde NC, Feschotte C. 2017. Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet* 18:71–86. <https://doi.org/10.1038/nrg.2016.139>
- Suh A. 2019. Genome size evolution: small transposons with large consequences. *Curr Biol* 29:R241–R243. <https://doi.org/10.1016/j.cub.2019.02.032>
- Ullastres A, Merenciano M, González J. 2021. Regulatory regions in natural transposable element insertions drive interindividual differences in response to immune challenges in *Drosophila*. *Genome Biol* 22:265. <https://doi.org/10.1186/s13059-021-02471-3>
- Sundaram V, Wysocka J. 2020. Transposable elements as a potent source of diverse cis-regulatory sequences in mammalian genomes. *Philos Trans R Soc Lond B Biol Sci* 375:20190347. <https://doi.org/10.1098/rstb.2019.0347>
- Etchegaray E, Naville M, Volff JN, Haftek-Terreau Z. 2021. Transposable element-derived sequences in vertebrate development. *Mob DNA* 12:1. <https://doi.org/10.1186/s13100-020-00229-5>
- Cosby RL, Judd J, Zhang R, Zhong A, Garry N, Pritham EJ, Feschotte C. 2020. Recurrent evolution of vertebrate transcription factors by transposase capture. *bioRxiv*. <https://doi.org/10.1101/2020.05.07.082677>
- Choi JY, Lee YCG, Betancourt A. 2020. Double-edged sword: the evolutionary consequences of the epigenetic silencing of transposable elements. *PLoS Genet* 16:e1008872. <https://doi.org/10.1371/journal.pgen.1008872>
- Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet* 9:397–405. <https://doi.org/10.1038/nrg2337>
- Slotkin RK, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* 8:272–285. <https://doi.org/10.1038/nrg2072>
- Almeida MV, Vernaz G, Putman ALK, Miska EA. 2022. Taming transposable elements in vertebrates: from epigenetic silencing to domestication. *Trends Genet* 38:529–553. <https://doi.org/10.1016/j.tig.2022.02.009>
- Quadrana L, Bortolini Silveira A, Mayhew GF, LeBlanc C, Martienssen RA, Jeddeloh JA, Colot V. 2016. The *Arabidopsis thaliana* mobilome and its impact at the species level. *Elife* 5:e15716. <https://doi.org/10.7554/eLife.15716>
- Carpentier MC, Manfroi E, Wei FJ, Wu HP, Lasserre E, Llauro C, Debladis E, Akakpo R, Hsing YI, Panaud O. 2019. Retrotranspositional landscape of Asian rice revealed by 3000 genomes. *Nat Commun* 10:24. <https://doi.org/10.1038/s41467-018-07974-5>
- Pereira D, Oggenfuss U, McDonald BA, Croll D. 2021. Population genomics of transposable element activation in the highly repressive genome of an agricultural pathogen. *Microb Genom* 7:000540. <https://doi.org/10.1099/mgen.0.000540>
- Slotkin RK, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* 8:272–285. <https://doi.org/10.1038/nrg2072>
- Mita P, Boeke JD. 2016. How retrotransposons shape genome regulation. *Curr Opin Genet Dev* 37:90–100. <https://doi.org/10.1016/j.gde.2016.01.001>
- Bucher E, Reinders J, Mirouze M. 2012. Epigenetic control of transposon transcription and mobility in *Arabidopsis*. *Curr Opin Plant Biol* 15:503–510. <https://doi.org/10.1016/j.pbi.2012.08.006>
- Gladyshev E. 2017. Repeat-induced point mutation (RIP) and other genome defense mechanisms in fungi. *Microbiol Spectr* 5. <https://doi.org/10.1128/microbiolspec.FUNK-0042-2017>
- Ramírez-Tejero JA, Cabanás CGL, Valverde-Corredor A, Mercado-Blanco J, Luque F. 2020. Epigenetic regulation of *Verticillium dahliae* virulence: does DNA methylation level play a role? *Int J Mol Sci* 21:5197. <https://doi.org/10.3390/ijms21155197>
- Torres DE, Thomma B, Seidl MF. 2021. Transposable elements contribute to genome dynamics and gene expression variation in the fungal plant pathogen *Verticillium dahliae*. *bioRxiv*. <https://doi.org/10.1093/gbe/evab135>
- Clairet C, Gay EJ, Porquier A, Blaise F, Marais C-L, Balesdent M-H, Rouxel T, Soyer JL, Fudal I. 2021. Regulation of effector gene expression as concerted waves in *Leptosphaeria maculans*: a two-players game. *bioRxiv*. <https://doi.org/10.1101/2021.12.15.472773>
- Schmidt SM, Houterman PM, Schreiber I, Ma L, Amyotte S, Chellappan B, Boeren S, Takken FLW, Rep M. 2013. MITES in the promoters of effector genes allow prediction of novel virulence genes in *Fusarium oxysporum*. *BMC Genomics* 14:1–21. <https://doi.org/10.1186/1471-2164-14-119>
- Zhang W, Huang J, Cook DE. 2021. Histone modification dynamics at H3K27 are associated with altered transcription of in planta induced genes in *Magnaporthe oryzae*. *PLoS Genet* 17:e1009376. <https://doi.org/10.1371/journal.pgen.1009376>
- Bao J, Chen M, Zhong Z, Tang W, Lin L, Zhang X, Jiang H, Zhang D, Miao C, Tang H, Zhang J, Lu G, Ming R, Norvienyeku J, Wang B, Wang X. 2017. PacBio sequencing reveals transposable elements as a key contributor to genomic plasticity and virulence variation in *Magnaporthe oryzae*. *Mol Plant* 10:1465–1468. <https://doi.org/10.1016/j.molp.2017.08.008>
- Jeon J, Choi J, Lee GW, Park SY, Huh A, Dean RA, Lee YH. 2015. Genome-wide profiling of DNA methylation provides insights into epigenetic regulation of fungal development in a plant pathogenic fungus, *Magnaporthe oryzae*. *Sci Rep* 5:8567. <https://doi.org/10.1038/srep08567>

32. Soyer JL, El Ghalid M, Glaser N, Ollivier B, Linglin J, Grandaubert J, Balesdent M-H, Connolly LR, Freitag M, Rouxel T, Fudal I. 2014. Epigenetic control of effector gene expression in the plant pathogenic fungus *Leptosphaeria maculans*. *PLoS Genet* 10:e1004227. <https://doi.org/10.1371/journal.pgen.1004227>
33. GourlieR, McDonaldM, HafezM, Ortega-PoloR, Low KE, AbbottDW, StrelkovSE, DaayfF, AboukhaddourR. 2022. Dissecting the *Pyrenophora tritici-repentis* (tan spot of wheat) pangenome. *bioRxiv*. <https://doi.org/10.1186/s12915-022-01433-w>
34. Moolhuijzen PM, See PT, Oliver RP, Moffat CS. 2018. Genomic distribution of a novel *Pyrenophora tritici-repentis* ToxA insertion element. *PLoS One* 13:e0206586. <https://doi.org/10.1371/journal.pone.0206586>
35. McDonald MC, Taranto AP, Hill E, Schwessinger B, Liu Z, Simpfordorfer S, Milgate A, Solomon PS, Di Pietro A. 2019. Transposon-mediated horizontal transfer of the host-specific virulence protein ToxA between three fungal wheat pathogens. *mBio* 10:e01515-19. <https://doi.org/10.1128/mBio.01515-19>
36. Badet T, Oggenfuss U, Abraham L, McDonald BA, Croll D. 2020. A 19-isolate reference-quality global pangenome for the fungal wheat pathogen *Zymoseptoria tritici*. *BMC Biol* 18:12. <https://doi.org/10.1186/s12915-020-0744-3>
37. Fouché S, Badet T, Oggenfuss U, Plissonneau C, Francisco CS, Croll D. 2020. Stress-driven transposable element de-repression dynamics and virulence evolution in a fungal pathogen. *Mol Biol Evol* 37:221–239. <https://doi.org/10.1093/molbev/msz216>
38. Meile L, Peter J, Puccetti G, Allassimone J, McDonald BA, Sánchez-Vallet A. 2020. Chromatin dynamics contribute to the spatiotemporal expression pattern of virulence genes in a fungal plant pathogen. *mBio* 11:e02343-20. <https://doi.org/10.1128/mBio.02343-20>
39. Meile L, Croll D, Brunner PC, Plissonneau C, Hartmann FE, McDonald BA, Sánchez-Vallet A. 2018. A fungal avirulence factor encoded in a highly plastic genomic region triggers partial resistance to septoria tritici blotch. *New Phytol* 219:1048–1061. <https://doi.org/10.1111/nph.15180>
40. Krishnan P, Meile L, Plissonneau C, Ma X, Hartmann FE, Croll D, McDonald BA, Sánchez-Vallet A. 2018. Transposable element insertions shape gene regulation and melanin production in a fungal pathogen of wheat. *BMC Biol* 16:78. <https://doi.org/10.1186/s12915-018-0543-2>
41. Omrane S, Audéon C, Ignace A, Duplaix C, Aouini L, Kema G, Walker A-S, Fillinger S. 2017. Plasticity of the MF51 promoter leads to multidrug resistance in the wheat pathogen *Zymoseptoria tritici*. *mSphere* 2:e00393-17. <https://doi.org/10.1128/msphere.00393-17>
42. Hartmann FE, Sánchez-Vallet A, McDonald BA, Croll D. 2017. A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. *ISME J* 11:1189–1204. <https://doi.org/10.1038/ismej.2016.196>
43. Feurtey A, Lorrain C, McDonald MC, Milgate A, Solomon PS, Warren R, Puccetti G, Scalliet G, Torriani SFF, Gout L, et al. 2023. A thousand-genome panel retraces the global spread and adaptation of a major fungal crop pathogen. *Nat Commun* 14:1059. <https://doi.org/10.1038/s41467-023-36674-y>
44. Oggenfuss U, Croll D. 2023. Recent transposable element bursts are associated with the proximity to genes in a fungal plant pathogen. *PLoS Pathog* 19:e1011130. <https://doi.org/10.1371/journal.ppat.1011130>
45. Möller M, Habig M, Lorrain C, Feurtey A, Hauelsen J, Fagundes WC, Alizadeh A, Freitag M, Stukenbrock EH, Krasileva K. 2021. Recent loss of the Dim2 DNA methyltransferase decreases mutation rate in repeats and changes evolutionary trajectory in a fungal pathogen. *PLoS Genet* 17:e1009448. <https://doi.org/10.1371/journal.pgen.1009448>
46. Goodwin SB, M'barek SB, Dhillon B, Wittenberg AHJ, Crane CF, Hane JK, Foster AJ, Van der Lee TAJ, Grimwood J, Aerts A, et al. 2011. Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. *PLoS Genet* 7:e1002070. <https://doi.org/10.1371/journal.pgen.1002070>
47. Torriani SFF, Stukenbrock EH, Brunner PC, McDonald BA, Croll D. 2011. Evidence for extensive recent Intron transposition in closely related fungi. *Curr Biol* 21:2017–2022. <https://doi.org/10.1016/j.cub.2011.10.041>
48. Bannister AJ, Kouzarides T. 2011. Regulation of chromatin by histone modifications. *Cell Res* 21:381–395. <https://doi.org/10.1038/cr.2011.22>
49. Martin C, Zhang Y. 2005. The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol* 6:838–849. <https://doi.org/10.1038/nrm1761>
50. Wang L, Sun Y, Sun X, Yu L, He Z, Huang J, Tian D, Hurst LD, Yang S. 2020. Repeat-induced point mutation in *Neurospora crassa* causes the highest known mutation rate and mutational burden of any cellular life. *Genome Biol* 21:142. <https://doi.org/10.1186/s13059-020-02060-w>
51. Abraham LN, Croll D. 2023. Genome-wide expression QTL mapping reveals the highly dynamic regulatory landscape of a major wheat pathogen. *bioRxiv*. <https://doi.org/10.1101/2023.07.14.549109>
52. Singh NK, Tralamazza SM, Abraham LN, Glauser G, Croll D. 2022. Genome-wide association mapping reveals genes underlying population-level metabolome diversity in a fungal crop pathogen. *BMC Biol* 20:224. <https://doi.org/10.1186/s12915-022-01422-z>
53. Li C, Zhang Q, Xia Y, Jin K. 2021. Manmra, a negative transcription regulator in nitrogen catabolite repression pathway, contributes to nutrient utilization, stress resistance, and virulence in entomopathogenic fungus *metarhizium acridum*. *Biology (Basel)* 10:1167. <https://doi.org/10.3390/biology10111167>
54. Torres DE, Oggenfuss U, Croll D, Seidl MF. 2020. Genome evolution in fungal plant pathogens: looking beyond the two-speed genome model. *Fungal Biol Rev* 34:136–143. <https://doi.org/10.1016/j.fbr.2020.07.001>
55. Yang WR, Ardeljan D, Pacyna CN, Payer LM, Burns KH. 2019. SQuIRE reveals locus-specific regulation of interspersed repeat expression. *Nucleic Acids Res* 47:27. <https://doi.org/10.1093/nar/gky1301>
56. Wu B, Macielog AI, Hao W. 2017. Origin and spread of spliceosomal introns: insights from the fungal clade *Zymoseptoria*. *Genome Biol Evol* 9:2658–2667. <https://doi.org/10.1093/gbe/evx211>
57. Wicker T, Gundlach H, Spannagl M, Uauy C, Borrill P, Ramírez-González RH, De Oliveira R, Mayer KFX, Paux E, Choulet F, International Wheat Genome Sequencing Consortium. 2018. Impact of transposable elements on genome structure and evolution in bread wheat. *Genome Biol* 19:103. <https://doi.org/10.1186/s13059-018-1479-0>
58. Komlusi J, Habig M, Stukenbrock EH. 2022. Repeat-induced point mutation and gene conversion coinciding with heterochromatin shape the genome of a plant pathogenic fungus. *bioRxiv*. <https://doi.org/10.1101/2022.11.30.518637>
59. Wells JN, Feschotte C. 2020. A field guide to eukaryotic transposable elements. *Annu Rev Genet* 54:539–561. <https://doi.org/10.1146/annurev-genet-040620-022145>
60. Tan BC, Chen Z, Shen Y, Zhang Y, Lai J, Sun SSM. 2011. Identification of an active new mutator transposable element in maize. *G3 (Bethesda)* 1:293–302. <https://doi.org/10.1534/g3.111.000398>
61. Baller JA, Gao J, Stamenova R, Curcio MJ, Voytas DF. 2012. A nucleosomal surface defines an integration hotspot for the *Saccharomyces cerevisiae* Ty1 retrotransposon. *Genome Res* 22:704–713. <https://doi.org/10.1101/gr.129585.111>
62. Mularoni L, Zhou Y, Bowen T, Gangadharan S, Wheelan SJ, Boeke JD. 2012. Retrotransposon Ty1 integration targets specifically positioned asymmetric nucleosomal DNA segments in tRNA hotspots. *Genome Res* 22:693–703. <https://doi.org/10.1101/gr.129460.111>
63. Gao X, Hou Y, Ebina H, Levin HL, Voytas DF. 2008. Chromodomains direct integration of retrotransposons to heterochromatin. *Genome Res* 18:359–369. <https://doi.org/10.1101/gr.7146408>
64. Spradling AC, Bellen HJ, Hoskins RA. 2011. *Drosophila* P elements preferentially transpose to replication origins. *Proc Natl Acad Sci U S A* 108:15948–15953. <https://doi.org/10.1073/pnas.1112960108>
65. Wei KHC, Mai D, Chatla K, Bachtrog D. 2022. Dynamics and impacts of transposable element proliferation in the *Drosophila nasuta* species group radiation. *Mol Biol Evol* 39:msac080. <https://doi.org/10.1093/molbev/msac080>
66. Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. 2005. Why highly expressed proteins evolve slowly. *Proc Natl Acad Sci U S A* 102:14338–14343. <https://doi.org/10.1073/pnas.0504070102>
67. Fouché S, Badet T, Oggenfuss U, Plissonneau C, Francisco CS, Croll D. 2019. Stress-driven transposable element de-repression dynamics in a fungal pathogen. *bioRxiv*. <https://doi.org/10.1101/633693>
68. Hassani MA, Oppong-Danquah E, Feurtey A, Tasdemir D, Stukenbrock EH. 2022. Differential regulation and production of secondary metabolites among isolates of the fungal wheat pathogen *Zymoseptoria*

- tritici*. Appl Environ Microbiol 88:e0229621. <https://doi.org/10.1128/aem.02296-21>
69. Derbyshire MC, Gohari AM, Mehrabi R, Kilaru S, Steinberg G, Ali S, Bailey A, Hammond-Kosack K, Kema GHJ, Rudd JJ. 2018. Phosphopantetheinyl transferase (Ppt)-mediated biosynthesis of lysine, but not siderophores or DHN melanin, is required for virulence of *Zymoseptoria tritici* on wheat. Sci Rep 8:17069. <https://doi.org/10.1038/s41598-018-35223-8>
 70. Domínguez M, Dugas E, Benchouaia M, Leduque B, Jiménez-Gómez JM, Colot V, Quadrona L. 2020. The impact of transposable elements on tomato diversity. Nat Commun 11:4058. <https://doi.org/10.1038/s41467-020-17874-2>
 71. Gluck-Thaler E, Ralston T, Konkel Z, Ocampos CG, Ganeshan VD, Dorrance AE, Niblack TL, Wood CW, Slot JC, Lopez-Nicora HD, Vogan AA. 2022. Giant starship elements mobilize accessory genes in fungal genomes. Mol Biol Evol 39:msac109. <https://doi.org/10.1093/molbev/msac109>
 72. Singh NK, Karisto P, Croll D. 2021. Population-level deep sequencing reveals the interplay of clonal and sexual reproduction in the fungal wheat pathogen *Zymoseptoria tritici*. Microb Genom 7:678. <https://doi.org/10.1099/mgen.0.000678>
 73. Perkins DD. 2006. How to choose and prepare media
 74. ScienceOpen. FastQC: a quality control tool for high throughput sequence data. Available from: <https://www.scienceopen.com/document?vid=de674375-ab83-4595-afa9-4c8aa9e4e736>. Retrieved 06 Oct 2022.
 75. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
 76. Linheiro RS, Bergman CM. 2012. Whole genome resequencing reveals natural target site preferences of transposable elements in *Drosophila melanogaster*. PLoS One 7:e30008. <https://doi.org/10.1371/journal.pone.0030008>
 77. Nelson MG, Linheiro RS, Bergman CM. 2017. McClintock: an integrated pipeline for detecting transposable element insertions in whole-genome shotgun sequencing data. G3 (Bethesda) 7:2763–2778. <https://doi.org/10.1534/g3.117.043893>
 78. R Core Team. 2021. R: a language and environment for statistical computing. R Foundation for Statistical Computing
 79. Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290. <https://doi.org/10.1093/bioinformatics/btg412>
 80. Xu S, Li L, Luo X, Chen M, Tang W, Zhan L, Dai Z, Lam TT, Guan Y, Yu G. 2022. Ggtree: a serialized data object for visualization of a phylogenetic tree and annotation data. iMeta 1. <https://doi.org/10.1002/imt2.56>
 81. Delaneau O, Ongen H, Brown AA, Fort A, Panousis NI, Dermitzakis ET. 2017. A complete tool set for molecular QTL discovery and analysis. Nat Commun 8:15452–15457. <https://doi.org/10.1038/ncomms15452>
 82. Grandaubert J, Bhattacharyya A, Stukenbrock EH. 2015. RNA-seq-based gene annotation and comparative genomics of four fungal grass pathogens in the genus *Zymoseptoria* identify novel orphan genes and species-specific invasions of transposable elements. G3 (Bethesda) 5:1323–1333. <https://doi.org/10.1534/g3.115.017731>
 83. Andrews S. 2010. FastQC: a quaFastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics
 84. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with bowtie 2. Nat Methods 9:357–359. <https://doi.org/10.1038/nmeth.1923>
 85. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303. <https://doi.org/10.1101/gr.107524.110>
 86. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
 87. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. 2021. Twelve years of SAMtools and BCFtools. Gigascience 10:giab008. <https://doi.org/10.1093/gigascience/giab008>
 88. Karisto P, Hund A, Yu K, Anderegg J, Walter A, Mascher F, McDonald BA, Mikaberidze A. 2018. Ranking quantitative resistance to septoria tritici blotch in elite wheat cultivars using automated image analysis. Phytopathology 108:568–581. <https://doi.org/10.1094/PHYTO-04-17-0163-R>
 89. Zhou X, Stephens M. 2012. Genome-wide efficient mixed-model analysis for association studies. Nat Genet 44:821–824. <https://doi.org/10.1038/ng.2310>
 90. Bland JM, Altman DG. 1995. Multiple significance tests: the Bonferroni method. BMJ 310:170. <https://doi.org/10.1136/bmj.310.6973.170>
 91. Korthauer K, Kimes PK, Duvall C, Reyes A, Subramanian A, Teng M, Shukla C, Alm EJ, Hicks SC. 2019. A practical guide to methods controlling false discoveries in computational biology. Genome Biol 20:118. <https://doi.org/10.1186/s13059-019-1716-1>