Dissecting the *Pyrenophora tritici-repentis* (tan spot of wheat) pangenome

Ryan Gourlie[a], Megan McDonald[b], Mohamed Hafez[a], Rodrigo Ortega-Polo[a], Kristin E. Low[a], D. Wade Abbott[a], Stephen E. Strelkov[c], Fouad Daayf[d] and Reem Aboukhaddour[a*]

[a]Agriculture and Agri-food Canada, Lethbridge, Alberta, Canada

[b]University of Birmingham, School of Biosciences, Edgbaston, Birmingham, United Kingdom

[c]University of Alberta, Faculty of Agricultural, Life, and Environmental Sciences, Edmonton, Alberta, Canada

[d]University of Manitoba, Faculty of Agricultural and Food Sciences, Winnipeg, Manitoba, Canada

*corresponding author: reem.aboukhaddour@agr.gc.ca

ABSTRACT

We sequenced the genome of a global collection (40 isolates) of the fungus *Pyrenophora tritici-repentis* (Ptr), a major foliar pathogen of wheat and model for the evolution of necrotrophic pathogens. Ptr exhibited an open-pangenome, with 43% of genes in the core set and 57% defined as accessory (present in only a subset of isolates), of which 56% were singleton genes (present in only one isolate). A clear distinction between pathogenic and non-pathogenic genomes was observed in size, gene content, and phylogenetic relatedness. Chromosomal rearrangements and structural organization, specifically around the effector coding genes, were explored further using the annotated genomes of two isolates sequenced by PacBio RS II and Illumina HiSeq. The Ptr genome exhibited major chromosomal rearrangements, including chromosomal fusion, translocation, and segment duplications. An intraspecies translocation of *ToxA*, the necrosis-inducing effector-coding gene, was facilitated within Ptr via a 143 kb '*Starship'* transposon (dubbed 'Horizon'). Additionally, *ToxB*, the gene encoding the chlorosis-inducing effector, was clustered as three copies on a 294 kb transposable element in a ToxB-producing isolate. *ToxB* and its carrying transposon were missing from the *ToxB* non-coding reference isolate, but the homolog *toxb* and the transposon were both present in another non-coding isolate. The Ptr genome also appears to exhibit a 'one-compartment' organization, but may still possess a 'two-speed genome' that is facilitated by copy-number variation as reported in other fungal pathosystems.

IMPORTANCE

Ptr is one of the most destructive wheat pathogens worldwide. Its genome is a mosaic of present and absent effectors, and serves as a model for examining the evolutionary processes behind the acquisition of virulence in necrotrophs and disease emergence. In this work, we took advantage of a diverse collection of pathogenic Ptr isolates with different global origins and applied short- and long-read sequencing technologies to dissect the Ptr genome. This study provides comprehensive insights into the Ptr genome and highlights its structural organization as an open pangenome with 'one-compartment'. In addition, we identified the potential involvement of transposable elements in genome expansion and the movement of virulence factors. The ability of effector-coding genes to shuffle across chromosomes on large transposons was illustrated by the intraspecies translocation of *ToxA* and the multi-copy *ToxB*. In terms of gene contents, the Ptr genome exhibits a large percentage of orphan genes, particularly in non-pathogenic or weakly-virulent isolates.

KEYWORDS

fungal wheat pathogen, pangenome, ToxA, ToxB, transposons, fungal evolution, necrotrophic effectors

INTRODUCTION

Pyrenophora tritici-repentis (Ptr) is an ascomycete fungus that causes tan spot, one of the most destructive foliar diseases of wheat worldwide (1, 2). The Ptr genome represents an exciting case to explore the evolution of virulence and the emergence of new diseases. Ptr has served as a model species for necrotrophic plant pathogens, as it has the ability to produce various necrotrophic effectors , previously known as host-selective toxins (2). To date, three effectors have been identified and are known virulence factors: ToxA is a necrosis-inducing effector, while ToxB and ToxC are chlorosis-inducing effectors (3). Additional effectors may be involved in pathogenicity and await characterization (4). ToxA and ToxB are ribosomally-synthesized proteins, and while ToxA is encoded by a single copy *ToxA* gene, ToxB is encoded by the multi-copy *ToxB* genes (3). Understanding of ToxC-coding gene(s) and its synthesis awaits further characterization (5, 6). In Ptr, the *ToxA* gene is present only in isolates that secrete the ToxA effector, and no homolog has been identified in non-producing isolates of Ptr. In contrast, *ToxB*

homologs, termed *toxb*, are present in some Ptr isolates that do not secrete the active form of the ToxB protein. Homologs of both *ToxA* and *ToxB* have been reported in related and distantly related fungal species (7-10).

Ptr isolates are grouped into eight different races (races 1 to 8) based on which combinations of effectors they secrete (3). In North America and Australia, Ptr isolates are mainly ToxA and ToxC-producers, with ToxB-producers being extremely rare or absent (4, 11). In regions encompassing the wheat centre of origin, like the Fertile Crescent and North Africa, all races and effectors are found even in under-surveyed regions (2, 4). ToxB-producing isolates were found to predominate in North Africa (4, 11, 12). The lack of isolates with *ToxB* in North America and Australia indicates divergent evolution of Ptr and its effectors worldwide. An independent origin of these effectors was suggested previously (7, 8, 13, 14). For example, *ToxA* was acquired by horizontal-gene-transfer (HGT) from a related fungal species (7, 13, 14), while *ToxB* (and its homologs) are believed to have evolved via vertical inheritance (7, 8, 13, 15).

The Ptr genome has been described as highly plastic, with evidence of major chromosomal rearrangements, including insertions, deletions, inversions, and translocations (13, 14, 16-18). The first full Ptr genome was released by the Broad Institute in 2007, based on a race 1 isolate (Pt-1C-BFP, abbreviated throughout as BFP) from the United States, and was assembled to the chromosome level with the aid of optical mapping (16). In addition, the full genomes of 11 Ptr isolates from North America and Australia have also been sequenced and are now publicly available (16, 17, 19). Of those, three (M4, V1, and DW5) were assembled from long-read sequences.

Previously, DNA transposons and long terminal repeat (LTR) retrotransposons were associated with genome-wide segmental duplications, insertions, and chromosomal fusions in Ptr (15). However, genomic regions with AT-rich content, which are associated with repeat-induced point mutations (RIP), were not reported at high frequencies in previously sequenced Ptr isolates from Australia and the USA (17). RIP is often associated with the “two-speed genome” model in a number of fungal pathogens; this model describes the pattern of physical genome compartmentation with repeat-rich gene-sparse regions and repeat-sparse gene-dense regions (20). Here, we sequenced 40 Ptr isolates of diverse origins and races, and performed a whole pangenome analysis with the addition of the reference isolate (BFP), for a total of 41 isolates. These isolates represent all known races including a newly identified North African atypical race that induces

necrosis but lacks the *ToxA* gene (4). In addition to the pangenome, we provide a detailed look at gene contents, phylogenetic relatedness, and effector gene movement. In this study, polished long-read assemblies representing the full genome of two isolates (I-73-1 and D308) were explored to define the structural organization and compartmentation of the Ptr genome by comparison with three previously published genomes (BFP, DW-5, and M4). I-73-1 is a race 8 isolate collected from Syria and represents the most complex race, producing all three known effectors (ToxA, ToxB, and ToxC). D308 is a race 3 isolate collected from Canada, and codes for ToxC only. This work provides a detailed comparative genomic analysis of Ptr for use by the research community on fungal pathogens.

RESULTS

Hybrid assemblies capture a better estimate for genome size

The *de novo* assemblies of Illumina short-reads indicated a variable genome size ranging from 34.1 Mb to 36.9 Mb, with an average of 34.82 Mb. The smallest assembly was for isolate T128-1 (34.12 Mb), which has been identified as an atypical isolate with a novel virulence phenotype (4). The largest assemblies were for isolates 92-171R5 (36.8 Mb, race 5) and G9-4 (36.97 Mb, race 4).

The hybrid assembly (PacBio RS II and Illumina Hi-Seq data) of isolate I-73-1 contained 39 contiguous sequences, with 11 chromosome-sized contigs greater than 2 Mb. Similar results were obtained with isolate D308, whose assembly contained 70 contiguous sequences, and 11 contigs greater than 1 Mb with an additional two smaller contigs representing the majority of the genome. Genome completeness was assessed by Benchmarking Universal Single-Copy Orthologs (BUSCO) using a set of conserved fungal genes, and all isolates (except G9-4) were assessed as >99% complete. The short-read assemblies of I-73-1 and D308 were ~5.3 Mb smaller than the long-read-based assemblies. Summary statistics for all assemblies are available in Table 1.

Ptr has an open pangenome with high accessory gene content and unique genes for non-pathogenic and weakly virulent isolates

On average, each isolate contained 13,071 predicted genes, and collectively, 522,848 predicted genes across all isolates were reported. After grouping by similarity, with a threshold of ≥90% (21), we were left with 23,454 unique gene clusters. In total, 10,159 genes were identified

as core (conserved across all isolates) (43% of total pangenome) (Figure 1A). The remaining 13,295 gene models (~57% of the total pangenome) were the accessory genome, with individual accessory sizes ranging from 2,661 to 3,424 genes. Within the accessory genome were 7,476 singletons (genes present in only one isolate) (56% of accessory genome; 32% of the total pangenome). Within the singletons, a large number 3,293 (44%) were present in isolates in the outgroup of the phylogeny (Figure 2A), including 90-2 (461 genes), 92-171R5 (1,598 genes), and G9-4 (1,234 genes). Additionally, isolate T128-1, which exhibits a novel necrotic phenotype (4), contained 54 unique genes.

Genes were binned into 'clusters' based on the presence or absence of genes to provide insights into gene gains and losses in Ptr (Supplemental File 1). Each cluster is indicative of how many isolates share a particular gene (e.g., cluster 2 shows genes shared by two isolates, cluster 3 by three isolates, etc.). Clusters 1 and 41 were not included, as these represent a singleton and the core gene set, respectively. The result of this clustering is groups of genes (grouped by how many isolates contain the gene) many of which are likely gene gains and losses throughout Ptr.

Isolates carrying the *ToxA* gene had significantly larger accessory genomes (mean of 2,987 genes) compared with isolates lacking *ToxA* (mean of 2,820 genes) based on a two-sample t-test (p = 0.002). No such relationship was found with *ToxB* or putative *ToxC*-carrying isolates. An ANOVA followed by Tukey's HSD test indicated that race 7 (*ToxA*, *ToxB*) and race 3 (*ToxC*) isolates also had significantly different accessory genome sizes (average of 3,162 and 2,779 genes, respectively; p = 0.04), with no other notable differences between races. Additionally, the total number of genes present in the pangenome increased steadily with each additional genome (Figure 1B), and the number of core genes appears to have reached a stable level after the addition of 38 genomes (Figure 1B). The continual increase of genes per genome and the large accessory genome (57%) are strong indications that Ptr has an open pangenome.

Ptr core protein and SNP phylogenies exhibit clustering based on necrotrophic effector secretion

A maximum-likelihood tree to estimate genetic relatedness among Ptr isolates was constructed using an alignment of the 10,159 core proteins, and showed that most isolates partitioned into two major clusters (Figure 2A). Cluster 1 contained 20 isolates, most of which were ToxA-producing Canadian isolates classified as races 1 and 2. Few isolates in cluster 1 were ToxB-producers; those that were ToxB-producers were from Azerbaijan and Algeria. Cluster 2

contained 19 isolates, most of which were ToxA non-producing races 3 and 5, which mostly originated in the Fertile Crescent and nearby regions. The Canadian isolates in cluster 2 were all classified as race 3 (ToxC producers). Three isolates (92-171R5, 90-2, and G9-4) clustered as an outgroup. 92-171R5 is the only race 5 isolate identified in Canada, and was described as weakly virulent (22). Isolates 90-2 and G9-4 were classified as non-pathogenic race 4 (12, 23).

Heirarchical set groupings based on accessory proteins revealed similar clustering as the core protein phylogeny (Figure 2B), with the only exceptions being in which major cluster the isolates 90-2 (race 4) and SW1-2 (race 1) were placed. These two isolates shared a high number of genes not present in other isolates, defining their placement within the hierarchical set. Clusters in the hierarchical sets were defined by the presence or absence of accessory genes and does not take into account any sequence variation in the homologs used to construct the protein phylogeny, but highlights how gene gains and losses can mirror mutation rates. Additional phylogenetic trees based on SNP data also closely match the core protein phylogeny and the hierarchical set trees in terms of isolate clustering, with the divergent isolates separating even further (Supplemental File 2A and 2B').

Predicted genes and effectors

Of the 23,454 unique proteins in the Ptr pangenome, only 46% were present in the Pfam functional database. Of the 10,159 core genes, 6,951 (69%) had some described functionality (e.g., domain, family, etc.). However, of the 13,295 accessory genes, only 3,782 (29%) were categorized, the remaining being hypothetical proteins. Core gene protein lengths were significantly greater than accessory proteins, with mean lengths of 482 aa and 241 aa, respectively, and medians of 411 aa and 156 aa (Supplemental File 3). Singleton proteins were marginally smaller than other accessory proteins, with a mean length of 232 aa (median 156 aa). A small number of proteins (108) were quite large, with lengths exceeding 2,000 aa, and six of those exceeding 5,000 aa; the largest protein was 9,750 aa in length. Many of these genes were annotated in previously published Ptr genomes.

Effectors play a major role in establishing infection in necrotrophic pathogens. In total, 729 potential effectors (3.1% of the total pangenome) were predicted in Ptr. Of these, 106 (14.5%) were present as core genes, and 626 (85%) were accessory genes, with 313 of the accessories being singletons (Figure 1A). Omission of the non-pathogenic race 4 isolates (90-2, G9-4, and T126-1)

revealed 584 unique effectors, with 127 genes in the core and 457 as accessory (198 singletons). Nearly half of the potential effectors were singletons present in the weakly virulent isolate 92-171R5.

CAZymes

We identified 305 unique CAZymes from 91 CAZyme families that are known to be active on cell wall polysaccharides. Of these, 281 (92%) were observed in all isolates; the number of proteins within the CAZyme families were predominantly conserved across all isolates (Supplementaal File 4). Although numbers were similar between isolates and were related at the sequence level, there may be functional differences that remain to be explored. CAZyme families associated with plant cell wall degradation and fungal phytopathogenesis were abundant (Supplemental File 4). These included GH5 cellulases and endoglucanases, GH43 arabinases and xylosidases, AA9 lytic cellulose monooxygenases, CE1 hemicellulases, and CE5 cutinases (24-29). Protein sequences within these families accounted for 31% of CAZymes identified. It should be noted that not all CAZymes are associated with pathogenesis, and many are likely involved in routine cellular activities, such as fungal cell wall remodeling and glycoprotein maturation.

Chromosomal structural organization

Full genome alignment between the hybrid assemblies of isolate I-73-1 (race 8) and the reference BFP (race 1) showed 11 contigs in I-73-1 of comparable size and content to the 11 chromosomes of the reference BFP. Alignment of D308 (race 3) indicated 10 contigs of comparable size, with an additional three smaller contigs completing a full alignment to BFP. BFP chromosomes 4 and 7 appeared to be homologous between all three isolates (contigs 12 and 9 in I-73-1 and contigs 5, 11, and 14 in D308) (Figure 3). Chromosomes 3, 5, 6, and 10 in BFP were also mostly homologous with contigs 2, 7, 17 and 6, respectively, in I-73-1 (Figure 3); the only notable exception was the *ToxA* containing region detailed further below (Figure 4A).

Large-scale chromosomal rearrangements such as chromosome fusions and inversions were observed between I-73-1 and BFP. This was illustrated clearly by the chr 1 fragmentation in I-73-1, whereas chr 1 is the largest chromosome in BFP (10.2 Mb). A significant section of chr 1 aligned with contig 18 in I-73-1, but the remaining sections were found to be distributed between four contigs: 1, 3, 16, and 20. Additionally, BFP chr 8 was split between I-73-1 contigs 18 and 20.

Approximately 30% of chr 3 (contig 2 in I-73-1) was inverted. Similar rearrangements on a smaller scale were observable in the alignment dotplot (Figure 3).

Similar to I-73-1, a chr 1 fragmentation was observed in D308. However, the fragmentation in the latter was less severe, with sections aligning to three contigs (D308 contigs 1, 2, and 15). There was a collinearity between D308 contigs 11 and 14 and chr 7 in BFP, and hence, these two contigs likely represent a single chromosome that would align fully with BFP chromosome 7 and I-73-1 contig 9. Furthermore, D308 contigs 7 and 25 fully aligned with BFP chr 2. The genome architectures of D308 and I-73-1 appear to exhibit higher collinearity between each other than to isolate BFP (Figure 3). We found similar types of rearrangements in alignments with M4 and DW-5 although they were fewer in number (not shown). There was no evidence for the presence of supernumerary chromosomes in I-73-1 or D308.

ToxA is present on a Starship class transposon and ToxB is present within a massive putative transposon

ToxA in BFP was located on chr 6 (2.8 Mb), but in I-73-1, *ToxA* was carried on contig 3 (4.0 Mb) which aligns with BFP chr 1 and 9. While BFP chr 6 was present in I-73-1 as contig 17, the contig lacked *ToxA* and a 143 Kb segment (Figure 4A). This validates the previously reported translocation of *ToxA* in I-73-1 (13). Edge analysis and review of gene content within the 143 kb segment indicated target site duplications (or short direct repeats) at the edges and a tyrosine recombinase (gene_09853; DUF3435 domain) at the 5' end, initially suggesting this element is a *crypton* (30). Recently, however, a new class of large mobile elements called *Starships* have been defined (31). An examination of other genes present within the transposon suggests that this element is actually a *Starship* with many similarities, including: the DUF3435 tyrosine recombinase 'captain', DUF3273 domains, ferric reductase, ankyrin repeats, heterokaryon incompatibility, NACHT nucleoside triphosphatase, and target site duplications (Figure 4B) (31). A full list of genes present are available in Supplemental File 5. The target site itself was identifiable in BFP chr1 (Figure 4B and Supplemental File 6), M4 (not shown), and D308 (not shown). We have named this new *Starship* 'Horizon'. Additionally, an alignment of the ToxhAT transposon, which facilitated the movement of *ToxA* between Ptr, *Parastagonospora nodorum*, and *Bipolaris sorokiniana* (14), showed that this smaller 14 Kb TE is nested within Horizon.

Four copies of *ToxB* were found in the I-73-1 long-read assembly; each copy contained a single 49 bp exon. Three copies were located within a 12 Kb region on contig 12 (homologous to chr 4 in BFP), and the fourth copy was present on a small 6 Kb contig (contig 13), which failed to assemble with the chromosome-sized contigs. The Australian isolate M4 (race 1) offered a better scaffolding quality than BFP to examine the 12 Kb *ToxB* region, and the alignment of I 73-1 contig 12 to M4 contig NQIK01000005 indicated an even larger 294 Kb gap around the region of *ToxB* (Figure 5). Examination of the edges of this gap revealed the presence of terminal inverted repeats in I-73-1 and a potential target site duplication in M4. This same 294 Kb transposon is present in D308 on contig 9, and although the *ToxB* gene is missing in D308, its inactive homolog *toxb* is present within the same transposon as a single copy (Supplemental File 7).

ToxB/toxb genes were found here to be located on an essential chromosome/contig in Ptr, as the same chromosome/contig is present in other isolates that do not carry the *ToxB* gene (BFP, D308 and M4). DW5 was reported to carry 10 copies of *ToxB* (32). Our BLAST output indicated 10 copies, but eight of these were on contigs ≤ 126 Kb, which may indicate misassembly. Only two copies of *ToxB* were on contigs of comparable size to chromosomes (CM025819 at 3.4 Mb and CM025824 at 2.2 Mb respectively). These contigs aligned with BFP chr 5 and 11 and I-73-1 contigs 7 and 16, respectively (Supplemental File 8). The locations of *ToxB* copies in DW-5 do not appear to have homology to the putative transposon in I-73-1 and D308, which may indicate additional transposon activity associated with *ToxB*.

The chromosomes and contigs that possess *ToxB*/*toxb* were fully (or nearly fully) present in all isolates examined, indicating they are essential in nature or contain essential genes for Ptr. Additionally, the presence of core and accessory genes and their ratios in isolates I-73-1 and D308 were examined, and for I-73-1 contig 12, the ratio of core to accessory genes was 3.4, while for D308 contig 9 the ratio was 3.8, supporting the essential nature of these chromosomes (Supplemental File 9).

Ptr exhibits a 'one-compartment' genome

The genome of many plant pathogens exhibits a two-compartment arrangement, one with AT-rich regions where effectors and TEs are embedded, and a second with GC-rich content (20). In order to evaluate if the Ptr genome exhibited compartmentalization, the intergenic distances (the number of nucleotides between genes) were measured and found to average ~4,600 bp in both I-

73-1 and D308. Density plots of the 5' and 3' intergenic distances showed a single hot spot at a similar size range (Figure 6). A scattering of genes in the upper right portion of the plot indicates that some genes are in gene-sparse regions, with intergenic distances as high as 100,000 bp in certain cases. The single hot spot (rather than two) is indicative of a 'one-compartment' genome, and although there are a number of genes in gene-sparse regions, their frequency does not support a 'two-compartment' genome. Additionally, analysis with RIPper showed that <2% of the I-73-1 and D308 genomes were affected by RIP, with <40 LRARs present in each.

Ptr genome and expansion by transposable elements

Transposable element (TE) content in the short read assemblies ranged from 6.4% to 10.2% of the assembled genomes (Figure 7A and Supplemental File 10). The highest TE content occurred in isolates 92-171R5 (10.2%) and G9-4 (9.8%). A large proportion of the TEs in 92-171R5 were classified as 'unknown' repetitive elements. In G9-4, however, the increased number of TEs were primarily Class I-LTRs. The high TE content in these two isolates also correlates with larger genome sizes relative to the other assemblies (Figure 7B). TE content in 12 isolates (excluding 92-171R5 and G9-4) exceeded 7%, and the remaining 26 isolates have TE content ranging between 6.4% and 7.0%. All assemblies contained significant numbers of 'unknown' repetitive elements. The program EDTA (Extensive de-novo TE Annotator) uses structural features to identify intact TEs at the beginning, and then classifies them into families based on coding features. RepeatModeler2 was used to identify repeats not initially reported, but due to a lack of homology and coding features, they were labelled as 'unknown' TEs.

The long-read assemblies of isolates I-73-1 and D308 showed a significantly larger number of transposable elements (>150%) (Supplemental File 11). TEs represented 6.8% and >18.3% for the short- and long-read assemblies, respectively. In both isolates, the increased TE content was due to a greater incidence of Class I transposons, primarily Copia and Gypsy elements. Additionally, many previously uncategorized transposons (labelled as 'unknown' in the short-reads), were reclassified with accurate transposon labels. Almost all groups of transposons identified in the short-read assemblies showed some increased incidence in the long-read assemblies, with the exceptions of the Tc1-Mariner, Helitrons, and PIF-Harbinger classes, which were consistent between assembly types.

DISCUSSION

In this study, we performed a global pangenome analysis of Ptr, an important foliar pathogen of wheat. In total, 41 Ptr genomes carrying various combinations of all known and unknown effectors were analysed. These isolates represent various geographical origins, extending from regions where wheat is of relatively recent introduction, to regions encompassing the centre of wheat origin. We identified major rearrangements at the chromosomal level among five assembled Ptr genomes, two of which were generated in this study (I-73-1 and D308) and three of which were previously described (BFP, M4, and DW5). These rearrangements included chromosomal fusions, segment inversions, and translocations. We showed that the Ptr genome appears to tolerate large-scale structural reorganizations, genome size variation, and has remarkable plasticity. Previously, a worldwide collection of Ptr isolates, some of which were included in this study, had shown independent chromosomal location of the virulence genes *ToxA* and *ToxB* and extensive genome plasticity (karyotypes), where aneuploidy, proliferation of repetitive DNA and transposon activities were suggested as driving mechanisms (13). Here, we showed plasticity in the Ptr genome is likely to be facilitated by the proliferation of TEs and the expansion of the accessory genome, in addition to the role of transposable elements in virulence gene shuffling.

Fungi are known to display high variability in their genomes, and the ability of plant pathogenic species to gain virulence genes embedded in a "pathogenicity island" and carried on supernumerary chromosomes that transfer horizontally as whole or in part is well described (33). In recent years, more reports have emerged on the ability of virulence genes to transfer on large transposons among plant pathogens (31). Overall, the ability of pathogenic fungi to evolve rapidly to invade their hosts or cause outbreaks is illustrated through their ability to recombine, mutate and shuffle their genetic components either vertically or horizontally. Traditionally, a 'two-compartment' genome has been associated with a 'two-speed' genome in pathogenic fungi. In this case, Ptr appears to possess a 'one-compartment' genome; however, this does not preclude the species from having a 'two-speed' genome driven by copy number variation and TEs not associated with compartmentalization (20), as we know *ToxB* is present as multiple copies in Ptr. Other effectors or genes associated with virulence may be present in multiple copies, but this remains to be explored.

The intraspecies translocation of ToxA in Ptr is facilitated by the Starship transposon 'Horizon'

ToxA is present in other fungal species such as *Parastagonospora nodorum* and *Bipolaris sorokiniana* (7, 14). The independent horizontal transfers between these species has been explored in detail, with the transfer being facilitated via a hAT transposon dubbed ToxhAT, which is 14 Kb in size (14). Previously, *ToxA* was found to be located on the same essential chromosome in all tested Ptr isolates with the exception of I-73-1, where it was shown to be translocated to a larger non-homologous chromosome, as indicated by pulse-field gel electrophoresis followed by southern hybridization (13). Although ToxhAT is present, we validate the translocation of *ToxA* in I-73-1 via ToxhAT, but nested on a much larger putative 143 Kb *Starship* transposon, which we have named 'Horizon'. This mobile element was integrated into an essential chromosome (contig 3; 4Mb) in I-73-1, corresponding to chr 9 in BFP. The *ToxA* coding gene in all Ptr isolates tested here has a very conserved sequence belonging to one haplotype (PtrH1) (34) (data not shown), and this supports the hypothesis of its recent integration into the Ptr species (7, 35).

The movement of *ToxA* in different types of mobile elements in Ptr and the nesting of transposons is indicative of this pathogen's ability to evolve and acquire virulence rapidly. The nesting of mobile elements with virulence factors has been linked to the evolution of pathogenicity in other species as well (14, 36, 37) and the association of TEs with effectors has been well documented in other fungal pathogens, most recently in another wheat pathogen *Zymospetoria tritici*, which is consistent with our findings (38). Intrachromosomal translocation of *ToxA* was also described in *B. sorokinaina* (14), and highlights the mobility of *ToxA* within Ptr and *B. sorokinaina* via transposon activity and/or genomic recombination. Multiple HGT events were previously suggested for ToxA and its 14 Kb surrounding region, the ToxhAT transposon (14). In this study we showed the nesting of ToxhAT in a larger mobile element Horizon which was not present in other ToxA containing isolates like BFP and M4. These findings add support to the idea that ToxhAT may have been transferred to Ptr multiple times, with one HGT event nesting ToxhAT within Horizon. Additionally, were Horizon to share sequence homology with other isolates or species in the future that would also support the hypothesis of multiple *ToxA* HGT events.

The ToxB gene is embedded on a large putative transposon in Ptr

For the first time, we revealed the possible movement of *ToxB* as a multi-copy gene clustered in race 8 isolate I-73-1 (contig 12), and the single copy *toxb* homolog on contig 5 in D308 race 3 on a large 294 Kb transposon potentially similar to that of *ToxA*, although an analysis that confirms this classification is required. No clear evidence of *ToxB* horizontal transfer has been found to date, but homologs of *ToxB* are present in closely related species (e.g., *P. bromi*, *Cochilobolus sativus*, *Alternaria alternata*, *Magnaporthe grisea*) (8). It has been suggested that *ToxB* was acquired vertically from a common ascomycete ancestor (8, 15). However, our discovery of *ToxB/toxb* on a large putative transposon inserted into an essential chromosome provides the new possibility of an alternative mechanism of acquisition via a horizontal transfer event older than that of *ToxA* (given the higher variability in *ToxB*/*toxb* reported sequences). Moreover, our analysis suggested a potential transposon(s) associated with *ToxB* in the previously published isolate DW-5, which provides further evidence of potential horizontal mobility of *ToxB*.

CAZymes are essential part of the Ptr genome

Ptr is a necrotrophic pathogen that posses the ability to directly penetrate the host epidermal cells soon after spore germination (39), and this penetration is likely facilitated by the fungus ability to secrete cell wall degrading enzymes. The CAZymes and CAZy families identified in the Ptr pangenome support that Ptr is adapted for plant cell wall degradation, as >30% of the total CAZome belongs to families involved in dismantling structural polysaccharides, such as cellulose, hemicelluloses, and the plant cuticle. Phytopathogenic fungi are known to deploy a number of CAZymes for invasion and infection (26, 28, 29). As an example, *Fusarium culmorum* has been shown to degrade cellulose, xylans, and pectins during invasion of wheat spike tissues (40). While previous work has linked the relatedness of CAZymes to fungal taxonomy (41), it is not known if there is functional specialization in different Ptr isolates. CAZomes may be tailored to accommodate unique structures in cell wall polysaccharides of different wheat varieties or plants growing in different geographic regions. Indeed, the diversification of AA9 structure, sequence, and expression levels in isolates of the phytopathogen *Rhizoctonia solani* has been proposed to be essential for the differential pathogenicity of these strains in rice and soybean (42), and the same may be true for Ptr and its hosts.

Ptr has an open pangenome with a high accessory gene content

In an open pangenome, each added genome increases the number of accessory genes, while the number of shared (core) genes decreases (43). In this work, we have shown that Ptr has an open pangenome. We also showed that Ptr, unlike other ascomycetes, contains a large proportion of accessory genes (57%) with a relatively small core gene set (43%). This is a significantly smaller core than the estimated 69% core previously reported in the pangenome of 11 Ptr isolates (17), and the 60% core estimated in the pangenome of the wheat pathogen *Z. tritici*, which was based on 19 isolates (44).

It has been suggested that plant pathogens possess accessory genomic elements related to pathogenicity (45). However, we showed that the Ptr pangenome possesses a higher number of accessory genes in the non-pathogenic and weakly virulent isolates; these isolates also exhibited a larger proportion of orphan (singleton) genes. Approximately 70% of singleton genes had unknown functions, but perhaps these genes are involved in adaptation for a lifestyle divergent to wheat pathogenesis (23). The open pangenome of Ptr may explain its ability to adapt to a wide variety of hosts, geographical regions, and environmental conditions. Despite its homothallic mating style, this pathogen evidently acquired virulence by horizontal transfer of a large segment of DNA from other fungal species (7, 14). The core genome is usually protected from high recombination and mutation rates in order to preserve its essential biological function, and the accessory genome is assumed to be under rapid evolution (45). It is still unclear how the accessory genome in Ptr evolves and why the non-pathogenic isolates exhibited higher contents of accessory genes and TEs.

The observed separation of isolates based on ToxA production in the core protein phylogeny and the accessory protein hierarchical sets may be due to the host specialization of Ptr between bread and durum wheat. A similar observation has been made before, using simple sequence repeats on a similar collection of isolates (12). The non-pathogenic isolates (except T-126-1 from Tunisia) were also distantly related to the rest of the isolates, and we observed a divergence in TE and effector content in these isolates, which is also an indicator of host specialization.

We also found variation in the genome size, particularly between pathogenic and non-pathogenic isolates. The genome size in the Dothideomycetes, based on the analysis of 101 species, varies 10-fold: from less than 17 Mb to more than 177 MB (46). Our Ptr genome size based on the

more accurate long read assemblies averaged at 39.8 Mb, which was very consistent with previous genome size estimates for long-read Ptr in previous studies (17, 19, 32). However, we showed that the non-pathogenic race 4 and weakly virulent race 5 isolates, G9-4 and 92-171R5 genome size were ~2 Mb larger than the average pathogenic isolates, and both isolates have an average of 3.1% more TE content than the other isolates. This clearly indicates an expansion of TEs in the genomes of these isolates. This larger genome size in non-pathogenic isolates was not expected, as previous reports of Ptr genome sizes indicated that pathogenic isolates have larger genomes, which was not consistent with our observations (17). The larger genomes reported in pathogenic races/species relative to their non-pathogenic counterparts has been attributed to the presence of more repetitive elements, and it was posited that the need of pathogenic species to evolve in an arm's race with its host could explain such genome expansion (47). Indeed, in some plant pathogens the acquisition of an entire supernumerary chromosome, often rich with repetitive sequences, by horizontal transfer is evident (48). However, not all pathogenic species follow this trend of a larger genome (35). A reduction in pathogen genome size may be an evolutionary mechanism to aid adaptation to specific environments or other niches, and could explain the recent specialization of pathogenic species from related generalist non-pathogens (49-52). Many pathogens adapt to exploit a limited number of species efficiently, and in some species, even a limited number of genotypes within a species (39) with rapid gene gains and losses previously linked to virulence adaptation in other studies (53, 54).

Nonetheless, while a larger genome size was found in the non-pathogenic race 4 isolate G9-4 and the weakly virulent race 5 isolate 92-171R5, this was not the case for the other race 4 isolates, 90-2 and T126-1, which had genome sizes similar to the pathogenic strains. As such, it is not easy to make a clear conclusion regarding genome size and pathogenicity in Ptr, and there is a need to explore additional non-pathogenic genomes. In previous studies, comparisons with Ptr non-pathogenic genomes were based on a single race 4 isolate (SD20-NP) (16, 17). Additionally, in many other ascomycetes, the number of non-pathogenic genomes studied is limited in comparison with pathogenic genomes (55). It is worth noting that non-pathogenicity in Ptr often has been assumed based an isolate's inability to cause disease symptoms on a limited number of host genotypes. However, what were assumed to be non-pathogenic race 4 isolates were recently found to cause extensive necrosis on specific durum wheat genotypes (56).

Conclusions

Collectively, this work highlights the high plasticity and potential adaptability of Ptr as a global wheat pathogen. The large-scale chromosomal rearrangements, open pangenome, and the extensive accessory gene and TEs content of its genome reflect its cosmopolitan nature. In addition, the nesting of the virulence genes *ToxA* and *ToxB* within multiple transposon types is a significant indication of the rapid evolutionary nature of this pathogen.

Methods and Materials

Isolates, DNA extraction, and sequencing

Isolate details are provided in Table 1. Virulence phenotypes were previously confirmed (4, 12, 13, 57). Isolates collected before 2016 and received from other labs were subjected to single spore isolates and their phenotypes confirmed on a wheat differential set as described by Aboukhaddour et al., (2013). Isolates were grown in 100 mL 1/4 concentration PDB and cultures incubated in a shaker (100 RPM) at room temperature (~25°C), without light, until the fungal mat covered the surface of the medium. Fungal mats were harvested and washed twice using autoclaved mili Q water. Mats were placed in whirl pack bags on dry ice until freeze-dried. Dried mats were stored at -20°C until DNA extraction (no longer than 14 days). Genomic DNA was extracted using a 'Genomic-tip 20/G' kit (Qiagen) and sequenced with 150 bp paired-end, 400 bp inserts at 100x coverage with Illumina HiSeq X. DNA from I-73-1 and D308 was extracted using a 'Genomic-tip 100/G' kit (Qiagen) and long-reads sequenced with PacBio RS II at 100x coverage. All sequencing was performed by the Centre d'expertise et de services Génome Québec (Montreal, Canada). The reference isolate BFP (accession GCA_000149985) (16) was included in the pangenome analysis and used for full genome alignments. Isolates M4 (GCA_003171515) (17) and DW5 (GCA_003231415) (32) were also included for *Tox* gene and transposon analysis.

De novo assemblies

Read quality was assessed with FASTQC (58) and poor quality reads filtered. Reads were filtered using the standard Kraken2 database (59) and any reads tagged as non-fungi were omitted. A subset of isolates (90-2, AB88-2, ASC1, AZ35-5, and I-72-1) were selected to test assembly program suitability using the Illumina reads. The assemblers were Shovill with SPAdes (60, 61), Shovill with MEGAHIT (61, 62), SOAPDenovo2 (63), and CLC Genomics Workbench 12

(Qiagen) all program arguments available in GitHub repository. QUAST (64) output and BUSCO scores were used to assess assembly quality. All remaining short-reads were assembled with Shovill/SPAdes. Long-reads were assembled with Flye (65) and polished with short-read data using Pilon (66). Completeness of the long-read assemblies was also assessed by the alignment of raw reads (Supplemental File 12).

Gene annotations

FunGAP (v1.0.1) is an annotation pipeline specifically for fungi (67). FunGAP makes use of many programs: RepeatModeler (68), RepeatMasker (69), HISAT (70), Trinity (71), Augustus (72), MAKER (73), BRAKER (74), InterProScan (75), and BLAST+ (76). MAKER and BRAKER utilize RNA-seq reads. RNA-seq reads from vegetative Ptr mycelia were retrieved from BioPlatforms Data Portal (https://data.bioplatforms.com/; Sample ID: 102.100.100.14350) (17). FunGAP provides a script to select a prediction model for Augustus, with *Botrytis cinerea* selected as the model in this case. Default settings provided by FunGAP were used to annotate all assemblies. After pangenome analysis (discussed below), representatives from the core and accessory gene clusters were used for functional annotation with Pfam v28.0 (77). All predicted gene sets were assessed for completeness using the Ascomycota BUSCO gene set (odb9) (78, 79).

Pangenome analysis

Pangloss is a pangenome pipeline designed for microbial eukaryotes like fungi (21). Panoct (80) is the primary program for calculating a pangenome. A custom bash script was used to modify *.gff3 files from FunGAP into *.attributes format required by PanOCT. Output from all-vs-all BLASTp ($E10^{-4}$) was used to determine if a gene belongs in the core or accessory genome. Pangloss does not distinguish singletons from the accessory genome, a custom bash script separated singletons. Binary gene presence or absence tables were parsed to generate figures showing the percentage of the genome comprised by core, accessory or singleton genes, total genes in the pangenome as new genomes were added, and the number of genes in accessory clusters. Statistics including t-tests, ANOVA and Tukey's HSD were performed with R (v3.4.3) in RStudio.

CAZymes and effectors

Phobius v1.01 (with --short) (81) was used to filter amino acid sequences for signal peptides and transmembrane domains. Sequences with signal peptides and without transmembrane domains were used as input for EffectorP-2.0 (82). Genes identified as potential effectors (probability >50%) were extracted from gene sets and used as input for Pangloss to assess presence/absence between isolates. Predicted protein sequences produced by FunGap from all isolates were annotated by dbCAN2 (83) and manually curated for selected GH, CE, PL, and AA families relevant for plant cell wall degradation from the CAZy database (24).

Phylogeny and accessory sets

Individual alignments for each of the 10,159 amino acid core orthologue clusters was performed using MUSCLE (v3.6) (84). Bash and python scripts were used to concatenate and combine the aligned core amino acid gene set for each isolate creating a single alignment. The core alignment was input for RAxML (85) to generate a Maximum Likelihood (ML) phylogeny using the PROTGAMMA model, 1,000 bootstrap replicates, and a starting seed of 10. Variant call files were generated via GATK HaplotypeCaller (86) using BFP as a reference. SNPs were converted to fasta format with vcf2phylip (87) and input into RAxML for a ML tree using GRTCAT model. The ML trees were visualized with FigTree (88) and the Tox/location content added manually. Binary data was filtered to include only accessory genes and was input for the R package HierarchicalSets with default settings (89).

Transposable element content

Transposable elements (TEs) were identified and categorized using EDTA with the higher sensitivity setting (--sensitive 1) (90), which utilizes several programs: LTRharvest (91), LTR_FINDER (92), LTR_FINDER_parrallel (93), LTR_retriever (94), and TIR-Learner (95), Generic Repeat Finder (96), HelitronScanner (97), and TEsorter (98). Output was aggregated and the total TE content per genome (stacked histogram) and TE content as a function of genome size (Mb) visualized.

Genome organization

Pair-wise full genome alignments of long-read assemblies: I-73-1, D308, BFP, M4, and DW-5 were performed with minimap2 (99) and visualized with dotPlotly (100). For chromosomes

with *ToxA* and *ToxB* putative transposons, syntenic blocks with a minimum length of 6,500 bp were generated using Sibelia (v3.0.7)(101), then visualized using Circos (v0.69-8)(102). Selected proteins within the putative transposons were modelled with Phyre2 (103). Intergenic distances were calculated from gff3 files using an adapted R-script (104) and plotted using an hexagonal heatmap of 2d bin counts geom_hex with 50 bins. Long-read genomes were uploaded to RIPper (105) to assess RIP and LRAR content. The core vs. accessory nature of contigs was assessed by alignment (i.e., presence in all isolates) and evaluation of the ratio of core to accessory genes, percentage of total genes, and genes per Mb on each contig.

Data availability

Isolate accession numbers are available in file Supplemental File 13. Raw data is available under NCBI BioProject PRJNA80319. Scripts and program settings used in this work are available at https://github.com/fungal-spore/Ptr-pangenome-paper.

AUTHOR CONTRIBUTIONS

RG performed most of the analyses including assemblies/annotations, pangenome, phylogenetics, TE content, chromosomal organization, effectors, and *ToxA* and *ToxB* transposon analysis. The first draft of this manuscript was prepared by RG, RA, and MM. MM aided and guided the transposon analysis, and project guidance. ROP aided with assemblies and analysis guidance. MH performed DNA extractions. KEL and DWA performed CAZyme work. SES and FD provided isolates and input. RA conceived/guided project, performed single-spore isolation for a number of isolates and confirmed phenotyping. All authors contributed to editing and improving the manuscript.

ORCID

RG: 0000-0003-3515-6737
MM: 0000-0002-9129-7393
MH: 0000-0001-9111-4983
ROP: 0000-0003-3411-923X
SS: 0000-0001-7069-9513
KEL: 0000-0003-1093-0044

DWA: 0000-0003-4893-5799

FD: 0000-0001-8912-0460

RA: 0000-0002-4040-9471

DEDICATION

This work is dedicated to the memory of the late Dr. Lakhdar Lamari, a passionate scientist and an exceptional teacher who had made significant contributions to the tan spot research community and the world of necrotrophic pathogens. Many of the isolates included in this study were collected by him during his trips to the Fertile Crescent, North Africa, and Caucuses regions.

ACKNOWLEDGEMENTS

We would like to thank AAFC BioCluster Team, Therese Despins (culture maintenance), and Kathryn Wilde (image formatting), Sana Kamel and Mejda Cherif (University of Carthage, Tunis; Tunisian isolates). Funding: Agriculture and Agri-Food Canada and Alberta Wheat Commission and Saskatchewan Wheat Development to RA. The funding bodies were not involved in the design of the experiments and collection, analysis, and interpretation of data, or in the writing of this manuscript.

604 **Tables**

Table 1. Summary statistics for 40 isolates of *Pyrenophora tritici-repentis* sequenced with Illumina HiSeq X and assembled with Shovill/SPAdes, and two isolates sequenced with PacBio and assembled with Flye/Pilon (table is sorted by race then isolate name).

Isolate	Race	NE[1]	Year	Location[2]	Host	Size (MB)	Contigs	GC (%)	TE (%)	N50 (kbp)	Genes	BUSCO (%)	Accessory (%)
ASC1	1	AC	1985	Manitoba	*T. aestivum*	34.78	6,495	50.9	6.9	65.5	13,124	99.3	22.5
I-33-1	1	AC	2001	Azerbaijan	*T. aestivum*	35.06	6,850	50.9	6.9	75.2	13,055	99.5	22.1
L3-1	1	AC	2016	Alberta	*T. durum*	34.93	6,666	50.9	6.9	77.2	13,115	99.4	22.4
L4-1	1	AC	2016	Alberta	*T. durum*	34.74	6,361	50.9	6.9	72.4	13,063	99.6	22.1
SW20-7	1	AC	2016	Saskatchewan	*T. durum*	35	6,568	50.9	7.2	76.2	13,116	99.4	22.5
SW2-1	1	AC	2016	Saskatchewan	*T. durum*	35.11	6,789	50.9	6.9	74.3	13,116	99.7	21.3
SW21-1	1	AC	2016	Saskatchewan	*T. durum*	34.62	6,282	50.9	6.8	74.3	12,965	99.4	21.6
SW21-7	1	AC	2016	Saskatchewan	*T. durum*	34.97	6,741	50.9	7	74.7	13,126	99.4	22.5
SW21-8	1	AC	2016	Saskatchewan	*T. durum*	34.96	6,631	50.9	6.9	75	13,073	99.4	22.2
SW7-5	1	AC	2016	Saskatchewan	*T. durum*	35.65	6,701	50.8	6.9	73.7	13,490	99.6	24.6
86-124	2	A	1986	Manitoba	*T. aestivum*	34.9	6,832	50.9	6.9	60.7	13,209	99.2	23
AB88-2	2	A	2010	Alberta	*T. aestivum*	34.83	6,465	50.9	6.9	75.3	13,126	99.5	22.5
L2-1	2	A	2016	Alberta	*T. durum*	34.96	6,923	50.9	7.1	72.9	13,130	99.4	22.5
SW1-2	2	A	2016	Saskatchewan	*T. durum*	35.08	6,599	50.9	7.1	71.2	13,365	99.4	24
SW15-1	2	A	2016	Saskatchewan	*T. durum*	34.83	6,272	50.9	6.6	78.8	13,201	99.7	22.9
T132-2	2	A	2017	Tunisia	*T. durum*	34.41	6,472	50.9	6.8	57.1	12,935	99.4	21.3
331-2	3	C	1985	Manitoba	*Triticum* sp.	34.44	6,828	51	6.9	55.6	12,909	99.5	21.2
D308	3	C	1985	Manitoba	*T. aestivum*	34.33	6,809	51	6.8	58.5	12,826	99.1	N/A[¥]
D308*	3	C	1985	Manitoba	*T. aestivum*	39.6	70	50.8	18.6	3,667	12,501	99.6	18.7
I-72-1	3	C	2001	Syria	*Triticum* sp.	34.35	6,734	51	6.8	58.6	13,011	99.5	21.8
I-72-7	3	C	2001	Syria	*Triticum* sp.	34.35	6,696	51	7	58.9	12,901	99.4	21.2
SC29-1	3	C	1999	Saskatchewan	*T. durum*	34.19	6,464	51	6.8	59	12,951	99.5	21.5
SW21-5	3	C	2016	Saskatchewan	*T. durum*	34.66	6,619	51	7.2	63.5	13,029	99.5	22
90-2	4	absent	1990	Canada	*T. aestivum*	35.22	3,818	50.8	7.2	225.9	12,909	99.5	21.3
G9-4	4	absent	2016	Alberta	wild grass	36.97	8,035	50.7	9.8	78.2	13,148	98.7	23

T126-1	4	absent	2017	Tunisia	*T. durum*	34.15	6,373	51	6.5	62.4	12,837	99.6	20.6
92-171R5	5	B	1992	Saskatchewan	*T. aestivum*	36.81	14,647	51.1	10.2	45.1	13,393	99.3	24.1
Alg3-24	5	B	1993	Algeria	*T. durum*	34.3	6,098	50.9	6.6	73.6	12,900	99.6	21.1
Alg4x-1	5	B	1993	Algeria	*T. durum*	35.71	8,072	50.9	7.5	71	13,193	99.4	22.9
I-17-2	5	B	2001	Azerbaijan	*T. durum*	34.25	6,555	51	7	62.5	12,820	99.6	20.7
I-34-5	5	B	2001	Azerbaijan	*Triticum* sp.	34.29	6,315	51	6.8	61.9	12,841	99.6	20.8
I-35-56	5	B	2001	Azerbaijan	*T. durum*	34.24	6,516	51	6.9	62.6	12,918	99.5	21.3
I-36-1	5	B	2001	Azerbaijan	*T. durum*	34.36	6,467	50.9	6.7	62.4	12,881	99.5	21
AlgH1	6	BC	1993	Algeria	*T. durum*	34.74	6,902	51	6.9	61.7	13,159	99.3	22.7
AZ35-5	7	AB	2001	Azerbaijan	*T. durum*	35.3	7,165	50.9	7.1	72.1	13,239	99.5	23.2
T176-2	7	AB	2017	Tunisia	*T. aestivum*	34.7	6,896	50.9	7.3	57.1	13,141	99.5	22.6
T181-1	7	AB	2017	Tunisia	*T. durum*	34.78	6,718	51.2	6.5	57.5	13,583	99.4	25.1
I-34-1	8	ABC	2001	Azerbaijan	*T. durum*	34.38	6,405	50.9	6.6	64.5	13,036	99.5	22
I-35-18	8	ABC	2001	Azerbaijan	*T. durum*	34.85	6,698	50.9	7	65.3	13,071	99.4	22.9
I-73-1	8	ABC	2001	Syria	*T. aestivum*	34.62	6,619	50.9	6.8	63.4	12,941	99.5	N/A[¥]
I-73-1*	8	ABC	2001	Syria	*T. aestivum*	39.9	39	50.8	18.3	3,647	12,744	99.6	20
T128-1	atypical	B	2017	Tunisia	*T. durum*	34.12	6,095	51	6.4	59.1	13,002	99.6	21.7

[1]NE= Necrotrophic effector; [2]Alberta, Saskatchewan, and Manitoba are provinces within Canada; *Long-read assemblies; [¥]not included in pangenome

605

Figure captions

Figure 1. The pangenome of *Pyrenophora-tritici repentis*. **a** Proportion of genes present in core, accessory (excluding singletons), and singleton sets for pangenome, CAZymes, and effectors. **b** Total number of unique genes and core genes in the pangenome of *Pyrenophora-tritici repentis* as genomes are added. Starting genome was always Pt-1C-BFP, with subsequent genomes added randomly, iterated x100, and visualized as boxplots.

Figure 2. Evolutionary relationships of isolates in the Ptr pangenome. **a** Maximum-likelihood phylogenetic tree of *Pyrenophora-tritici repentis* based on aligned and concatenated proteins present in the core genome (total 10,159 genes). Isolate names are followed by their race (R1 to R8) designation, necrotrophic effector(s) produced (A,B, and/or C), and their location of origin (isolates collected within Canada show province names; isolates from outside Canada show country of origin). Presence of *ToxA* (blue circles) and *ToxB* (orange circles), confirmed via BLAST, and presence of *ToxC* (grey circles) determined by phenotyping. **b** Hierarchical sets of *Pyrenophora tritici-repentis* accessory pangenome, size represents number of genes shared by isolates in each cluster, where inclusion in a cluster is based on any given isolate overlapping with the horizontal bars.

Figure 3. Full genome alignments of newly sequenced long-read assemblies of *Pyrenophora tritici-repentis* isolates I73-1 (Race 8) and D308 (Race 3) to the reference isolate Pt-1C-BFP (Race 1). **a** Pt-1C-BFP and I73-1; **b** Pt-1C-BFP and D308; **c** I73-1 and D308.

Figure 4. Evidence of translocation of *ToxA* containing region between two races of *Pyrenophora tritici-repentis*. **a** The *ToxA* containing chromosome 6 of isolate Pt-1C-BFP (race 1) is compared with the *ToxA* containing contig 3 and contig 17 of isolate I73-1 (race 8). Pt-1C-BFP chromosome 6 aligns almost completely to I73-1 contig 17, with the exception of the *ToxA* containing region, which aligns to contig 3. The majority of contig 3 aligns to Pt-1C-BFP chromosome 1 or chromosome 9 (Figure 3). **b** schematic of the 'Horizon' Starship class transposon.

Figure 5. Circular alignment of contig 12 from race 8 isolates I-73-1 and contig NQIK01000005 from race 1 isolates M4. A large 294 Kb region which contains three copies of the *ToxB* (black arrow) is visible, this section did not align with any other contig of M4. The edges of the 294 Kb region revealed terminal inverted repeats suggesting transposon activity.

Figure 6. Intergenic distances of all genes in *Pyrenophora tritici-repentis* of: **a** isolate D308; **b** isolate I-73-1. The 3' intergenic length (x-axis) is the distance (bp) from the 3' end of current gene to the 5' end of next, and the 5' intergenic length (y-axis) is the distance from the 3' end of the previous gene to the 5' end of the current gene.

Figure 7. Transposable element content and genome size variation of *Pyrenophora tritici-repentis* from short-read assemblies. **a** Contribution of TEs (%) to total genome size across 40 isolates. **b** Contribution of each TE class to genome size for individual isolates.

Supplemental File Captions

S01. Genes were clustered based on the number of isolates in which they were present (e.g. genes in cluster 2 are present in two isolates, genes in cluster 3 are present in three isolates, etc.). Genes

in low clusters may represent recently gained genes, as only a few isolate contain them, while genes in high clusters may represent recently lost genes, as most isolates contain them. Clusters 1 and 41 were omitted as they represent singletons and the core gene set respectively.

S02. Maximum likelihood phylogenies created by RAxML based on SNP data. **a** ML tree with all isolates; **b** ML tree with the divergent outgroup omitted to aid reading of the other branches.

S03. Number of proteins of a given length in the core gene set (green), the accessory set not including singletons (red), and singleton gene set of the *Pyrenophora tritici-repentis* pangenome.

S04. Curated output from dbCAN for grouping and counting the number of genes in specific CAZyme families.

S05. Annotation of genes present within the 143 kb *Starship* transposon 'Horizon' present in the race 8 isolate I-73-1.

S06. Truncated MAUVE alignment of BFP chr1 and I-73-1 contig 3 which shows the target insertion site and short direct repeats of putative *ToxA* transposon.

S07. Circular alignment of contig 5 from race 3 isolate D308, contig 12 from race 8 isolates I-73-1, and contig NQIK01000005 from race 1 isolates M4. A large 294 Kb region which contains three copies of the *ToxB* (black arrow) is visible which aligns with a section in D308 which contains a single copy of the inactive *toxb* (grey arrow).

S08. Circular alignment of *ToxB* carrying DW-5 contigs (CM025819.1 and CM025824.1) to matching Pt-1C-BFP chromosomes (chr 5 and 11 respectively). Sections containing *ToxB* (black arrows) do not appear to share homology with each other or the reference chromosomes indicating possible transposon activity.

S09. Number of core, accessory, and singleton genes on individual contigs of the long-read assemblies of I-73-1 and D308.

S10. Table summary of transposable element content for each isolate of Ptr.

S11. The percentage of transposable elements annotated by EDTA in short-read vs. long-read assemblies of isolates I-73-1 and D308.

S12. Alignment of raw read data to long-read assemblies; **a** I-73-1; **b** D308.

S13. List of GenBank accession numbers associated with the genomes generated by this study and of the genomes downloaded and used for comparison.

REFERENCES

1. Savary S, Willocquet L, Pethybridge SJ, Esker P, McRoberts N, Nelson A. 2019. The global burden of pathogens and pests on major food crops. Nature Ecology and Evolution 3:430-439.

697 2. Aboukhaddour R, Hafez M, Strelkov S, Fernandez MR. 2021. Tan spot under the lenses of plant
698 pathologists. *In* Oliver RP (ed), Achieving durable resistance in cereals. Burliegh Dodds,
699 Cambridge, United Kingdom.
700 3. Lamari L, Strelkov S, Yahyaoui A, Orabi J, Smith R. 2003. The identification of two new races of
701 *Pyrenophora tritici-repentis* from the host center of diversity confirms a one-to-one relationship
702 in tan spot of wheat. Phytopathology 93:391-396.
703 4. Kamel S, Cherif M, Hafez M, Despins T, Aboukhaddour R. 2019. *Pyrenophora tritici-repentis* in
704 Tunisia: race structure and effector genes. Frontiers in Plant Science 10:1562.
705 5. Effertz R, Meinhardt SW, Anderson J, Jordahl J, Francl L. 2002. Identification of a chlorosis-
706 inducing toxin from *Pyrenophora tritici-repentis* and the chromosomal location of an
707 insensitivity locus in wheat. Phytopathology 92:527-533.
708 6. Shi GS, Kariyawasam G, Liu S, Leng Y, Zhong S, Ali S, Moolhuijzen P, Moffat C, Rasmussen JB,
709 Friesen TL. 2022. A conserved hypothetical gene is required but not sufficient for Ptr ToxC
710 production in *Pyrenophora tritici-repentis*. Molecular Plant-Microbe Interactions.
711 7. Friesen TL, Stukenbrock EH, Liu Z, Meinhardt S, Ling H, Faris JD, Rasmussen JB, Solomon PS,
712 McDonald BA, Oliver RP. 2006. Emergence of a new disease as a result of interspecific virulence
713 gene transfer. Nature Genetics 38:953-956.
714 8. Andrie RM, Schoch CL, Hedges R, Spatafora JW, Ciuffetti LM. 2008. Homologs of *ToxB*, a host-
715 selective toxin gene from *Pyrenophora tritici-repentis*, are present in the genome of sister-
716 species *Pyrenophora bromi* and other members of the Ascomycota. Fungal Genetics and Biology
717 45:363-377.
718 9. Lu S, Turgeon BG, Edwards MC. 2015. A *ToxA*-like protein from *Cochliobolus heterostrophus*
719 induces light-dependent leaf necrosis and acts as a virulence factor with host selectivity on
720 maize. Fungal Genetics and Biology 81:12-24.
721 10. Strelkov S, Kowatsch R, Ballance G, Lamari L. 2005. Characterization of the *ToxB* gene from
722 North African and Canadian isolates of *Pyrenophora tritici-repentis*. Physiological and Molecular
723 Plant Pathology 67:164-170.
724 11. Aboukhaddour R, Turkington TK, Strelkov SE. 2013. Race structure of *Pyrenophora triciti-repentis*
725 (tan spot of wheat) in Alberta, Canada. Canadian Journal of Plant Pathology 35:256-268.
726 12. Aboukhaddour R, Cloutier S, Lamari L, Strelkov S. 2011. Simple sequence repeats and diversity of
727 globally distributed populations of *Pyrenophora tritici-repentis*. Canadian Journal of Plant
728 Pathology 33:389-399.
729 13. Aboukhaddour R, Cloutier S, Ballance G, Lamari L. 2009. Genome characterization of
730 *Pyrenophora tritici-repentis* isolates reveals high plasticity and independent chromosomal
731 location of *ToxA* and *ToxB*. Molecular Plant Pathology 10:201-212.
732 14. McDonald MC, Taranto AP, Hill E, Schwessinger B, Liu Z, Simpfendorfer S, Milgate A, Solomon
733 PS. 2019. Transposon-mediated horizontal transfer of the host-specific virulence protein ToxA
734 between three fungal wheat pathogens. MBio 10.
735 15. Ciuffetti LM, Manning VA, Pandelova I, Faris JD, Friesen TL, Strelkov SE, Weber GL, Goodwin SB,
736 Wolpert TJ, Figueroa M. 2014. *Pyrenophora tritici-repentis*: a plant pathogenic fungus with
737 global impact, p 1-39, Genomics of plant-associated fungi: monocot pathogens. Springer.
738 16. Manning VA, Pandelova I, Dhillon B, Wilhelm LJ, Goodwin SB, Berlin AM, Figueroa M, Freitag M,
739 Hane JK, Henrissat B. 2013. Comparative genomics of a plant-pathogenic fungus, *Pyrenophora*
740 *tritici-repentis*, reveals transduplication and the impact of repeat elements on pathogenicity and
741 population divergence. G3: Genes Genomes Genetics 3:41-63.
742 17. Moolhuijzen P, See PT, Hane JK, Shi G, Liu Z, Oliver RP, Moffat CS. 2018. Comparative genomics
743 of the wheat fungal pathogen *Pyrenophora tritici-repentis* reveals chromosomal variations and
744 genome plasticity. BMC Genomics 19:279.

745 18. McDonald MC, Ahren D, Simpfendorfer S, Milgate A, Solomon PS. 2018. The discovery of the
746 virulence gene *ToxA* in the wheat and barley pathogen *Bipolaris sorokiniana*. Molecular Plant
747 Pathology 19:432-439.
748 19. Moolhuijzen P, See PT, Moffat CS. 2019. A new PacBio genome sequence of an Australian
749 *Pyrenophora tritici-repentis* race 1 isolate. BMC Res Notes 12:1-5.
750 20. Torres DE, Oggenfuss U, Croll D, Seidl MF. 2020. Genome evolution in fungal plant pathogens:
751 looking beyond the two-speed genome model. Fungal Biology Reviews.
752 21. McCarthy CG, Fitzpatrick DA. 2019. Pangloss: A Tool for Pan-Genome Analysis of Microbial
753 Eukaryotes. Genes 10:521.
754 22. Strelkov SE, Lamari L, Sayoud R, Smith RB. 2002. Comparative virulence of chlorosis-inducing
755 races of *Pyrenophora tritici-repentis*. Canadian Journal of Plant Pathology 24:29-35.
756 23. Wei B, Despins T, Fernandez MR, Strelkov SE, Ruan Y, Graf R, Aboukhaddour R. 2021. Race
757 distribution of *Pyrenophora tritici-repentis* in relation to ploidy level and susceptibility of durum
758 and winter bread wheat. Canadian Journal of Plant Pathology 43:582-598.
759 24. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-
760 active enzymes database (CAZy) in 2013. Nucleic Acids Research 42:D490-D495.
761 25. Rafiei V, Vélëz H, Tzelepis G. 2021. The Role of Glycoside Hydrolases in Phytopathogenic Fungi
762 and Oomycetes Virulence. International Journal of Molecular Sciences 22:9359.
763 26. Jagadeeswaran G, Veale L, Mort AJ. 2021. Do lytic polysaccharide monooxygenases aid in plant
764 pathogenesis and herbivory? Trends in Plant Science 26:142-155.
765 27. Nakamura AM, Nascimento AS, Polikarpov I. 2017. Structural diversity of carbohydrate
766 esterases. Biotechnology Research and Innovation 1:35-51.
767 28. Zhao Z, Liu H, Wang C, Xu J-R. 2013. Comparative analysis of fungal genomes reveals different
768 plant cell wall degrading capacity in fungi. BMC Genomics 14:1-15.
769 29. Davies K, De Lorono I, Foster S, Li D, Johnstone K, Ashby A. 2000. Evidence for a role of cutinase
770 in pathogenicity of *Pyrenopeziza brassicae* on brassicas. Physiological & Molecular Plant
771 Pathology 57:63-75.
772 30. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M,
773 Panaud O. 2007. A unified classification system for eukaryotic transposable elements. Nature
774 Reviews Genetics 8:973-982.
775 31. Gluck-Thaler E, Ralston T, Konkel Z, Ocampos CG, Ganeshan VD, Dorrance AE, Niblack TL, Wood
776 CW, Slot JC, Lopez-Nicora HD, Vogan AA. 2021. Giant *Starship* elements mobilize accessory
777 genes in fungal genomes. BioRxiv doi:10.1101/2021.12.13.472469 %J
778 bioRxiv:2021.12.13.472469.
779 32. Moolhuijzen P, See PT, Moffat CS. 2020. PacBio genome sequencing reveals new insights into
780 the genomic organisation of the multi-copy *ToxB* gene of the wheat fungal pathogen
781 *Pyrenophora tritici-repentis*. BMC Genomics 21:1-12.
782 33. Vanheule A, Audenaert K, Warris S, van de Geest H, Schijlen E, Höfte M, De Saeger S, Haesaert
783 G, Waalwijk C, van der Lee T. 2016. Living apart together: crosstalk between the core and
784 supernumerary genomes in a fungal plant pathogen. BMC Genomics 17:1-18.
785 34. Hafez M, Despins T, Nakajima K, Aboukhaddour R. 2022. Identification of a novel *ToxA* haplotype
786 of *Pyrenophora tritici-repentis* from Japan. Phytopathology.
787 35. Friesen TL, Ali S, Klein KK, Rasmussen JB. 2005. Population Genetic Analysis of a Global
788 Collection of *Pyrenophora tritici-repentis*, Causal Agent of Tan Spot of Wheat. Phytopathology
789 95:1144-50.
790 36. Mat Razali N, Cheah BH, Nadarajah K. 2019. Transposable elements adaptive role in genome
791 plasticity, pathogenicity and evolution in fungal phytopathogens. International Journal of
792 Molecular Sciences 20:3597.

793 37. Dhillon B, Gill N, Hamelin RC, Goodwin SB. 2014. The landscape of transposable elements in the
794 finished genome of the fungal wheat pathogen *Mycosphaerella graminicola*. BMC Genomics
795 15:1-17.
796 38. Lorrain C, Feurtey A, Möller M, Haueisen J, Stukenbrock E. 2021. Dynamics of transposable
797 elements in recently diverged fungal pathogens: lineage-specific transposable element content
798 and efficiency of genome defenses. G3: Genes Genomes Genetics 11:jkab068.
799 39. Aboukhaddour R, Strelkov SE. 2016. Exploring de novo specificity: the *Pyrenophora tritici-*
800 *repentis*-barley interaction. Plant Pathology 65:1347-1357.
801 40. Kang Z, Buchenauer H. 2000. Ultrastructural and cytochemical studies on cellulose, xylan and
802 pectin degradation in wheat spikes infected by *Fusarium culmorum*. Journal of Phytopathology
803 148:263-275.
804 41. Barrett K, Jensen K, Meyer AS, Frisvad JC, Lange L. 2020. Fungal secretome profile categorization
805 of CAZymes by function and family corresponds to fungal phylogeny and taxonomy: Example
806 *Aspergillus* and *Penicillium*. Scientific Reports 10:1-12.
807 42. Yamamoto N, Wang Y, Lin R, Liang Y, Liu Y, Zhu J, Wang L, Wang S, Liu H, Deng Q. 2019.
808 Integrative transcriptome analysis discloses the molecular basis of a heterogeneous fungal
809 phytopathogen complex, *Rhizoctonia solani* AG-1 subgroups. Scientific Reports 9:1-14.
810 43. Richard G-F. 2020. Eukaryotic pangenomes, p 253-291. *In* Tettelin H, Medini D (ed), The
811 Pangenome. Springer, Gewerbestrasse, Switzerland.
812 44. Badet T, Oggenfuss U, Abraham L, McDonald BA, Croll D. 2020. A 19-isolate reference-quality
813 global pangenome for the fungal wheat pathogen *Zymoseptoria tritici*. BMC Biology 18:1-18.
814 45. Croll D, McDonald BA. 2012. The accessory genome as a cradle for adaptive evolution in
815 pathogens. PLoS Pathogens 8:e1002608.
816 46. Haridas S, Albert R, Binder M, Bloem J, LaButti K, Salamov A, Andreopoulos B, Baker S, Barry K,
817 Bills G. 2020. 101 *Dothideomycetes* genomes: a test case for predicting lifestyles and emergence
818 of pathogens. Studies in Mycology 96:141-153.
819 47. Kelkar YD, Ochman H, Evolution. 2012. Causes and consequences of genome expansion in fungi.
820 Genome Biology 4:13-23.
821 48. Ma L-J, Van Der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, Di Pietro A, Dufresne M,
822 Freitag M, Grabherr M, Henrissat B. 2010. Comparative genomics reveals mobile pathogenicity
823 chromosomes in *Fusarium*. Nature 464:367-373.
824 49. de Man TJ, Stajich JE, Kubicek CP, Teiling C, Chenthamara K, Atanasova L, Druzhinina IS,
825 Levenkova N, Birnbaum SS, Barribeau SM. 2016. Small genome of the fungus *Escovopsis weberi*,
826 a specialized disease agent of ant agriculture. Proceedings of the National Academy of Sciences
827 113:3567-3572.
828 50. Ochman H, Moran NA. 2001. Genes lost and genes found: evolution of bacterial pathogenesis
829 and symbiosis. Science 292:1096-1099.
830 51. Lawrence JG. 2005. Common themes in the genome strategies of pathogens. Curr Opin Genet
831 Dev 15:584-588.
832 52. Nagendran S, Hallen-Adams HE, Paper JM, Aslam N, Walton JD. 2009. Reduced genomic
833 potential for secreted plant cell-wall-degrading enzymes in the ectomycorrhizal fungus *Amanita*
834 *bisporigera*, based on the secretome of *Trichoderma reesei*. Fungal Genetics and Biology 46:427-
835 435.
836 53. Badet T, Croll D. 2020. The rise and fall of genes: origins and functions of plant pathogen
837 pangenomes. Current Opinion in Plant Biology 56:65-73.
838 54. Fouché S, Plissonneau C, Croll D. 2018. The birth and death of effectors in rapidly evolving
839 filamentous pathogen genomes. Current Opinion in Microbiology 46:34-42.

840 55. Aylward J, Steenkamp ET, Dreyer LL, Roets F, Wingfield BD, Wingfield MJ. 2017. A plant
841 pathology perspective of fungal genome sequencing. IMA Fungus 8:1-15.
842 56. Guo J, Shi G, Kalil A, Friskop A, Elias E, Xu SS, Faris JD, Liu Z. 2020. *Pyrenophora tritici-repentis*
843 race 4 isolates cause disease on tetraploid wheat. Phytopathology 110:1781-1790.
844 57. Wei B, Moscou MJ, Sato K, Gourlie R, Strelkov S, Aboukhaddour R. 2020. Identification of a locus
845 conferring dominant susceptibility to *Pyrenophora tritici-repentis* in barley. Frontiers in Plant
846 Science 11:158.
847 58. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Babraham
848 Bioinformatics, Babraham Institute, Cambridge, United Kingdom.
849 59. Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. Genome
850 Biology 20:257.
851 60. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI,
852 Pham S, Prjibelski AD. 2012. SPAdes: a new genome assembly algorithm and its applications to
853 single-cell sequencing. Journal of Computational Biology 19:455-477.
854 61. Seemann T. Shovill. https://githubcom/tseemann/shovill.
855 62. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for
856 large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674-
857 1676.
858 63. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y. 2012. SOAPdenovo2: an
859 empirically improved memory-efficient short-read de novo assembler. Gigascience 1:2047-217X-
860 1-18.
861 64. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome
862 assemblies. Bioinformatics 29:1072-1075.
863 65. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using
864 repeat graphs. Nature Biotechnology 37:540-546.
865 66. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman
866 J, Young SK. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and
867 genome assembly improvement. PLoS One 9:e112963.
868 67. Min B, Grigoriev IV, Choi IG. 2017. FunGAP: Fungal Genome Annotation Pipeline using evidence-
869 based gene model evaluation. Bioinformatics 33:2936-2937.
870 68. Smit AF, Hubley R. 2008. RepeatModeler Open-1.0.
871 69. Smit AFJhwro. 2004. Repeat-Masker Open-3.0.
872 70. Kim D, Langmead B, Salzberg SLJNm. 2015. HISAT: a fast spliced aligner with low memory
873 requirements. 12:357-360.
874 71. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L,
875 Raychowdhury R, Zeng QJNb. 2011. Trinity: reconstructing a full-length transcriptome without a
876 genome from RNA-Seq data. 29:644.
877 72. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern BJNar. 2006. AUGUSTUS: ab initio
878 prediction of alternative transcripts. 34:W435-W439.
879 73. Cantarel BL, Korf I, Robb SM, Parra G, Ross E, Moore B, Holt C, Alvarado AS, Yandell MJGr. 2008.
880 MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes.
881 18:188-196.
882 74. Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. 2019. Whole-genome annotation with BRAKER, p
883 65-95, Gene prediction. Springer.
884 75. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez RJNar. 2005.
885 InterProScan: protein domains identifier. 33:W116-W120.
886 76. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool.
887 Journal of Molecular Biology 215:403-410.

77. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL. 2004. The Pfam protein families database. Nucleic Acids Research 32:D138-D141.
78. Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness, p 227-245, Gene Prediction. Springer.
79. 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.
80. Fouts DE, Brinkac L, Beck E, Inman J, Sutton G. 2012. PanOCT: automated clustering of orthologs using conserved gene neighborhood for pan-genomic analysis of bacterial strains and closely related species. Nucleic Acids Research 40:e172-e172.
81. Käll L, Krogh A, Sonnhammer EL. 2005. An HMM posterior decoder for sequence feature prediction that includes homology information. Bioinformatics 21:i251-i257.
82. Sperschneider J, Dodds PN, Gardiner DM, Singh KB, Taylor JM. 2018. Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. Molecular Plant Pathology 19:2094-2110.
83. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Research 46:W95-W101.
84. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32:1792-1797.
85. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312-1313.
86. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 20:1297-1303.
87. Ortiz E. 2019. vcf2phylip v2. 0: convert a VCF matrix into several matrix formats for phylogenetic analysis. https://doiorg/105281/zenodo 2540861.
88. Rambaut A. 2007. FigTree, a graphical viewer of phylogenetic trees. https://githubcom/rambaut/figtree/.
89. Pedersen TL. 2017. Hierarchical sets: analyzing pangenome structure through scalable set visualizations. Bioinformatics 33:1604-1612.
90. Ou S, Su W, Liao Y, Chougule K, Agda JR, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T. 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome biology 20:1-18.
91. Ellinghaus D, Kurtz S, Willhoeft UJBb. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. 9:1-14.
92. Xu Z, Wang HJNar. 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. 35:W265-W268.
93. Ou S, Jiang NJMD. 2019. LTR_FINDER_parallel: parallelization of LTR_FINDER enabling rapid identification of long terminal repeat retrotransposons. 10:1-3.
94. Ou S, Jiang NJPp. 2018. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. 176:1410-1422.
95. Su W, Gu X, Peterson TJMp. 2019. TIR-Learner, a new ensemble method for TIR transposable element annotation, provides evidence for abundant new transposable elements in the maize genome. 12:447-460.
96. Shi J, Liang CJPp. 2019. Generic Repeat Finder: a high-sensitivity tool for genome-wide de novo repeat detection. 180:1803-1815.

97. Xiong W, He L, Lai J, Dooner HK, Du CJPotNAoS. 2014. HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. 111:10263-10268.
98. Zhang R-G, Wang Z-X, Ou S, Li G-YJb. 2019. TEsorter: lineage-level classification of transposable elements using conserved protein domains.800177.
99. Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094-3100.
100. Porten T. 2018. dotPlotly. https://githubcom/tpoorten/dotPlotly.
101. Minkin I, Patel A, Kolmogorov M, Vyahhi N, Pham S. Sibelia: a scalable and comprehensive synteny block generation tool for closely related microbial genomes, p 215-229. *In* (ed), Springer,
102. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. Genome Research 19:1639-1645.
103. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols 10:845-858.
104. Frantzeskakis L, Kracher B, Kusch S, Yoshikawa-Maekawa M, Bauer S, Pedersen C, Spanu PD, Maekawa T, Schulze-Lefert P, Panstruga R. 2018. Signatures of host specialization and a recent transposable element burst in the dynamic one-speed genome of the fungal barley powdery mildew pathogen. BMC Genomics 19:1-23.
105. Van Wyk S, Harrison CH, Wingfield BD, De Vos L, van Der Merwe NA, Steenkamp ET. 2019. The RIPper, a web-based tool for genome-wide quantification of Repeat-Induced Point (RIP) mutations. PeerJ 7:e7447.

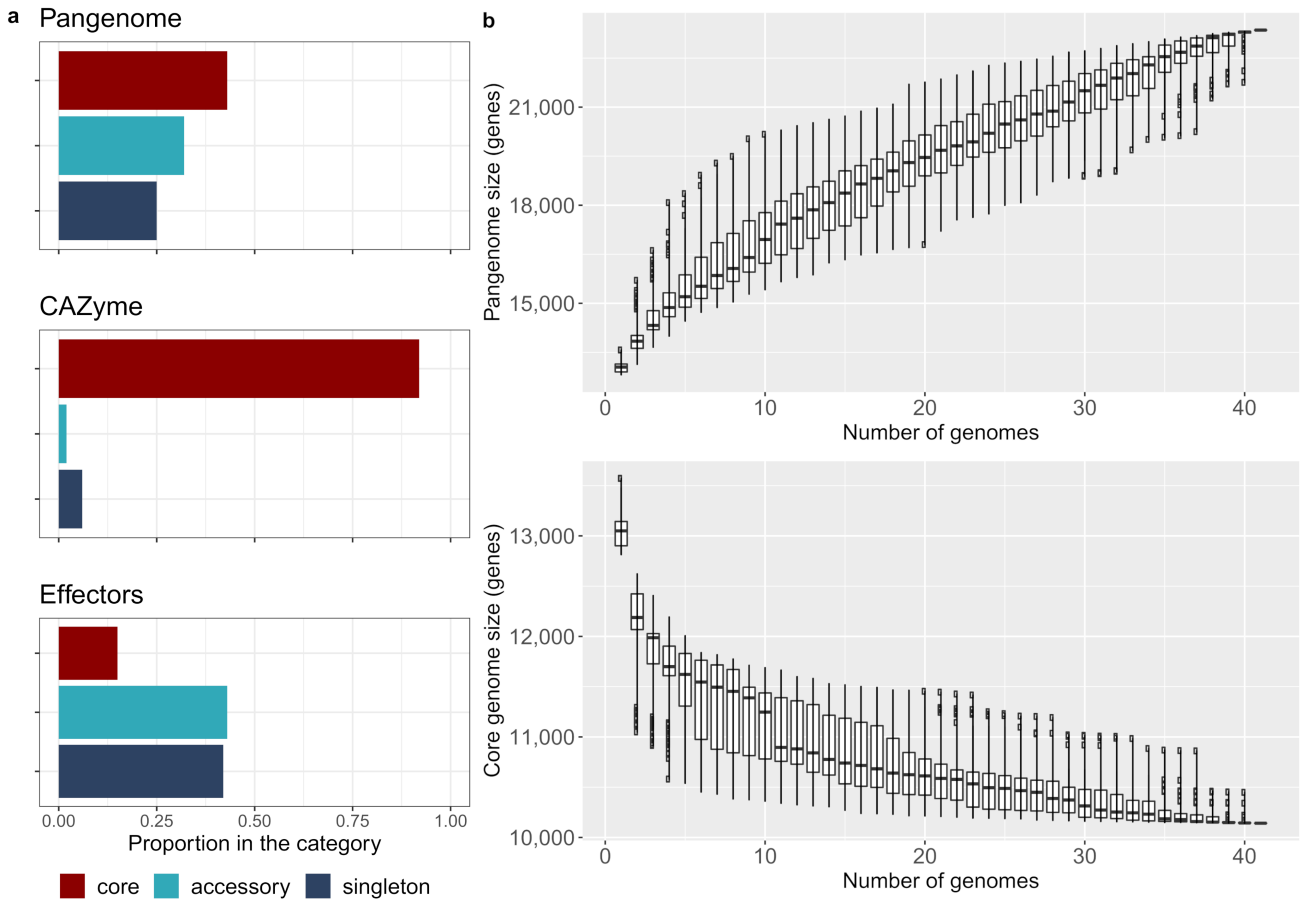

Figure 1. The pangenome of *Pyrenophora tritici-repentis*. **a** Proportion of genes present in core, accessory (excluding singletons), and singleton sets for the pangenome, CAZymes, and effectors. **b** Total number of unique genes and core genes in the pangenome of *Pyrenophora tritici-repentis* as genomes are added. Starting genome was always Pt-1C-BFP with subsequent genomes added randomly (x100).

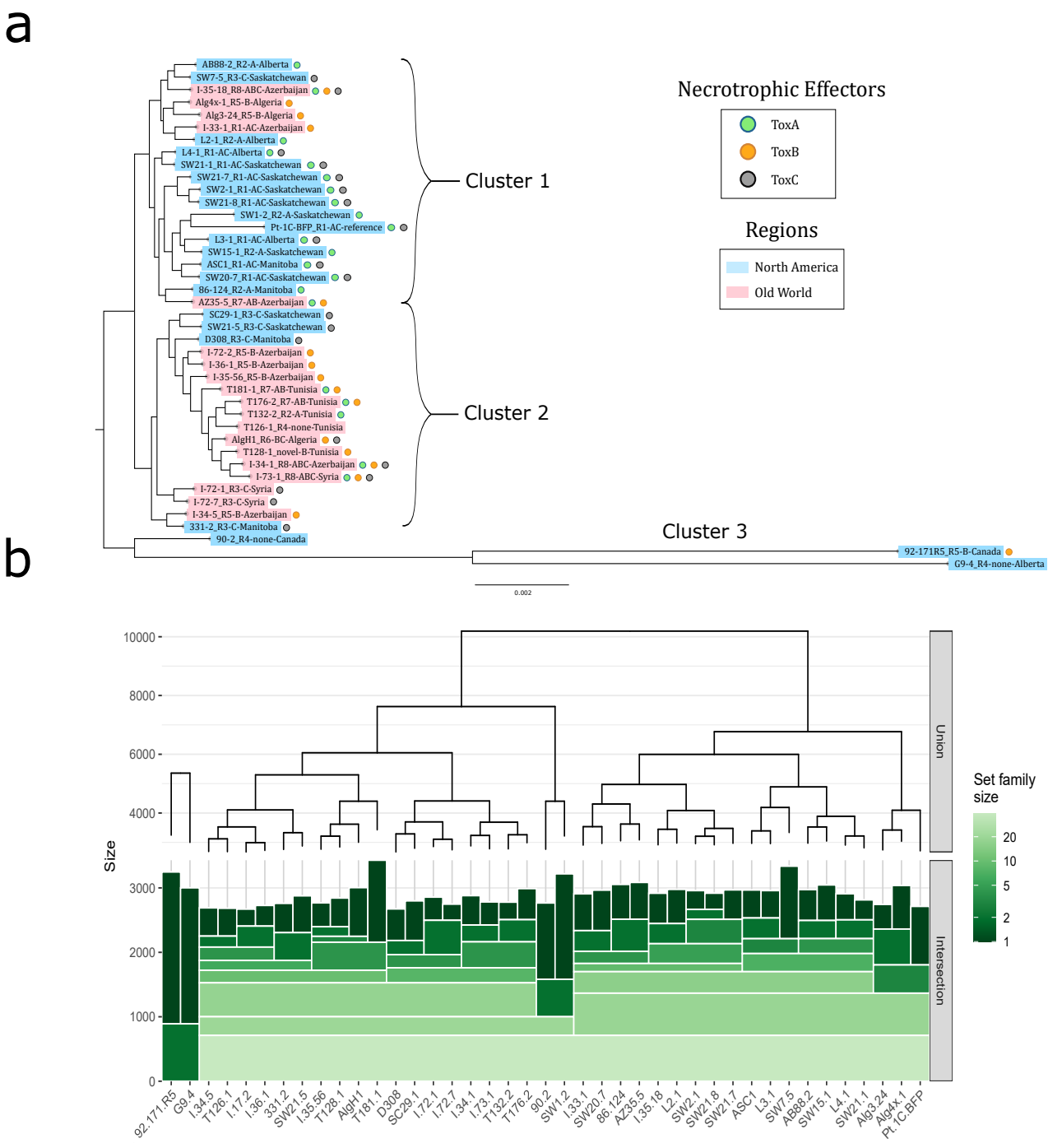

Figure 2. Evolutionary relatioships of isolates in the Ptr pangenome. **a** Maximum-likelihood phylogenetic tree of *Pyrenophora tritici-repentis* base on aligned and concatenated proteins present in the core genome (10,159 genes). Isolate names are followed by their race (R1 to R8) designation, necrotrophic effector(s) produced (A, B, and/or C), and their location of origin (isolates collected within Canada show province names; isolates outside Canada show country of origin). Presence of ToxA (blue circles) and ToxB (orange circles) confirmed via BLAST, and presence of ToxC (grey circles) determined by phenotyping. Isolates were color coded by region with North America representing Canada and the United States and Old World representing Caucuses, North Africa, and Fertile Crescent. **b** Hierarchical sets of *Pyrenophora tritici-repentis* accessory pangenome, size represents number of genes shared by isolates in each cluster, where inclusion in a cluster is based on any given isolate overlapping with the horizontal bars.

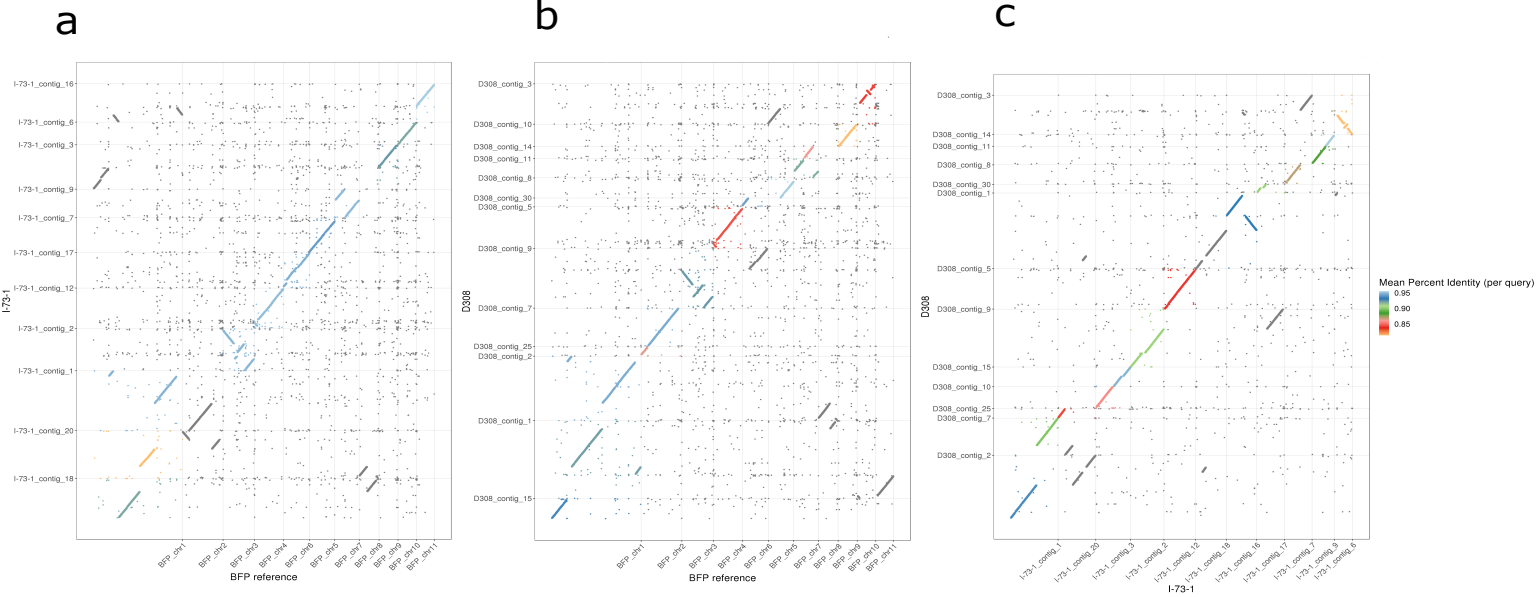

Figure 3. Full genome dotplot alignments of newly sequenced long-read assemblies of *Pyrenophora tritici-repentis* isolates I-73-1 (race 8) and D308 (race 3) to the reference isolate Pt-1C-BFP (race 1). **a** Pt-1C-BFP and I-73-1; **b** Pt-1C-BFP and D308; **c** I-73-1 and D308.

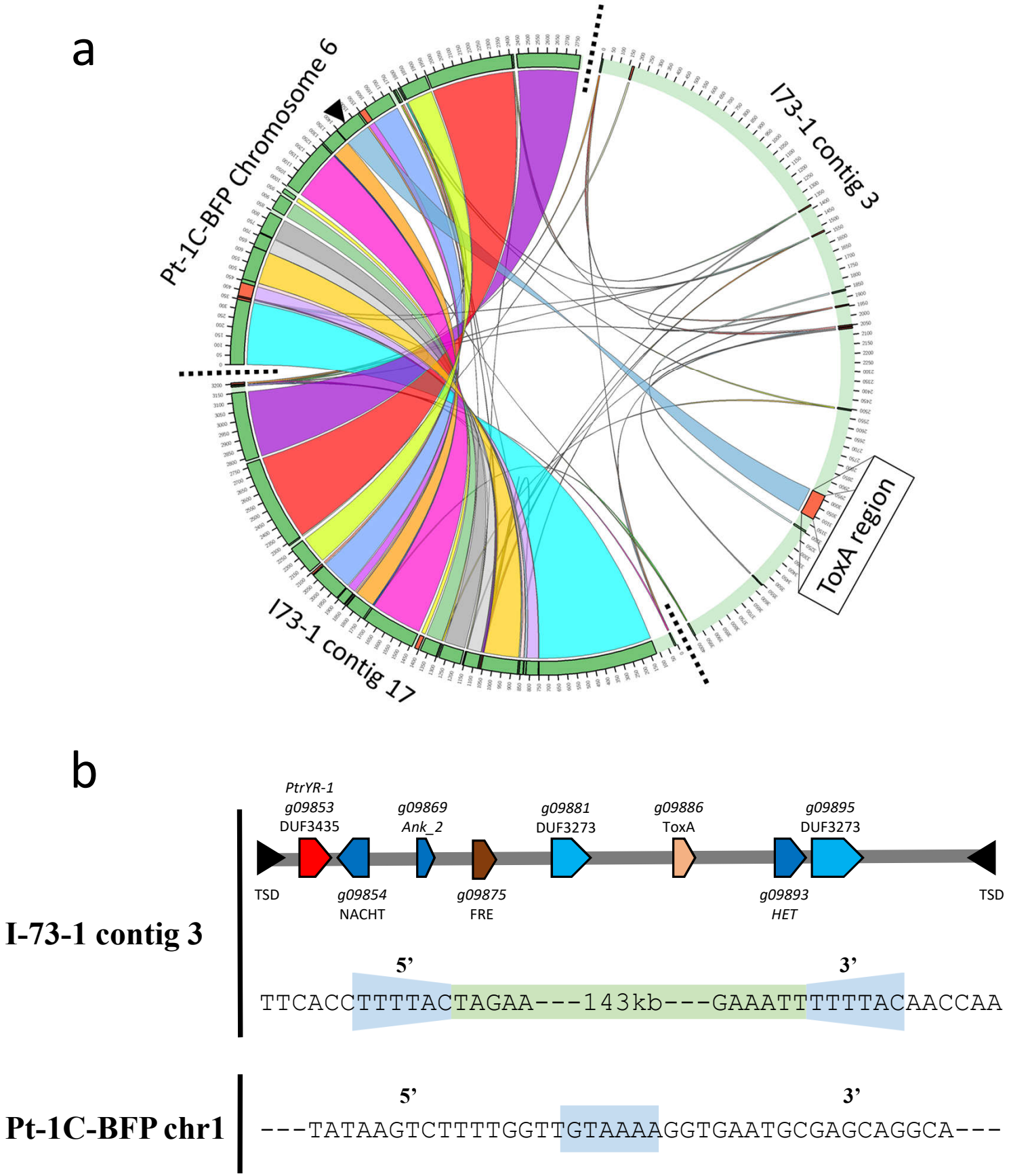

Figure 4. Evidence of translocation of *ToxA* containing region between two races of *Pyrenophora tritici-repentis*. **a** The *ToxA* containing chromosome 6 of isolate Pt-1C-BFP (race 1) is compared with the *ToxA* containing contig 3 and contig 17 of isolate I73-1 (race 8). Pt-1C-BFP chromosome 6 aligns almost completely to I73-1 contig 17, with the exception of the *ToxA* containing region, which aligns to contig 3. The majority of contig 3 aligns to Pt-1C-BFP chromosome 1 or chromosome 9 (Figure 3). **b** schematic of the 'Horizon' Starship class transposon.

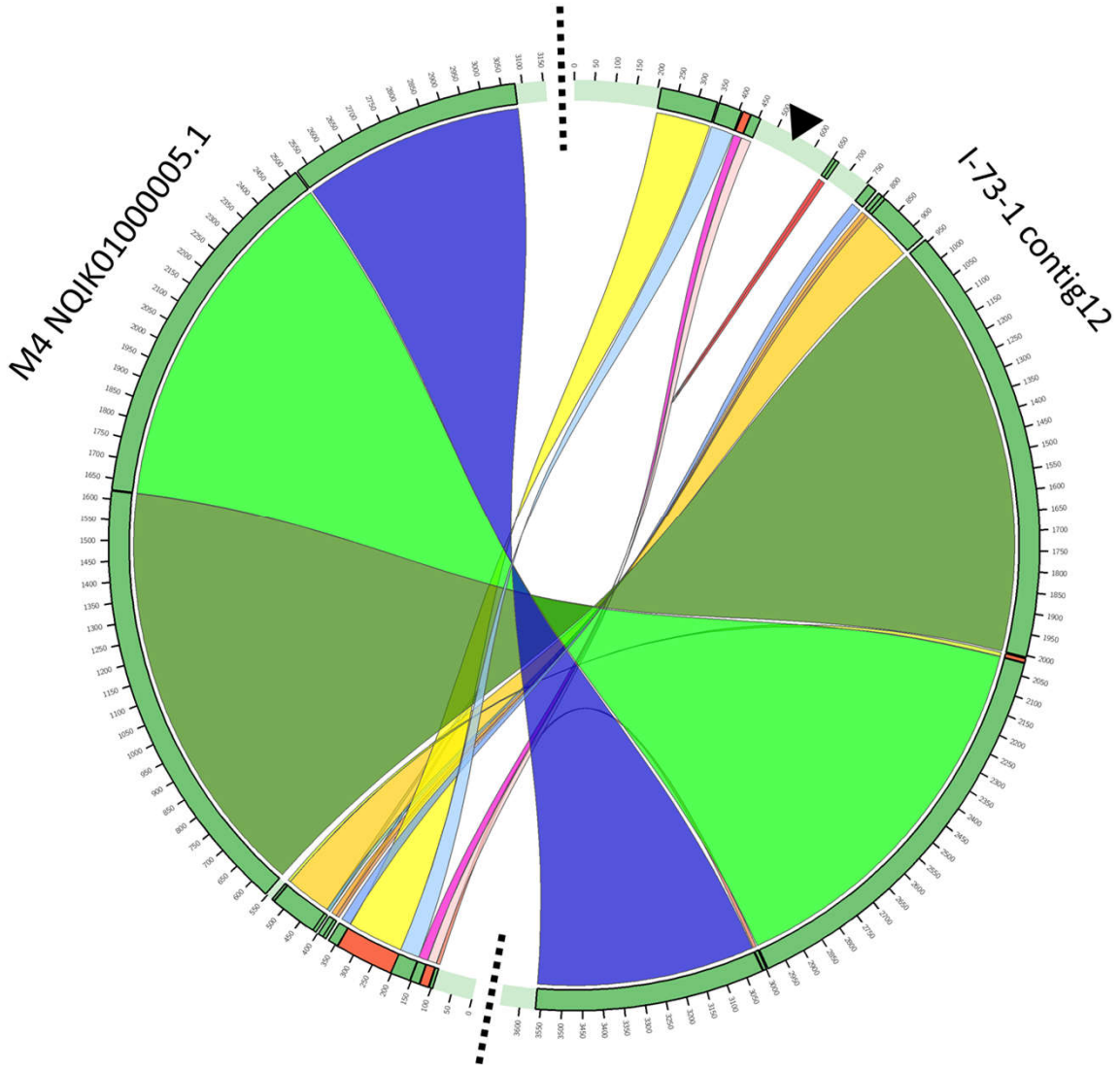

Figure 5. Circular alignment of contig 12 from race 8 isolates I-73-1 and contig NQIK01000005 from race 1 isolate M4. A large 294 Kb region which contains three copies of the *ToxB* (black arrow) is visible, this section did not align with any other contig of M4. The edges of the 294 Kb region revealed terminal inverted repeats suggesting transposon activity.

Figure 6. Intergenic distances of all genes in *Pyrenophora tritici-repentis* of: **a** isolate D308; **b** isolate I-73-1. The 3' intergenic legnth (x-axis) is the distance (bp) from the 3' end of the current gene to the 5' end of the next, and the 5' intergenic lngth (y-axis) is the distance from the 3' end of the previous gene to the 5' end of the current gene.

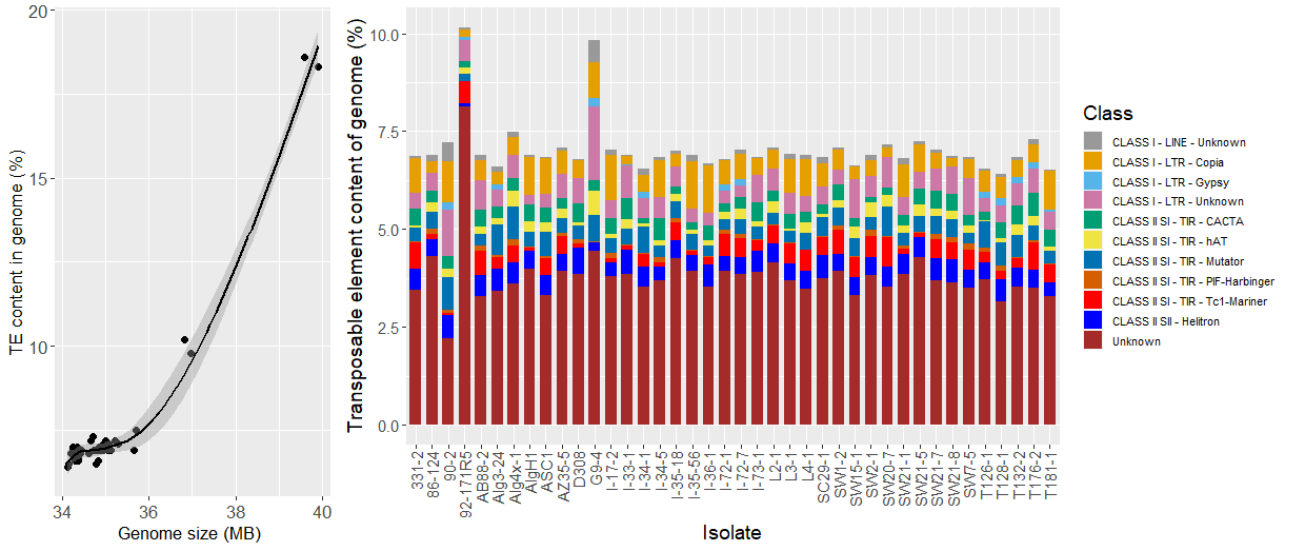

Figure 7. Transposable element content and genome size variation of *Pyrenophora tritici-repentis* from short-read assemblies. **a** Contribution of TEs (%) to total genome size across 40 isolates; **b** Contribution of each TE class to genome size for individual isolates.