AMERICAN SOCIETY FOR MICROBIOLOGY | mBio®

RESEARCH ARTICLE
Ecological and Evolutionary Science

Transposon-Mediated Horizontal Transfer of the Host-Specific Virulence Protein ToxA between Three Fungal Wheat Pathogens

Megan C. McDonald,[a] Adam P. Taranto,[a] Erin Hill,[a] Benjamin Schwessinger,[a] Zhaohui Liu,[b] Steven Simpfendorfer,[c] Andrew Milgate,[d] Peter S. Solomon[a]

[a]Division of Plant Sciences, Research School of Biology, The Australian National University, Canberra, Australia
[b]Department of Plant Pathology, North Dakota State University, Fargo, North Dakota, USA
[c]NSW Department of Primary Industries, Tamworth Agricultural Institute, Tamworth, NSW, Australia
[d]NSW Department of Primary Industries, Wagga Wagga Agricultural Institute, Wagga Wagga, NSW, Australia

ABSTRACT Most known examples of horizontal gene transfer (HGT) between eukaryotes are ancient. These events are identified primarily using phylogenetic methods on coding regions alone. Only rarely are there examples of HGT where noncoding DNA is also reported. The gene encoding the wheat virulence protein ToxA and the surrounding 14 kb is one of these rare examples. *ToxA* has been horizontally transferred between three fungal wheat pathogens (*Parastagonospora nodorum*, *Pyrenophora tritici-repentis*, and *Bipolaris sorokiniana*) as part of a conserved ~14 kb element which contains coding and noncoding regions. Here we used long-read sequencing to define the extent of HGT between these three fungal species. Construction of near-chromosomal-level assemblies enabled identification of terminal inverted repeats on either end of the 14 kb region, typical of a type II DNA transposon. This is the first description of *ToxA* with complete transposon features, which we call ToxhAT. In all three species, ToxhAT resides in a large (140-to-250 kb) transposon-rich genomic island which is absent in isolates that do not carry the gene (annotated here as *toxa*$^{-}$). We demonstrate that the horizontal transfer of ToxhAT between *P. tritici-repentis* and *P. nodorum* occurred as part of a large (~80 kb) HGT which is now undergoing extensive decay. In *B. sorokiniana*, in contrast, ToxhAT and its resident genomic island are mobile within the genome. Together, these data provide insight into the noncoding regions that facilitate HGT between eukaryotes and into the genomic processes which mask the extent of HGT between these species.

IMPORTANCE This work dissects the tripartite horizontal transfer *of ToxA*, a gene that has a direct negative impact on global wheat yields. Defining the extent of horizontally transferred DNA is important because it can provide clues to the mechanisms that facilitate HGT. Our analysis of *ToxA* and its surrounding 14 kb suggests that this gene was horizontally transferred in two independent events, with one event likely facilitated by a type II DNA transposon. These horizontal transfer events are now in various processes of decay in each species due to the repeated insertion of new transposons and subsequent rounds of targeted mutation by a fungal genome defense mechanism known as repeat induced point mutation. This work highlights the role that HGT plays in the evolution of host adaptation in eukaryotic pathogens. It also increases the growing body of evidence indicating that transposons facilitate adaptive HGT events between fungi present in similar environments and hosts.

KEYWORDS horizontal transfer, transposon, fungal wheat pathogen, adaptive evolution, ToxA, fungal pathogen, wheat pathogen

Citation McDonald MC, Taranto AP, Hill E, Schwessinger B, Liu Z, Simpfendorfer S, Milgate A, Solomon PS. 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.

Editor Antonio Di Pietro, Universidad de Córdoba

Address correspondence to Megan C. McDonald, megan.mcdonald@anu.edu.au, or Peter S. Solomon, peter.solomon@anu.edu.au.

Received 12 June 2019
Accepted 16 August 2019
Published 10 September 2019

Horizontal gene transfer (HGT) is a mechanism whereby DNA from unrelated organisms is transferred between the organisms in a non-Mendelian fashion (1). In proteobacteria, HGT is thought to have occurred in over 75% of all protein families, making HGT one of the most important tools facilitating adaptation to stressful environments (2, 3). This propensity to share DNA between species has been attributed to many human health issues, such as the rapid rise and spread of antibiotic resistance in hospitals (4, 5). In eukaryotes, HGT was once thought to be a rare event and therefore not an important contributor to environmental adaptation. However, numerous studies have now shown that HGT between eukaryotes plays a very important role in adaptation, especially in the case of microbes that colonize a common host (6–8).

Among eukaryotic microbes, fungi are often used for kingdom-wide studies of adaptation, due to their relatively small genome size, importance in human and plant disease, and applications in food and biotechnology (6, 8, 9). Domesticated fungi, particularly those used in food production, are now being used as model organisms to understand the genetic basis of adaptation (10–12). On an evolutionary time scale, these organisms have been subjected to a short but intense period of selection, which has had dramatic effects on their preferred carbon and nitrogen sources, secondary metabolite production, and many other physiological traits (12, 13). One emerging theme from these studies is that organisms which are common contaminants of the food-making process are often donors of the genes that provide fitness advantages in these specialized environments. The reported HGT events are extensive and involve tens of thousands of bases of DNA, which remain over 90% identical between very distantly related species (10, 11). These HGTs contain both coding and noncoding regions which are stably integrated into the core nuclear genomes of the recipient species (10, 11). While the original reports suggested that these regions were important for adaptation to the domestic environment, the fitness advantage conferred by these genes had to be demonstrated in follow-up studies performed with knockout strains (13, 14).

Rapid adaptation via HGT is not restricted to domesticated species, but there exist very few described instances where the horizontally transferred DNA is integrated into the core nuclear genome and remains highly identical outside coding regions. One standout example is the virulence gene *ToxA* and the surrounding 11 to 12 kb, which to date has been reported in three fungal wheat pathogens: *Parastagonospora nodorum*, *Pyrenophora tritici-repentis*, and *Biopolaris sorokiniana* (13–16). While all three species belong to the same fungal order, the Pleosporales, they are distant relatives, with several million years separating their speciation (15, 16). Similar to the domesticated fungi discussed above, this HGT event is hypothesized to be extremely recent, as the average pairwise nucleotide identity across this 12 kb region remains greater than 92% (17). The *ToxA* gene itself remains identical between *P. tritici-repentis* and *B. sorokiniana* and only three nucleotides different from between *B. sorokiniana* and *P. nodorum* (17). The fitness advantage that *ToxA* confers has been demonstrated experimentally, whereby the presence of *ToxA* in a fungal isolate leads to faster development of necrotic lesions on wheat leaves (17, 18). This virulence function is genotype specific, as *ToxA* causes necrosis only on wheat lines that carry the susceptibility gene called *Tsn1* (19–21). In the absence of *Tsn1*, all three fungal species can still infect wheat due to the presence of other virulence genes (17, 19, 20).

Though *ToxA* confers a strong fitness advantage, this HGT event is not a fixed insertion and persists in all three pathogen populations as a presence/absence polymorphism (17, 21–23). The size of this presence/absence polymorphism has yet to be fully characterized. The frequency of *ToxA* in different field populations around the world also varies dramatically, ranging from 6% to 97% in different pathogen field populations (22). The selective forces that increase the frequency of *ToxA* in some fungal populations and decrease it in others remain unknown. Results of studies that examined whether there was a positive correlation between the frequency of *ToxA* in fields planted with *Tsn1* (susceptible) wheat cultivars were inconclusive (24). For ease of

TABLE 1 Summary of genome assembly statistics for each isolate assembled in this study

Species and isolate	Raw data	ToxA genotype	Assembly size (Mb)[a]	No. of contigs[b]	No. of contigs with both telomeres	No. of contigs with 1 telomere	BUSCO score[c]
B. sorokiniana							
CS10	PacBio	ToxA$^+$	36.9	22	14	3	98.9
CS27	Nanopore + Illumina	*toxa*$^-$	35.2	19	NA[d]	NA	98.9
WAI2406	Nanopore	ToxA$^+$	36.9	21	NA	NA	72.4
WAI2411	Nanopore	ToxA$^+$	36.2	21	NA	NA	69.8
P. nodorum							
SN15	PacBio	ToxA$^+$	37.3	26	18	6	99.1
Sn79-1087	Nanopore + Illumina	*toxa*$^-$	34.7	23	13	8	92.4

[a]Size in millions of base pairs.
[b]Nuclear contigs only.
[c]Percentage of 1,313 ascomycete single-copy orthologs found in assembly.
[d]NA, not assessed.

reading, here we use the notation *ToxA*$^+$ for isolates that contain the gene and *toxa*$^-$ for isolates that do not carry the gene.

Despite detailed knowledge on the molecular function of *ToxA* and its prevalence in fungal pathogen populations throughout the world, we still do not know the origins of this important virulence gene or the mechanisms that facilitated its transfer and stable integration into the genomes of these three pathogen species. There is clear evidence that *ToxA* is embedded in an AT-rich, repeat-dense region of the genome in all three species. AT richness in these portions of the genome is driven by a fungus-specific genome defense process known as the repeat induced point mutation (RIP) (25). RIP targets repeated sequences as small as 155 bp, mutating C:G to T:A, which introduces early stop codons in repeated DNA sequences (25–28). This mechanism is hypothesized to have evolved in some phyla of fungi to stop the spread of transposons or other self-copying elements within their genomes (26).

ToxA and its highly conserved flanking DNA provide a unique opportunity to dissect the integration of horizontally transferred DNA into the nuclear genomes of three fungal pathogens. To define the location and extent of each HGT event, we used long-read DNA sequencing to generate near-complete genome assemblies for several representatives from two of the three species, in addition to several other published assemblies (29, 30). We performed extensive *de novo* annotation of the repeat families in all three fungal species and manually annotated the region surrounding the *ToxA* gene. These assemblies and repeat annotations resolve the genomic context in which the virulence gene is located and provide insights into potential mechanisms of HGT as well as into the history of horizontal transfer events.

RESULTS

Long-read sequencing reveals a conserved type II DNA transposon. The genomic location of *ToxA* has been best described in *P. tritici-repentis*, where two long-read assemblies place this gene in the middle of chromosome 06 (supercontig1.4) (23, 31). Several long-read assemblies have also been generated for *ToxA*$^+$ *P. nodorum* isolates Sn4 and Sn2000, where *ToxA* is found on chromosome 08 in both isolates (29). In addition to these publicly available assemblies, we sequenced *ToxA*$^+$ isolates *P. nodorum* SN15 and *B. sorokiniana* CS10 (original isolate name, BRIP10943) with seven PacBio SMRT cells each, resulting in approximately 500,000 reads with average read lengths of 10.6 kb and 9.4 kb, respectively. In addition to the two SMRT assemblies, we resequenced an additional four isolates with the Oxford Nanopore MinION. This included two *toxa*$^-$ isolates, *P. nodorum* isolate Sn79-1087 and *B. sorokiniana* isolate CS27 (original isolate name, BRIP27492a), as well as two additional *ToxA*$^+$ *B. sorokiniana* isolates, WAI2406 and WAI2411. All isolates were *de novo* assembled using long-read data only. Short-read Illumina data were used to "polish" the Nanopore *de novo* assemblies of CS27 and Sn79-1087. A complete list of all isolates used in this study and assembly quality indicators is provided in Table 1. Genome assembly accession num-

bers and additional information about the isolates are provided at https://github.com/megancamilla/Transposon-Mediated-transfer-of-ToxA/tree/master/S1_GenomeStats. *B. sorokiniana* chromosomes were ordered from largest to smallest and named based on the PacBio assembly of isolate CS10. Our *P. nodorum* contigs were named based on synteny alignments to the assemblies published recently by Richards et al. (29).

Assembly quality was assessed using the Benchmarking Universal Single-Copy Orthologs (BUSCO) tool, which identifies fragmented, duplicated, and missing genes from *de novo* assemblies (32, 33). The score reported in Table 1 is the percentage of complete genes found in a set of 1,313 BUSCO genes from the Ascomycota odb9 data set. The number of complete genes is used as a proxy to estimate total-genome completeness (32). The assembly completeness scores were much lower for Nanopore-only assemblies in isolates WAI2406 and WAI2411, where no short-read data were available for genome correction. For isolates CS27 and Sn79-1087, available short-read data allowed correction of the assemblies, such that the proportion of complete BUSCO genes found exceeded 90%. Both PacBio assemblies, without short-read data, generated a BUSCO score above 98% (Table 1). For both the CS10 and SN15 PacBio assemblies, the 6-bp telomeric repeat (TAACCC) was found on the ends of several contigs (summarized in Table 1). We could not identify telomeric repeats in the Nanopore assemblies for the *B. sorokiniana* isolates. However, we were able to identify many contigs with telomeric repeats for the assembly of *P. nodorum* isolate Sn79-1087 (Table 1).

We have previously reported the presence of >92% identical 12 kb region shared between all three species (17). This aligns with the original publication from Friesen et al. (20), in which an 11 kb element was reported to be conserved between *P. tritici-repentis* and *P. nodorum*. In both *P. nodorum* SN15 and *B. sorokiniana* CS10, the chromosomes that contain *ToxA* were assembled completely, with telomeric repeats on both ends. A self-alignment of this region in *B. sorokiniana* CS10 revealed intact, terminal inverted repeats (TIRs) separated by 14.3 kb (Fig. 1A). TIRs are structural features of type II DNA transposons of the order "TIR," which are required for excision by transposases (34). These TIRs were not identified in previous studies (17, 20), which we explore further below. The aligned TIRs consist of 74 bp and are ~92% identical (Fig. 1B). Here, we refer to *ToxA* and the accompanying noncoding and coding DNA enclosed within these TIRs as "ToxhAT." This name reflects the historical association of *ToxA* with the neighboring type II hAT-like transposase gene (20).

We annotated the coding regions within ToxhAT in *B. sorokiniana* isolate CS10 with both long-read and short-read RNA sequencing (RNA-seq) data (see Fig. S1 in the supplemental material). This annotation plus the self-aligned sequence revealed eight genes, three inverted repeats (IRs), and two additional internal TIRs (Fig. 1C). Three annotated genes, *CS10_08.708*, *CS10_08.709*, and *CS10_08.714*, contain no known protein-coding domains. Excluding *ToxA*, the remaining four genes had conserved domains, as identified by NCBI's conserved domain database search tool (Fig. 1C). One gene contained a major facilitator superfamily (MFS) domain (NCBI accession no. cl28910), which in yeast was shown to be a proton-coupled transporter of dipeptides and tripeptides (35). The largest coding DNA sequence (CDS) within ToxhAT contained two known protein domains, with a Patatin domain at the N terminus (accession no. cd07216) followed by tetratricopeptide repeats (TPR; pfam13424) at the C terminus. In fungi, proteins that contain these domains are recognized as members of the NOD-like receptor (NLR) family (36). Only a limited number of these proteins have been functionally studied in fungi, but they are considered to be broadly involved in self-recognition and immunity (36, 37). The fourth gene was flanked by its own set of TIRs and contained a helix-turn-helix (HTH) DNA binding domain (accession no. cl04999). This structure indicated that this CDS likely represents a nested type II transposase within ToxhAT (Fig. 1C). This indicates that ToxhAT is a composite of at least two DNA transposons. Fragments of the eight open reading frames (ORFs) are also found in the other two species, *P. nodorum* and *P. tritici-repentis* (Fig. S2); however, the 3′ end of ToxhAT is invaded by unique sequences in these two species (Fig. S2).

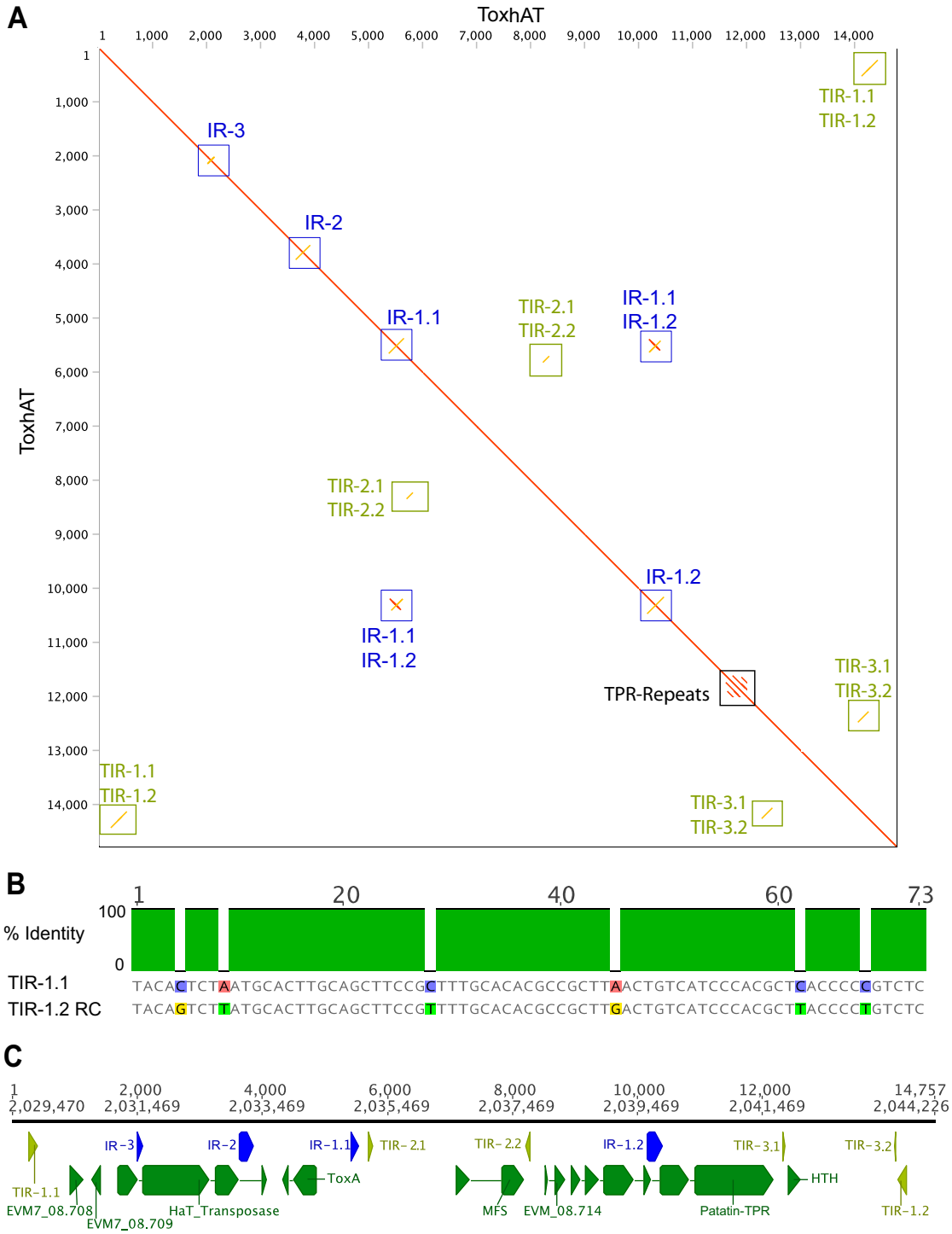

FIG 1 Characterization of ToxhAT in *B. sorokiniana* isolate CS10. (A) Self-alignment of ToxhAT drawn as a dot plot. The red line down the center indicates a 1-to-1 alignment; yellow lines show inverse alignments. Terminal inverted repeats (TIRs) and inverted repeats (IRs) are boxed. The TPRs are short tandem repeats found in the gene with the Patatin domain. (B) Alignment of 74-bp TIR1.1 and the reverse complement (RC) of TIR1.2. Gray bases indicates aligned positions that are identical between the two sequences. (C) Manual annotation of coding regions within ToxhAT showing each open reading frame (green), inverted repeat (blue), and TIR (light green).

Using *B. sorokiniana* as a guide, we were able to identify remnants of the ToxhAT TIRs in both *P. tritici-repentis* and *P. nodorum*. The 5′ TIR in *P. tritici-repentis* remains largely intact, whereas the 3′ TIR is enriched in the C-to-T and G-to-A transitions characteristic of RIP (Fig. S2). In *P. nodorum*, both the 5′ and 3′ TIR are enriched in RIP

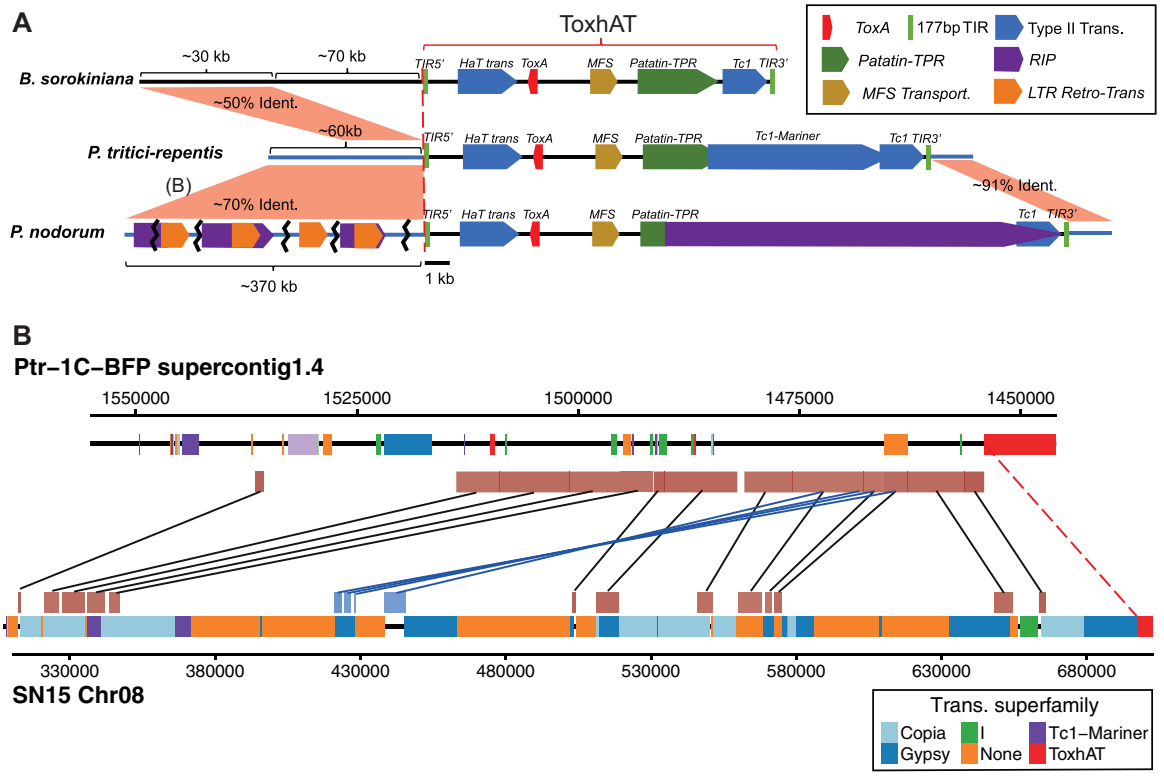

FIG 2 (A) Overview of the ToxhAT in all three pathogen species. All features drawn to the right of the red dashed line are drawn to scale as indicated by the black scale bar at the bottom. Features to the left of the red dashed line are not drawn to scale; relative sizes are indicated with brackets. The opaque red rectangles drawn between the chromosomes outside the TIRs show regions of synteny as indicated by whole-chromosome alignment. The approximate percentage of nucleotide identity is indicated within the red shading. "(B)" in part A indicates the region shared between *P. nodorum* and *P. tritici-repentis*, which is drawn in part. (B) Closeup view of whole-chromosome alignment between *P. nodorum* and *P. tritici-repentis*. Chromosomes are drawn as thick black lines with positions of annotated transposons shown in colored blocks. Transposons are classified into superfamiles as indicated by the legend. The additional opaque red/blue boxes appearing above or below the chromosomes represent nucleotide regions that are >70% identical and were identified by whole-chromosome alignment with LASTZ. Black lines connect syntentic blocks aligned in the same direction, while blue lines connect inverted syntenic blocks. Trans, transposon.

mutations, which, without prior knowledge of the TIR location in *B. sorokiniana*, would be impossible to identify. There were additional unique sequence insertions inside ToxhAT in *P. tritici-repentis* and *P. nodorum* (Fig. 2; see also Fig. S2). Manual annotation of this unique sequence showed *P. tritici-repentis* 1C-BFP has a 3′ insertion of a type II DNA transposon sequence, a Tc1-mariner-like sequence, which interrupts ToxhAT, separating the 5′ TIR from the 3′ TIR by ~20.2 kb. In *P. nodorum* SN15, ToxhAT is interrupted by a different transposon which resembles a type I long-terminal-repeat (LTR) retrotransposon. The exact identity of this transposon was difficult to determine due to the presence of extensive RIP-like mutations in this sequence. This insertion separates the ToxhAT TIRs in *P. nodorum* by ~25.6 kb (Fig. S2). Despite these additional insertions, the TIRs identified in *B. sorokiniana* were found to be present, though heavily mutated, in the other two species, indicating a common evolutionary origin for ToxhAT.

ToxhAT was transferred in two independent HGT events. Whole-chromosome alignments (WCA) between the ToxhAT-containing chromosomes of *P. nodorum* and *P. tritici-repentis* revealed DNA with >70% sequence identity beyond the boundaries set by ToxhAT TIRs (Fig. 2A; see also Fig. S2). Pairwise alignments of these regions revealed that almost all polymorphisms were RIP-like (Fig. S3). Excluding the RIP-like mutations, the sequence identity between the two species is nearly 100%. Ten genes annotated in *P. tritici-repentis* 1C-BFP, PTRG-04890 to PTRG-04909, are all present in *P. nodorum* SN15 upstream of the 5′ ToxhAT TIR. In *P. nodorum*, however, each of these was a pseudogene due to RIP and therefore has not been annotated in any assembly (Fig. S3) (21, 29).

TABLE 2 Summary of both partial and full-length ToxhATs identified by the repeat annotation pipeline in each isolate

Species and isolate	n[b]	Max length[c]	Min length[c]	Avg length[d]	Med length[d]
B. sorokiniana					
CS10	42	14,079	28	831	251
CS27[a]	22	1,552	26	437	179
WAI2406	35	14,082	28	935	225
WAI2411	41	14,072	28	891	250
P. tritici-repentis					
PTR1C-BPF	29	20,213	32	1,081	188
P. nodorum					
SN15	18	11,153	54	1,219	295
Sn79-1087[a]	8	537	99	266	261

[a]*toxa*$^-$ isolate.
[b]*n*, total number of ToxhAT annotations identified by the repeat annotation pipeline within the genome.
[c]Maximum (Max) or minimum (Min) length of ToxhAT in base pairs identified in the genome.
[d]Average (Avg) or median (Med) length of ToxhAT in base pairs identified in the genome.

Furthermore, 5 of these 10 genes, PTRG-04891 to PTRG-04895, are duplicated and found in inverse orientation within the *P. nodorum* SN15 assembly (Fig. 2B, blue boxes). In *P. tritici-repentis* 1C-BFP, these 10 genes are on a contiguous piece of DNA that extends 61.2 kb upstream of the ToxhAT 5′ TIR. The total length of nearly identical sequence shared between these two species is ~80 kb, which includes 61.2 kb upstream and 1.7 kb downstream of ToxhAT. In *P. nodorum* SN15, the 61.2 kb shared with *P. tritici-repentis* 1C-BFP is present but highly fragmented across chromosome 08, spanning nearly ~370 kb (Fig. 2B). These data demonstrate that a specific HGT event occurred between *P. tritici-repentis* and *P. nodorum* that included ToxhAT and a large segment of surrounding DNA.

WCA between *P. tritici-repentis* and *B. sorokiniana* also revealed ~30 kb outside ToxhAT that was ~50% identical and partially overlapped the DNAs shared between *P. tritici-repentis and P. nodorum* (Fig. 2A [inclusive of genes PTRG-04892 to PTRG-04900]). This indicated that a region outside ToxhAT was potentially also horizontally transferred between *B. sorokiniana* and *P. tritici-repentis*. However, a pairwise alignment of this region showed no evidence of extensive RIP-like mutations that could account for the pairwise nucleotide differences observed between these two species (Fig. S4). Furthermore, we could identify the same region in *toxa*$^-$ isolate *B. sorokiniana* ND90Pr as well as in other closely related *Bipolaris* spp. (Fig. S5). We conclude that this region was not part of a horizontal transfer of ToxhAT into *B. sorokiniana* and instead represents a region of synteny between the two species.

Individual components of ToxhAT are found in other Pleosporales. After careful manual annotation of ToxhAT, we conducted full *de novo* repeat prediction and annotation with the REPET pipeline (38, 39). The total proportion of each genome annotated as repeats is shown in Fig. S6. The nonredundant repeat library generated by REPET included the manually annotated ToxhAT transposon from *B. sorokiniana* CS10, named DTX-comp_CS10_RS_00, and a second nearly full-length version from *P. nodorum*, named DTX-incomp-chim_SN2000-L-B14-Map1. These two sequences were used to identify all instances of ToxhAT within each genome listed in Table 1 and in *P. tritici-repentis* 1C-BFP. A total of 195 instances of ToxhAT were annotated in these seven isolates, 183 (~94%) of which we were able to successfully align to the CS10 ToxhAT sequence (Fig. S7). This alignment showed distinct areas within ToxhAT that were found in high abundance within these genomes, primarily overlapping *CS10_08.708*, *CS10_08.709*, and the Patatin-like gene (Fig. S7). A summary of the total number of identified ToxhAT instances is presented in Table 2. These data show that most of the annotations represent fragments, with the median length ranging from 176 to 295 bp, representing approximately 1% of the total length of ToxhAT.

The large number of partial ToxhAT annotations in *toxa*$^-$ isolates suggested that some regions may be repetitive elements, independent of ToxhAT. To investigate this further, we performed tBLASTn queries using the NCBInr database and the Dothideomycetes genomes available at JGI MycoCosm. In both searches, the hAT transposase, MFS transporter, and Patatin-TPR genes had over 500 partial hits with an E value of less than 1e−10 (Fig. S8A and B). Within the JGI Dothideomycetes database search, *CS10_08.708* had 139 hits, *CS10_08.709* had 85 hits, and the *Tc1* transposase had 263 hits (E value, <1e−10). The small gene *CS10_08.714* had the lowest number of hits, with only four in total across both databases (excluding known instances of ToxhAT). Hits for the *ToxA* gene itself represented mostly distant (<50% identical) homologs, previously described as *ToxA*-like or ToxA* in various *Bipolaris* spp. (40). A short summary of the top hits from both database searches is presented in Table 3.

The hits with the highest identity were in a single *Alternaria alternata* strain, NBRC 8984 (GenBank accession no. AB525198.1), for two contiguous open reading frames, *CS10_08.708* and *CS10_08.709*. The pairwise identity for these two genes exceeded 90%, and the genes are colocalized in *A. alternata* NBRC 8984. This strain was also the highest-identity hit in NCBI for the hAT-transposase. *A. alternata* is a well-known plant pathogen that has a broad host range and is also a member of the Pleosporales. In the JGI database, a fungal isolate collected from leaf litter, *Didymella exigua* CBS 183.55 v1.0, also produced colocalized hits for *CS10_08.708* and *CS10_08.709* with identity greater than 85%, indicating that these two predicted genes may represent a single repetitive unit. This species is also classified in the Pleosporales (41). The third species identified with >90% sequence identity for the hAT transposase was *Decospora gaudefroyi*, again classified in the Pleosporales as a salt-tolerant marine fungus (42). Overall, the large number and breadth of hits across different fungal species confirmed our hypothesis that the individual coding regions of ToxhAT are part of repetitive elements broadly found within the Pleosporales. This indicates a common evolutionary origin of these repetitive coding regions within this fungal Order.

The presence/absence polymorphism of *ToxA* is much larger than the extent of HGT. We showed above that the extent of shared DNA differ between different pairwise comparisons of the three species harboring ToxhAT. This does not address the unknown size of the presence/absence polymorphism maintained in these species. To investigate this, the homologous *toxa*$^-$ and *ToxA*$^+$ chromosomes were aligned (Fig. 3). For *P. tritici-repentis*, where no long-read data for a *toxa*$^-$ isolate were available, short reads from isolate DW5 were aligned against the assembly of *P. tritici-repentis* 1C-BFP. The large spike in coverage for DW5 within the deleted region corresponds to a 7 kb TIR transposon, most likely from the Tc1-Mariner superfamily, which is found near ToxhAT (DTX-incomp-chim_Ptr-L-B62-Map1_reversed). Both of the chromosomes containing *ToxA* in isolates *P. nodorum* SN15 and *B. sorokiniana* CS10 contain telomeric repeats and are complete chromosomes. This shows that the absence polymorphism spans several thousand kilobases in all three species (Fig. 3). Using the last known homologous regions from the whole-chromosome alignments, the absence alleles in *B. sorokiniana* CS27, *P. nodorum* Sn79-1087, and *P. tritici-repentis* DW5 were estimated to span ~239 kb, ~467 kb, and ~150 kb, respectively.

To compare the size of the presence/absence polymorphism of *ToxA* with the sizes of two other well characterized necrotrophic effectors, we examined the location of *SnTox3* and *SnTox1* in *P. nodorum* SN15 (43). These two effectors also exist as a presence/absence polymorphism in this species but have no known history of HGT. Sn*Tox3* is the last annotated gene on chromosome 11 in isolate SN15, and the absence polymorphism approximately corresponds to the 7 kb tail of this chromosome. This absence encompasses annotated SN15 genes SNOG_08981 (*Tox3*) to SNOG-08984 (Fig. S9A). The end of chromosome 11 in the assembly of Sn79-1087 (which lacks *SnTox3*) contains telomeric repeats (data not shown), which suggests that the absence of the 7 kb is not due to an incompletely assembled chromosome. The absence allele of *SnTox1* is even smaller, spanning ~3 kb on chromosome 6 of *SN15*. At this locus, there is a unique insertion of ~1.3 kb which is present only in Sn79-1087 (Fig. S9B).

TABLE 3 Summary of top BLAST hits excluding known examples of ToxhAT in *P. tritici-repentis*, *P. nodorum*, and *B. sorokiniana*[a]

Query	Sequence length (aa)	Database	Genome/ BLAST description	Class[c]/ order	Location or accession no.	% identity	% query coverage	E value	Bit score
CS10_ 08.708	84	JGI	*Didymella exigua* CBS 183.55 v1.0[b]	Pleosporales	Didex1|scaffold_87: 8242–8451	85.7	83.3	9.89E−36	305
			Didymella exigua ArDII	Pleosporales	Ascra1|scaffold_164: 46459–46707	71.1	98.8	9.28E−33	284
			Lizonia empirigonia CBS 542.76 v1.0	Dothideomycetes[c]	Lizem1|scaffold_85: 18354–18605	77.1	72.6	2.60E−34	251
		NCBI	*Alternaria alternata* DNA, AMT genes region, strain: NBRC 8984[b]	Pleosporales	AB525198.1	90.5	100	3.23E−34	129
			Bipolaris maydis clone FNFP145-M02, complete sequence	Pleosporales	AC277024.1	67.3	65.5	4.28E−13	69.7
			Bipolaris maydis clone FNFP209-G23, complete sequence	Pleosporales	AC277374.1	67.3	65.5	4.28E−13	69.7
CS10_ 08.709	55	JGI	*Didymella exigua* CBS 183.55 v1.0[b]	Pleosporales	Didex1|scaffold_87: 7965–8131	90.7	78.2	2.63E−27	201
			Phoma multirostrata 7a v1.0	Pleosporales	Phomu1|scaffold_34: 319170–385826	81.4	78.2	7.06E−24	183
			Lizonia empirigonia CBS 542.76 v1.0	Dothideomycetes[c]	Lizem1|scaffold_39: 422775–422939	69.1	100	1.10E−18	177
		NCBI	*Alternaria alternata* DNA, AMT genes region, strain: NBRC 8984[b]	Pleosporales	AB525198.1	95.3	78.2	3.48E−18	82.4
			Bipolaris maydis ATCC 48331 hypothetical protein mRNA	Pleosporales	XM_014219256.1	69.6	100	3.85E−14	69.7
			Bipolaris maydis clone FNFP145-M02, complete sequence	Pleosporales	AC277024.1	69.6	100	1.04E−13	69.7
CS10_hAT	647	JGI	*Decorospora gaudefroyi* v1.0	Pleosporales	Decga1|scaffold_269: 17873–19837	90.6	56.1	0.00E+00	1653
			Ophiobolus disseminans CBS 113818 v1.0	Pleosporales	Ophdi1|scaffold_7: 902547–1437652	87.9	56.1	0.00E+00	1614
			Pyrenophora tritici-repentis 1C-BFP	Pleosporales	Pyrtr1|supercontig_1.22: 2970–154460	83.2	56.1	0.00E+00	1496
		NCBI	*Alternaria alternata* DNA, AMT genes region, strain: NBRC 8984	Pleosporales	AB525198.1	53.5	88.3	0.00E+00	612
			Alternaria alternata TLS-S1-3 transposase pseudogene sequence	Pleosporales	AB236735.1	44.4	88.1	6.14E−164	507
			Alternaria alternata TLS-S1-2 transposase pseudogene sequence	Pleosporales	AB236734.1	44.4	88.1	6.14E−164	507
CS10_ToxA	178	JGI	*Cochliobolus heterostrophus* C5 v2.0	Pleosporales	CocheC5_3|scaffold_2: 2350132–2350380	54.6	24.7	1.44E−16	126
			Cochliobolus heterostrophus C4 v1.0	Pleosporales	CocheC4_1|scaffold_114: 2535–2783	54.6	24.7	1.50E−16	126

(Continued on next page)

TABLE 3 (Continued)

Query	Sequence length (aa)	Database	Genome/ BLAST description	Class[c]/ order	Location or accession no.	% identity	% query coverage	E value	Bit score
			Cochliobolus heterostrophus C5 v2.0	Pleosporales	CocheC5_3|scaffold_2: 2350132–2350380	36.8	21.3	1.44E−16	85
		NCBI	*Bipolaris maydis* strain C4 ToxA-like protein (TOXA) mRNA, complete	Pleosporales	KJ664925.1	43.1	78.7	4.92E−26	105
			Bipolaris maydis ATCC 48331 hypothetical protein mRNA	Pleosporales	XM_014216967.1	43.1	78.7	5.78E−26	105
			Bipolaris maydis strain C5 ToxA-like protein (TOXA) mRNA, complete	Pleosporales	KJ664924.1	43.1	78.7	9.98E−26	105
MFS	202	JGI	*Setosphaeria turcica* NY001 v2.0	Pleosporales	Settur3|scaffold_53: 4328–83573	70.3	55.0	5.35E−51	419
			Setosphaeria turcica Et28A v2.0	Pleosporales	Settu3|scaffold_6: 884922–885254	70.3	55.0	1.48E−50	419
			Trematosphaeria pertusa CBS 122368 v1.0	Pleosporales	Trepe1|scaffold_8: 1918893–1919171	71.0	46.0	6.11E−44	372
		NCBI	*Setosphaeria turcica* Et28A hypothetical protein partial mRNA	Pleosporales	XM_008031524.1	68.1	57.4	1.35E−46	168
			Parastagonospora nodorum SN15 hypothetical protein (SNOG_12844),	Pleosporales	XM_001803010.1	64.9	46.5	1.22E−33	132
			Parastagonospora nodorum isolate Sn2000 chromosome 18 sequence	Pleosporales	CP022843.1	64.9	46.5	2.13E−33	133
CS10_ 08.714	85	JGI	*Alternaria alternata* SRC1lrK2f v1.0	Pleosporales	Altal1|scaffold_52: 12656–12787	77.3	51.8	1.70E−15	170
			Ophiobolus disseminans CBS 113818 v1.0	Pleosporales	Ophdi1|scaffold_44: 27494–27639	70.3	43.5	1.46E−11	128
			Ophiobolus disseminans CBS 113818 v1.0	Pleosporales	Ophdi1|scaffold_44: 27494–27639	81.8	12.9	1.46E−11	48
		NCBI	*Alternaria alternata* hypothetical protein partial mRNA	Pleosporales	XM_018535381.1	77.3	51.8	2.41E−14	67.4
			Alternaria alternata FabD/ lysophospholipase-like protein mRNA	Pleosporales	XM_018526331.1	77.3	51.8	3.66E−12	67
Patatin	936	JGI	*Setomelanomma holmii* CBS 110217 v1.0	Pleosporales	Setho1|scaffold_339: 2–8483	79.6	17.3	0.00E+00	685
			Cochliobolus victoriae FI3 v1.0	Pleosporales	Cocvi1|scaffold_110: 673–4261	77.1	18.2	0.00E+00	680
			Clohesyomyces aquaticus v1.0	Pleosporales	Cloaq1|scaffold_160: 91821–95443	77.8	17.3	0.00E+00	677
		NCBI	*Alternaria alternata* FabD/ lysophospholipase-like protein mRNA	Pleosporales	XM_018526331.1	73.2	100	0.00E+00	1383
			Bipolaris victoriae FI3 hypothetical protein partial mRNA	Pleosporales	XM_014696900.1	73.2	92.3	0.00E+00	1348

(Continued on next page)

TABLE 3 (Continued)

Query	Sequence length (aa)	Database	Genome/ BLAST description	Class[c]/ order	Location or accession no.	% identity	% query coverage	E value	Bit score
			Parastagonospora nodorum SN15 hypothetical protein (SNOG_12454)	Pleosporales	XM_001802625.1	70.3	93.9	0.00E+00	1258
Tc1	75	JGI	*Phoma tracheiphila* IPT5 v1.0	Pleosporales	Photr1|scaffold_21: 57505–57729	62.7	100	3.31E−24	228
			Phoma tracheiphila IPT5 v1.0	Pleosporales	Photr1|scaffold_93: 22667–22891	61.3	100	5.80E−23	219
			Setomelanomma holmii CBS 110217 v1.0	Pleosporales	Setho1|scaffold_126: 700–924	52.0	100	3.45E−19	192
		NCBI	*Rasamsonia emersonii* CBS 393.64 hypothetical protein partial	Eurotiomycetes[c]	XM_013470328.1	40.0	100	4.95E−11	60.5
			Rasamsonia emersonii CBS 393.64 hypothetical protein partial	Eurotiomycetes[c]	XM_013469195.1	40.0	100	5.27E−11	60.5
			Rasamsonia emersonii CBS 393.64 hypothetical protein partial	Eurotiomycetes[c]	XM_013473453.1	44.9	92.0	5.61E−11	59.3

[a]Only the top 3 BLAST hits are shown sorted according to the bit score, which is a measure of pairwise sequence similarity. aa, amino acids.
[b]Data indicate hits of interest with >85% pairwise nucleotide identity and >78% query coverage.
[c]Class data are indicated for species outside the Dothideomycetes or for which the order is unknown.

These data demonstrate that the absence allele of the horizontally transferred *ToxA* gene is much larger than absence alleles of other known effectors and highlight potential genome instability after HGT events.

Evidence of mobility of ToxhAT in *B. sorokiniana*. The finding of intact TIRs in *B. sorokiniana* CS10 suggested that ToxhAT in this species may remain mobile. To investigate this, we resequenced two additional *ToxA*$^+$ isolates of *B. sorokiniana* (WAI2406 and WAI2411). ToxhAT was found in different genomic locations in the two genomes compared to isolate CS10, where ToxhAT is located near the end of chromosome 08 (Fig. 4A). For WAI2406, ToxhAT and the surrounding ~200 kb of repeat-rich DNA were found imbedded in the middle of chromosome 01 (Fig. 4A to C). This has led to an increase in the size of WAI2406's chromosome 01, which is ~4.0 Mbp compared to CS10's PacBio assembled chromosome 01 size of ~3.8 Mbp. To confirm that this translocation was not a misassembly, we aligned the corrected Nanopore reads from Canu to both the CS10 and WAI2406 assemblies (Fig. 4B and C). These reads aligned, with a slope of 1, to chromosome 01 in isolate WAI2406, with single reads clearly spanning the breakpoints on both sides of the translocation. These same reads from isolate WAI2406 did not align well to chromosome 08 in isolate CS10.

For isolate WAI2411, ToxhAT assembled to a small contig (~776 kb) that has homology to chromosome 02 and chromosome 08 of the PacBio assembled CS10 (data not shown). While it remains unclear if this small scaffold is a part of chromosome 02 or chromosome 08, the flanking DNA on either side of ToxhAT was conserved but shuffled in order (Fig. 4D). ToxhAT was found to be inverted and in a different position from that in chromosome 08 in CS10. The breakpoints of the inversion were precisely from TIR to TIR of ToxhAT. Again, we aligned the corrected nanopore reads to isolate WAI2411 (self) and isolate CS10. The self-alignment showed reads that clearly crossed the breakpoints of the inversion/transposition (Fig. 4E); however, these same reads did not map contiguously to the CS10 assembly (Fig. 4F). As a secondary confirmation, we used BLASTn to identify all reads that contained *ToxA* and generated a multiple-sequence alignment (Fig. S10). Twenty-nine single-molecule reads were aligned that spanned ToxhAT and continued into both flanking regions. These flanking regions

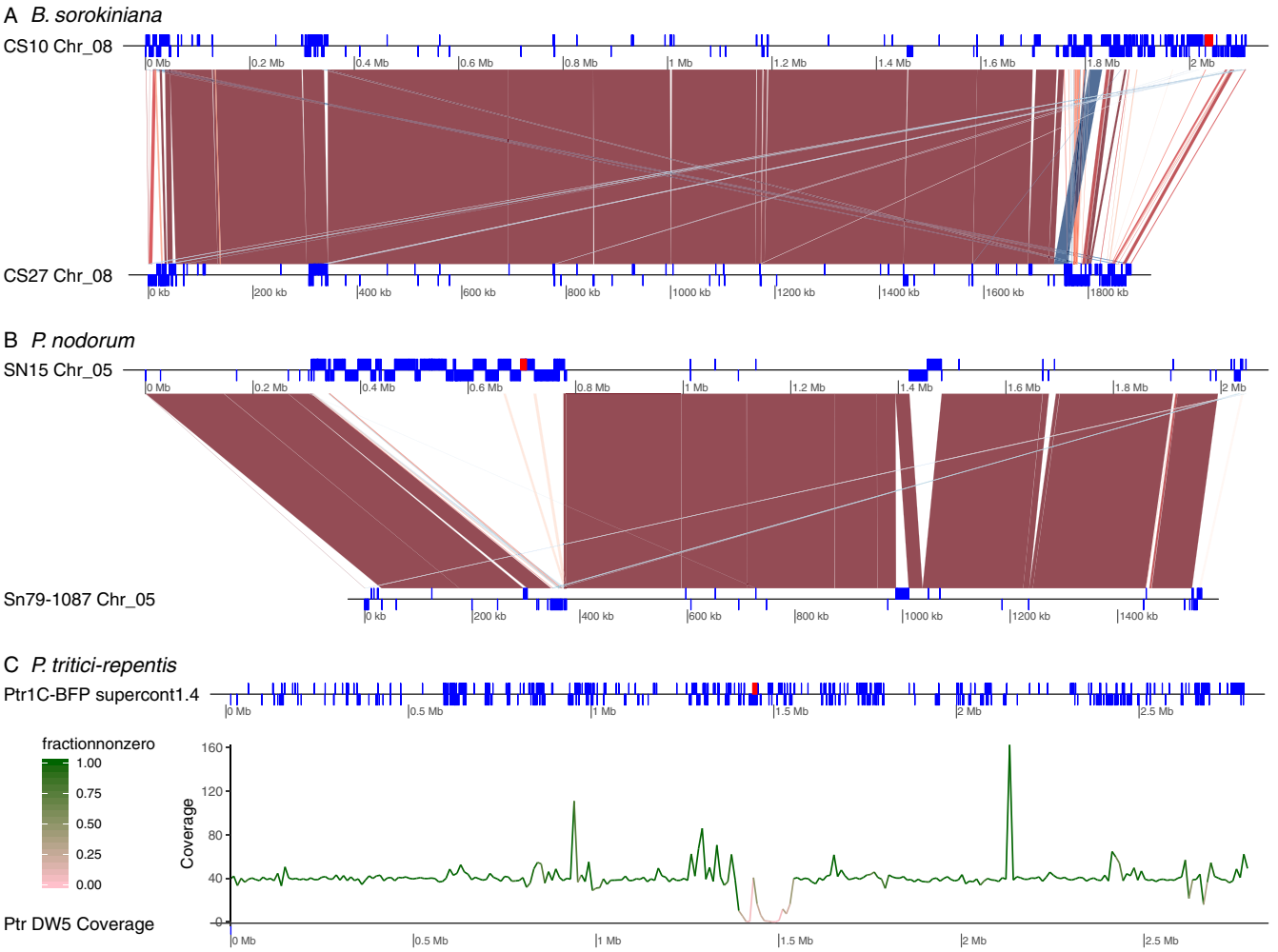

FIG 3 Genomic context of the ToxhAT-containing region (red box) in each of the three species in comparison to *toxa*$^{-}$ isolates. (A) Lastz alignment of the homologous chromosome between *B. sorokiniana ToxA*$^{+}$ isolate CS10 and *toxa*$^{-}$ isolate CS27. Blue blocks drawn on the chromosome maps (black lines) represent the location of annotated transposons within each genome. Red ribbons drawn between the two isolates represent syntenic alignments found in Lastz that showed more than 70% identity and were greater than 2 kb in length. Blue ribbons drawn between the two isolates show inversions between the two genomes. (B) Lastz alignment of the homologous chromosome between *P. nodorum ToxA*$^{+}$ isolate SN15 and *toxa*$^{-}$ isolate Sn79-1087. The coloring scheme is the same as that used in panel A. (C) Isolate *Ptr*1C-BFP with repeat regions shown in the blue blocks along the chromosome (black line). The average Illumina coverage for 10 kb windows across the chromosome is indicated at the bottom. The color of the line corresponds to the proportion of bases within the 10 kb window that had nonzero coverage.

aligned well to WAI2411's contig 17, confirming the inversion of ToxhAT precisely at the TIRs (Fig. 4E). We postulated that this inversion could represent an active transposition event, in which case there might be signature target-site duplications (TSDs) made by the transposase. However, neither this inversion nor any other instance of ToxhAT in any sequenced isolate was flanked by identifiable target-site duplications.

DISCUSSION

As shown in previous studies, the *ToxA* gene and surrounding noncoding DNA are highly conserved between these three species (17, 20, 30). We extend this knowledge by defining flanking TIRs that give this region the structural features of a type II DNA transposon, defined here as ToxhAT. These TIRs and the enclosed 14.3 kb are conserved in all three fungal species, indicating that ToxhAT had a common evolutionary origin prior to HGT. The finding of homologous DNA shared between *P. nodorum* and *P. tritici-repentis* outside these TIRs indicates that the HGT event between these two species did not involve ToxhAT alone but included ~63 kb of flanking DNA. In contrast, homology between these two species and *B. sorokiniana* breaks precisely at the ToxhAT

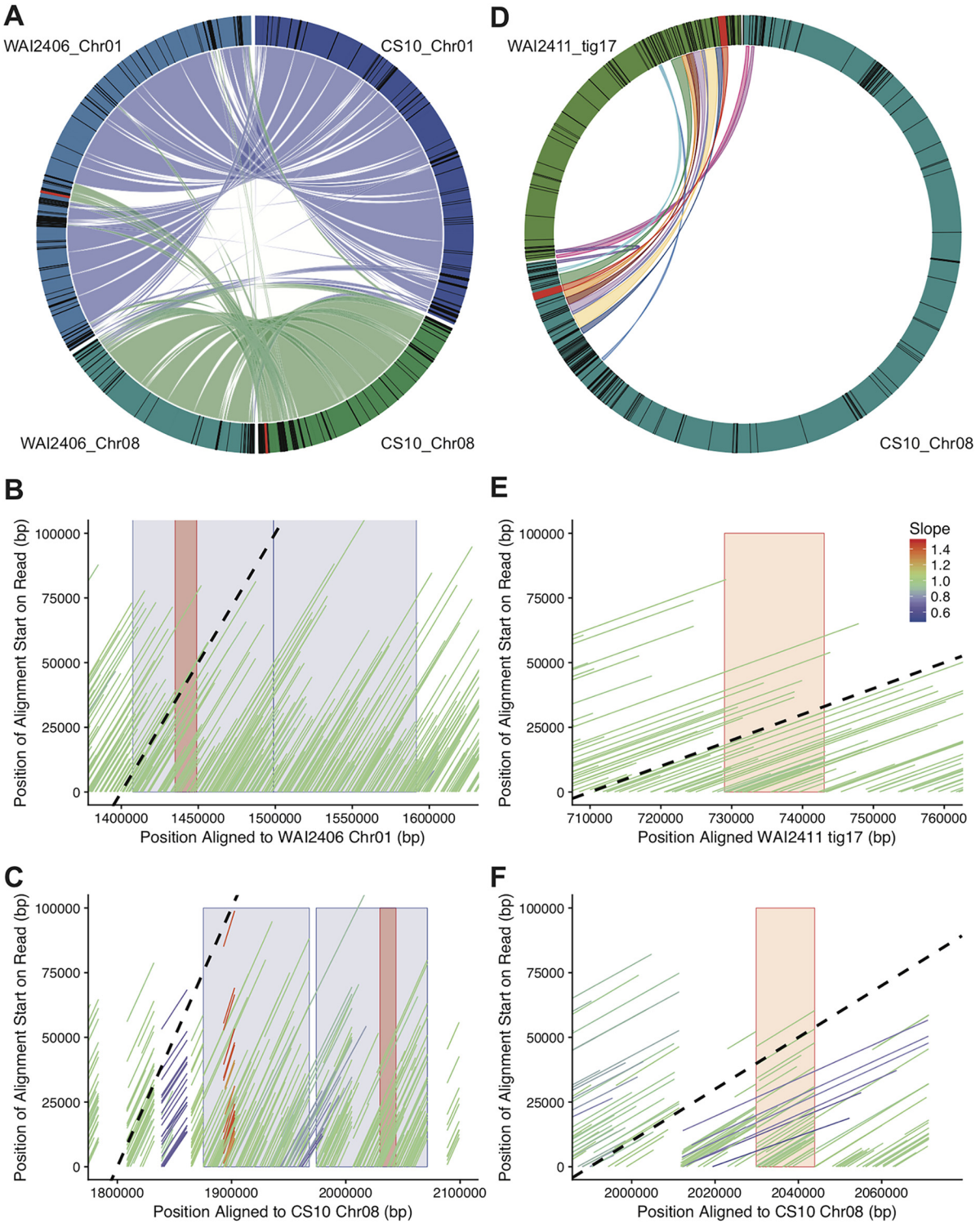

FIG 4 ToxhAT within *B. sorokiniana* is mobile in two distinct ways. (A) Alignment of chromosomes 01 and 08 of CS10 against WAI2406 homologous chromosomes. The ToxhAT (red line/box drawn on outer circle) is located on Chr01 in WAI2406 along with a large region of repeat-rich DNA (black boxes in outer circle). (B) The corrected and trimmed WAI2406 Nanopore reads aligned to the *de novo* version of itself. The black dotted line shows a slope value of 1, which indicates that the read aligned base per base to the chromosome shown on the *x* axis. Reads with a slope value different from 1 are reads that have been mapped discontinuously (i.e., with large insertions or deletions). The blue blocks show the translocated DNA, found in chromosome 08 in CS10 but in chromosome 01 in WAI2406. The red box indicates the position of ToxhAT. (C) The data correspond to the same reads as those described for panel B but represent alignment to chromosome 08 of CS10. Note that the read alignment is not continuous and breaks at the translocation edges. (D) Alignment of chromosome 08 from CS10 against tig17 from WAI2411. (E) The data are presented as described for panel B but represent isolate WAI2411. The red block shows the position of ToxhAT. (F) The data are presented as described for panel C but represent isolate WAI2411. Note that no reads with a slope value of 1 extend beyond the ToxhAT itself (red box).

TIRs. Within *B. sorokiniana*, there is strong evidence that ToxhAT and the surrounding repeat-rich DNA remain mobile within the genome. This finding contrasts with two previously published long-read assemblies of *P. nodorum* and *P. tritici-repentis*, which show *ToxA* and the surrounding horizontally transferred (HT) DNA on the same chromosome (23, 29, 30). The individual coding regions found within ToxhAT appear to be part of repetitive elements in other Dothideomycetes. However, the breadth of hits gives no indication of whether any single species could have assembled ToxhAT as a unit before horizontal transfer between these three wheat pathogens.

ToxhAT is structured like a type II DNA transposon, but mobility may be facilitated by recombination. By identifying conserved TIRs in all three pathogens, we describe a unit of DNA that has the structural features of a type II DNA TIR transposon. In *B. sorokiniana* isolate WAI2411, ToxhAT itself is inverted precisely at the TIRs. This inversion bounded precisely by these structural features suggested that this putative transposon may remain active. However, extensive searches of the flanking DNA in all three species did not reveal TSDs typical of other TIR transposases (44, 45). TSDs are created by a transposase when it cuts at its target site, usually a short sequence ranging in length from 2 to 11 bases (34). Most transposases make an uneven cut which, after DNA repair, leads to a duplication of the target site on either side of the inserted transposon (34). The absence of TSDs in WAI2411 suggests that the inversion seen in isolate WAI2411 was not facilitated by an active transposition event. An alternative mechanism that could explain the inversion is intrachromosomal recombination between these structural features (25). In the absence of TSDs, we consider this mechanism a strong alternative to explain the movement of ToxhAT in WAI2411.

While WAI2411 showed a relatively small genomic rearrangement, the other resequenced *B. sorokiniana* isolate, WAI2406, contained a large segmental movement of 200 kb from one chromosome to another. Similar interchromosomal translocations were observed previously in the fungal plant pathogens *Verticillium dahliae*, *Magnaporthe oryzae*, and *Colletotrichum higginsianum* (46–48). While all three studies demonstrated that translocations occur in regions where transposons are prevalent, only the study by Faino et al. (47) in the asexual species *V. dahliae* was able to show homologous transposons/DNA sequence at the translocation breakpoints. These breakpoints in *V. dahliae* implicate homologous recombination as the mechanism underpinning genome plasticity in this species (47). In *M. oryzae* and *C. higginsianum*, transposons are associated with translocation events, but it remains unclear if these translocations occurred over regions with high sequence identity (46, 48). Similarly, in *B. sorokiniana*, whole-chromosome alignment of chromosome 08 and chromosome 01 in isolate WAI2406 did not show high sequence identity in the regions outside the translocation breakpoints on either chromosome. However, the regions near these breakpoints are enriched in transposon annotations. A key difference between the three fungal species examined in this study and *Verticillium* is their sexual life cycle. Meiotic recombination could obscure the precise translocation boundaries and is also a prerequisite for RIP. We postulate that in sexual fungal species, AT-rich regions, with otherwise limited sequence identity, may undergo "homologous" interchromosomal recombination events like those observed in *B. sorokiniana*. Supporting this hypothesis is a recent study on *Epichloë festucue* that leveraged Hi-C data to build a contact map of DNA within the nucleus (49). The Hi-C sequencing technique shows the frequency at which different regions of a genome interact with each other in the nucleus (50). For example, DNA fragments that are physically close to each other, on the same chromosome, have a higher frequency of interaction than sequences on different chromosomes (49). A study by Winter et al. showed that AT-rich regions, usually in the form of heavily RIPed transposon islands, on different chromosomes had significantly higher interactions with each other than non-AT-rich regions (49). These data suggest that in sexual fungi, where RIP is active, AT-rich islands are associated with each other in three-dimensional (3D) space. We postulate that AT-rich regions in *B. sorokiniana*, like those in *E. festucue*, also associate with each other within the nucleus, which provides the physical proximity required for interchromosomal recombination events.

The opportunity to examine interchromosomal transfer of ToxhAT extends beyond *B. sorokiniana*. Chromosomal movement of the *ToxA* gene was also observed in *P. tritici-repentis* (51). In that study, pulsed-field gel electrophoresis followed by Southern hybridization was used to show that *ToxA* was found on chromosomes of different sizes in *P. tritici-repentis*. Going further, the authors showed that, in at least one isolate, *ToxA* was on a chromosome different from the *ToxA* location in reference isolate 1C-BFP (51). While that study probed for the *ToxA* gene alone, we consider it likely that ToxhAT or a larger chromosomal segment was translocated, similarly to what was observed in *B. sorokiniana*. Further long-read assembly coupled with Hi-C data from both *P. tritici-repentis* and *B. sorokiniana*, ideally from a sexual cross of two previously sequenced isolates, is required to systematically reconstruct the level of sequence identity or other genome features that facilitate interchromosomal translocations.

ToxhAT resides in an accessory genomic region in all three species. The analysis of the syntenic relationship between *ToxA*$^+$ and *toxa*$^-$ chromosomes within each species showed that the absence of ToxhAT is coincident with the absence of large (>100 kb) chromosomal segments. The DNA composition of these regions fits well within the definition of the "lineage-specific" or "accessory" regions described in pathogenic fungi, where virulence genes are found nested within transposon-rich regions of the genome (52–55). This genome structure is hypothesized to facilitate the rapid adaptation of fungal pathogenicity genes, often referred to as the "two-speed" or "two-compartment" genome (56, 57). These data further show that the absence polymorphism does not coincide with the exact boundaries of HGT. This is most clearly seen in *P. nodorum*, where fragments of the horizontally transferred DNA are scattered across a 300 kb region. However, the entire region, extending well beyond the horizontally transferred fragments, is missing from *toxa*$^-$ isolate Sn79-1087. This leads to an interesting issue concerning whether the horizontal acquisition of ToxhAT precipitated the expansion of transposons within this region. Our data for *SnTox3* and *SnTox1* in *P. nodorum* suggest that this may be the case, whereby the absence alleles for these well-characterized effectors span only a few kilobases. Unfortunately, similar comparisons were not possible in *B. sorokiniana* due to a lack of known effectors or in *P. tritici-repentis*, where the only other known effector, *ToxB*, is present in multiple copies in the genome (58). Intrachromosomal recombination is also a possible mechanism to generate these large absence polymorphisms. In many model organisms, such as *Drosophila*, yeast, and human cell lines, large segmental deletions were previously found to be facilitated by ectopic recombination between tandemly arrayed repeat sequences (59, 60). This mechanism is particularly interesting in the context of HGT, as these recombination events often result in the formation of circular extrachromosomal DNA (61, 62).

The origins of ToxhAT remain obscure. Since the discovery of *ToxA* in the genome of *P. nodorum*, the evolutionary origin of this gene has been a topic of debate (20, 23, 63, 64). To date, *P. nodorum* remains the species with the highest known *ToxA* sequence diversity. This diversity underpins the prevailing hypothesis that *ToxA* has had the most time to accumulate mutations and therefore has resided in the genome of this organism longest (20, 63). The discovery of *ToxA* in *B. sorokiniana* and the characterization of the conserved 74 bp TIRs in all three species strongly suggest that ToxhAT had a single evolutionary origin in all three species. The ToxhAT TIRs in *B. sorokiniana* define the exact boundaries of the HGT event, where sequence identity with the other two species abruptly ends. In contrast, the HGT between *P. tritici-repentis* and *P. nodorum* included ToxhAT and 63 kb of flanking DNA. In *P. tritici-repentis* 1C-BFP, this flanking DNA remains contiguous; in *P. nodorum*, however, this same DNA is fragmented and partially duplicated across a 370 kb island of RIPed transposable elements (TEs). Together, these data suggest that ToxhAT was horizontally transferred in two separate events both with and without flanking DNA.

We propose two opposing models to explain the HGT of ToxhAT between the three species. In the first model, we assume a population genetic perspective where the

longer the DNA has resided in the genome, the more fragmented and dispersed the HT DNA becomes. In this model, *P. nodorum* would be the donor of ToxhAT along with the flanking DNA to *P. tritici-repentis*. This is based on our observation that the flanking DNA outside ToxhAT is highly fragmented and duplicated in *P. nodorum*, indicating this species has had more time to accumulate these changes. This model assumes that the *P. nodorum* donor isolate once had a contiguous stretch of DNA as observed in *P. tritici-repentis*. In this model, we also postulate that ToxhAT was most recently acquired by *B. sorokiniana* due to its relatively compact form and conservation of structural features. Again, for this model to hold, we must assume that an intact form of ToxhAT exists in the population of *P. nodorum* or *P. tritici-repentis* that could act as a donor to *B. sorokiniana*. Our second model assumes that changes can accumulate more rapidly in transposon-rich regions and therefore sequence diversity in these regions is not a good indication of evolutionary time. In this model, the intact version of ToxhAT observed in *B. sorokiniana* would represent an ancestral version. This is the minimal unit of HT DNA, bounded by conserved TIRs. In this model, we propose that the first HGT event was from *B. sorokiniana* to *P. tritici-repentis*, based on the identical *ToxA* sequences that they share. Then, in a second HGT event, *P. tritici-repentis* would have been the donor of the large segment of DNA inclusive of ToxhAT to *P. nodorum*. The HT DNA flanking ToxhAT was then subject to rapid decay and duplication in *P. nodorum*, due to its proximity to transposable elements. This model does not provide a good explanation of why the rapid decay is not also observed in the other two species.

In order to differentiate between these two models, we suggest long-read sequencing across the broadest possible sample times and locations for both *ToxA*$^+$ and *toxa*$^-$ isolates in all three species. The first goal of this sequencing would be to systematically characterize the sequence diversity of ToxhAT variants all three species. The presence of any "intact" ToxhAT sequences coupled with overall higher ToxhAT sequence diversity in isolates of *P. nodorum* and *P. tritici-repentis* would support the first model. Second, we suggest searching the sequence of *toxa*$^-$ isolates for the HT DNA flanking ToxhAT in *P. nodorum* and *P. tritici-repentis*. For example, if this region were present in a *toxa*$^-$ *P. nodorum* isolate, this would identify *P. nodorum* as the donor of ToxhAT and its flanking DNA to *P. tritici-repentis*. Determining the direction of transfer between *B. sorokiniana* and the other two species is more challenging, given the overall high sequence conservation of ToxhAT. Support for the model where *B. sorokiniana* harbors the ancestral ToxhAT could come from sequence data that show that ToxhAT in this species has a level of sequence diversity similar to that seen with the other two species.

While we have presented two models above that describe the history of HGT between these species alone, results of the BLAST searches conducted on the coding regions annotated in *B. sorokiniana* indicate that there are other species which may harbor highly identical components of ToxhAT. One standout isolate is *A. alternata* strain NBRC8984, which carries two genes that are 90% and 95% identical to *CS10_08.708* and *CS10_08.709*, respectively. Similarly to their arrangement in ToxhAT, these two coding sequences also neighbor each other in *A. alternata* NBRC8984. These hits within ToxhAT are by far the most similar sequences found in a species that is not reported to carry ToxA. These high-identity hits also included some nonpathogenic and marine species also found within the Pleosporales. Collectively, the coding regions within ToxhAT had hundreds of hits across species representing several hundred million years of evolution. However, none of these coding regions have been characterized as repetitive or classified in a transposon family. Despite the ancient evolutionary history of transposons, the vast majority of described DNA transposons with TIRs are classified into only 10 superfamilies (34, 65). Our detailed characterization of ToxhAT highlights an opportunity to characterize novel transposon superfamilies in nonmodel fungi.

Towards a mechanism; flanking noncoding DNA provides clues. The tBLASTn results, coupled with a detailed structural characterization of ToxhAT, suggest that it is

a mosaic of repetitive coding regions. We propose that ToxhAT was transferred horizontally as, or by, a transposon, with the fitness advantage of *ToxA* fixing these HGT events in three wheat-infecting species. Similarly, the horizontally transferred regions in the cheese-making *Penicillium* spp. were flanked by unusual *i* non-LTR retrotransposons (13). In the present study, it remained unclear whether ectopic recombination, rather than active transposition, is the mechanism that integrates the HT DNA into the recipient genome. Horizontal transfer of transposons (HTT) has been widely reported in eukaryotes since the early discovery of P elements in *Drosophila* (66, 67). The literature on this topic, however, seems to clearly divide HGT from HTT as two separate phenomena, the latter being much more common (68–70). Recent reports of the HTT between insects has used noncoding regions flanking horizontally transferred genes to demonstrate that a viral parasite, with a broad insect host range, is the vector for the horizontally transferred DNA (71). This report highlights how insights from noncoding regions can bring these studies closer to a mechanistic understanding of the HGT event (71, 72). Other studies which report HGT of secondary metabolite clusters into and between fungal species often rely on phylogenetic methods performed using coding regions alone to detect these events (73–76). While these studies focus on the biological significance of the coding regions, clues to a possible mechanism may remain in the surrounding noncoding DNA.

One limitation encountered with early Illumina-based genome assemblies was the inability to correctly assemble highly repetitive regions. Here we demonstrated with two long-read sequencing technologies that it is possible to assemble very large repetitive regions heavily affected by RIP. These assemblies allow us to look at the noncoding and transposon-rich regions that may be important for the movement or integration of HT DNA. Further population-scale long-read sequencing will enable further refinement of our understanding of the role that transposons play in facilitating adaptive gene transfer.

MATERIALS AND METHODS

Fungal culture and DNA extraction. Fungal cultures were grown on V8-potato dextrose agar (PDA) media at 22°C under conditions of a 12-h light/dark cycle (77). Cultures ranging in age from 5 to 10 days were scraped from the surface of agar plates into water by the use of a sterile razor blade. These harvested cultures were freeze-dried for 48 h to remove all water. High-molecular-weight (HMW) DNA was extracted using a C-TAB phenol-chloroform method modified from Fulton et al. (1995) (78). Full details of our protocol, including gel images of the final DNA, are available at https://doi.org/10.17504/protocols.io.hadb2a6. DNA size was assessed by pulsed-field electrophoresis and DNA purity by examining 260/280 and 230/280 UV absorbance ratios on a NanoDrop spectrophotometer (Thermo Scientific, USA). The total quantity of DNA was measured using a Qubit fluorometer (Life Technologies, USA).

Genome sequencing and assembly. (i) PacBio DNA sequencing. Isolates SN15 (*P. nodorum*) and CS10 (*B. sorokiniana*; BRIP10943) were sequenced using Pacific Biosciences SMRT sequencing. Genomic P6 SMRT cell library preparations (15 to 20 kb) were made at the Ramaciotti Centre for Genomics (University of New South Wales [UNSW] Sydney, Sydney, Australia). Each library was sequenced on 7 SMRT cells by the use of P6-C4 chemistry on the PacBio RSII instrument (Pacific Biosciences, USA). Libraries were size selected on a Blue Pippen from 15-50 kb. Each isolate genome was assembled *de novo* using Canu v1.5 and a minimum read length of 2,000 bp (79). The Canu parameter "genomeSize" was set to 41 Mb for isolate SN15 and 35 Mb for isolate CS10. Canu assemblies were further corrected using SMRT Analysis package v2.3.0. First, the raw PacBio reads were mapped to the *de novo* assembly using blaser with the following settings: --seed = 1 --minAccuracy = 0.75 --minLength = 500 --forQuiver --algorithmOptions='-minMatch 12 -bestn 1 -minPctIdentity 65.0 –hitPolicy=randombest'. The resulting bam file was used as input for Quiver to call a new consensus sequence with the following settings: makeVcf=True, makeBed=True, enableMapQVFilter=T, minConfidence = 40, minCoverage-10, diploidMode=False (80).

(ii) Nanopore DNA sequencing. Resequencing of *B. sorokiniana* isolates CS27 (BRIP27), WAI2406, and WAI2411 and *P. nodorum* isolate Sn79-1087 was performed on Oxford Nanopore's MinION sequencer. R9.4 flow cells were used for sequencing, and a SQK-LSK08 1D library kit was used to prepare the libraries according to the manufacturer's protocol. All DNA samples were purified using Agencourt AMPure beads prior to starting the 1D library preparation (Beckman Coulter, Inc., CA, USA). Genomes were assembled with Canu v1.5 or v1.6 with a minimum read length of 5 kb (79). *De novo* genome assemblies were corrected using the trimmed reads output from Canu. Trimmed reads were mapped to the genome with Minimap2 followed by correction with Racon (81). The output consensus sequence from Racon was used as the input for additional corrections steps performed iteratively up to five times. Sn79-1087 and CS27 assemblies were further refined using Pilon software; this correction was performed

iteratively up to five times (82). Sn79-1087 Illumina data were taken from Syme et al. (21) and BRIP27 Illumina from McDonald et al. (17).

RNA sequencing to aid annotation of long-read assemblies. *B. sorokiniana* isolate CS10 was cultured for 10 days in a range of liquid growth media, including V8-juice broth (potato dextrose broth [PDB]) (77); PDB supplemented with the epigenetic modifier 5-azacytidine (150 μM) (83); minimal media (84); wheat extract-supplemented minimal media; and Fries 3 and Fries 3 modified media (85). Mycelia were harvested, and total RNA was extracted with a Zymo Research fungal/bacterial RNA miniprep kit. RNA quality assessment (Agilent Bioanalyzer), library preparation (strand-specific TruSeq v3), and Illumina RNA sequencing (MiSeq; 150-bp single-end reads) were performed at the Ramaciotti Centre for Genomics (UNSW Sydney, Sydney, Australia).

Long-read RNA-seq data were also generated using Oxford Nanopore's MinION sequencer. Total RNA extracted from mycelia cultured in Fries 3 medium was enriched for mRNA by the use of NEBNext oligo(dT) magnetic beads and concentrated by the use of Agencourt RNAClean XP magnetic beads (Beckman Coulter, Inc., CA, USA). RNA-seq and cDNA-PCR libraries were generated directly with SQK-RNA001 and SQK-PCS108 library preparation kits, respectively, and sequenced with R9.4 flow cells. Reads were base called with Albacore v2.0.2, quality filtered with Nanofilt, and subjected to error correction using the CS10 genome sequence with Proovread (default settings) (86). Error-corrected reads were filtered for reads corresponding to full-length transcripts using SQANTI (run with default settings) (87).

Annotation of long-read assemblies. Illumina RNA-seq data were used for gene prediction in both CS10 and CS27 *B. sorokiniana* isolates. Reads were trimmed with Trimmomatic v0.32 (parameters: -phred 33, ILLUMINACLIP TruSeq3- SE.fa:2:30:10, SLIDINGWINDOW:4:20, LEADING:20, TRAILING:20, and MINLEN: 75) and aligned to the genomes using STAR v2.5 (parameters: --alignIntronMin 10, --alignIntronMax 1000, --twoPassMode Basic), and then transcript assembly was performed using StringTie v1.3.3 (default parameters, except for --f 0.3). StringTie transcripts were filtered for high-quality ORFs by the use of TransDecoder v5.0.2 (88–90).

Transcripts and aligned Illumina RNA-seq reads from all culture conditions were pooled for gene prediction. Pooled transcripts were used for gene prediction with CodingQuarry v2.0 (self-training Pathogen mode; default parameters) (91). Aligned reads and protein sequences from *P. nodorum* were used as evidence for gene prediction by BRAKER v2.0 (nondefault parameters --fungus, --prg gth) and GeMoMa v4.33 (default parameters) (92, 93). Gene predictions were combined with a nonredundant (nr) set of reviewed fungal Uniprot protein sequences and high-quality StringTie transcripts using EVidenceModeler (EVM), which generated a weighted consensus of all predicted gene models (Haas et al., 2008) (94). Evidence sources were weighted accordingly as follows: CodingQuarry > BRAKER ≥ GeMoMa > StringTie transcripts > nr fungal proteins. An assembly of MinION RNA-seq reads concordant with EVM gene models and StringTie transcripts was generated using PASA; this was used to update the EVM models (e.g., correcting intron-exon boundaries) and to annotate untranscribed regions (UTRs) (94). The completeness of these gene models was assessed using the BUSCO ascomycete database (32, 33). Complete gene annotations for *B. sorokiniana* isolates CS10 and CS27 are available in File S2 at https://github.com/megancamilla/Transposon-Mediated-transfer-of-ToxA/tree/master/S2_Bipolaris_Gff3.

Annotation of transposons. Transposons were identified *de novo* using the TEdenovo pipeline distributed as part of the REPET package v2.5 (38, 39). Long-read assemblies from the following species and isolates were used for *de novo* discovery: *P. nodorum* isolates Sn2000 and Sn4, assembled by Richards et al. (29), and SN15 (this study); *P. tritici-repentis* 1-C-BFP; and *B. sorokiniana* isolate CS10. Repbase v20.05 was used as the reference transposon database. TEs from each genome were combined into a common database according to the parameters set in Tedenovo_six_dnLibTEs_90_92.cfg. After combining the TEs, we manually added the coordinates of ToxhAT and named this TE "DTX-comp_CS10_RS_00." Finally, TEs were annotated in each genome listed in Table 1 with the common TE database using the TEannot pipeline and settings available in the TEannot.cfg file available at https://github.com/megancamilla/Transposon-Mediated-transfer-of-ToxA. Transposons were automatically classified into three-letter codes based on the Wicker nomenclature (34, 95). (REPET pipeline configuration and run scripts, plus the repeat annotation files for each genome used in this study, can be found online in File S3 at https://github.com/megancamilla/Transposon-Mediated-transfer-of-ToxA/tree/master/S3_REPET_Files. Inverted repeats and TIRs flanking the *ToxA* gene and surrounding DNA were identified in Geneious with the Dotplot (Self) viewing tool, which is based on the EMBOSS 6.5.7 tool dotmatcher (http://emboss.sourceforge.net/apps/release/6.5/emboss/apps/dotmatcher.html) (96). The specific settings required to reproduce the line plot shown in Fig. 1 are as follows: Reverse complement=yes, Score Matrix=exact, window size = 100, Threshold = 75, and Tile Size = 1000.

Whole-chromosome alignments. Initial whole-genome alignments (WGA) were conducted using Lastz v1.02.00 or Mauve as implemented in Geneious v.9.1.8 (95–97). To obtain a clean gff or tab delimited file of high-identity segments, we developed mimeo, which parses the alignment output of LASTZ (97–99). mimeo and a description of its full features can found at https://github.com/Adamtaranto/mimeo. WGA of *ToxA*$^+$ and *toxa*$^-$ isolates was performed with LASTZ as implemented in mimeo with the following settings: mimeo-map --minIdt 60 --minLen 100 --maxtandem 40 --writeTRF. Candidate regions of transfer between the three species were also identified by WGA performed with LASTZ as described above using mimeo-map --minIdt 70 --minLen 100. All alignments were inspected manually in Geneious v9.1.8. Chromosomal alignments were plotted in R v3.5.2 using the genoPlotR package (100). The corrected and trimmed nanopore reads generated by Canu were used for realignment to the *de novo* assemblies. These reads were mapped to the assembly with Minimap2 with the following settings: minimap2 -x map-ont -a. The output pairwise alignment file (paf) was modified in R for plotting. The complete R_markdown document, which includes all of the code needed to reproduce

Fig. 2, 3B, and 4, can be found online in File S4 at https://github.com/megancamilla/Transposon-Mediated-transfer-of-ToxA/tree/master/S4_Fig_Rmkds.

BLAST searches. All open reading frames identified within ToxhAT in isolate CS10 were translated using Geneious v9.18. These amino acid sequences were used as queries in tBLASTn searches on the NCBInr database and at JGI MycoCosm with the following settings: Blosum62 Matrix, Gap Costs: Existence 11 extension 1, e-value max 1e-10, max hits = 500 (last accessed 10 November 2018) (https://genome.jgi.doe.gov/programs/fungi/index.jsf) (101).

Data availability. Individual isolate accession numbers are listed in File S1 at https://github.com/megancamilla/Transposon-Mediated-transfer-of-ToxA/tree/master/S1_GenomeStats. All raw sequencing data are available under NCBI BioProject PRJNA505097. The *P. nodorum* SN15 Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under accession number SSHU00000000. Version SSHU01000000 is described in this paper. The *P. nodorum* Sn79-1087 Whole Genome Shotgun project has been deposited under accession numbers CP039668 to CP039689. The Whole Genome shotgun project and accession numbers for *B. sorokiniana* isolates are as follows: for CS10, SRZH00000000; for CS27, SRZG00000000; for WAI2406, SRZF00000000; for WAI2411, SRZE00000000. Transposon annotations and CS10 and CS27 gene annotations are available at https://github.com/megancamilla/Transposon-Mediated-transfer-of-ToxA.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mBio.01515-19.

FIG S1, PDF file, 0.1 MB.
FIG S2, PDF file, 1.3 MB.
FIG S3, PDF file, 2.3 MB.
FIG S4, PDF file, 0.4 MB.
FIG S5, PDF file, 0.03 MB.
FIG S6, PDF file, 0.02 MB.
FIG S7, PDF file, 1.3 MB.
FIG S8, PDF file, 2.2 MB.
FIG S9, PDF file, 0.5 MB.
FIG S10, PDF file, 1.5 MB.

ACKNOWLEDGMENTS

M.C.M. acknowledges The Sun Foundation's Peer Prize for Women in Science for support to sequence additional *ToxA* isolates. E.H. acknowledges The Grains and Research Development Corporation (project UHS11002). M.C.M., A.M., S.S., and P.S.S. also acknowledge The Grains and Research Development Corporation for the collection of isolates (projects DAN00203 and DAN00177).

REFERENCES

1. Keeling PJ, Palmer JD. 2008. Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet 9:605–618. https://doi.org/10.1038/nrg2386.
2. Dagan T, Artzy-Randrup Y, Martin W. 2008. Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. Proc Natl Acad Sci U S A 105:10039–10044. https://doi.org/10.1073/pnas.0800679105.
3. Kloesges T, Popa O, Martin W, Dagan T. 2011. Networks of gene sharing among 329 proteobacterial genomes reveal differences in lateral gene transfer frequency at different phylogenetic depths. Mol Biol Evol 28:1057–1074. https://doi.org/10.1093/molbev/msq297.
4. Wintersdorff von CJH, Penders J, van Niekerk JM, Mills ND, Majumder S, van Alphen LB, Savelkoul PHM, Wolffs P. 2016. Dissemination of antimicrobial resistance in microbial ecosystems through horizontal gene transfer. Front Microbiol 7:173.
5. Warnes SL, Highmore CJ, Keevil CW. 2012. Horizontal transfer of antibiotic resistance genes on abiotic touch surfaces: implications for public health. mBio 3:e00489-12. https://doi.org/10.1128/mBio.00489-12.
6. Wisecaver JH, Slot JC, Rokas A. 2014. The evolution of fungal metabolic pathways. PLoS Genet 10:e1004816. https://doi.org/10.1371/journal.pgen.1004816.
7. Richards TA, Soanes DM, Jones MDM, Vasieva O, Leonard G, Paszkiewicz K, Foster PG, Hall N, Talbot NJ. 2011. Horizontal gene transfer facilitated the evolution of plant parasitic mechanisms in the oomycetes. Proc Natl Acad Sci U S A 108:15258–15263. https://doi.org/10.1073/pnas.1105100108.
8. Marcet-Houben M, Gabaldón T. 2010. Acquisition of prokaryotic genes by fungal genomes. Trends Genet 26:5–8. https://doi.org/10.1016/j.tig.2009.11.007.
9. Shen X-X, Opulente DA, Kominek J, Zhou X, Steenwyk JL, Buh KV, Haase MAB, Wisecaver JH, Wang M, Doering DT, Boudouris JT, Schneider RM, Langdon QK, Ohkuma M, Endoh R, Takashima M, Manabe R-I, Čadež N, Libkind D, Rosa CA, DeVirgilio J, Hulfachor AB, Groenewald M, Kurtzman CP, Hittinger CT, Rokas A. 2018. Tempo and mode of genome evolution in the budding yeast subphylum. Cell 175:1533–1545.e20. https://doi.org/10.1016/j.cell.2018.10.023.
10. Novo M, Bigey F, Beyne E, Galeote V, Gavory F, Mallet S, Cambon B, Legras J-L, Wincker P, Casaregola S, Dequin S. 2009. Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. Proc Natl Acad Sci U S A 106:16333–16338. https://doi.org/10.1073/pnas.0904673106.
11. Cheeseman K, Ropars J, Renault P, Dupont J, Gouzy J, Branca A, Abraham A-L, Ceppi M, Conseiller E, Debuchy R, Malagnac F, Goarin A, Silar P, Lacoste S, Sallet E, Bensimon A, Giraud T, Brygoo Y. 2014. Multiple recent horizontal transfers of a large genomic region in cheese making fungi. Nat Commun 5:2876. https://doi.org/10.1038/ncomms3876.
12. Borneman AR, Desany BA, Riches D, Affourtit JP, Forgan AH, Pretorius IS, Egholm M, Chambers PJ. 2011. Whole-genome comparison reveals

novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*. PLoS Genet 7:e1001287. https://doi.org/10.1371/journal.pgen.1001287.

13. Ropars J, Rodríguez De La Vega RC, Lopez-Villavicencio M, Gouzy J, Sallet E, Dumas E, Lacoste S, Debuchy R, Dupont J, Branca A, Giraud T. 2015. Adaptive horizontal gene transfers between multiple cheese-associated fungi. Curr Biol 25:2562–2569. https://doi.org/10.1016/j.cub.2015.08.025.
14. Marsit S, Mena A, Bigey F, Sauvage F-X, Couloux A, Guy J, Legras J-L, Barrio E, Dequin S, Galeote V. 2015. Evolutionary advantage conferred by an eukaryote-to-eukaryote gene transfer event in wine yeasts. Mol Biol Evol 32:1695–1707. https://doi.org/10.1093/molbev/msv057.
15. Ohm RA, Feau N, Henrissat B, Schoch CL, Horwitz BA, Barry KW, Condon BJ, Copeland AC, Dhillon B, Glaser F, Hesse CN, Kosti I, LaButti K, Lindquist EA, Lucas S, Salamov AA, Bradshaw RE, Ciuffetti L, Hamelin RC, Kema GHJ, Lawrence C, Scott JA, Spatafora JW, Turgeon BG, De Wit P, Zhong S, Goodwin SB, Grigoriev IV. 2012. Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen *Dothideomycetes* fungi. PLoS Pathog 8:e1003037. https://doi.org/10.1371/journal.ppat.1003037.
16. Grandaubert J, Lowe RGT, Soyer JL, Schoch CL, Van de Wouw AP, Fudal I, Robbertse B, Lapalu N, Links MG, Ollivier B, Linglin J, Barbe V, Mangenot S, Cruaud C, Borhan H, Howlett BJ, Balesdent M-H, Rouxel T. 2014. Transposable element-assisted evolution and adaptation to host plant within the *Leptosphaeria maculans-Leptosphaeria biglobosa* species complex of fungal pathogens. BMC Genomics 15:891. https://doi.org/10.1186/1471-2164-15-891.
17. McDonald MC, Ahren D, Simpfendorfer S, Milgate A, Solomon PS. 2018. The discovery of the virulence gene *ToxA* in the wheat and barley pathogen *Bipolaris sorokiniana*. Mol Plant Pathol 19:432–439. https://doi.org/10.1111/mpp.12535.
18. Friesen TL, Holmes DJ, Bowden RL, Faris JD. 2018. *ToxA* is present in the U.S. *Bipolaris sorokiniana* population and is a significant virulence factor on wheat harboring *Tsn1*. Plant Dis 102:2446–2452. https://doi.org/10.1094/PDIS-03-18-0521-RE.
19. Ciuffetti LM, Tuori RP, Gaventa JM. 1997. A single gene encodes a selective toxin causal to the development of tan spot of wheat. Plant Cell 9:135–144. https://doi.org/10.1105/tpc.9.2.135.
20. Friesen TL, Stukenbrock EH, Liu Z, Meinhardt S, Ling H, Faris JD, Rasmussen JB, Solomon PS, McDonald BA, Oliver RP. 2006. Emergence of a new disease as a result of interspecific virulence gene transfer. Nat Genet 38:953–956. https://doi.org/10.1038/ng1839.
21. Syme RA, Hane JK, Friesen TL, Oliver RP. 2013. Resequencing and comparative genomics of *Stagonospora nodorum*: sectional gene absence and effector discovery. G3 (Bethesda) 3:959–969. https://doi.org/10.1534/g3.112.004994.
22. McDonald MC, Oliver RP, Friesen TL, Brunner PC, McDonald BA. 2013. Global diversity and distribution of three necrotrophic effectors in *Phaeosphaeria nodorum* and related species. New Phytol 199:241–251. https://doi.org/10.1111/nph.12257.
23. Manning VA, Pandelova I, Dhillon B, Wilhelm LJ, Goodwin SB, Berlin AM, Figueroa M, Freitag M, Hane JK, Henrissat B, Holman WH, Kodira CD, Martin J, Oliver RP, Robbertse B, Schackwitz W, Schwartz DC, Spatafora JW, Turgeon BG, Yandava C, Young S, Zhou S, Zeng Q, Grigoriev IV, Ma L-J, Ciuffetti LM. 2013. Comparative genomics of a plant-pathogenic fungus, *Pyrenophora tritici-repentis*, reveals transduplication and the impact of repeat elements on pathogenicity and population divergence. G3 (Bethesda) 3:41–63. https://doi.org/10.1534/g3.112.004044.
24. Oliver RP, Rybak K, Solomon PS, Ferguson-Hunt M. 2009. Prevalence of ToxA-sensitive alleles of the wheat gene *Tsn1* in Australian and Chinese wheat cultivars. Crop Pasture Sci 60:348–352. https://doi.org/10.1071/CP08259.
25. Selker EU. 1990. Premeiotic instability of repeated sequences in *Neurospora crassa*. Annu Rev Genet 24:579–613. https://doi.org/10.1146/annurev.ge.24.120190.003051.
26. Hane JK, Oliver RP. 2008. RIPCAL: a tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences. BMC Bioinformatics 9:478. https://doi.org/10.1186/1471-2105-9-478.
27. Gladyshev E, Kleckner N. 2014. Direct recognition of homology between double helices of DNA in *Neurospora crassa*. Nat Commun 5:3509. https://doi.org/10.1038/ncomms4509.
28. Gladyshev E, Kleckner N. 2017. Recombination-independent recognition of DNA homology for repeat-induced point mutation. Curr Genet 63:389–400. https://doi.org/10.1007/s00294-016-0649-4.
29. Richards JK, Wyatt NA, Liu Z, Faris JD, Friesen TL. 2018. Reference quality genome assemblies of three *Parastagonospora nodorum* isolates differing in virulence on wheat. G3 (Bethesda) 8:393–399. https://doi.org/10.1534/g3.117.300462.
30. Moolhuijzen P, See PT, Hane JK, Shi G, Liu Z, Oliver RP, Moffat CS. 2018. Comparative genomics of the wheat fungal pathogen *Pyrenophora tritici-repentis* reveals chromosomal variations and genome plasticity. BMC Genomics 19:279. https://doi.org/10.1186/s12864-018-4680-3.
31. 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.
32. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol 35:543–548. https://doi.org/10.1093/molbev/msx319.
33. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351.
34. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH. 2007. A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973–982. https://doi.org/10.1038/nrg2165.
35. Perry JR, Basrai MA, Steiner HY, Naider F, Becker JM. 1994. Isolation and characterization of a *Saccharomyces cerevisiae* peptide transport gene. Mol Cell Biol 14:104–115. https://doi.org/10.1128/mcb.14.1.104.
36. Dyrka W, Lamacchia M, Durrens P, Kobe B, Daskalov A, Paoletti M, Sherman DJ, Saupe SJ. 2014. Diversity and variability of NOD-like receptors in fungi. Genome Biol Evol 6:3137–3158. https://doi.org/10.1093/gbe/evu251.
37. Heller J, Clavé C, Gladieux P, Saupe SJ, Glass NL. 2018. NLR surveillance of essential SEC-9 SNARE proteins induces programmed cell death upon allorecognition in filamentous fungi. Proc Natl Acad Sci U S A 115:E2292–E2301. https://doi.org/10.1073/pnas.1719705115.
38. Quesneville H, Bergman CM, Andrieu O, Autard D, Nouaud D, Ashburner M, Anxolabehere D. 2005. Combined evidence annotation of transposable elements in genome sequences. PLoS Comp Biol 1:166–175. https://doi.org/10.1371/journal.pcbi.0010022.
39. Flutre T, Duprat E, Feuillet C, Quesneville H. 2011. Considering transposable element diversification in *de novo* annotation approaches. PLoS One 6:e16526. https://doi.org/10.1371/journal.pone.0016526.
40. Lu S, Gillian Turgeon B, Edwards MC. 2015. A ToxA-like protein from *Cochliobolus heterostrophus* induces light-dependent leaf necrosis and acts as a virulence factor with host selectivity on maize. Fung Genet Biol 81:12–24. https://doi.org/10.1016/j.fgb.2015.05.013.
41. Aveskamp MM, de Gruyter J, Woudenberg JHC, Verkley GJM, Crous PW. 2010. Highlights of the *Didymellaceae*: a polyphasic approach to characterise *Phoma* and related pleosporalean genera. Stud Mycol 65:1–60. https://doi.org/10.3114/sim.2010.65.01.
42. Inderbitzin P, Kohlmeyer J, Volkmann-Kohlmeyer B, Berbee ML. 2002. *Decorospora*, a new genus for the marine ascomycete *Pleospora gaudefroyi*. Mycologia 94:651–659. https://doi.org/10.1080/15572536.2003.11833193.
43. McDonald MC, Solomon PS. 2018. Just the surface: advances in the discovery and characterization of necrotrophic wheat effectors. Curr Opin Microbiol 46:14–18. https://doi.org/10.1016/j.mib.2018.01.019.
44. Rubin E, Lithwick G, Levy AA. 2001. Structure and evolution of the hAT transposon superfamily. Genetics 158:949–957.
45. Daboussi M-J, Capy P 2003. Transposable elements in filamentous fungi. Annu Rev Microbiol 57:275–299. https://doi.org/10.1146/annurev.micro.57.030502.091029.
46. 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 Z. 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.
47. Faino L, Seidl MF, Shi-Kunne X, Pauper M, van den Berg GCM, Wittenberg AHJ, Thomma B. 2016. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. Genome Res 26:1091–1100. https://doi.org/10.1101/gr.204974.116.
48. Tsushima A, Gan P, Kumakura N, Narusaka M, Takano Y, Narusaka Y, Shirasu K. 2019. Genomic plasticity mediated by transposable elements in the plant pathogenic fungus *Colletotrichum higginsianum*. Gen Biol Evol 11:1487–1500. https://doi.org/10.1093/gbe/evz087.

49. Winter DJ, Ganley ARD, Young CA, Liachko I, Schardl CL, Dupont P-Y, Berry D, Ram A, Scott B, Cox MP. 2018. Repeat elements organise 3D genome structure and mediate transcription in the filamentous fungus *Epichloe festucae*. PLoS Genet 14:e1007467. https://doi.org/10.1371/journal.pgen.1007467.
50. Belton J-M, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. 2012. Hi–C: a comprehensive technique to capture the conformation of genomes. Methods 58:268–276. https://doi.org/10.1016/j.ymeth.2012.05.001.
51. Aboukhaddour R, Cloutier S, Ballance GM, Lamari L. 2009. Genome characterization of *Pyrenophora tritici-repentis* isolates reveals high plasticity and independent chromosomal location of *ToxA* and *ToxB*. Mol Plant Pathol 10:201–212. https://doi.org/10.1111/j.1364-3703.2008.00520.x.
52. Ma L-J, van der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, Di Pietro A, Dufresne M, Freitag M, Grabherr M, Henrissat B, Houterman PM, Kang S, Shim W-B, Woloshuk C, Xie X, Xu J-R, Antoniw J, Baker SE, Bluhm BH, Breakspear A, Brown DW, Butchko RAE, Chapman S, Coulson R, Coutinho PM, Danchin EGJ, Diener A, Gale LR, Gardiner DM, Goff S, Hammond-Kosack KE, Hilburn K, Hua-Van A, Jonkers W, Kazan K, Kodira CD, Koehrsen M, Kumar L, Lee Y-H, Li L, Manners JM, Miranda-Saavedra D, Mukherjee M, Park G, Park J, Park S-Y, Proctor RH, Regev A, Ruiz-Roldan MC, et al. 2010. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. Nature 464:367–373. https://doi.org/10.1038/nature08850.
53. de Jonge R, Bolton MD, Kombrink A, van den Berg GCM, Yadeta KA, Thomma B. 2013. Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. Gen Res 23:1271–1282. https://doi.org/10.1101/gr.152660.112.
54. Croll D, McDonald BA. 2012. The accessory genome as a cradle for adaptive evolution in pathogens. PLoS Pathog 8:e1002608. https://doi.org/10.1371/journal.ppat.1002608.
55. 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.
56. Dong S, Raffaele S, Kamoun S. 2015. The two-speed genomes of filamentous pathogens: waltz with plants. Curr Opin Genet Dev 35:57–65. https://doi.org/10.1016/j.gde.2015.09.001.
57. Frantzeskakis L, Kusch S, Panstruga R. 2019. The need for speed: compartmentalized genome evolution in filamentous phytopathogens. Mol Plant Pathol 20:3–7. https://doi.org/10.1111/mpp.12738.
58. Martinez JP, Oesch NW, Ciuffetti LM. 2004. Characterization of the multiple-copy host-selective toxin gene, *ToxB*, in pathogenic and nonpathogenic isolates of *Pyrenophora tritici-repentis*. Mol Plant Microbe Interact 17:467–474. https://doi.org/10.1094/MPMI.2004.17.5.467.
59. Kuttler F, Mai S. 2007. Formation of non-random extrachromosomal elements during development, differentiation and oncogenesis. Semin Cancer Biol 17:56–64. https://doi.org/10.1016/j.semcancer.2006.10.007.
60. González J, Petrov DA. 2012. Evolution of genome content: population dynamics of transposable elements in flies and humans. Methods Mol Biol 855:361–383. https://doi.org/10.1007/978-1-61779-582-4_13.
61. Moller HD, Parsons L, Jorgensen TS, Botstein D, Regenberg B. 2015. Extrachromosomal circular DNA is common in yeast. Proc Natl Acad Sci U S A 112:E3114–E3122. https://doi.org/10.1073/pnas.1508825112.
62. Mourier T. 2016. Transposable elements and circular DNAs. Mob Genet Elements 6:e1240748. https://doi.org/10.1080/2159256X.2016.1240748.
63. Stukenbrock EH, McDonald BA. 2007. Geographical variation and positive diversifying selection in the host-specific toxin *SnToxA*. Mol Plant Pathol 8:321–332. https://doi.org/10.1111/j.1364-3703.2007.00396.x.
64. Ciuffetti LM, Manning VA, Pandelova I, Betts MF, Martinez JP. 2010. Host-selective toxins, Ptr ToxA and Ptr ToxB, as necrotrophic effectors in the *Pyrenophora tritici-repentis*-wheat interaction. New Phyt 187:911–919. https://doi.org/10.1111/j.1469-8137.2010.03362.x.
65. Han M-J, Xu H-E, Zhang H-H, Feschotte C, Zhang Z. 2014. Spy: a new group of eukaryotic DNA transposons without target site duplications. Genome Biol Evol 6:1748–1757. https://doi.org/10.1093/gbe/evu140.
66. Clark JB, Silva JC, Kidwell MG. 2002. Evidence for horizontal transfer of P transposable elements, p 161–171. *In* Horizontal Gene Transfer. Elsevier, Philadelphia, PA.
67. Daniels SB, Peterson KR, Strausbaugh LD, Kidwell MG, Chovnick A. 1990. Evidence for horizontal transmission of the P transposable element between *Drosophila* species. Genetics 124:339–355.
68. Panaud O. 2016. Horizontal transfers of transposable elements in eukaryotes: the flying genes. C R Biol 339:296–299. https://doi.org/10.1016/j.crvi.2016.04.013.
69. Martin WF. 2017. Too much eukaryote LGT. BioEssays 39:1700115. https://doi.org/10.1002/bies.201700115.
70. Gogarten JP, Townsend JP. 2005. Horizontal gene transfer, genome innovation and evolution. Nat Rev Microbiol 3:679–687. https://doi.org/10.1038/nrmicro1204.
71. Gilbert C, Chateigner A, Ernenwein L, Barbe V, Bezier A, Herniou EA, Cordaux R. 2014. Population genomics supports baculoviruses as vectors of horizontal transfer of insect transposons. Nat Commun 5:3348. https://doi.org/10.1038/ncomms4348.
72. Gilbert C, Peccoud J, Chateigner A, Moumen B, Cordaux R, Herniou EA. 2016. Continuous influx of genetic material from host to virus populations. PLoS Genet 12:e1005838. https://doi.org/10.1371/journal.pgen.1005838.
73. Garcia-Vallve S, Romeu A, Palau J. 2000. Horizontal gene transfer of glycosyl hydrolases of the rumen fungi. Mol Biol Evol 17:352–361. https://doi.org/10.1093/oxfordjournals.molbev.a026315.
74. Slot JC, Rokas A. 2011. Horizontal transfer of a large and highly toxic secondary metabolic gene cluster between fungi. Curr Biol 21:134–139. https://doi.org/10.1016/j.cub.2010.12.020.
75. Campbell MA, Rokas A, Slot JC. 2012. Horizontal transfer and death of a fungal secondary metabolic gene cluster. Gen Biol Evol 4:289–293. https://doi.org/10.1093/gbe/evs011.
76. Marcet-Houben M, Gabaldón T. 2016. Horizontal acquisition of toxic alkaloid synthesis in a clade of plant associated fungi. Fung Genet Biol 86:71–80. https://doi.org/10.1016/j.fgb.2015.12.006.
77. Condon BJ, Leng Y, Wu D, Bushley KE, Ohm RA, Otillar R, Martin J, Schackwitz W, Grimwood J, MohdZainudin N, Xue C, Wang R, Manning VA, Dhillon B, Tu ZJ, Steffenson BJ, Salamov A, Sun H, Lowry S, LaButti K, Han J, Copeland A, Lindquist E, Barry K, Schmutz J, Baker SE, Ciuffetti LM, Grigoriev IV, Zhong S, Turgeon BG. 2013. Comparative genome structure, secondary metabolite, and effector coding capacity across *Cochliobolus* pathogens. PLoS Genet 9:e1003233. https://doi.org/10.1371/journal.pgen.1003233.
78. Fulton TM, Chunwongse J, Tanksley SD. 1995. Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Mol Biol Rep 13:207–209. https://doi.org/10.1007/BF02670897.
79. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116.
80. Chaisson MJ, Tesler G. 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. BMC Bioinformatics 13:238. https://doi.org/10.1186/1471-2105-13-238.
81. Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res 27:737–746. https://doi.org/10.1101/gr.214270.116.
82. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.
83. Yakasai AA, Davison J, Wasil Z, Halo LM, Butts CP, Lazarus CM, Bailey AM, Simpson TJ, Cox RJ. 2011. Nongenetic reprogramming of a fungal highly reducing polyketide synthase. J Am Chem Soc 133:10990–10998. https://doi.org/10.1021/ja204200x.
84. Solomon PS, Lee RC, Wilson T, Oliver RP. 2004. Pathogenicity of *Stagonospora nodorum* requires malate synthase. Mol Microbiol 53:1065–1073. https://doi.org/10.1111/j.1365-2958.2004.04178.x.
85. Liu ZH, Faris JD, Meinhardt SW, Ali S, Rasmussen JB, Friesen TL. 2004. Genetic and physical mapping of a gene conditioning sensitivity in wheat to a partially purified host-selective toxin produced by *Stagonospora nodorum*. Phytopathology 94:1056–1060. https://doi.org/10.1094/PHYTO.2004.94.10.1056.
86. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics 34:2666–2669. https://doi.org/10.1093/bioinformatics/bty149.
87. Tardaguila M, la Fuente de L, Marti C, Pereira C, Pardo-Palacios FJ, del Risco H, Ferrell M, Mellado M, Macchietto M, Verheggen K, Edelmann M, Ezkurdia I, Vazquez J, Tress M, Mortazavi A, Martens L, Rodriguez-Navarro S, Moreno-Manzano V, Conesa A. 2018. SQANTI: extensive characterization of long-read transcript sequences for quality control in

full-length transcriptome identification and quantification. Genome Res 28:396–411. https://doi.org/10.1101/gr.222976.117.
88. 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.
89. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21. https://doi.org/10.1093/bioinformatics/bts635.
90. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol 33:290–295. https://doi.org/10.1038/nbt.3122.
91. Testa AC, Hane JK, Ellwood SR, Oliver RP. 2015. CodingQuarry: highly accurate hidden Markov model gene prediction in fungal genomes using RNA-seq transcripts. BMC Genomics 16:170. https://doi.org/10.1186/s12864-015-1344-4.
92. Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: unsupervised RNA-seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics 32:767–769. https://doi.org/10.1093/bioinformatics/btv661.
93. Keilwagen J, Hartung F, Paulini M, Twardziok SO, Grau J. 2018. Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. BMC Bioinformatics 19:189. https://doi.org/10.1186/s12859-018-2203-5.
94. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Robin CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol 9:R7. https://doi.org/10.1186/gb-2008-9-1-r7.
95. Hoede C, Arnoux S, Moisset M, Chaumier T, Inizan O, Jamilloux V, Quesneville H. 2014. PASTEC: an automatic transposable element classification tool. PLoS One 9:e91929. https://doi.org/10.1371/journal.pone.0091929.
96. Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16:276–277. https://doi.org/10.1016/S0168-9525(00)02024-2.
97. Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, Hardison RC, Haussler D, Miller W. 2003. Human-mouse alignments with BLASTZ. Gen Res 13:103–107. https://doi.org/10.1101/gr.809403.
98. Harris RS. 2007. Improved pairwise alignment of genomic DNA. PhD thesis. Pennsylvania State University, State College, PA.
99. Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27:573. https://doi.org/10.1093/nar/27.2.573.
100. Guy L, Kultima JR, Andersson S. 2010. genoPlotR: comparative gene and genome visualization in R. Bioinformatics 26:2334–2335. https://doi.org/10.1093/bioinformatics/btq413.
101. Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402. https://doi.org/10.1093/nar/25.17.3389.