ON THE EVOLUTION OF *PYRENOPHORA TRITICI-REPENTIS*

RYAN ALLAN RUSSELL GOURLIE
Master of Science, University of Guelph, 2018

A doctoral thesis submitted
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY
in
BIOMOLECULAR SCIENCE - BIOINFORMATICS

Department of Biological Sciences
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

ON THE EVOLUTION OF *PYRENOPHORA TRITICI-REPENTIS*

RYAN GOURLIE

Date of Defence: April 10th, 2025

Dr. Dmytro Yevtushenko	Associate Professor	PhD
Dr. Reem Aboukhaddour	Research Scientist	PhD
Thesis Co-Supervisors		
Dr. Megan McDonald	Assistant Professor	PhD
Thesis Examination Committee Member		
Dr. Athanasios Zovoilis	Associate Professor	PhD
Thesis Examination Committee Member		
Dr. Emile Gluck-Thaler	Assistant Professor	PhD
External Examiner		
University of Wisconsin-Madison		
Madison, Wisconsin, United States of America		
Dr. Theresa Burg	Professor	PhD
Chair, Thesis Examination Committee		

DEDICATION

For my family, past, present, and future.

ABSTRACT

Pyrenophora tritici-repentis (Ptr) is a globally distributed and economically important plant pathogen causing tan spot, a destructive foliar disease of wheat. The relatively recent emergence of Ptr provides a unique opportunity to better understand how necrotrophic pathogens evolve in modern agricultural systems. In this thesis, a variety of genomic and bioinformatic techniques were employed to examine the genomes of a large set of geographically diverse isolates with representatives from all of the established virulence races. These races contain different combinations of Ptrs three primary necrotrophic effectors: ToxA, ToxB, and ToxC. Comparison of total gene content between isolates found that Ptr possesses an open-pangenome, with a high accessory gene content (57%), and significant differences between pathogenic and non-pathogenic races. Putative effectors were found to be primarily accessory in nature, while carbohydrate-active enzymes associated with plant cell wall degradation were conserved. Gene distances and evolutionary rates suggest that the genomic architecture of Ptr is that of a 'one-compartment' genome despite having a high proportion of transposable elements (~18 to 25% of the genome). Significant chromosomal rearrangements were observed between isolates including the translocation of *ToxA* and the presence of *ToxB* within an accessory region. The translocation of *ToxA,* along with a 143 kbp region, was facilitated by the Starship transposon *Horizon.* Within *Horizon*, was nested the *ToxhAT* transposon which was responsible for the horizontal transfer of *ToxA* into Ptr. The accessory region which contains *ToxB* contained several Starship cargo genes but lacked other defining Starship characteristics, namely the requisite tyrosine recombinase. This region may be an accessory chromosomal arm, an ancient derelict Starship, or the beginning of genome compartmentalization. Detailed examination and extensive manual alignments between *ToxB* containing isolates revealed the presence of a Helitron-like transposable element,

ToxB-HLE, which is replicating the *ToxB* gene in some isolates. Furthermore, two independent *Copia* retrotransposon insertions were found to be responsible for the disruption of *ToxB* thereby creating two of the known inactive *toxb* haplotypes. Finally, a genome wide association study revealed single-nucleotide polymorphisms significantly associated with the ToxC phenotype (-log(p) = 5.5). The nature of ToxC is unknown, and this analysis provides a number of candidate genes which may be involved in ToxC biosynthesis, secretion, or regulation. This body of work represents one of the most extensive examinations of the Ptr genome and its necrotrophic effectors preformed to-date and has shown that the pathogen exhibits high genome plasticity through not only through rearrangements but gene gains and losses as well. It also shows that transposons play a significant role in the evolution and adaptability of this global plant pathogen by mobilizing, replicating, and disrupting virulence genes.

CONTRIBUTION OF AUTHORS

Chapter 2 - Ryan Gourlie performed most of the analyses including assemblies/annotations, pangenome, phylogenetics, transposon content, chromosomal organization, effectors, and *ToxA* and *ToxB* transposon analysis. The first draft of this manuscript was prepared by Ryan Gourlie, Reem Aboukhaddour, and Megan McDonald. Megan McDonald aided and guided the transposon analysis and contributed to project guidance. Rodrigo Ortega Polo aided with assemblies and analysis guidance. Mohamed Hafez performed DNA extractions. Kristin Low and Wade Abott performed CAZyme work. Steven Strelkov and Fouad Daayf provided isolates. Reem Aboukhaddour conceived and guided the project, performed single-spore isolation for a number of isolates, and confirmed phenotyping. All authors contributed to editing and improving the manuscript. The authors read and approved the final manuscript.

Chapter 3 - Ryan Gourlie performed all analyses and created the first draft of the manuscript. Ryan Gourlie and Mohamed Hafez performed DNA extractions. Megan McDonald aided with data interpretation, provided scripts for gggenomes (adapted by Ryan Gourlie), and provided significant structural improvements to the manuscript. Ryan Gourlie and Reem Aboukhaddour conceived the project. Reem Aboukhaddour confirmed isolate phenotypes and provided funding support. All authors contributed to editing and improving the manuscript. The authors read and approved the final manuscript.

Chapter 4 - Ryan Gourlie performed all analyses. Megan McDonald provided input on methodology and scripting. Reem Aboukhaddour conceived the project. Ryan Gourlie and Reem Aboukhaddour drafted the manuscript. All authors contributed to improving the manuscript.

PREFACE

The main body of work in this thesis is presented in manuscript style. Research chapters are self contained from published or to be published works, each containing their own abstract, introduction, methods, results, and discussion. These research chapters are preceded by a general introduction and are followed by a summarizing chapter which brings the research together into a cohesive body of work.

Chapter 2 is published as: Gourlie, R., McDonald, M., Hafez, M., Ortega-Polo, R., Low, K.E., Abbott, D.W., Strelkov, S.E., Daayf, F., and Aboukhaddour, R., 2022. The pangenome of the wheat pathogen *Pyrenophora tritici-repentis* reveals novel transposons associated with necrotrophic effectors ToxA and ToxB. *BMC Biology*, *20*(1), p.239. https://doi.org/10.1186/s12915-022-01433-w

Chapter 3 is submitted to New Phytologist as: Gourlie, R., McDonald, M., Hafez, and Aboukhaddour, R., 2025. The virulence gene *ToxB* is both amplified and disrupted by transposons in the wheat pathogen *Pyrenophora tritici-repentis*.

Chapter 4 has not been submitted at the time of writing but will be submitted to the Canadian Journal of Plant Pathology as: Gourlie, R., McDonald, M., and Aboukhaddour, R., 2025. Genome-wide association study reveals SNPs associated with the ToxC phenotype in *Pyrenophora tritici-repentis*.

ACKNOWLEDGEMENTS

No PhD is earned without the support of others, and though the degree may have my name, this accomplishment is shared with many who deserve recognition.

First and foremost I express my immense gratitude to Dr. Reem Aboukhaddour for encouraging and supporting my pursuits. Without her, this would not have been possible.

Many thanks to Dr. Dmytro Yevtushenko for being my supervisor and providing constructive feedback throughout my degree.

A special thank you to Dr. Megan McDonald who has helped guide my learning in genomics and for inviting me to visit her lab. I always look forward to our discussions and hope to continue collaborations well into the future.

I would like to thank Dr. Mohamed Hafez, who has provided so much support and encouragement over the years. My gratitude extends to many other members of the lab, too numerous to list, who have contributed in various ways.

Thank you to Rodrigo Ortega Polo and the Bioinformatics Support Unit at AAFC who helped in many ways to improve my knowledge and facilitate my research.

Though he was not involved in this work, I would like to thank Dr. Tom Hsiang, my MSc advisor. Without his guidance and mentorship during some of my most difficult years I would not be where I am today.

To my co-authors, thank you for your contributions to our publications; it is not only your efforts, but your insights, additions, and improvements which have led to quality manuscripts.

Financial contributions which allowed me to attend conferences, meetings, and pay tuition were made by the University of Lethbridge, the Government of Alberta (through Alberta Graduate Excellence and Alberta Innovates scholarships), and the British Society of Plant Pathology, all of whom I sincerely thank for supporting me.

My research was funded in part by: Agriculture and Agri-food Canada, Alberta Wheat Commission, and Saskatchewan Wheat Development. Thank you for supporting research in plant pathology.

And finally, thank you to my family and friends for your constant unconditional love and support.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANOVA	Analysis of Variance
ATIR	Asymmetrical Inverted Repeats
BFP	Reference isolate Pt-1C-BFP
BLAST	Basic Local Alignment Search Tool
bp; kbp; Mb	base pair; kilo base pair; mega base pair
BUSCO	Benchmarking Universal Single-Copy Ortholog
CAZyme	Carbohydrate-Active Enzyme
Chr	Chromosome
CNV	Copy Number Variation
DNA	Deoxyribonucleic Acid
DUF	Domain of Unknown Function
GAG	Retrotransposon Capsid Protein
GO	Gene Ontology
GWAS	Genome Wide Association Study
HLE	Helitron-Like Element
HST	Host Selective Toxin
INT	Integrase
LTR	Long Terminal Repeat
LTS	Left-Terminal Sequence
ORF	Open Reading Frame
Pol	Retrovirus-related Polyprotein
Ptr	*Pyrenophora tritici-repentis*
qPCR	Quantitative Polymerase Chain Reaction
RC	Reverse Complimented
RT	Reverse Transcriptase
RTS	Right-Terminal Sequence
SJ	Sequence Junction
SNP	Single Nucleotide Polymorphism
TE	Transposable Element
TID	Tandem Inverted Repeat
TIR	Terminal Inverted Repeat
TSD	Target Site Duplication

CHAPTER 1: GENERAL INTRODUCTION

1.1 AN OVERVIEW OF *PYRENOPHORA TRITICI-REPENTIS*

Pyrenophora tritici-repentis (Died.) Drechs. (anamorph *Drechslera tritici-repentis* (Died.) Shoem.) (referred to as Ptr) is a destructive fungal plant pathogen which primarily infects bread wheat (*Triticum aestivum*) and durum wheat (*T. turgidum* sp. *durum*) causing the foliar disease tan spot. Initially identified in 1902 from wild grass (*Agropyron repens*), it was first described as *Pleospora trichostoma* (Diedicke, 1902; Wolf et al., 1998). Later the species was renamed as *Helminthosporium tritici-repentis*, then updated to *Pyrenophora tritici-repentis* to reflect changing fungal classification and the species pathogenicity on wheat (Diedicke, 1902; Drechsler, 1923; Mitra, 1934; Shoemaker, 1962). Ptr is an Ascomycete, in the fungal order Pleosporales within the class Dothidiomycetes. The Pleosporales order is home to many other important fungal pathogens including other leaf spot pathogens which are often confused with Ptr, such as *Parastagonospora nodorum* and *Bipolaris sorokiniana*. Along with *Zymoseptoria tritici* (another Dothidiomycete), these pathogens make up what is commonly referred to as the 'leaf spot pathogen complex' (Figueroa et al., 2018). These different pathogens can often be isolated from the same plant, and in some cases, the same lesion.

As a leaf-spot pathogen, Ptr attacks foliar tissue, causing necrotic and/or chlorotic spots on susceptible hosts. These symptoms are the result of secreted proteins and metabolites known as necrotrophic effectors (syn. host-selective toxins), which interact with the host in an inverse gene-for-gene manner. There are three main effectors which define the current race system of Ptr: ToxA, ToxB, and ToxC which cause necrosis, chlorosis, and an alternate form of chlorosis, respectively (Figure 1.1) (Strelkov and Lamari, 2003; Aboukhaddour et al., 2013). Isolates may produce one, two, three, or none

of these effectors resulting in eight defined races (Figure 1.1). Details of each effector are discussed further in section 1.2.

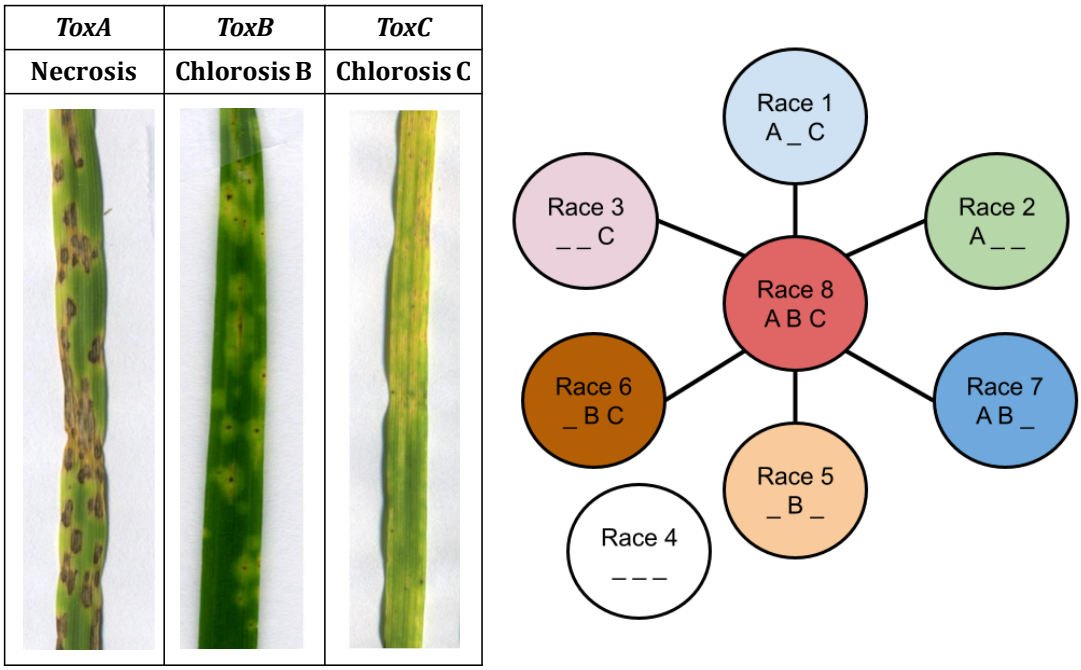


Figure 1.1 The necrotrophic effectors of *Pyrenophora tritici-repentis*; left: symptoms caused by the necrotrophic effectors ToxA, ToxB, and ToxC (images credited to Aboukhaddour Lab); right: current race definitions of Ptr based on the presence or absence of three primary necrotrophic effectors. Race figure is adapted from Strelkov and Lamari (2003) and Aboukhaddour et al., (2013).

The life-cycle of Ptr is similar to that of other Ascomycete pathogens, utilizing both sexual and asexual modes of reproduction. The primary inoculum source for infections in a new growing season is ascospores (Moffat and Santana, 2018; Oliver, 2024). Ascospores are created via sexual reproduction and are released from small black pseudothecia which form on crop stubble and straw (Hosford, 1971). Ptr is a homothallic (i.e. self fertilizing) species which eliminates the need to find compatible mating types,

ensuring sexual reproduction can occur each season (Turgeon, 1998). As ascospores are produced at ground level, initial infection most often occurs on lower leaves during the tillering stage if conditions are favorable (Wolf et al., 1998; Moffat and Santana, 2018; Oliver, 2024). As the season progresses, disease spreads upwards through the canopy when there is sufficient moisture and temperatures are not extreme (18 to 28°C), though Ptr is able to grow and cause disease in a wider temperature range (Da Luz and Bergstrom, 1986; Wolf et al., 1998; Moffat and Santana, 2018; Oliver, 2024). Asexual spores, conidia, emerge from colonized tissue and are formed on the top of conidiophores. These conidia are carried by wind and serve as the secondary inoculum which infects new hosts throughout the growing season. After host senescence, sexual reproduction occurs. This creates new pseudothecia, allowing Ptr to overwinter in crop stubble and restart the infection cycle the following spring (Wolf et al., 1998; Moffat and Santana, 2018; Oliver, 2024).

Foliar fungicides are commonly used to manage tan spot, particularly in varieties of wheat which are known to be susceptible. Crop scouting combined with weather forecasting can inform growers of optimal spray timings for their fields, although tan spot can easily be confused for other leaf-spot diseases. In continuous cereal rotations fungicides may need to be applied early in the season to deal with the high primary inoculum loads associated with continuous cereal production. Since the corresponding susceptibility targets for the Tox effectors are known, selecting varieties which lack these genes is an effective way of mitigating tan spot, although resistant lines are limited (Lamari et al., 2005). Breaking the disease cycle through crop rotation can reduce inoculum loads as well. Planting wheat crops near previously infected fields, particularly

if those fields are upwind, should be avoided, as condia have been shown to travel vast distances through air currents (Wolf et al., 1998). Additionally, managing crop residues through tilling or burning can be effective, but must be balanced against other agronomic considerations such as soil health which can be negatively impacted by such practices.

Wheat is one of the most cultivated crops on the planet and provides an enormous amount of calories to the world's population (Erenstein et al., 2022). Grain yields can be adversely affected by various external sources such as a lack of irrigation, poor soil health, and other environmental conditions. However, the major continuous drivers of yield reduction are biological - meaning weeds, insects, and disease. As one of our oldest crops, wheat is subject to a wide variety of devastating pathogens. Ptr has been observed across the globe, including areas in which wheat is not grown, making the species a worldwide problem for wheat production (Lamari et al., 2005a; Strelkov and Lamari, 2003; Savary et al., 2019; Kamel et al., 2019). A recent analysis on the global impact of pathogens on various crop yields found that diseases account for a 21.5% reduction in wheat output (Savary et al., 2019). In the Americas, tan spot accounts for approximately 5% of wheat losses in North America and 7% in South America, making it one of the most damaging diseases on the continent (Savary et al., 2019). The tan spot pathogen is present in most regions where wheat is cultivated, however, the impact of Ptr is more limited in regions where other leaf-spot pathogens are dominant. If left uncontrolled and conditions are ideal severe infections have reportedly reduced yields by up to 50% (Rees et al., 1982).

1.2 NECROTROPHIC EFFECTORS AND THE TOX GENES OF *P. TRITICI-REPENTIS*

Necrotrophic effectors, historically referred to as host selective toxins (HSTs), are considered pathogenicity or virulence factors which act in a host-specific manner (Wolpert et al., 2002; Petrov et al., 2018). These compounds are generally only effective in hosts which can be infected by the pathogen. Necrotrophic effectors follow an inverse gene-for-gene interaction model which is in contrast to the traditional gene-for-gene model. In the gene-for-gene model, established by Flor when working with a rust pathogen, a plant resistance gene (R gene) recognizes a corresponding pathogen avirulence gene (Avr gene) (Flor, 1971; Strelkov & Lamari, 2003). Most R genes have since been found to be dominant and encode for receptors containing nucleotide binding sites and/or leucine rich repeats (NBS-LRRs). In the inverse interaction, necrotrophic effectors, which are small metabolites or proteins, interact with the host via a dominant susceptibility factor, directly causing disease symptoms (Wolpert et al., 2002; Friesen et al., 2008; Virdi et al., 2016). Purified effectors are able to induce symptoms in susceptible hosts, even at low concentrations, and produce no symptoms on resistant hosts even at high concentrations. In systems where the effectors are known, purification of the effectors is a useful tool in breeding programs (i.e. effector assisted breeding) (Walton & Panaccione, 1993; Oliver, 2024).

The existence of necrotrophic effectors has been known since the 1930's when it was suggested such compounds were involved with black spot of pear caused by the widespread pathogen *Alternaria alternata* (Tanaka, 1933; reviewed in Tsuge et al., 2013). This hypothesis was later proven by the discovery of many necrotrophic effectors in various *Alternaria* species which contribute to their host specificity (e.g. AM-toxin, AF-

toxin, AK-toxin, ACT-toxin, ACR-toxin, AAL-toxin, AT-toxin, brassicicolin A) (reviewed in Tsuge et al., 2013; Petrov et al., 2018). Various necrotrophic effectors (secondary metabolites and proteinaceous) have been discovered in other species including: several *Cochliobolus* species (T-toxin, HC-toxin, victorin, ChToxA) (Kono et al., 1980; Walton et al., 1982; Wolpert et al., 1985; Lu et al., 2015), *Bipolaris sacchari* (HS-toxin) (Macko et al., 1983), Didymella maydis (*Pyronellaea zeae-maydis*) (PM-toxin) (Kono et al., 1980), *Periconia circinate* (PC-toxin) (Macko et al., 1992), *Leptosphaeria maculans* (phomalide, depsilairdin, maculansin) (Ward et al., 1992; Pedras et al., 2004; Pedras and Yu, 2008), *Corynespora cassiicola* (cassiicolin) (Barthe et al., 2007), and of course the leaf-spot pathogens *Parastagonospora nodorum* (ToxA, SnTox1, SnTox2, SnTox3, SnTox4, SnTox5, SnTox6 SnTox7) (Liu et al., 2004; 2009; Friesen et al., 2006; 2007; 2012; Abeysekara et al., 2009; Gao et al., 2015; Shi et al., 2015), *Bipolaris sorokiniana* (ToxA) (McDonald et al., 2018) and *Pyrenophora tritici-repentis* (ToxA, ToxB, ToxC) (Ballance et al., 1989; Strelkov et al., 1999; Effertz et al., 2002). The movement, duplication, and horizontal transfer of necrotrophic effectors have significant consequences. Movement of effectors intra-specifically can place effectors into isolates with other adaptive traits such, as thermal adaptation, which may help expand the regions in which the pathogen can thrive. Duplication of effectors can directly increase virulence through dosage effects and can lead to evolution of new effectors through divergence. Horizontal transfers are of particular concern as the mere presence of a necrotrophic effector may be enough to alter the host range of an existing pathogen or potentially convert a non-pathogenic species into a pathogen.

1.2.1 ToxA

ToxA has been the most widely studied necrotrophic effector in Ptr and was the first proteinaceous necrotrophic effector characterized in fungi (Ballance et al., 1989; reviewed in Aboukhaddour et al., 2023). The 13.2 KDa protein causes necrosis in wheat by interaction with the transmembrane domain protein TaNHL10 (NDR/HIN1-like protein), however, it only induces cell death in cultivars which carry the gene *Tsn1* (Tuori et al., 2000; Manning et al., 2007; Tai et al., 2007; Faris et al., 2010; Lu et al., 2014; Dagvadorj et al., 2022). ToxA is encoded by the single copy gene, *ToxA*, which has been identified in other fungal pathogens: *Parastagonospora nodorum*, *Parastagonospora pseudonodorum*, and *Bipolaris sorokiniana* (Friesen et al., 2006; McDonald et al., 2013; McDonald et al., 2018). Previous work has shown that the *ToxA* gene was horizontally transferred between these pathogens, (Friesen et al., 2006; McDonald et al., 2017). In the case of Ptr, it is likely that the transfer of *ToxA* coincided with the species transition into a global pathogen sometime around the 1940's (Lamari et al., 2005a; Friesen et al., 2006). The horizontal transfer was facilitated via a hAT transposon (*ToxhAT*), and is one of the first reported cases of such a transfer in fungi (McDonald et al., 2019). The example of *ToxA* highlights the role that mobile DNA plays in pathogen emergence and shows the need for continued exploration between transposons and pathogen evolution. Indeed, very recently *ToxA* was found embedded within another type of transposon, the extremely large Starship *Sanctuary* in the species *B. sorokiniana* (Bucknell et al., 2024). There is also new evidence showing that *ToxA* has begun to evolve in Ptr, with novel haplotypes being identified in countries like Japan and in the other *ToxA* containing species like *P. nodorum* (Hafez et al., 2022; Aboukhaddour et al., 2023). Although ToxA is the most

studied of the Ptr effectors, continued research is required to track the evolution and mobility of this important virulence factor.

1.2.2 ToxB

ToxB is the second effector identified in the tan spot pathosystem and is another proteinaceous effector which is 6.6 KDa in size (Lamari et al., 1995; Lamari et al., 2003; Strelkov et al., 1999; Martinez et al., 2001). Similar to ToxA, the corresponding host interaction target is known and has been identified as *Tsc2* (Faris et al., 2013). The ToxB protein also interacts with the chloroplast and causes chlorosis through the light-dependent degradation of chlorophyll (Strelkov et al., 1999; Kim et al., 2010). Unlike *ToxA*, *ToxB* is a multi-copy gene with higher disease symptoms correlated with higher copy numbers (Strelkov et al. 2005; Amaike et al. 2008; Aboukhaddour et al. 2012). An inactive isoform of the gene, dubbed *toxb*, is often found within the genomes of non-producers (Strelkov and Lamari, 2003; Martinez et al., 2001; Hafez et al., 2024). Furthermore, homologs of *ToxB* have been found in the related species *P. bromi*, *P. seminiperda*, *P. teres* as well as 15 more distantly related species such *Bipolaris oryzae*, *Magnaporthe grisea*, and various *Colletotrichum* species (Andrie et al., 2008; Hafez et al., 2024). ToxB protein sequences in these distant species shared homology as low as 31% identity but shared key structural similarities indicating potentially active isoforms (Hafez et al., 2024). It has been traditionally thought that Ptr acquired *ToxB* through vertical inheritance but this hypothesis has not been confirmed or truly explored. With the recent identification of homologs in more distantly related species the probability that *ToxB* has moved horizontally is increased. Additionally, a comprehensive comparison of

ToxB-containing isolates with chromosome resolution assemblies has never been conducted to determine how *ToxB* has become a multi-copy gene.

1.2.3 ToxC

The third primary effector of Ptr is ToxC, whose existence was known prior to ToxB as races 1 and 3 are predominant in North America and Australia (Lamari and Bernier, 1991; Lamari, et al., 1991). Unlike ToxA and ToxB however, the identity of ToxC has since remained a mystery. The first work which attempted to uncover the nature of ToxC determined it was not proteinaceous like ToxA and ToxB. Rather, ToxC was likely to be a polar, non-ionic metabolite with a low-molecular weight (Effertz et al., 2002). Aside from this single work, very little has been done to identify ToxC. Recently however, through the use of segregating Ptr populations (artificially created as Ptr is homothallic) a gene was identified which may play a role in ToxC synthesis (Shi et al, 2022). Described as *ToxC1*, knock-out experiments showed the gene was potentially necessary but not sufficient by itself to produce ToxC symptoms (Shi et al., 2022). Interestingly, the host gene *Tsc1* was also recently identified, whose presence plays a role in disease development (Effertz et al. 2002; Liu et al. 2020; Running et al. 2022). Renewed interest in ToxC suggests that the open question of 'What is ToxC?' is now tractable with modern genomics techniques.

1.2.4 PTR RACE AND TOX GENE DISTRIBUTION

While Ptr has a global presence, the Tox effectors are not equally represented in the various wheat growing regions of the world. For the past few decades, race testing combined with PCR testing has painted a clear picture of effector distribution in Ptr across the globe. In Canada, the Ptr population is predominantly composed of races 1

(ToxA and C) and 2 (ToxA), with races 3 (ToxC) and 4 (no toxin production) having also been isolated (Aboukhaddour et al. 2021; Guo et al. 2020; Wei et al. 2020). *ToxB*-containing races like race 5 (ToxB) are extremely rare, and races 6 (ToxB and C), 7 (ToxA and B), and 8 (ToxA, B, and C) have not been isolated in Canada (Aboukhaddour et al. 2021; Guo et al. 2020; Wei et al. 2020). Some isolates have been found with the *toxb* homolog however (Aboukhaddour et al. 2021; Guo et al. 2020; Wei et al. 2020; Hafez et al., 2024). This race structure is mirrored in the United States. In Australia, only races 1 (AC) and 2 (A) have been found. The *ToxB*-containing races appear to be restricted to North Africa and the Fertile Crescent (i.e. Mid-East and Caucasus regions) (reviewed in Kamel et al., 2019). Additionally, it is known that ToxA, ToxB, and ToxC are not the only effectors produced by Ptr. For example, some isolates from Tunisia were found to cause necrosis on some wheat cultivars but those isolates lack ToxA (Kamel et al., 2019). The race structure of Ptr populations continues to be monitored and provide clues into how Ptr has spread in the past as well as how Ptr and its Tox genes are evolving with respect to the different environmental conditions and host genotypes. Investigations into other potential Ptr effectors is also an essential component of current and future work.

1.3 DRIVERS OF PATHOGEN EVOLUTION

Pathogens have been attacking plants and co-evolving alongside them for millenia. However, with the advent of agriculture some 10,000 years ago, a unique selection pressure began as humans began to selectively domesticate crops. The relatively recent, high-intensity, uniform agricultural systems of modern society have placed an immensely strong selection pressure to enhance virulence and promote the emergence of

new pathogens (Stukenbrock and McDonald, 2008; Möller and Stuckenbrock 2017; Croll and McDonald, 2017). Various processes and mechanisms contribute to pathogen evolution and adaptability including, but not limited to: transposons, genome architecture, gene duplications, and horizontal gene transfers (Stukenbrock and McDonald, 2008; Möller and Stuckenbrock 2017). Each of these mechanisms are briefly discussed below, along with select examples which provide evidence for the roles of each in plant pathogen evolution.

1.3.1 TRANSPOSONS

Transposons, or transposable elements, are sections of genetic material that are mobile within or between genomes. These elements are ubiquitous in the genetic code of both prokaryotes and eukaryotes, and with few exceptions, are often abundant, with most genomes containing thousands of transposons (Schabel et al., 2009; Lander et al., 2011; Canapa et al., 2015). In plants, transposons account for a wide range of the total genomic content, from ~10% in Arabidopsis (Arabidopsis Genome Institute, 2000) to up to 85% in maize (Schabel et al., 2009). In fungi and animals, a wide range of transposon content has also been observed, with transposon content less than 1% in metazoans, ~20% many fungi with some species as high as ~90% (Badet et al., 2020; Gupta et al., 2023), and up to ~45% in humans and other mammals (Lander et al., 2011; Canapa et al., 2015). Transposons can vary greatly in size, from hundreds of base pairs to the extremely large Starships which can be >100,000 base pairs (Arkhipova et al., 2019; Gluck-Thaler et al., 2021).

Transposons were discovered by the cytogeneticist Barbara McClintock, who revealed that their movement was associated with colour changes in maize kernels

(McClintock, 1948; 1950). This foundational discovery spurred a growing body of research that has led to important conceptual shifts in evolutionary biology and genetics. Previously, transposons were commonly considered “junk DNA” that simply acted as spacers between genes, the truly important sections of DNA. Upon their discovery, transposons were considered “selfish” (Doolittle and Sapienza 1980; Orgel and Crick 1980), thought to be fulfilling no function besides their own self-replication at the expense of their host. We now know that transposons play many roles, and over time have contributed greatly to adaptation and evolution in all branches of life. Moreover, transposons are now considered by many to be their own evolutionary units with parasitic lifestyles and whose history predates the emergence of eukaryotes (Jeffares et al., 1998; Schulmann and Wicker, 2013).

1.3.1.1 CONSEQUENCES OF TRANSPOSONS

Any given transposon insertion has the ability to alter the carrier phenotype, potentially driving adaptation and carrier fitness. Ultimately, the exact location of an insertion determines how, and to what level, phenotypes are altered and varies depending on whether a promoter region, intron, exon, or some other region is interrupted. While some transposons have a low insertion specificity, and will therefore randomly insert into host genomes, others are biased towards specific genomic regions or sequences (Sultana et al., 2017). Such specificity may reduce the probability of causing a detrimental impact on the host and thereby improve the chances of their own proliferation (Sultana et al., 2017). Similar to any mutation, the impacts of transposon insertion and proliferation on carrier fitness can broadly be categorized as neutral, beneficial, or deleterious. Furthermore, although transposons by definition are mobile within genomes, for certain

groups, only a fraction are actually capable of mobilization. This is due to truncation of the sequences necessary to mobilize, because of host defence mechanisms (e.g. methylation preventing transposase transcription, heterochromatin formation, etc.), or due to a disruption or lack of the necessarily molecular machinery (i.e they are non-autonomous).

Insertion within a gene may or may not alter the gene function, depending on whether insertion occurs within an intron or exon. As introns are non-coding and are excised during normal gene expression, insertion into these regions may not alter the final product, so long as the transposon is fully spliced out during transcription and there is no frame shift that could result in exonization, gene truncation, or alternative splicing of coding regions (Dubin et al., 2018). Conversely, insertion within an exon is likely to cause drastic changes to gene function due to gene truncation and/or exonization. Gene truncation occurs when a transposon introduces a stop codon and prematurely aborts transcription. Outcomes can include incomplete transcripts or alternative gene splicing, both of which modify the intended amino acid sequence and consequently the resultant protein product. Exonization, in contrast to gene truncation, occurs when a new exon is introduced into a gene. This phenomenon alters transcript sequences and can therefore create defective proteins or proteins with novel functions. Additionally, transposons have the ability to create genes *de novo* through the creation or shuffling of open-reading frames (ORFs) and/or regulatory elements (Carvunis et al., 2012).

While transposon insertion within a gene can have a substantial influence, so can an insertion near a gene, due to alteration of regulatory sequences. Particularly, introduction into the promoter, enhancer, or silencer regions can cause complete

suppression or significant alteration of gene expression (Sultana et al., 2017; Dubin et al., 2018). Promoter regions are located upstream of their corresponding gene and control gene expression by providing a site for enzymes such as RNA polymerase to initiate transcription. Promoter sequences can be complex and any sequence alterations can completely suppress or severely alter protein expression profiles, ultimately having a variable suite of effects on the host. Enhancer and silencer regions serve as binding sites for transcription factors, either to increase or decrease gene expression, respectively. Changes in these regions can result in the same suite of responses as alterations to the promoter region (Dubin et al., 2018).

The potential effects of transposons are not limited to the direct alteration of genetic sequences and/or the functioning of regulatory elements. Transposon proliferation can interfere with genomic repair mechanisms that are intrinsic in all organisms, resulting in a phenomenon known as genome instability (Eichler and Sankoff, 2003; Chadha and Sharma, 2014; Bhat et al., 2022). Genome instability results in increased mutation rates, chromosomal rearrangements, and aneuploidy (abnormal chromosome number). Increased chromosomal rearrangement (i.e. translocations, inversions, duplications, deletions) in particular have been linked to transposons via an increase in potential break points and homologous regions during DNA replication (Wöstemeyer and Kreibich, 2002; Chan and Kolodner, 2011; Faino et al., 2016; Bourque et al., 2018). Similarly, aneuploidy can be a consequence of chromosome content imbalances originating from the disproportionate translocation of mobile elements (Yang et al., 2022). Gene duplications can also be caused by transposon proliferation due to two main mechanisms: gene-capture, where duplication occurs during transposon replication, or via unequal

cross overs caused by additional homologous sequences on which recombination can act (Reams and Roth, 2015). Gene duplication is discussed further in section 1.3.3.

1.3.1.2 TRANSPOSON INFLUENCES IN PLANT PATHOGENS

The effects of transposons in plant pathogens has been explored across a variety of species. There is a growing body of work to suggest transposons play multiple roles in altering pathogenicity and virulence through gene disruptions, silencing, duplications, and mobilization. For example, in the tomato pathogen *Fusarium oxysporum* f.sp *lycopersici* a new virulence race was created through the deactivation of the *AVR1* gene. *AVR1* is recognized by many host cultivars, and insertion of a *hAT* transposon within an exon resulted in the disruption of *AVR1* and subsequent non-recognition of the pathogen by the corresponding host R gene (Inami et al., 2012). A similar gain-of-virulence occurred in *Magnaporthe oryzae* via disruption of the avirulence gene *AvrPi9*, again resulting in a loss of recognition (Wu et al., 2015). In the widely studied potato pathogen *Phytophthora infestans*, effector silencing was shown to occur when multiple transposon insertions occurred near effector genes resulting in the formation of heterochromatin (Whisson et al., 2012). Heterochromatin formation is a defense mechanism employed by eukaryotes to prevent transposon activity through DNA packing, which prevents activity by restricting access to the DNA from necessary enzymes like transposases. In this case, the effector genes were unable to be expressed, resulting in an observable decrease in virulence.

A reduction in spore potential constitutes a major reduction in fitness, particularly in pathogens where spores are used as the major infecting propagule. In *Zymoseptoria tritici*, the multi-copy gene *REP9-1*, which negatively affects pycnidia (sexual spore bearing structure) production, was shown to be present on and replicating via the DNA transposon *REP9* with evidence of active mobility (Wang et al., 2021). Spore production

was also affected in the rice blast fungus *Magnaporthe oryzae,* with the insertion of a MGL LINE element in the conidiophore (asexual spore bearing structure) controlling gene *ACR1*/*medA*. This resulted in decreased appressorium (infection structure) formation, decreased pathogenicity, and altered conidiogenesis (asexual spore production) (Nishimura et al., 2000).

Chemical controls are widely used to mitigate damage caused by pathogens, and fungicide resistance is a major issue in modern agriculture. In *Blumeriella jaapi* (cherry leaf spot), resistance to the sterol demethylation inhibitor (DMI) class of fungicides was conferred by over-expression of the target gene *CYP51* through insertion of LINE-like elements upstream of the coding sequence (Ma et al., 2006). Over-expression of the target limits the efficacy of fungicides as the processes which were meant to be inhibited are still able to take place due to over abundance. Similar insertions which alter expression, leading to DMI resistance, have been observed in both *Venturia inequalis* (apple scab) and *Penicillium digitatum* (green rot) (Hamamoto et al., 2000; Schnabel and Jones, 2001). Many more examples can be found in the literature and it is clear that transposons can directly affect a wide variety of pathogen systems like virulence, spore production, and chemical sensitivity. However, the role transposons play in shaping pathogen evolution is even deeper as they are an integral part of defining genomic architecture in pathogen genomes (Faino et al., 2016).

1.3.2 GENOME ARCHITECTURE AND THE 'TWO-SPEED' MODEL

How genes are arranged within a genome can have significant impacts. It has been known for some time that genes with related functions are often clustered together (Walton, 2000; Keller et al., 2005; Hoffmeister & Keller, 2007). Indeed the 'selfish

clustering' hypothesis posits that genes with linked functions which are clustered give the individual genes increased fitness due to the additive selective advantage (Lawrence and Roth, 1996; Walton, 2000; Slot, 2017). This may be due in part to coordinated gene regulation, co-selection, and co-adaptation (Gacek and Strauss, 2012; Slot, 2017). This is particularly true for those involved in metabolic synthesis pathways (i.e. metabolic gene clusters (MGCs)) (Rokas et al., 2018).

A wide variety of MGCs in fungi have been studied due to their important roles in fungal biology and ecology. For example, MGCs for the synthesis of the antibiotic penicillin or the breakdown of substrates like nitrates and sugars (Diez et al., 1990; Johnstone et al., 1990; Slot and Rokas, 2010). Many MGCs in fungi have been found, but the most pertinent are those of necrotrophic effectors. Many of the identified necrotrophic effectors in pathogenic fungi are secondary metabolites. Biosynthesis of these toxic metabolites requires multiple genes generally consisting of enzymes, regulators, transporters, and primary synthesis genes, which catalyze the key step. It has been found that the production of aflatoxins, trichothecenes, and T-toxin (among others) are mediated by MGCs (Yu et al., 2004; Brown et al., 2004). The grouping of genes in pathogen genomes is not limited to only MGCs, however.

In bacterial pathogens, the concept of 'pathogenicity islands' (PAIs) has been used to explain the organizational structure and crowding of virulence related genes within and near mobile units like plasmids, insertion sequences (IS), and prokaryotic transposons like '*Tn*'s (Hacker and Kaper, 2000). A similar structural model to PAIs has been developed in fungal (and oomycete) pathogens, dubbed the 'two-speed' or 'two-compartment' genome (Raffaele et al., 2010; Dong et al., 2015). In this model, virulence

genes, accessory genes, transposons, and other repetitive elements are clustered together while core conserved genes are in regions with few repeats or transposable elements (Raffaele et al., 2010; Dong et al., 2015). The two regions, a.k.a compartments, also have different rates of mutation and rearrangements, with the accessory and transposon-dense regions having increased instances of both (Raffaele et al., 2010; Dong et al., 2015). Such an architecture has wide-reaching implications in pathogens that possess them including mobilization, duplication, and alteration of virulence genes, genes associated with adaptation to specific environmental stressors (e.g. heat tolerance, heavy metal tolerance) or hosts, and MGCs.

Originally the 'two-compartment' genome was dubbed the 'two-speed' genome. However, it is well established that regardless of genomic architecture, pathogens have conserved, slowly evolving core genes as well as rapidly evolving virulence genes, and thus all have 'two speeds'. Not all fungal pathogens conform to the 'two-compartment' model. In species with 'one-compartment' genomes, there are strong links between transposable elements and virulence genes (Frantzeskakis et al., 2019; Torres et al., 2020). It is hypothesized that in pathogens with 'one-compartment' genomes, rapid evolution can be attributed to high numbers of copy-number variation, often induced by transposable elements (Frantzeskakis et al., 2019; Torres et al., 2020). Furthermore, there is another type of compartmentalization which can create a 'two-speed' genome in which accessory chromosomes (syn. dispensable or mini-chromosomes) harbour large numbers of virulence genes or other adaptive traits similar to plasmids in prokaryotes (Frantzeskakis et al., 2019).

1.3.3 GENE DUPLICATION

Gene duplication is a major driver of evolution and adaptation by acting as a source of novel genetic material as the additional copies are free to diverge, often with at least one copy maintaining the original function (Nei et al., 1969; Holland et al., 1999; Hastings et al., 2009; Ohno et al., 2013; Sacristán et al., 2021). Ancient duplicated genes which have since diverged are often referred to as paralogous, and together are grouped into gene families (Zhang et al., 2003). Changes in gene family sizes can reflect adaptations to specific circumstances. For example, expansion of gene families such as carbohydrate-active enzymes (CAZymes), which are utilized by pathogens to degrade cell structures like cell walls, has been linked to host range changes in the *Colletotrichum* genus (Baronchelli et al., 2016). Gene duplication is also theorized to be a contributor to novel effector formation (Fouché et al., 2018; Sacristán et al., 2021). In the widespread fungal pathogen *Rhizoctonia solani,* pathogenicity gene duplication has been shown to be mediated by transposons with paralogs diverging into new functions (Francis et al., 2023). Differential expression of paralogs during host infection confirmed novel effector production as well (Francis et al., 2023).

Duplications can occur for a variety of reasons (e.g. non-homologous end joining, break-fusion-bridge, retroposition, whole genome duplication, etc.) which can produce segments of differing size (Ohno et al., 2013; Reams and Roth, 2015). As described, MGCs consist of clustered genes with related processes, and therefore duplication of MGC containing regions potentially provides an entire pathway on which selection can act. Different pathotypes of *Alternaria alternata* for example, produce different necrotrophic effectors (e.g. AM-toxin, AF-toxin, etc.) which cause disease on different

hosts (Tsuge et al., 2013). The evolution of these toxins is linked with complex duplications and recombination events of MGCs likely involving transposons (Tsuge et al., 2013). Similarly, the production of trichothecenes in *Fusarium* sp. is linked to a previous duplication of a biosynthetic gene cluster followed by gene losses and functional divergence (Ward et al., 2002; Brown et al., 2004; Proctor et al., 2018; Rokas et al., 2018). Such duplications support the hypothesis that duplications can result in novel virulence factors which contribute to pathogen success.

Examining copy-number variation (CNV) across a genome has been difficult in the past due to collapsing of short-reads during assembly and other issues with genome resolution. Advances in sequencing technology and improved genome assemblies have allowed systematic analysis of CNVs across a variety of genomes. In fungi, CNVs are linked to adaptive traits like heavy metal tolerance and contribute to genome plasticity in response to stress (Bazzicalupo et al., 2020; Vande Zande et al., 2023). In the plant pathogen *Rhynchosporim commune*, CNVs have been linked to azole resistance, CAZymes, and effectors (Stalder et al, 2023). In *Z. tritici*, CNVs contribute to climatic adaptation, and in *Melampsora lini* (flax rust) CNVs of *Avr* genes have been shown to disrupt host recognition (Figueroa et al., 2020; Tralamazza et al., 2024).

1.3.4 HORIZONTAL GENE TRANSFER

Horizontal gene transfers (HGTs) refer to the passing of genetic material through non-Mendelian mechanisms. Such transfers commonly occur in prokaryotes through mobile elements like plasmids, however, the extent to which HGTs have shaped eukaryotic biology is poorly understood. There is a growing body of evidence to suggest that HGTs have played an important role in eukaryotic evolution, including that of

pathogenic fungi. Horizontal transfers between fungi likely occur during anastomosis of hyphae or germ tubes which creates a cell-to-cell cytoplasmic connection through which DNA can move (Soanes and Richards, 2014). Rather than transferring circular plasmids as in bacteria, the largest horizontal transfers in pathogenic fungi are facilitated via accessory chromosomes (syn. dispensable chromosomes), mini-chromosomes, chromosomal arms, or Starship transposons (Bucknell et al., 2023; Wang et al., 2023; Barragan et al., 2024; Peck et al., 2024; Urquhart et al., 2024; Habig et al., 2025). Such significant transfers are often followed by large gene losses due to purifying selection (Kellis et al., 2004; Bloome et al., 2006; Wolf and Koonin, 2013). However, some genes acquired through HGT are likely to experience strong positive selection, in particular when the gene provides immediate benefit. The transfer of *ToxA* into the Ptr genome is a prime example of this phenomenon as the acquisition provided a large fitness advantage when infecting its primary host (Liu et al., 2006; Friesen et al., 2006; McDonald et al., 2018; 2019).

Most smaller HGT events appear to be driven by transposons, as seen with *ToxA* and *ToxhAT*. For example, the effector gene *Ave1* from fungal pathogen *Verticillium dahlia* was identified in the bacterial pathogen *Xanthomonas axonopodis,* as well as in other fungal pathogens *Fusarium oxysporum*, *Colletotrichum higginsianum*, and *Cercospora beticola* (De Jonge et al., 2012; Shi-Kunne et al., 2019). *Ave1* was found to be associated with the *Ty1-Copia* class of retrotransposons and phylogenetic analysis combined with gene distributions suggest horizontal movement of *Ave1* between these species (Nembaware et al., 2004; De Jong et al., 2012; Faino et al., 2016). *Ave1* may not always provide a virulence advantage when present, and so its gain or loss will confer

positive or negative fitness depending on the host genotype. For example, in tomato, the resistance gene that recognizes *Ave1* is *Vei1*, and upon detection of the effector, initiates an immune response (De Jonge et al., 2012). In tomato populations with *Vei1* the acquisition of *Ave1* confers a disadvantage to the pathogen, while in populations lacking *Vei1*, the pathogen has acquired a new virulence tool (De Jonge et al., 2012). The ability of pathogens to gain and lose virulence factors is a key component of their adaptability.

'Crinkling-and-necrosis' (CRN) effectors have been studied for some time as a highly conserved protein family present in both eukaryotic and prokaryotic plant pathogens (Zhang et al., 2016; Amaro et al., 2017). Although there is divergence between the eukaryotic and bacterial CRN families, phylogenetic analysis has suggested the horizontal movement of CRNs into eukaryotes has occurred multiple times (Zhang et al., 2016). Recipients of these transfers include both pathogenic fungi and oomycetes. It appears that CRNs themselves may have once acted as transposases and their functional shift to act as effectors is a notable example of exaptation (Zhang et al., 2016). Similar to the *Ave1* effector above, the addition of CRNs to a pathogens genome provides an advantage when hosts lack the appropriate genes to respond.

As Ptr has experienced at least one confirmed HGT event in its history, it is worth considering that it may have experienced others in the past, either as the recipient or the donor. For example, *ToxB* is traditionally thought to have been acquired via traditional vertical inheritance, but it is entirely possible that *ToxB* was acquired through an ancient HGT event. The extent to which HGTs have shaped the evolutionary trajectory of plant pathogens is an open question and the leaf-spot disease complex is an ideal system in

which to study not only this phenomenon but all the aforementioned drivers of pathogen evolution.

1.4 OBJECTIVES

During the course of a comparative genomics project, the trajectory of analysis and research objectives can shift as new knowledge and insights are gleaned. Having stated that, the primary objectives were to sequence and annotate a large number of *Pyrenophora tritici-repentis* genomes in order to explore the species genetic diversity. In particular, comparative analysis between isolates with different virulence race profiles and isolates from different geographical regions. Finally, as it was previously established that *ToxA* was mobile and that *ToxB* is present in multiple copies, an assessment on the potential mobility and duplication mechanisms of the Ptr Tox genes was a high priority.

As work progressed, more specific objectives were created:

1) Sequence a large set of Ptr isolates and create a pangenome;

2) Determine if the Ptr genome has a one or two-compartment architecture;

3) Annotate the transposable element content of the Ptr genome;

4) Assess the relatedness of different Ptr races from diverse geographical backgrounds;

5) Sequence isolates which were known to have the *ToxA* gene on different chromosomes using long-reads and determine by what mechanism *ToxA* was translocated;

6) Sequence a large number of isolates containing *ToxB* and discover the mechanism by which the *ToxB* gene replicates;

7) Utilize the large amount of sequence data generated to attempt a genome-wide association study to correlate single nucleotide polymorphisms with the ToxC phenotype.

1.5 REFERENCES

Abeysekara, N.S., Friesen, T.L., Keller, B. and Faris, J.D., 2009. Identification and characterization of a novel host–toxin interaction in the wheat–*Stagonospora nodorum* pathosystem. *Theoretical and Applied Genetics*, **120**: 117-126.

Aboukhaddour, R., Kim, Y.M. and Strelkov, S.E., 2012. RNA-mediated gene silencing of *ToxB* in *Pyrenophora tritici-repentis*. *Molecular Plant Pathology*, **13**: 318-326.

Aboukhaddour, R., Turkington, T.K. and Strelkov, S.E., 2013. Race structure of *Pyrenophora triciti*-repentis (tan spot of wheat) in Alberta, Canada. *Canadian Journal of Plant Pathology*, **35**: 256-268.

Aboukhaddour, R., Hafez, M., Strelkov, S.E. and Fernandez, M.R., 2021. Tan spot disease under the lenses of plant pathologists. In *Achieving durable disease resistance in cereals*, 589-621. Burleigh Dodds Science Publishing.

Aboukhaddour, R., Hafez, M., McDonald, M., Moffat, C.S., Navathe, S., Friesen, T.L., Strelkov, S.E., Oliver, R.P., Tan, K.C., Liu, Z. and Moolhuijzen, P.M., 2023. A revised nomenclature for *ToxA* haplotypes across multiple fungal species. *Phytopathology*, **113**: 1180-1184.

Amaike, S., Ozga, J.A., Basu, U. and Strelkov, S.E., 2008. Quantification of *ToxB* gene expression and formation of appressoria by isolates of *Pyrenophora tritici-repentis* differing in pathogenicity. *Plant Pathology*, **57**: 623-633.

Amaro, T.M., Thilliez, G.J., Motion, G.B. and Huitema, E., 2017. A perspective on CRN proteins in the genomics age: evolution, classification, delivery and function revisited. *Frontiers in Plant Science*, **8**: 99.

Andrie, R. M., Schoch, C. L., Hedges, R., Spatafora, J. W., and Ciuffetti, L. M. 2008. Homologs of *ToxB*, a host-selective toxin gene from *Pyrenophora tritici-repentis*, are present in the genome of sister-species *Pyrenophora bromi* and other members of the Ascomycota. *Fungal Genetics and Biology*, **45**: 363-377.

Arabidopsis Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**: 796-815.

Arkhipova I.R., Yushenova I.A. and Angert E. 2019. Giant transposons in eukaryotes: Is bigger better? *Genome Biology and Evolution,* **11**: 906–18.

Badet, T., Oggenfuss, U., Abraham, L., McDonald, B.A. and Croll, D., 2020. A 19-isolate reference-quality global pangenome for the fungal wheat pathogen *Zymoseptoria tritici*. *BMC Biology*, **18**: 1-18.

Ballance, G.M., Lamari, L. and Bernier, C.C., 1989. Purification and characterization of a host-selective necrosis toxin from *Pyrenophora tritici-repentis*. *Physiological and Molecular Plant Pathology*, **35**: 203-213.

Barragan, A.C., Latorre, S.M., Malmgren, A., Harant, A., Win, J., Sugihara, Y., Burbano, H.A., Kamoun, S. and Langner, T., 2024. Multiple horizontal mini-chromosome transfers drive genome evolution of clonal blast fungus lineages. *Molecular Biology and Evolution*, **41**, 164.

Baroncelli, R., Amby, D.B., Zapparata, A., Sarrocco, S., Vannacci, G., Le Floch, G., Harrison, R.J., Holub, E., Sukno, S.A., Sreenivasaprasad, S. and Thon, M.R., 2016. Gene family expansions and contractions are associated with host range in plant pathogens of the genus *Colletotrichum*. *BMC Genomics*, **17**: 1-17.

Barthe, P., Pujade-Renaud, V., Breton, F., Gargani, D., Thai, R., Roumestand, C. and De Lamotte, F., 2007. Structural analysis of cassiicolin, a host-selective protein toxin from *Corynespora cassiicola*. *Journal of Molecular Biology*, **367**: 89-101.

Bazzicalupo, A.L., Ruytinx, J., Ke, Y.H., Coninx, L., Colpaert, J.V., Nguyen, N.H., Vilgalys, R. and Branco, S., 2020. Fungal heavy metal adaptation through single nucleotide polymorphisms and copy-number variation. *Molecular Ecology*, **29**: 4157-4169.

Bhat A, Ghatage T, Bhan S, Lahane GP, Dhar A, Kumar R, Pandita RK, Bhat KM, Ramos KS, Pandita TK. 2022. Role of transposable elements in genome stability: implications for health and disease. *International Journal of Molecular Science*, **23**:7802.

Blomme, T., Vandepoele, K., De Bodt, S., Simillion, C., Maere, S. and Van de Peer, Y., 2006. The gain and loss of genes during 600 million years of vertebrate evolution. *Genome Biology*, **7**: 1-12.

Bourque, G., Burns, K.H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H.L., Macfarlan, T.S. and Mager, D.L., 2018. Ten things you should know about transposable elements. *Genome Biology*, **19**: 1-12.

Brown, D.W., Dyer, R.B., McCormick, S.P., Kendra, D.F. and Plattner, R.D., 2004. Functional demarcation of the *Fusarium* core trichothecene gene cluster. *Fungal Genetics and Biology*, **41**: 454-462.

Bucknell, A., Wilson, H., Goncalves dos Santos, K.C., Simpfendorfer, S., Milgate, A., Germain, H., Solomon, P., Bentham, A. and McDonald, M., 2024. Sanctuary: A Starship transposon facilitating the movement of the virulence factor *ToxA* in fungal wheat pathogens. *bioRxiv*: 2024-03.

Canapa, A., Barucca, M., Biscotti, M.A., Forconi, M. and Olmo, E., 2015. Transposons, genome size, and evolutionary insights in animals. *Cytogenetic and Genome Research*, **147**: 217-239.

Carvunis, A.R., Rolland, T., Wapinski, I., Calderwood, M.A., Yildirim, M.A., Simonis, N., Charloteaux, B., Hidalgo, C.A., Barbette, J., Santhanam, B. and Brar, G.A., 2012. Proto-genes and de novo gene birth. *Nature*, **487**: 370-374.

Chadha, S. and Sharma, M., 2014. Transposable elements as stress adaptive capacitors induce genomic instability in fungal pathogen *Magnaporthe oryzae*. *PLoS One*, **9:** e94415.

Chan, J.E. and Kolodner, R.D., 2011. A genetic and structural study of genome rearrangements mediated by high copy repeat *Ty1* elements. *PLoS Genetics*, **7**: e1002089.

Croll, D. and McDonald, B.A., 2017. The genetic basis of local adaptation for pathogenic fungi in agricultural ecosystems. *Molecular Ecology*, **26**: 2027-2040.

Dagvadorj, B., Outram, M.A., Williams, S.J. and Solomon, P.S., 2022. The necrotrophic effector ToxA from *Parastagonospora nodorum* interacts with wheat NHL proteins to facilitate Tsn1-mediated necrosis. *The Plant Journal*, **110**: 407-418.

De Jonge, R., Peter van Esse, H., Maruthachalam, K., Bolton, M.D., Santhanam, P., Saber, M.K., Zhang, Z., Usami, T., Lievens, B., Subbarao, K.V. and Thomma, B.P., 2012. Tomato immune receptor *Ve1* recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. *Proceedings of the National Academy of Sciences*, **109**: 5110-5115.

Diedicke H, 1902. Uber den Zusammenhang zwischen Pleospora—und *Helminthosporium*—Arten. *Centralblatt für Bakteriologie und Parasitenkunde Jena Abt* **9**: 317–329.

Díez, B., Gutiérrez, S., Barredo, J.L., van Solingen, P., Van Der Voort, L.H. and Martin, J.F., 1990. The cluster of penicillin biosynthetic genes. Identification and characterization of the *pcbAB* gene encoding the alpha-aminoadipyl-cysteinyl-valine synthetase and linkage to the *pcbC* and *penDE* genes. *Journal of Biological Chemistry*, **265**: 16358-16365.

Dong, S., Raffaele, S. and Kamoun, S., 2015. The two-speed genomes of filamentous pathogens: waltz with plants. *Current Opinion in Genetics & Development*, **35**: 57-65.

Doolittle, W.F. and Sapienza, C., 1980. Selfish genes, the phenotype paradigm and genome evolution. *Nature*, **284**: 601-603.

Drechsler C, 1923. Some graminicolous species of *Helminthosporium*. *International Journal of Agricultral Research*, **24**: 641-740.

Dubin, M.J., Scheid, O.M. and Becker, C., 2018. Transposons: a blessing curse. *Current Opinion in Plant Biology*, **42**: 23-29.

Eichler, E.E. and Sankoff, D., 2003. Structural dynamics of eukaryotic chromosome evolution. S*cience*, **301**: 793-797.

Effertz, R.J., Meinhardt, S.W., Anderson, J.A., Jordahl, J.G. and Francl, L.J., 2002. Identification of a chlorosis-inducing toxin from *Pyrenophora tritici-repentis* and the chromosomal location of an insensitivity locus in wheat. *Phytopathology*, **92**: 527-533.

Erenstein, O., Jaleta, M., Mottaleb, K.A., Sonder, K., Donovan, J. and Braun, H.J., 2022. Global trends in wheat production, consumption and trade. In *Wheat improvement: food security in a changing climate* (pp. 47-66). Cham: Springer International Publishing.

Faino, L., Seidl, M.F., Shi-Kunne, X., Pauper, M., van den Berg, G.C., Wittenberg, A.H. and Thomma, B.P., 2016. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. *Genome Research*, **26**: 1091-1100.

Faris, J.D., Zhang, Z., Lu, H., Lu, S., Reddy, L., Cloutier, S., Fellers, J.P., Meinhardt, S.W., Rasmussen, J.B., Xu, S.S. and Oliver, R.P., 2010. A unique wheat disease resistance-like gene governs effector-triggered susceptibility to necrotrophic pathogens. *Proceedings of the National Academy of Sciences*, **107**: 13544-13549.

Faris, J.D., Liu, Z. and Xu, S.S., 2013. Genetics of tan spot resistance in wheat. *Theoretical and Applied Genetics*, **126**: 2197-2217.

Figueroa, M., Hammond-Kosack, K.E. and Solomon, P.S., 2018. A review of wheat diseases—a field perspective. *Molecular Plant Pathology*, **19**: 1523-1536.

Figueroa, M., Dodds, P.N. and Henningsen, E.C., 2020. Evolution of virulence in rust fungi—multiple solutions to one problem. *Current Opinion in Plant Biology*, **56**: 20-27.

Flor HH, 1971. Current status of the gene-for-gene concept. *Annual Revie Phytopathology*, **9**: 175-296.

Fouché, S., Plissonneau, C. and Croll, D., 2018. The birth and death of effectors in rapidly evolving filamentous pathogen genomes. *Current Opinion in Microbiology*, **46**: 34-42.

Francis, A., Ghosh, S., Tyagi, K., Prakasam, V., Rani, M., Singh, N.P., Pradhan, A., Sundaram, R.M., Priyanka, C., Laha, G.S. and Kannan, C., 2023. Evolution of pathogenicity-associated genes in *Rhizoctonia solani* AG1-IA by genome duplication and transposon-mediated gene function alterations. *BMC Biology*, **21**: 1-19.

Frantzeskakis, L., Kusch, S. and Panstruga, R., 2018. The need for speed: compartmentalized genome evolution in filamentous phytopathogens. *Molecular Plant Pathology*, **20**: 3.

Friesen, T.L., Stukenbrock, E.H., Liu, Z., Meinhardt, S., Ling, H., Faris, J.D., Rasmussen, J.B., Solomon, P.S., McDonald, B.A. and Oliver, R.P., 2006. Emergence of a new disease as a result of interspecific virulence gene transfer. *Nature Genetics*, **38**: 953-956.

Friesen, T.L., Meinhardt, S.W. and Faris, J.D., 2007. The *Stagonospora nodorum*-wheat pathosystem involves multiple proteinaceous host-selective toxins and corresponding host sensitivity genes that interact in an inverse gene-for-gene manner. *The Plant Journal*, **51**: 681-692.

Friesen, T.L., Faris, J.D., Solomon, P.S. and Oliver, R.P., 2008. Host-specific toxins: effectors of necrotrophic pathogenicity. *Cellular Microbiology*, **10**: 1421-1428.

Friesen, T.L., Chu, C., Xu, S.S. and Faris, J.D., 2012. *SnTox5–Snn5*: a novel *Stagonospora nodorum* effector–wheat gene interaction and its relationship with the *SnToxA–Tsn1* and *SnTox3–Snn3–B1* interactions. *Molecular Plant Pathology*, **13**: 1101-1109.

Gacek, A. and Strauss, J., 2012. The chromatin code of fungal secondary metabolite gene clusters. *Applied Microbiology and Biotechnology*, **95**: 1389-1404.

Gao, Y., Faris, J.D., Liu, Z., Kim, Y.M., Syme, R.A., Oliver, R.P., Xu, S.S. and Friesen, T.L., 2015. Identification and characterization of the *SnTox6-Snn6* interaction in the *Parastagonospora nodorum*–wheat pathosystem. *Molecular Plant-Microbe Interactions*, **28**: 615-625.

Gluck-Thaler, E., Ralston, T., Konkel, Z., Ocampos, C.G., Ganeshan, V.D., Dorrance, A.E., Niblack, T.L., Wood, C.W., Slot, J.C., Lopez-Nicora, H.D. and Vogan, A.A., 2022. Giant starship elements mobilize accessory genes in fungal genomes. *Molecular Biology and Evolution*, **39**: p.msac109.

Guo, J., Shi, G., Kalil, A., Friskop, A., Elias, E., Xu, S.S., Faris, J.D. and Liu, Z., 2020. *Pyrenophora tritici-repentis* Race 4 Isolates cause disease on tetraploid wheat. *Phytopathology*, **110**: 1781-1790.

Gupta, Y.K., Marcelino-Guimarães, F.C., Lorrain, C., Farmer, A., Haridas, S., Ferreira, E.G.C., Lopes-Caitar, V.S., Oliveira, L.S., Morin, E., Widdison, S. and Cameron, C., 2023. Major proliferation of transposable elements shaped the genome of the soybean rust pathogen *Phakopsora pachyrhizi*. *Nature Communications*, **14**: 1835.

Habig, M., Patneedi, S.K., Stam, R. and De Fine Licht, H.H., 2025. Horizontal transfer of accessory chromosomes in fungi–a regulated process for exchange of genetic material?. *Heredity*: 1-7.

Hacker, J. and Kaper, J.B., 2000. Pathogenicity islands and the evolution of microbes. *Annual Reviews in Microbiology*, **54**: 641-679.

Hafez, M., Despins, T., Nakajima, K. and Aboukhaddour, R., 2022. Identification of a novel *ToxA* haplotype of *Pyrenophora tritici-repentis* from Japan. *Phytopathology*, **112**: 1597-1602.

Hafez, M., Gourlie, R., McDonald, M., Telfer, M., Carmona, M.A., Sautua, F.J., Moffat, C.S., Moolhuijzen, P.M., See, P.T. and Aboukhaddour, R., 2024. Evolution of the *ToxB* gene in *Pyrenophora tritici-repentis* and related species. *Molecular Plant-Microbe Interactions*, **37**: 327-337.

Hamamoto, H., Hasegawa, K., Nakaune, R., Lee, Y.J., Makizumi, Y., Akutsu, K. and Hibi, T., 2000. Tandem repeat of a transcriptional enhancer upstream of the sterol 14α-demethylase gene (*CYP51*) in *Penicillium digitatum*. *Applied and Environmental Microbiology*, **66**: 3421-3426.

Hastings, P. J., Lupski, J. R., Rosenberg, S. M., and Ira, G. 2009. Mechanisms of change in gene copy number. *Nature Reviews Genetics*, **10**: 551– 564.

Hoffmeister, D. and Keller, N.P., 2007. Natural products of filamentous fungi: enzymes, genes, and their regulation. *Natural Product Reports*, **24**: 393-416.

Holland, P.W., 1999, October. Gene duplication: past, present and future. In *Seminars in Cell & Developmental Biology*, **10**: 541-547. Academic Press.

Hosford Jr, R.M., 1971. A form of *Pyrenophora trichostoma* pathogenic to wheat and other grasses. *Phytopathology*, **61**: 28-32.

Inami, K., Yoshioka-Akiyama, C., Morita, Y., Yamasaki, M., Teraoka, T. and Arie, T., 2012. A genetic mechanism for emergence of races in *Fusarium oxysporum* f. sp. *lycopersici*: inactivation of avirulence gene *AVR1* by transposon insertion. *PLoS One* **7**: e44101

Jeffares, D.C., Poole, A.M. and Penny, D., 1998. Relics from the RNA world. *Journal of Molecular Evolution*, **46**: 18-36

Johnstone, I.L., McCabe, P.C., Greaves, P., Gurr, S.J., Cole, G.E., Brow, M.A.D., Unkles, S.E., Clutterbuck, A.J., Kinghorn, J.R. and Innis, M.A., 1990. Isolation and characterisation of the *crnA-niiA-niaD* gene cluster for nitrate assimilation in *Aspergillus nidulans*. *Gene*, **90**: 181-192.

Kamel, S., Cherif, M., Hafez, M., Despins, T. and Aboukhaddour, R., 2019. *Pyrenophora tritici–repentis* in Tunisia: race structure and effector genes. *Frontiers in Plant Science*, **10**: 1562.

Keller, N.P., Turner, G. and Bennett, J.W., 2005. Fungal secondary metabolism—from biochemistry to genomics. *Nature Reviews Microbiology*, **3**: 937-947.

Kellis M., Birren B. W. and Lander E. S. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cereviseae*. *Nature*, **428**: 617–624.

Kim, Y.M., Bouras, N., Kav, N.N. and Strelkov, S.E., 2010. Inhibition of photosynthesis and modification of the wheat leaf proteome by Ptr *ToxB*: a host-specific toxin from the fungal pathogen *Pyrenophora tritici-repentis*. *Proteomics*, **10**: 2911-2926.

Kono, Y., Takeuchi, S., Kawarada, A., Daly, J.M. and Knoche, H.W., 1980. Structure of the host-specific pathotoxins produced by *Helminthosporium maydis*, race T. *Tetrahedron Letters*, **21**: 1537-1540.

Lamari, L. and Bernier, C.C., 1991. Genetics of tan necrosis and extensive chlorosis in tan spot of wheat caused by *Pyrenophora tritici-repentis*. *Phytopathology*, **81**: 1092-1095.

Lamari, L., Bernier, C.C. and Smith, R.B., 1991. Wheat genotypes that develop both tan necrosis and extensive chlorosis in response to isolates of *Pyrenophora tritici-repentis*. *Plant Disease*, **75**: 121-122.

Lamari, L., Sayoud, D., Boulif, M., and Bernier, C. C. 1995. Identification of a new race in *Pyrenophora tritici-repentis*: Implications for the current pathotype classification system. *Canadian Journal of Plant Pathology*. **17**: 312-318.

Lamari, L., Strelkov, S.E., Yahyaoui, A., Orabi, J. and Smith, R.B., 2003. The identification of two new races of *Pyrenophora tritici-repentis* from the host center of diversity confirms a one-to-one relationship in tan spot of wheat. *Phytopathology*, **93**: 391-396.

Lamari, L., Strelkov, S.E., Yahyaoui, A., Amedov, M., Saidov, M., Djunusova, M. and Koichibayev, M., 2005a. Virulence of *Pyrenophora tritici-repentis* in the countries of the Silk Road. *Canadian Journal of Plant Pathology*, **27**: 383-388.

Lamari, L., McCallum, B.D. and Depauw, R.M., 2005b. Forensic pathology of Canadian bread wheat: the case of tan spot. *Phytopathology*, **95**: 144-152.

Lander, E.S., 2011. Initial impact of the sequencing of the human genome. *Nature*, **470**: 187-197.

Lawrence, J.G. and Roth, J.R., 1996. Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics*, **143**: 1843-1860.

Liu, Z.H., Faris, J.D., Meinhardt, S.W., Ali, S., Rasmussen, J.B. and Friesen, T.L., 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.

Liu, Z., Faris, J.D., Oliver, R.P., Tan, K.C., Solomon, P.S., McDonald, M.C., McDonald, B.A., Nunez, A., Lu, S., Rasmussen, J.B. and Friesen, T.L., 2009. *SnTox3* acts in effector triggered susceptibility to induce disease on wheat carrying the *Snn3* gene. *PLoS Pathogens*, **5:** e1000581.

Liu, Y., Salsman, E., Wang, R., Galagedara, N., Zhang, Q., Fiedler, J.D., Liu, Z., Xu, S., Faris, J.D. and Li, X., 2020. Meta-QTL analysis of tan spot resistance in wheat. *Theoretical and Applied Genetics*, **133**: 2363-2375.

Lu, S., Faris, J.D., Sherwood, R., Friesen, T.L. and Edwards, M.C., 2014. A dimeric PR-1-type pathogenesis-related protein interacts with *ToxA* and potentially mediates ToxA-induced necrosis in sensitive wheat. *Molecular Plant Pathology*, **15**: 650-663.

Lu, S., Turgeon, B.G. and Edwards, M.C., 2015. A ToxA-like protein from *Cochliobolus heterostrophus* induces light-dependent leaf necrosis and acts as a virulence factor with host selectivity on maize. *Fungal Genetics and Biology*, **81**: 12-24.

Da Luz, W.C. and Bergstrom, G.C., 1986. Effect of temperature on tan spot development in spring wheat cultivars differing in resistance. *Canadian Journal of Plant Pathology*, **8**: 451-454.

Ma, Z., Proffer, T.J., Jacobs, J.L. and Sundin, G.W., 2006. Overexpression of the 14α-demethylase target gene (*CYP51*) mediates fungicide resistance in *Blumeriella jaapii*. *Applied and Environmental Microbiology*, **72**: 2581-2585.

Manning, V.A., Hardison, L.K. and Ciuffetti, L.M., 2007. Ptr ToxA interacts with a chloroplast-localized protein. *Molecular Plant-Microbe Interactions*, **20**: 168-177.

Macko, V., Acklin, W., Hildenbrand, C., Weibel, F. and Arigoni, D., 1983. Structure of three isomeric host-specific toxins from *Helminthosporium sacchari*. *Experientia*, **39**: 343-347.

Macko, V., Stimmel, M.B., Wolpert, T.J., Dunkle, L.D., Acklin, W., Bänteli, R., Jaun, B. and Arigoni, D., 1992. Structure of the host-specific toxins produced by the fungal pathogen *Periconia circinata*. *Proceedings of the National Academy of Sciences*, **89**: 9574-9578.

Martinez, J. P., Ottum, S. A., Ali, S., Francl, L. J., and Ciuffetti, L. M. 2001. Characterization of the *ToxB* gene from *Pyrenophora tritici-repentis*. *Molecular Plant-Microbe Interactions,* **14**: 675-677

McClintock, B., 1948. Mutable loci in maize. *Carnegie Institution of Washington,* **46**: 155-169

McClintock, B., 1950. The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences*, **36**: 344-355

McDonald, M.C., Oliver, R.P., Friesen, T.L., Brunner, P.C. and McDonald, B.A., 2013. Global diversity and distribution of three necrotrophic effectors in *Phaeosphaeria nodorum* and related species. *New Phytologist*, **199**: 241-251.

McDonald, M.C., Ahren, D., Simpfendorfer, S., Milgate, A. and Solomon, P.S., 2018. The discovery of the virulence gene *ToxA* in the wheat and barley pathogen *Bipolaris sorokiniana*. *Molecular Plant Pathology*, **19**: 432-439.

Mitra M, 1934. A leaf spot disease of wheat caused by *Helminthosporium tritici-repentis Died. Indian Journal of Agricultral Science,* **4**: 692.

Moffat, C.S. and Santana, F.M., 2018. Diseases affecting wheat: tan spot. In *Integrated disease management of wheat and barley* (pp. 115-128). Burleigh Dodds Science Publishing.

Möller, M. and Stukenbrock, E.H., 2017. Evolution and genome architecture in fungal plant pathogens. *Nature Reviews Microbiology*, **15**: 756-771.

Nei, M., 1969. Gene duplication and nucleotide substitution in evolution. *Nature*, **221**: 40-42.

Nembaware, V., Seoighe, C., Sayed, M. and Gehring, C., 2004. A plant natriuretic peptide-like gene in the bacterial pathogen *Xanthomonas axonopodis* may induce hyper-hydration in the plant host: a hypothesis of molecular mimicry. *BMC Evolutionary Biology*, **4**: 1-7.

Nishimura, M., Hayashi, N., Jwa, N.S., Lau, G.W., Hamer, J.E. and Hasebe, A., 2000. Insertion of the LINE retrotransposon MGL causes a conidiophore pattern mutation in *Magnaporthe grisea*. *Molecular Plant-Microbe Interactions*, **13**: 892-894.

Ohno, S., 2013. Evolution by gene duplication. Springer Science & Business Media.

Oliver, R. P., 2024. Control of plant diseases, in *Agrios' Plant Pathology*. (Eds). Oliver, R. P, Ralph, H., Del Ponte, E. M., and Di Petrio A. London, UK: Academic Press: 279-316.

Orgel, L.E. and Crick, F.H., 1980. Selfish DNA: the ultimate parasite. *Nature*, **284**: 604-607.

Pedras, M.S.C., Chumala, P.B. and Quail, J.W., 2004. Chemical mediators: the remarkable structure and host-selectivity of depsilairdin, a sesquiterpenic depsipeptide containing a new amino acid. *Organic Letters*, **6**: 4615-4617.

Pedras, M.S.C. and Yu, Y., 2008. Structure and biological activity of maculansin A, a phytotoxin from the phytopathogenic fungus *Leptosphaeria maculans*. *Phytochemistry*, **69**: 2966-2971.

Petrov, V., Qureshi, M.K., Hille, J. and Gechev, T., 2018. Occurrence, biochemistry and biological effects of host-selective plant mycotoxins. *Food and Chemical Toxicology*, **112**: 251-264.

Proctor, R.H., McCormick, S.P., Kim, H.S., Cardoza, R.E., Stanley, A.M., Lindo, L., Kelly, A., Brown, D.W., Lee, T., Vaughan, M.M. and Alexander, N.J., 2018. Evolution of structural diversity of trichothecenes, a family of toxins produced by plant pathogenic and entomopathogenic fungi. *PLoS Pathogens*, **14**: e1006946.

Raffaele, S., Farrer, R.A., Cano, L.M., Studholme, D.J., MacLean, D., Thines, M., Jiang, R.H., Zody, M.C., Kunjeti, S.G., Donofrio, N.M. and Meyers, B.C., 2010. Genome evolution following host jumps in the Irish potato famine pathogen lineage. *Science*, **330**: 1540-1543.

Rees, R.G., Platz, G.J. and Mayer, R.J., 1982. Yield losses in wheat from yellow spot: comparison of estimates derived from single tillers and plots. *Australian Journal of Agricultural Research*, **33**: 899-908.

Reams, A.B. and Roth, J.R., 2015. Mechanisms of gene duplication and amplification. *Cold Spring Harbor Perspectives in Biology*, **7**: a016592.

Rokas, A., Wisecaver, J.H. and Lind, A.L., 2018. The birth, evolution and death of metabolic gene clusters in fungi. *Nature Reviews Microbiology,* **16**: 731-744.

Running, K.L., Momotaz, A., Kariyawasam, G.K., Zurn, J.D., Acevedo, M., Carter, A.H., Liu, Z. and Faris, J.D., 2022. Genomic analysis and delineation of the tan spot susceptibility locus *Tsc1* in wheat. *Frontiers in Plant Science*, **13**: 793925.

Sacristán, S., Goss, E.M. and Eves-van den Akker, S., 2021. How do pathogens evolve novel virulence activities?. *Molecular Plant-Microbe Interactions*, **34**: 576-586.

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

Schnabel, G. and Jones, A.L., 2001. The 14α-demethylasse (*CYP51A1*) gene is overexpressed in *Venturia inaequalis* strains resistant to myclobutanil. *Phytopathology*, **91**: 102-110.

Schnabel, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A. and Minx, P., 2009. The B73 maize genome: complexity, diversity, and dynamics. *Science*, **326**: 1112-1115.

Schulman, A.H. and Wicker, T., 2013. A field guide to transposable elements. *Plant Transposons and Genome Dynamics in Evolution*, **15**: 40.

Shi, G., Friesen, T.L., Saini, J., Xu, S.S., Rasmussen, J.B. and Faris, J.D., 2015. The wheat *Snn7* gene confers susceptibility on recognition of the *Parastagonospora nodorum* necrotrophic effector SnTox7. *The Plant Genome*, **8**: 2015-02.

Shi, G., Kariyawasam, G., Liu, S., Leng, Y., Zhong, S., Ali, S., Moolhuijzen, P., Moffat, C.S., Rasmussen, J.B., Friesen, T.L. and Faris, J.D., 2022. A conserved hypothetical gene is required but not sufficient for Ptr ToxC production in *Pyrenophora tritici-repentis*. *Molecular Plant-Microbe Interactions*, **35**: 336-348.

Shi-Kunne, X., van Kooten, M., Depotter, J.R., Thomma, B.P. and Seidl, M.F., 2019. The genome of the fungal pathogen *Verticillium dahliae* reveals extensive bacterial to fungal gene transfer. *Genome Biology and Evolution*, **11**: 855-868.

Shoemaker RA, 1962. Drechslera Ito. *Canadian Journal of Botany*, **40**: 809-36.

Slot, J.C. and Rokas, A., 2010. Multiple GAL pathway gene clusters evolved independently and by different mechanisms in fungi. *Proceedings of the National Academy of Sciences*, **107**: 10136-10141.

Slot, J.C., 2017. Fungal gene cluster diversity and evolution. *Advances in Genetics*, **100**: 141-178.

Stalder, L., Oggenfuss, U., Mohd-Assaad, N. and Croll, D., 2023. The population genetics of adaptation through copy number variation in a fungal plant pathogen. *Molecular Ecology*, **32**: 2443-2460.

Strelkov, S.E., Lamari, L. and Ballance, G.M., 1999. Characterization of a host-specific protein toxin (Ptr ToxB) from *Pyrenophora tritici-repentis*. *Molecular Plant-Microbe Interactions*, **12**: 728-732.

Strelkov, S.E. and Lamari, L., 2003. Host–parasite interactions in tan spot [*Pyrenophora tritici-repentis*] of wheat. *Canadian Journal of Plant Pathology*, **25**: 339-349.

Strelkov, S.E., Kowatsch, R.F., Ballance, G.M. and Lamari, L., 2005. Characterization of the *ToxB* gene from North African and Canadian isolates of *Pyrenophora tritici-repentis*. *Physiological and Molecular Plant Pathology*, **67**: 164-170.

Stukenbrock, E.H. and McDonald, B.A., 2008. The origins of plant pathogens in agro-ecosystems. *Annual Review Phytopathology*, **46**: 75-100.

Sultana T., Zamborlini A., Cristofari G. and Lesage P., 2017. Integration site selection by retroviruses and transposable elements in eukaryotes. *Nature Reviews Genetics*, **18**: 292–308.

Tai, Y.S., Bragg, J. and Meinhardt, S.W., 2007. Functional characterization of ToxA and molecular identification of its intracellular targeting protein in wheat. *American Journal of Plant Physiology*, **2:** 76–89.

Tanaka S (1933) Studies on black spot disease of the Japanese pear (*Pyrus serotina Rehd.*). *Mem Coll Agric Kyoto Imp Univ,* **28**: 1–31.

Torres, D.E., Oggenfuss, U., Croll, D. and Seidl, M.F., 2020. Genome evolution in fungal plant pathogens: looking beyond the two-speed genome model. *Fungal Biology Reviews*, **34**: 136-143.

Tralamazza, S.M., Gluck-Thaler, E., Feurtey, A. and Croll, D., 2024. Copy number variation introduced by a massive mobile element facilitates global thermal adaptation in a fungal wheat pathogen. *Nature Communications*, **15**: 5728.

Tsuge, T., Harimoto, Y., Akimitsu, K., Ohtani, K., Kodama, M., Akagi, Y., Egusa, M., Yamamoto, M. and Otani, H., 2013. Host-selective toxins produced by the plant pathogenic fungus *Alternaria alternata*. *FEMS Microbiology Reviews*, **37**: 44-66.

Tuori RP, Wolpert TJ, Ciuffetti LM. 2000. Heterologous expression of functional Ptr *ToxA*. *Molecular Plant-Microbe Interactions,* **13**: 456–464.

Turgeon, B.G., 1998. Application of mating type gene technology to problems in fungal biology. *Annual Review of Phytopathology*, **36**: 115-137.

Vande Zande, P., Zhou, X. and Selmecki, A., 2023. The dynamic fungal genome: polyploidy, aneuploidy and copy number variation in response to stress. *Annual Review of Microbiology*, **77**: 341-361.

Virdi, S.K., Liu, Z., Overlander, M.E., Zhang, Z., Xu, S.S., Friesen, T.L. and Faris, J.D., 2016. New insights into the roles of host gene-necrotrophic effector interactions in governing susceptibility of durum wheat to tan spot and Septoria nodorum Blotch. *G3: Genes, Genomes, Genetics*, **6**: 4139-50.

Walton, J.D., Earle, E.D. and Gibson, B.W., 1982. Purification and structure of the host-specific toxin from *Helminthosporium carbonum* race 1. *Biochemical and Biophysical Research Communications*, **107**: 785-794.

Walton, J.D. and Panaccione, D.G., 1993. Host-selective toxins and disease specificity: perspectives and progress. *Annual Review of Phytopathology*, **31**: 275-303.

Walton, J.D., 2000. Horizontal gene transfer and the evolution of secondary metabolite gene clusters in fungi: an hypothesis. *Fungal Genetics and Biology*, **30**: 161-171.

Wang, C., Milgate, A.W., Solomon, P.S. and McDonald, M.C., 2021. The identification of a transposon affecting the asexual reproduction of the wheat pathogen *Zymoseptoria tritici*. *Molecular Plant Pathology*, **22**: 800-816.

Wang, H., Huang, R., Ren, J., Tang, L., Huang, S., Chen, X., Fan, J., Li, B., Wang, Q., Hsiang, T. and Liu, H., 2023. The evolution of mini-chromosomes in the fungal genus *Colletotrichum*. *MBio*, **14**: e00629-23.

Ward, D.E., Vázquez, A. and Pedras, M.S.C., 1999. Probing host-selective phytotoxicity: synthesis and biological activity of phomalide, isophomalide, and dihydrophomalide. *The Journal of Organic Chemistry*, **64**: 1657-1666.

Ward, T.J., Bielawski, J.P., Kistler, H.C., Sullivan, E. and O'Donnell, K., 2002. Ancestral polymorphism and adaptive evolution in the trichothecene mycotoxin gene cluster of phytopathogenic *Fusarium*. *Proceedings of the National Academy of Sciences*, **99**: 9278-9283.

Wei, B., Despins, T., Fernandez, M.R., Strelkov, S.E., Ruan, Y., Graf, R. and Aboukhaddour, R., 2021. Race distribution of *Pyrenophora tritici-repentis* in relation to ploidy level and susceptibility of durum and winter bread wheat. *Canadian Journal of Plant Pathology*, **43**: 582-598.

Whisson, S.C., Vetukuri, R.R., Avrova, A.O. and Dixelius, C., 2012. Can silencing of transposons contribute to variation in effector gene expression in *Phytophthora infestans*?. *Mobile Genetic Elements*, **2**: 110-114.

Wolf, E.D., Effertz, R.J., Ali, S. and Francl, L.J., 1998. Vistas of tan spot research. *Canadia Journal of Plant Pathology*, **20**: 349-370.

Wolf, Y.I. and Koonin, E.V., 2013. Genome reduction as the dominant mode of evolution. *Bioessays*, **35**: 829-837.

Wolpert, T.J., Macko, V., Acklin, W., Jaun, B., Seibl, J., Meili, J. and Arigoni, D., 1985. Structure of victorin C, the major host-selective toxin from *Cochliobolus victoriae*. *Experientia*, **41**: 1524-1529.

Wolpert, T.J., Dunkle, L.D. and Ciuffetti, L.M., 2002. Host-selective toxins and avirulence determinants: what's in a name?. *Annual Review of Phytopathology*, **40**: 251-285.

Wöstemeyer, J. and Kreibich, A., 2002. Repetitive DNA elements in fungi (Mycota): impact on genomic architecture and evolution. *Current Genetics*, **41**: 189-198.

Wu, J., Kou, Y., Bao, J., Li, Y., Tang, M., Zhu, X., Ponaya, A., Xiao, G., Li, J., Li, C. and Song, M.Y., 2015. Comparative genomics identifies the *Magnaporthe oryzae* avirulence effector AvrPi9 that triggers Pi9-mediated blast resistance in rice. *New Phytologist*, **206**: 1463-1475.

Yang, H., Shi, X., Chen, C., Hou, J., Ji, T., Cheng, J. and Birchler, J.A., 2022. Genomic imbalance modulates transposable element expression in maize. *Plant Communications*, **4**: 100467

Yu, J., Chang, P.K., Ehrlich, K.C., Cary, J.W., Bhatnagar, D., Cleveland, T.E., Payne, G.A., Linz, J.E., Woloshuk, C.P. and Bennett, J.W., 2004. Clustered pathway genes in aflatoxin biosynthesis. *Applied and Environmental Microbiology*, **70**: 1253-1262.

Zhang, J., 2003. Evolution by gene duplication: an update. *Trends in Ecology & Evolution*, **18**: 292-298.

Zhang, D., Burroughs, A.M., Vidal, N.D., Iyer, L.M. and Aravind, L., 2016. Transposons to toxins: the provenance, architecture and diversification of a widespread class of eukaryotic effectors. *Nucleic Acids Research*, **44**: 3513-3533.

CHAPTER 2: THE PANGENOME OF THE WHEAT PATHOGEN *PYRENOPHORA TRITICI-REPENTIS* REVEALS NOVEL TRANSPOSONS ASSOCIATED WITH NECROTROPHIC EFFECTORS *TOXA* AND *TOXB*

2.0 ABSTRACT

In fungal plant pathogens, genome rearrangements followed by selection pressure for adaptive traits have facilitated the co-evolutionary arms race between hosts and their pathogens. *Pyrenophora tritici-repentis* (Ptr) has emerged recently as a foliar pathogen of wheat worldwide and its populations consist of isolates that vary in their ability to produce combinations of different necrotrophic effectors. These effectors play vital roles in disease development. Here, we sequenced the genomes of a global collection (40 isolates) of Ptr to gain insights into its gene content and genome rearrangements.

A comparative genome analysis revealed an open pangenome, with an abundance of accessory genes (~ 57%) reflecting Ptr's adaptability. 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 effector coding genes, were detailed using long-read assemblies (PacBio RS II) generated in this work in addition to previously assembled genomes. We also discovered the involvement of large mobile elements associated with Ptr's effectors: *ToxA*, the gene encoding for the necrosis effector, was found as a single copy within a 143-kb Starship transposon (dubbed *Horizon*) with a clearly defined target site and target site duplications. *Horizon* was located on different chromosomes in different isolates, indicating mobility, and the previously described *ToxhAT* transposon (responsible for horizontal transfer of *ToxA*) was nested within this newly identified Starship. Additionally, *ToxB*, the gene encoding the chlorosis effector, was clustered as three copies

on a 294-kb element, which is likely a different putative Starship (dubbed *Icarus*) in a ToxB-producing isolate. *ToxB* and its putative transposon were missing from the *ToxB* non-coding reference isolate, but the homolog *toxb* and 'Icarus' were both present in a different non-coding isolate. This suggests that *ToxB* may have been mobile at some point during the evolution of the Ptr genome which is contradictory to the current assumption of *ToxB* vertical inheritance. Finally, the genome architecture of Ptr was defined as 'one-compartment' based on calculated gene distances and evolutionary rates.

These findings together reflect on the highly plastic nature of the Ptr genome which has likely helped to drive its worldwide adaptation and has illuminated the involvement of giant transposons in facilitating the evolution of virulence in Ptr.

2.1 BACKGROUND

Pyrenophora tritici-repentis (Ptr) is an ascomycete fungus that causes tan spot, one of the most destructive foliar diseases of wheat worldwide (Savary et al., 2019; Aboukhaddour et al., 2021). Tan spot was not considered a significant disease of wheat until the 1970s, and from an evolutionary perspective, this recent emergence offers an intriguing case to gain insights into how a weak pathogen rose to a global threat in a short time span (Aboukhaddour et al., 2021). The Ptr genome is a mosaic of present and absent effectors and hence serves as a model for examining the evolutionary processes behind the acquisition of virulence in necrotrophs and how disease emergence develops.

In a fungal species, the genomic variation in contents and structural organization reflects its evolution and the selection pressures driving its diversification and speciation. Genome rearrangements, such as chromosomal polymorphisms, duplications, deletions,

inversions, and fusions are very common in fungi. These rearrangements in plant pathogens can facilitate the selection for adaptive traits that enhance pathogenic abilities on a certain host in a certain niche, as the genome plasticity often will lead to the gain or loss of functions in the co-evolution arms race between the pathogen and its host. In this study, we took advantage of a diverse collection of Ptr isolates representing all pathotypes, from diverse global origins, and applied short- and long-read sequencing technologies to dissect the Ptr genome. The aim is to gain a better understanding of Ptr genome plasticity, to answer how a fungal pathogen like Ptr has evolved to have a mosaic of several effectors in different geographical regions, to explore the role of transposons in genome expansion or compartmentalization, and to define at the genome level, in a meticulous way, the link between virulence evolution and transposons. Ultimately, to release useful genomic resources of well-phenotyped isolates to the research community to be utilized to advance our knowledge on disease emergence.

Ptr has served as a model species for necrotrophic plant pathogens, as it produces various necrotrophic effectors, previously known as host-selective toxins that play essential roles in disease development (Aboukhaddour et al., 2021). 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 (Lamari et al., 2003). Additional effectors may be involved in pathogenicity and await characterization (Kamel et al., 2019). ToxA and ToxB are ribosomally synthesized proteins, and while ToxA is encoded by a single-copy *ToxA* gene, ToxB is encoded by multi-copy *ToxB* genes (Lamari et al., 2003). ToxC is a putative secondary metabolite, and identification of the ToxC-coding gene(s) and its synthesis awaits further characterization (Effertz et

al., 2002; Shi et al., 2022). 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 (Lamari et al., 2003). *ToxA* has been found in Ptr and in other related species, whereas *ToxB* has been reported only in Ptr, but its inactive homologs, termed *toxb*, were identified in Ptr and other closely related and distantly related fungal species (Strelkov et al., 2005; Friesen et al., 2006; Andrie et al., 2008; Lu et al., 2015).

Ptr isolates are designated into eight different races (races 1 to 8) based on which combinations of these three effectors they secrete (Lamari et al., 2003). There is an unexplained correlation between Ptr ability to secrete various combinations of effectors and its geographical origins, and in North America and Australia, Ptr isolates are mainly ToxA and ToxC producers, with ToxB producers being extremely rare or absent (Aboukhaddour et al., 2013; Kamel et al., 2019), whereas, 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 (Kamel et al., 2019; Aboukhaddour et al., 2021), with ToxB producers predominating North Africa (Aboukhaddour et al., 2009; Aboukhaddour et al., 2011; Kamel et al., 2019). This variation in the presence/absence of effectors suggests a divergent evolution of Ptr and its effectors worldwide.

An independent origin of these effectors was suggested previously (Friesen et al., 2006; Andrie et al., 2008; Aboukhaddour et al., 2009; McDonald et al., 2019). In Ptr, *ToxA* and *ToxB* were never localized to the same chromosome (Aboukhaddour et al., 2009). Yet, *ToxA* was found to reside on a homolog chromosome of essential nature, meaning a chromosome is present in all isolates regardless if the isolate possesses *ToxA*

or not. An exception was reported in a race 8 isolate, I-73–1, collected from the Fertile Crescent in Syria. In this isolate, *ToxA* was present on a non-homolog and larger chromosome, which indicates a possible translocation of *ToxA* in the Ptr species (Aboukhaddour et al., 2009). The ability of *ToxA* to transfer horizontally between species and into Ptr has been demonstrated before, and in addition to Ptr, *ToxA* was found in two related species, *Parastagonospora nodorum* and *Bipolaris sorokiniana* (Friesen et al., 2006; McDonald et al., 2019). *ToxA* was horizontally transferred via a 14-kb type II DNA transposon named *ToxhAT*, and multiple horizontal-gene-transfer events including regions surrounding the *ToxhAT* were likely involved in these species (McDonald et al., 2019). *ToxhAT* appears to still be active in *B. sorokiniana* but has likely decayed in *P. nodorum* and may no longer be functional (McDonald et al., 2019). Unlike the horizontal transfer of *ToxA*, the *ToxB* gene is believed to be vertically inherited from ancestral species. This was indicated by the presence of *ToxB* homologs or *ToxB*-like genes in multiple fungal orders (Ploesporales, Dothideomycetes, and Sordariomycetes) (Ciuffetti et al., 2014). In this study, we performed a detailed analysis to dissect the regions surrounding *ToxA* and *ToxB* in Ptr.

In addition to the mobility of the effector genes and the surrounding chromosomal segments, the Ptr genome as a whole has been described as highly plastic, with evidence of major chromosomal rearrangements, including insertions, deletions, inversions, and translocations (Aboukhaddour et al., 2009; Manning et al., 2013; Moolhuijzen et al., 2018; McDonald et al., 2018; 2019). 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 USA, and was assembled to the chromosome level with the aid of optical

mapping (Manning et al., 2013). In addition, the full genomes of 11 Ptr isolates from North America and Australia have also been sequenced and are now publicly available (Manning et al., 2013; Moolhuijzen et al., 2018; 2019). Of those, three (M4, V1, and DW5) were assembled from long-read sequences. The role of DNA transposons and long terminal repeat (LTR) retrotransposons in genome-wide segmental duplications, insertions, and chromosomal fusions in Ptr was investigated (Ciuffetti et al., 2014). 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 (Moolhuijzen et al., 2018). RIP is often associated with the "two-speed genome" model in a number of fungal pathogens; this model describes a pattern of physical genome compartmentation with repeat-rich gene-sparse regions and repeat-sparse gene-dense regions with different evolutionary rates (Raffaele et al., 2010; Dong et al., 2015; Torres et al., 2020). 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 atypical race that induces necrosis but lacks the *ToxA* gene (Kamel et al., 2019). In addition to the pangenome, we provide a detailed look at gene contents, phylogenetic relatedness, and effector gene movement. Furthermore, the polished long-read assemblies representing the full genome of two isolates (I-73–1 a race 8 isolate from Syria and produces all three effectors, and therefore known as the most complex race so far in Ptr, and D308, a Canadian race 3 isolate that produces only the ToxC, and carries inactive homolog of *ToxB*) were explored. The structural organization and genome compartmentation in these two isolates were analysed

and in comparison with three previously published genomes (BFP and DW-5 from USA and M4 from Australia).

2.2 RESULTS

2.2.1 HYBRID ASSEMBLIES CAPTURE A BETTER ESTIMATE OF GENOME SIZE

The de novo assemblies of Illumina short-reads indicated a variable genome size ranging from 34.1 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 (Kamel et al., 2019). 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 HiSeq 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, likely due to the collapsing of repetitive regions and duplicated transposons. Summary statistics for all assemblies are available in Table 2.1.

Table 2.1 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

2.2.2 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 amino acid sequence similarity with a similarity threshold of ≥ 90% ($E10^{-4}$) and local synteny (conserved gene neighbourhood scores) via Pangloss (McCarthy et al., 2019), we were left with 23,454 groups of unique homologous genes. It must be noted that Pangloss will not group more than one gene per isolate into a gene cluster, and therefore, the following statistics may include genes with multiple copies. In total, 10,159 genes were identified as core (conserved across all isolates) (43% of total pangenome) (Figure 2.1a). The remaining 13,295 gene models (~ 57% of the total pangenome) were the accessory genome, with individual isolates carrying 2661 to 3424 accessory genes. Within the accessory genome were 7330 singletons (genes present in only one isolate) (55% of the accessory genome; 31% of the total pangenome) which were parsed via a custom script. Within the singletons, a large number 3293 (44%) were present only in isolates from the outgroup of the phylogeny (Figure 2.1a), including 90–2 (435 genes), 92-171R5 (1598 genes), and G9-4 (1204 genes). Additionally, isolate T128-1, which exhibits a novel necrotic phenotype (Kamel et al., 2019), contained 54 unique genes. Post-Pangloss BLAST searches of singletons suggest a small number of singletons (~ 6%) are present in multiple copies (90% identity over 80% of query length) which suggests the number of singletons is over-estimated.

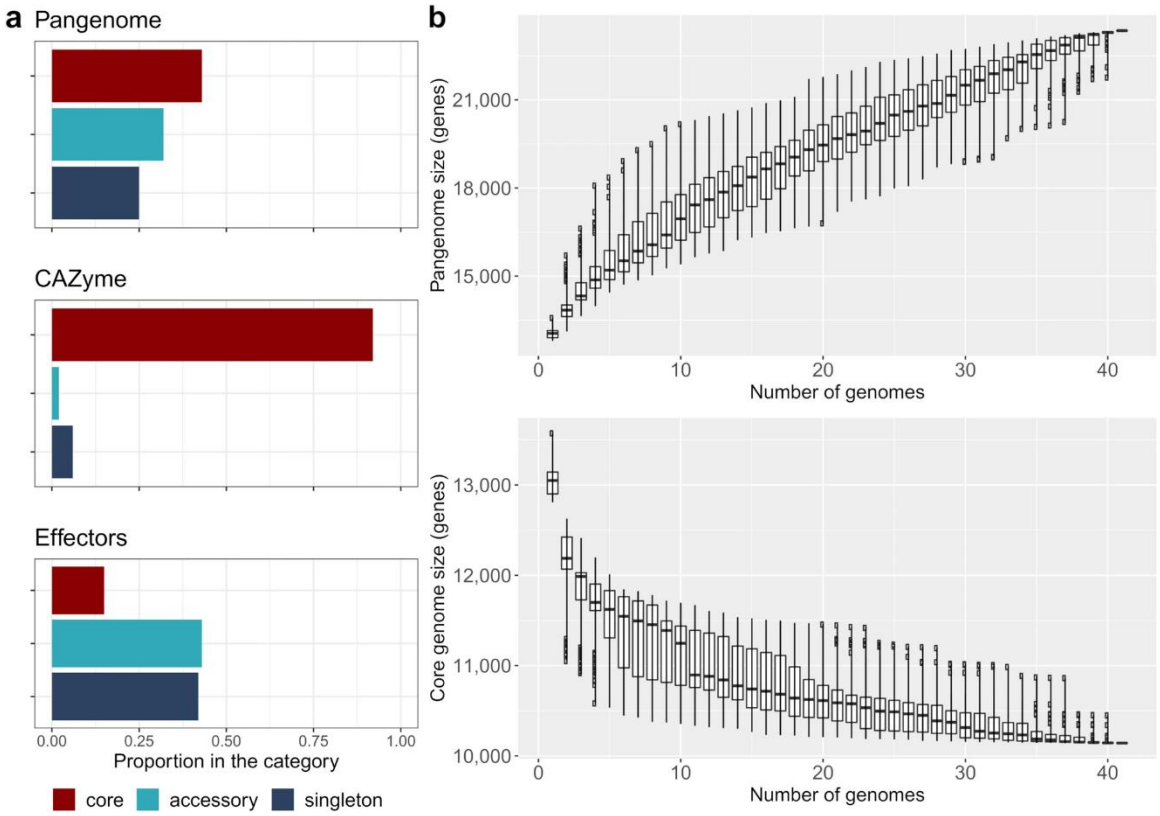


Figure 2.1 The pangenome of *Pyrenophora tritici-repentis*. **a** Proportion of genes present in the core, accessory (excluding singletons), and singleton sets for pangenome, CAZymes, and effectors. **b** Total number of unique genes and core genes within the constructed pangenome of *Pyrenophora tritici-repentis* as genomes are added. Starting genome was always Pt-1C-BFP, with subsequent genomes added randomly, iterated × 100, and visualized as boxplots.

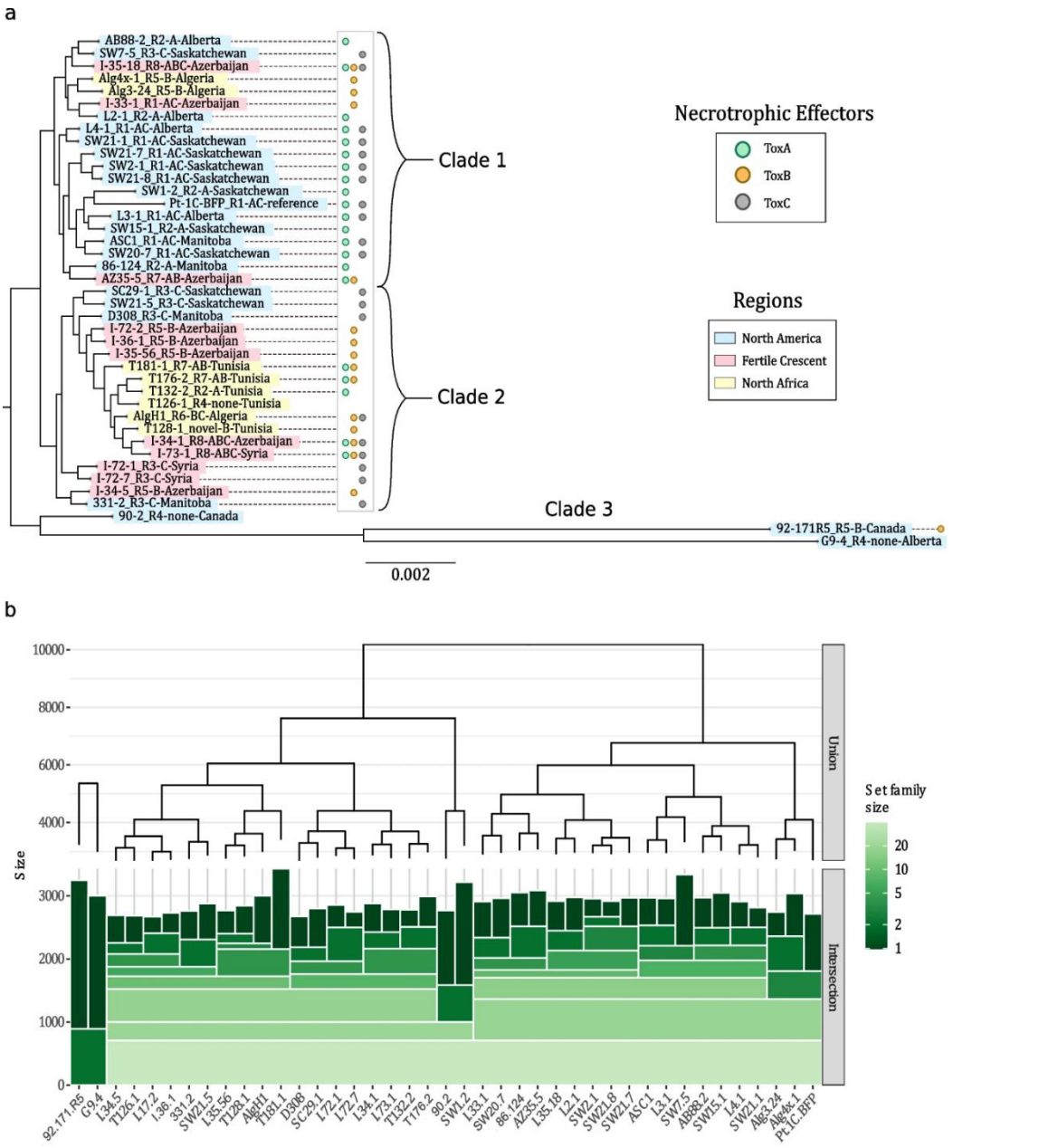


Figure 2.2 Evolutionary relationships of isolates from the constructed pangenome of *Pyrenophora tritici-repentis*. **a** Maximum-likelihood phylogenetic tree of *Pyrenophora tritici-repentis* based on aligned and concatenated proteins present in the core genome (total 10,159 genes). Isolate names are 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* (green 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 genes, set family size represents the number of genes shared by isolates in each group, where inclusion in a group is based on any given isolate overlapping with the horizontal bars.

Genes were binned into sets based on the presence or absence of genes to provide insights into gene gains and losses in Ptr (Additional file 2.1). Each set is indicative of how many isolates share a particular gene (e.g. set 2 shows genes shared by two isolates, set 3 by three isolates, etc.). Sets 1 and 41 were not included, as these represent the singleton and core gene set, respectively. The result of this sub-setting is groups of genes (based on 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 2987 genes) compared with isolates lacking *ToxA* (mean of 2820 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 3162 and 2779 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. It is worth noting as well, that when the three outlier isolates (90–2, 92-171R5, and G9-4) are removed from the analysis, the genome curve is similar (not shown) and still indicates Ptr possesses an open pangenome.

2.2.3 PTR CORE PROTEIN AND SNP PHYLOGENIS 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 clades supported by monophyletic branches (Figure 2.2a). Clade 1 contained 20 isolates, most of which were ToxA-producing Canadian isolates classified as races 1 and 2. Few isolates in clade 1 were ToxB producers; those that were ToxB producers were from Azerbaijan and Algeria. Clade 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 clade 2 were all classified as race 3 (ToxC producers). Three isolates (92-171R5, 90–2, and G9-4) clustered as an outgroup defining clade 3. 92-171R5 is the only race 5 isolate identified in Canada and was described as weakly virulent (Strelkov et al., 2002). Isolates 90–2 and G9-4 were classified as non-pathogenic race 4 (Aboukhaddour et al., 2011; Wei et al., 2021).

Heirarchical set groupings based on accessory proteins revealed a similar structure as the core protein phylogeny (Figure 2.2b), with the only exceptions being in which major clades 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 (horizontal bars) were defined by the presence or absence of accessory genes and do not take into account any sequence variation in the homologs used to construct the core protein phylogeny, but highlight that Ptr isolates in this study were mainly distributed in two clades: one is

ToxA-producing isolates, with few exceptions, and a second contains mainly ToxA non-producing isolates of races 3 and 5. These race 3 and 5 isolates were mainly isolated from durum (tetraploid) wheat, whereas the ToxA producers were mainly collected from bread wheat (hexaploid). This may reflect on the genetic relatedness accumulated in relation to the host ploidy adaptation. Additional phylogenetic trees based on SNP data also closely match the core protein phylogeny and the hierarchical set trees in terms of branching, with the divergent isolates separating even further (Additional file 2.2).

2.2.4 PREDICTED GENES AND EFFECTORS

Of the 23,454 gene clusters in the Ptr pangenome, only 46% were present in the Pfam functional database. Of the 10,159 core genes, 6951 (69%) had some described functionality (e.g. domain, family, etc.). However, of the 13,295 accessory genes, only 3782 (29%) were annotated with a functional domain, 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. Singleton proteins were marginally smaller than other accessory proteins, with a mean length of 232 aa (median 156 aa). A small number of proteins (Kelley et al., 2015) were quite large, with lengths exceeding 2000 aa, and six of those exceeding 5000 aa; the largest protein was 9750 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 2.1a). 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.

2.2.5 CAZYMES ARE HIGHLY CONSERVED IN PTR

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 was predominantly conserved across all isolates (Figure 2.1a). 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 (Additional file 2.3). These included GH5 cellulases and endoglucanases, GH43 arabinases and xylosidases, AA9 lytic cellulose monooxygenases, CE1 hemicellulases, and CE5 cutinases (Davies et al., 2000; Lombard et al., 2013; Zhao et al., 2013; Nakamura et al., 2017; Jagadeeswaran et al., 2021; Rafiei et al., 2021). 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 remodelling and glycoprotein maturation.

2.2.6 EXTENSIVE CHROMOSOMAL STRUCTURAL REORGANIZATION IN THE PTR GENOME

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 appear to be co-linear between all three isolates (contigs 12 and 9 in I-73–1 and contigs 5, 11, and 14 in D308) (Figure 2.3). 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 2.3); the only notable exception was the *ToxA* containing region detailed further below (Figure 2.4a).

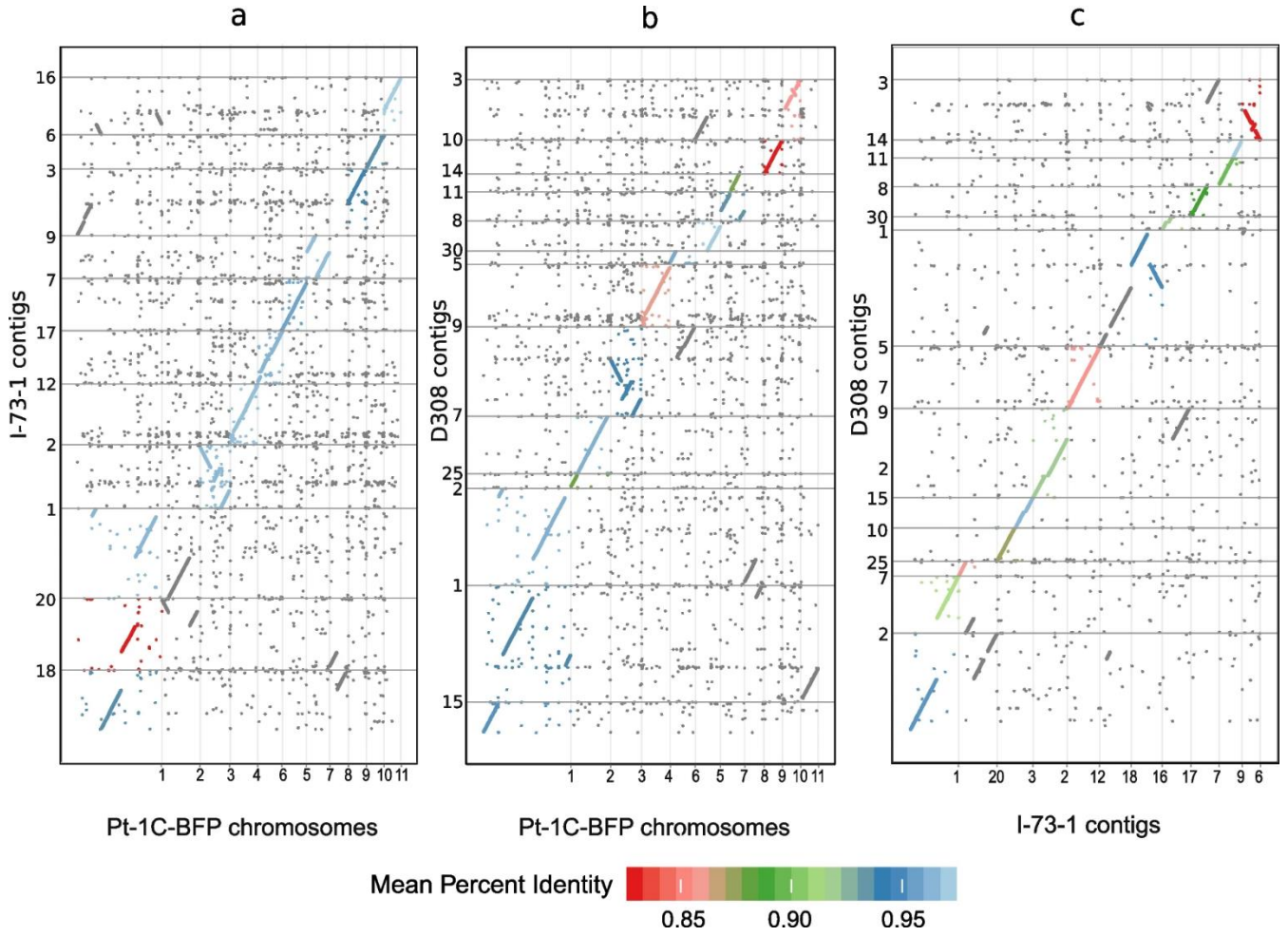


Figure 2.3 Full genome alignments of newly sequenced long-read assemblies of *Pyrenophora tritici-repentis* isolates I-73–1 (race 8) and D308 (race 3) to the reference isolate Pt-1C-BFP (race 1). **a** Pt-1C-BFP and I-73–1; **b** Pt-1C-BFP and D308; **c** I-73-1 and D308.

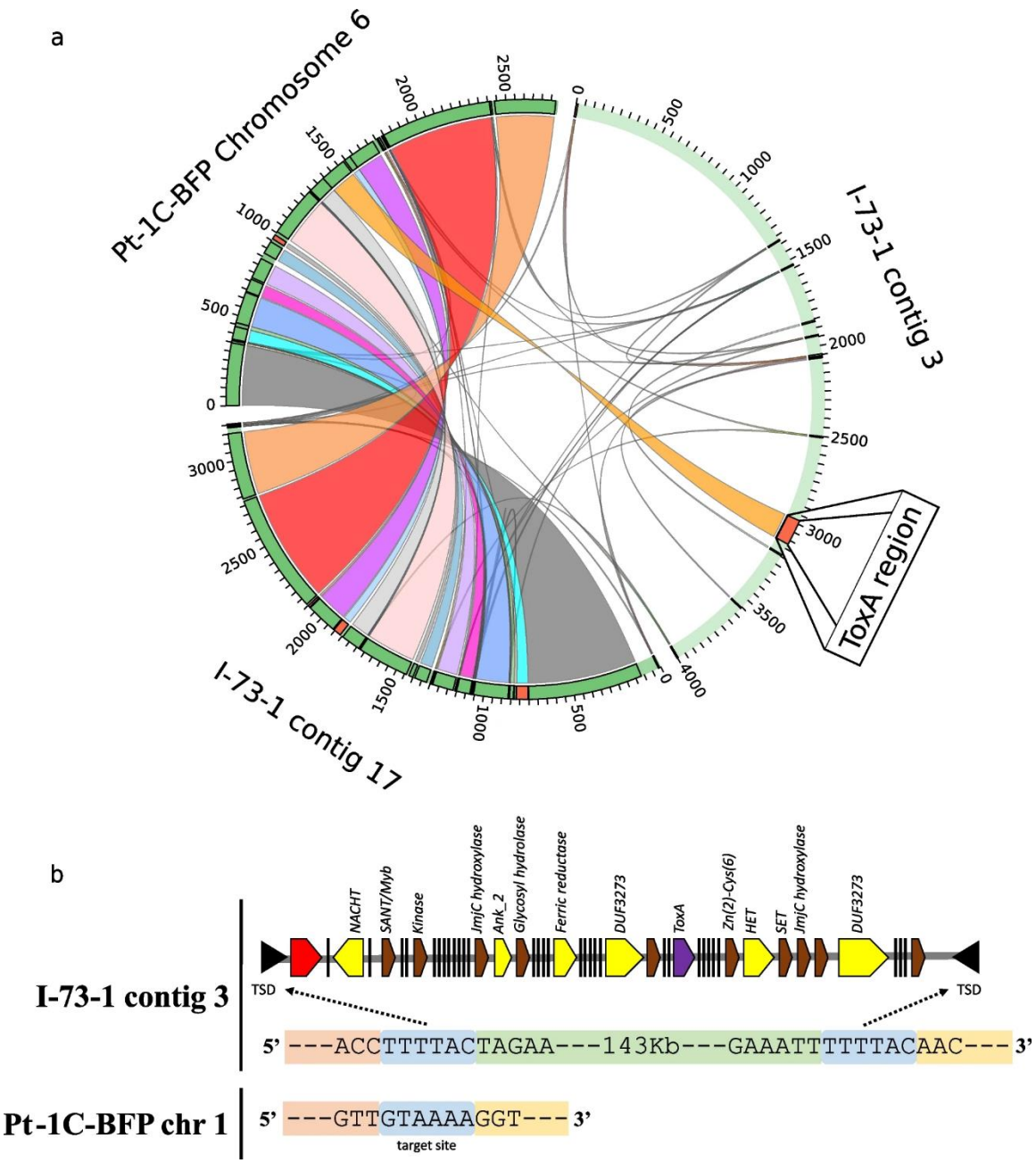


Figure 2.4 Evidence of a likely translocation of the *ToxA* containing region between two races of *Pyrenophora tritici-repentis* facilitated Starship transposon *Horizon*. **a** The *ToxA* containing chromosome 6 of isolate Pt-1C-BFP (race 1) is compared with the *ToxA* containing contig 3 and contig 17 of isolate I-73–1 (race 8). Pt-1C-BFP chromosome 6 aligns almost completely to I-73–1 contig 17, with the exception of the *ToxA* containing region, which aligns to contig 3. The majority of contig 3 aligns to either Pt-1C-BFP chromosome 1 or chromosome 9 (Figure 3). **b** Schematic of the *Horizon* Starship (red gene = tyrosine recombinase; yellow = known Starship cargo genes; purple = *ToxA* gene; brown = annotated gene; black bar = gene with no known function).

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 2.3).

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 2.3). 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.

2.2.7 *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 2.4a). This validates the previously reported translocation of *ToxA* in I-73–1 (Aboukhaddour et al., 2009). 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 (Wicker et al., 2007). Recently, however, a new class of large mobile elements called Starships have been defined (Gluck-Thaler et al., 2021). An examination of other genes present within the transposon suggests that this element is also 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 2.4b) (Gluck-Thaler et al., 2021). A full list of genes present is available in Additional file 2.4. The target site itself was identifiable in BFP chr1 (Figure 2.4b), 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* (McDonald et al., 2019), showed that this smaller 14-kb transposon 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 2.5a). Examination of the edges of this gap revealed the presence of terminal inverted repeats (TIRS) in I-73–1 and a potential target site duplication in M4 (Figure 2.5b). This same 294-kb transposon with the same TIRs 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 (Figure 2.5b and Additional file 2.5a). While the DUF3435 tyrosine recombinase 'captain' was absent in both I-73–1 and D308, the heterokaryon incompatibility gene was present in I-73–1 while DUF3273 and patatlin-like phospholipase genes were present in D308. Additionally, DDE endonucleases (D:aspartic acid; E:glutamic acid) commonly present in class II DNA transposases were present in both isolates along with many other genes associated with class I RNA intermediate transposons (Figure 2.5b). These findings taken together suggest this region may be either disabled or a type of Starship which makes use of something other than a tyrosine recombinase. We have called this putative Starship *Icarus*.

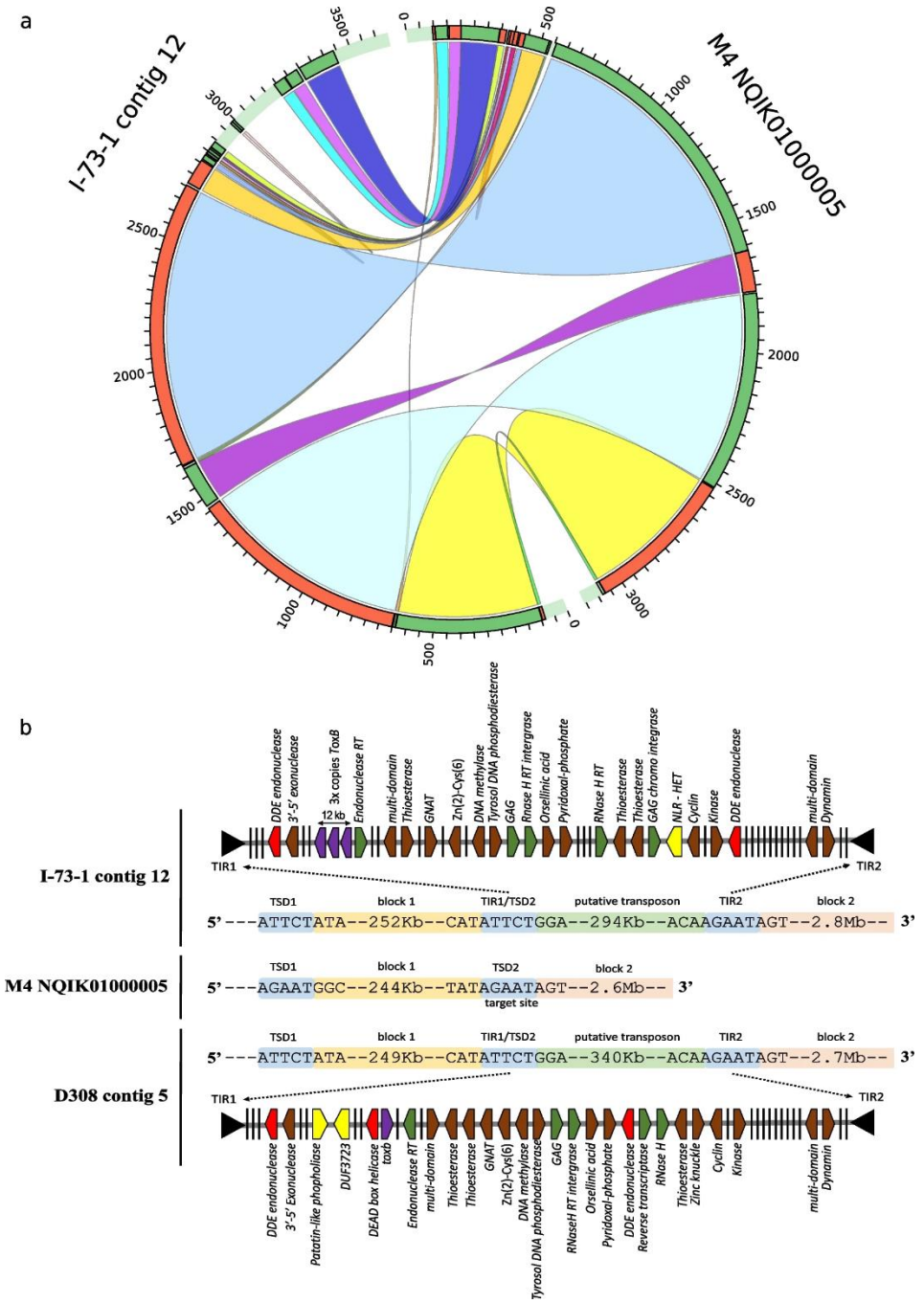


Figure 2.5 Evidence of putative Starship transposon *Icarus* associated with *ToxB* in *Pyrenophora tritici-repentis* **a** Circular alignment of contig 12 from race 8 isolates I-73–1 and contig NQIK01000005 from race 1 isolates M4. A large 294-kb region which contains three copies of the *ToxB* (black arrow) is visible, and 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. **b** Schematic of *Icarus* Starship (red genes = associated with type II DNA transposons; green = associated with type I RNA transposons; yellow = known Starship cargo genes; purple = *ToxB/toxb*; brown = annotated gene; black bar = gene with no known function).

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 and M4). DW5 was reported to carry 10 copies of *ToxB* (Moolhuijzen et al., 2020). 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 (Additional file 2.5b). The locations of *ToxB* copies in DW-5 do not appear to have co-linearity 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 (Additional file 2.6).

2.2.8 PTR EXHIBITS A 'ONE-COMPARTMENT' GENOME

The genome of many plant pathogens exhibits a two-compartment arrangement, one with AT-rich gene-sparse regions (GSR) where effectors and TEs are embedded, and a second with gene-dense GC-rich regions (GDR) (Raffaele et al., 2010; Dong et al., 2015; Torres et al., 2020). 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 ~ 4600 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 2.6). 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. Genes with the highest and lowest intergenic distances (90th percentiles; labelled GSR set and GDR set respectively; Additional file 2.7) were selected for evolutionary rate analysis by comparison of non-synonymous (dN) and synonymous (dS) substitutions. The I-73–1 GDR set contained 163 genes and the GSR set contained 134 genes, and of these, the majority (> 88% for each set) had dN/dS ratios < 1 indicating negative selection pressure. Comparison of dN/dS means from the GDR and GSR sets revealed no difference in evolutionary rates between the sets ($p = 0.6207$) (Figure 2.6). Similar results were found with D308 (Additional file 2.8). Additionally, analysis with RIPper showed that < 2% of the I-73–1 and D308 genomes were affected by RIP, with < 40 large RIP affected regions (LRARs) present in each.

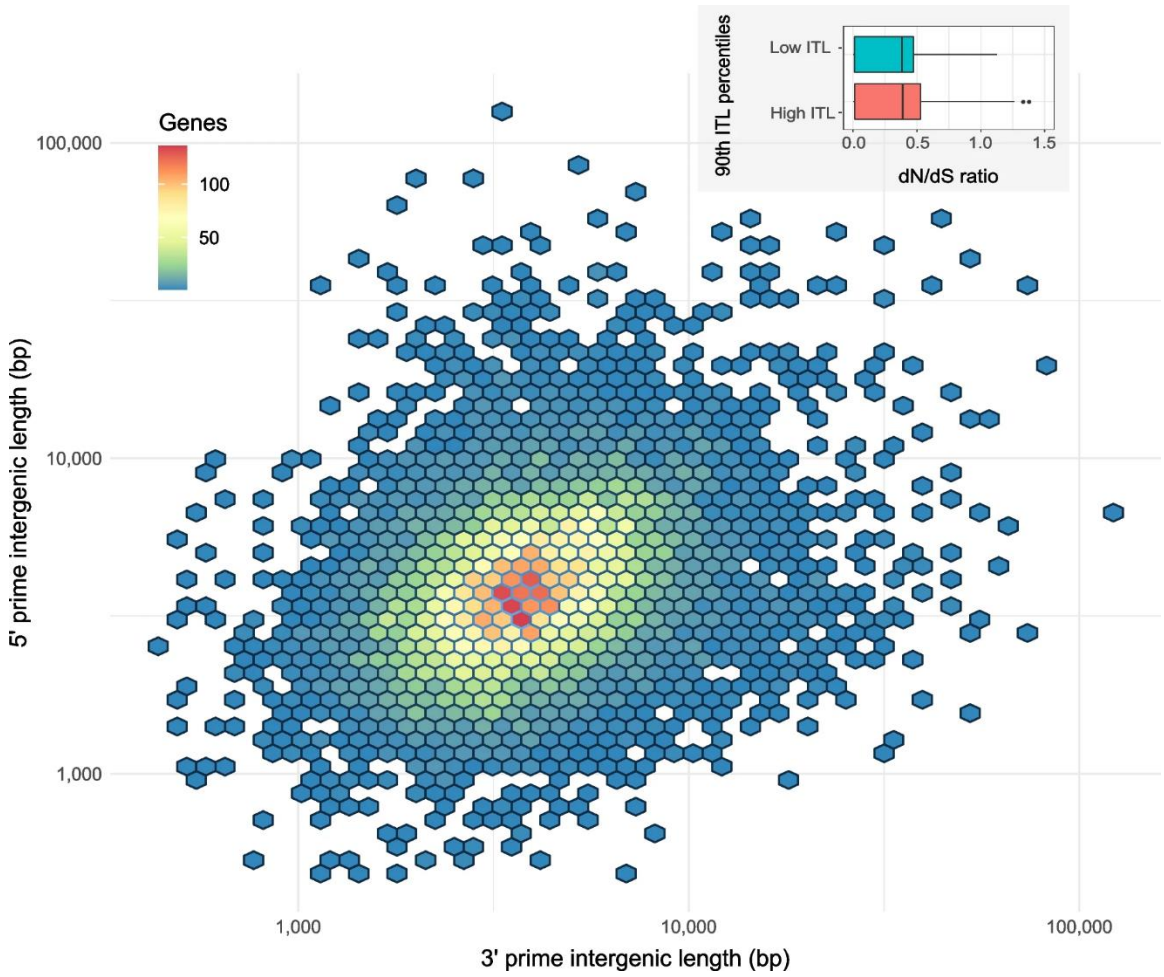


Figure 2.6 Intergenic distances of all genes in *Pyrenophora tritici-repentis* isolate I-73–1. The 3′ intergenic length (*x*-axis) is the distance (bp) from the 3′ end of the current gene to the 5′ end of next, and the 5′ intergenic length (*y*-axis) is the distance from the 3′ end of the previous gene to the 5′ end of the current gene. Inset (top right) shows dN/dS ratios for genes from the 90th ITL percentiles (i.e. genes furthest from others and genes nearest to others; outliers omitted for clarity.

2.2.9 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 2.7a). 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 was primarily

class I-LTRs. The high TE content in these two isolates also correlates with larger genome sizes relative to the other assemblies (Figure 2.7b). 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.

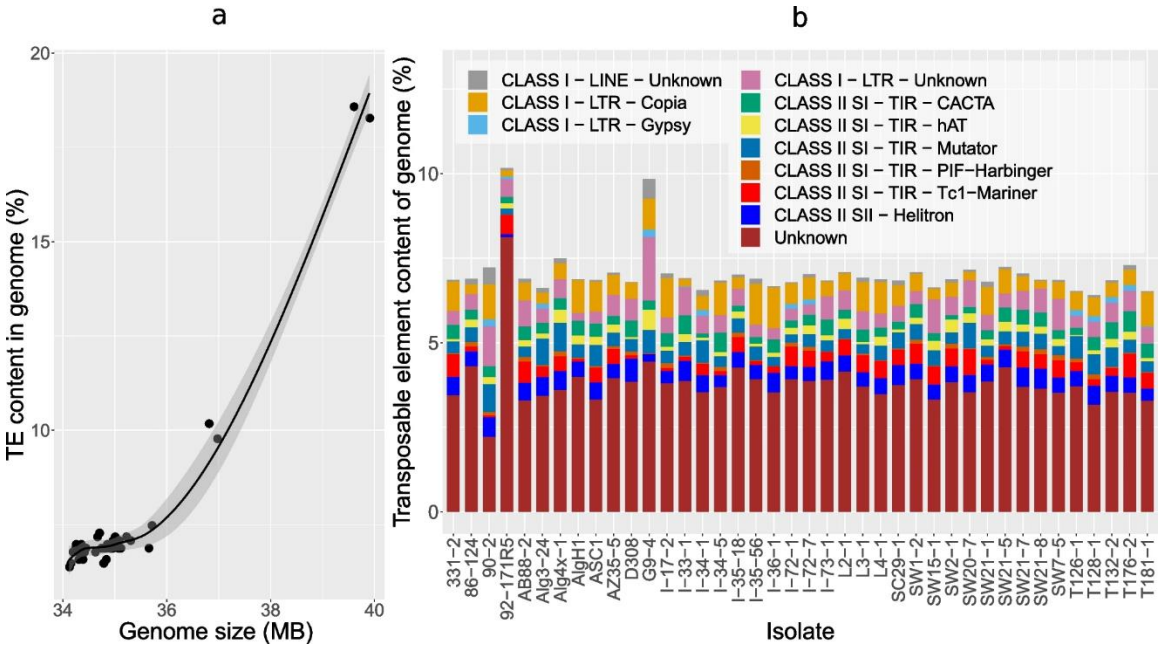


Figure 2.7 Transposable element content and genome size variation of *Pyrenophora tritici-repentis* from short-read assemblies. **a** Contribution of TEs (%) to total genome size across 40 isolates. **b** Contribution of each TE class to genome size for individual isolates.

The long-read assemblies of isolates I-73–1 and D308 showed a significantly larger number of transposable elements (> 150%) when compared to short-read assemblies of the same isolates. 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.

2.3 DISCUSSION

In this study, we performed a global pangenome analysis of Ptr, an important foliar pathogen of wheat. In total, 40 newly sequenced genomes were added and 41 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 the 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 (Aboukhaddour et al., 2009). 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 (Vanheule et al., 2016). Recently, more reports have emerged on the ability of virulence genes to transfer on large transposons among plant pathogens (Gluck-Thaler et al., 2021). 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 (Torres et al., 2020), 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.

2.3.1 THE MOBILITY 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* (Friesen et al., 2006; McDonald et al., 2019). The independent horizontal transfers between these species have been explored in detail, with the transfer being facilitated via a hAT transposon dubbed *ToxhAT*, which is 14 kb in size (McDonald et al., 2019). 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 (Aboukhaddour et al., 2009). Although *ToxhAT* is present, we validate the location of *ToxA* in I-

73–1 within *ToxhAT*, but also 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; 4 Mb) 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) (Hafez et al., 2022) (data not shown), and this supports the hypothesis of its recent integration into the Ptr species (Friesen et al., 2005; 2006). This new class of transposable elements, *Starships*, are extremely large cargo-carrying elements associated with the movement of diverse gene sets such as biosynthetic clusters, virulence factors (e.g. *ToxA*, *Ago1*, and *OCH1*) (Bates et al., 2006; Habig et al., 2021), heavy metal resistances, and metalloreductases. Starships were identified in a broad set of Ascomycete fungi sharing a common ancestor approximately 400 million years ago. There is also evidence for the involvement of Starships in at least two horizontal-gene transfer events (Gluck-Thaler et al., 2021).

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 (Dhillon et al., 2014; McDonald et al., 2019; Mat Razali et al., 2019) 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 (Lorrain et al., 2021). Intrachromosomal translocation of *ToxA* was also described in *B. sorokinaina* (McDonald et al., 2019) 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 (McDonald et al., 2019). 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 support that *ToxhAT* may have been transferred to Ptr multiple times, with one HGT event nesting *ToxhAT* within *Horizon*.

2.3.2 THE *TOXB* GENE IS EMBEDDED ON A LARGE PUTATIVE TRANSPOSON

For the first time, we showed the possible movement of *ToxB* as a multi-copy gene on a large 294-kb putative Starship transposon dubbed *Icarus*. While *Icarus* appears to lack the tyrosine recombinase 'captain', genes with DDE endonucleases are present where the captain is normally positioned. DDE endonucleases are associated DNA transposons transposases and are known to bind to TIRs which were identifiable at the edges of *Icarus* (Figure 2.5b) (Wicker et al., 2007; Nesmelova et al., 2010). This could indicate an alternative transposase associated with the mobility of *Icarus*. Alternatively, there appears to be a number of introgressed class I RNA transposons suggesting this putative Starship is rather old, having had the time to accumulate other transposons and expanded in size. The DDE endonucleases may also indicate the introgression of class II DNA transposons, and not be associated with Starship mobility. Combined with the fact that different Starship cargo genes are present within I-73–1 compared to D308 (NLR – HET, and DUF3723/phospholipase respectively) suggests that *Icarus* has been disabled. No clear evidence of *ToxB* horizontal transfer has been found to date, but homologs of *ToxB* are present in closely and distantly related species (e.g. *P. bromi*, *Cochilobolus sativus*, *Alternaria alternata*, *Magnaporthe grisea*) (Andrie et al., 2008). It has been suggested that *ToxB* was acquired vertically from a common ascomycete ancestor (Andrie et al., 2008; Ciuffetti et al., 2014). However, our discovery of *ToxB/toxb* on a large, potentially ancient, 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 limited analysis of DW-5 suggests a possible transposon associated with *ToxB* in that isolate. We found no co-linearity between the region containing *ToxB* in DW-5 and the reference isolates BFP or M4 and found no co-linearity with the putative transposon *Icarus* from I-73–1 provides additional support of potential mobility of *ToxB* within Ptr or between species.

2.3.3 CAZYMES ARE AN ESSENTAIL PART OF THE PTR GENOME

Ptr is a necrotrophic pathogen that possesses the ability to directly penetrate the host epidermal cells soon after spore germination (Aboukhaddour et al., 2016), 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 (Davies et al., 2000; Zhao et al., 2013; Jagadeeswaran et al., 2021). As an example, *Fusarium culmorum* has been shown to degrade cellulose, xylans, and pectins during invasion of wheat spike tissues (Kang et al., 2000). While previous work has linked the relatedness of CAZymes to fungal taxonomy (Barrett et al., 2020), 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 (Yamamoto et al., 2019), and the same may be true for Ptr and its hosts.

2.3.4 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 core genes decreases (Richard, 2020). 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 (Moolhuijzen et al., 2018), and the 60% core estimated in the pangenome of the wheat pathogen *Z. tritici*, which was based on 19 isolates (Badet et al., 2020). Our method of analysis assigns at most a single gene per isolate into a gene cluster (Fouts et al., 2012; McCarthy et al., 2019); additionally, although the pipeline attempts to account for gene truncation, it is not always successful. For these two reasons, the number of accessory genes may be over-estimated relative to these other studies which used different grouping algorithms.

It has been suggested that plant pathogens possess accessory genomic elements related to pathogenicity (Croll and McDonald, 2012). 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 (Wei et al., 2021). 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 the horizontal transfer of a large segment of DNA from other fungal species (Friesen et al., 2006; McDonald et al., 2019). 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 (Croll and McDonald, 2012). 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 (Aboukhaddour et al., 2011). 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 tenfold: from less than 17 Mb to more than 177 Mb (Haridas et al., 2020). 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 (Moolhuijzen et al., 2018; 2019; 2020). 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 (Moolhuijzen et al., 2018). The larger genomes reported in pathogenic races/species relative to their non-pathogenic counterparts have 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 (Kelkar and Ochman, 2012). Indeed, in some plant pathogens, the acquisition of an entire supernumerary chromosome, often rich with repetitive sequences, by horizontal transfer is evident (Ma et al., 2010). However, not all pathogenic species follow this trend of a larger genome (Friesen et al., 2005). 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 (Ochman and Moran, 2001; Lawrence, 2005; Nagendran et al., 2009; de Man et al., 2016). Many pathogens adapt to exploit a limited number of species efficiently, and in some species, even a limited number of genotypes within a species (Aboukhaddour et al., 2016) with rapid gene gains and losses previously linked to virulence adaptation in other studies (Badet and Croll, 2020; Fouché et al., 2018).

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) (Manning et al., 2013; Moolhuijzen et al., 2018). Additionally, in many other ascomycetes, the number of non-pathogenic genomes studied is limited in comparison with pathogenic genomes (Aylward et al., 2017). It is worth noting that non-pathogenicity in Ptr often has been assumed based on 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 (Guo et al., 2020).

2.3.5 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 and the contribution of transposons to virulence evolution and disease emergence.

2.4 METHODS

2.4.1 ISOLATES, DNA EXTRACTION, AND SEQUENCING

Isolate details are provided in Table 2.1. Virulence phenotypes were previously confirmed (Aboukhaddour et al., 2009; 2011; Kamel et al., 2019; Wei et al., 2020). 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 100 × 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 100 × coverage. All sequencing was performed by the Centre d'expertise et de services Génome Québec (Montreal, Canada). The reference isolate BFP (accession GCA_000149985) (Manning et al., 2013) was included in the pangenome analysis and used for full genome alignments. Isolates M4 (GCA_003171515) (Moolhuijzen et al., 2018) and DW5 (GCA_003231415) (Moolhuijzen et al., 2020) were also included for *Tox* gene and transposon analysis.

2.4.2 DE NOVO ASSEMBLIES

Read quality was assessed with FASTQC (Andrews, 2010) and poor-quality reads filtered. Reads were filtered using the standard Kraken2 database (Woods and Langmead, 2019) 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 (Bankevich et al., 2012; Seemann, 2020), Shovill with MEGAHIT (Li et al., 2015), SOAPDenovo2 (Luo et al., 2012), and CLC Genomics Workbench 12 (Qiagen) all program arguments available in the GitHub repository. QUAST (Gurevich et al., 2013) 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 (Kolmogorov et al., 2019) and polished with short-read data using Pilon (Walker et al., 2014). Completeness of the long-read assemblies was also assessed by the alignment of raw reads (Additional file 2.9).

2.4.3 GENE ANNOTATIONS

FunGAP (v1.0.1) is an annotation pipeline specifically for fungi (Min et al., 2017). FunGAP makes use of many programs: RepeatModeler (Smit and Hubley, 2008), RepeatMasker

(Smit, 2004), HISAT (Kim et al., 2015), Trinity (Grabherr et al., 2011), Augustus (Stanke et al., 2006), MAKER (Cantarel et al., 2008), BRAKER (Hoff et al., 2019), InterProScan (Quevillon et al., 2005), and BLAST + (Altschul et al., 1990). 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) (Moolhuijzen et al., 2018). 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 (Bateman et al., 2004). All predicted gene sets were assessed for completeness using the Ascomycota BUSCO gene set (odb9) (Simão et al., 2015; Seppey et al., 2019).

2.4.4 PANGENOME ANALYSIS

Pangloss is a pangenome pipeline designed for microbial eukaryotes like fungi (McCarthy et al., 2019). Panoct (Fouts et al., 2012) 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. Singletons were run through Pangloss for a second iteration. 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.

2.4.5 CAZYMES AND EFFECTORS

Phobius v1.01 (with –short) (Käll et al., 2005) 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 (Sperschneider et al., 2018). Genes identified as potential effectors (probability > 50%) were extracted from gene sets and used as input for Pangloss to assess the presence/absence between isolates. Predicted protein sequences produced by FunGap from all isolates were annotated by dbCAN2 (Zhang et al., 2018) and manually curated for selected GH, CE, PL, and AA families relevant for plant cell wall degradation from the CAZy database (Lombard et al., 2014).

2.4.6 PHYLOGENY AND ACCESSORY SETS

Individual alignments for each of the 10,159 amino acid core orthologue clusters were performed using MUSCLE (v3.6) (this may include multi-copy genes which would have been treated individually) (Edgar, 2004). 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 (Stamatakis, 2014) to generate a maximum likelihood (ML) phylogeny using the PROTGAMMA model, 1000 bootstrap replicates, and a starting seed of 10. Variant call files were generated via GATK HaplotypeCaller (McKenna et al., 2010) using BFP as a reference. SNPs were converted to fasta format with vcf2phylip (Ortiz, 2019) and input into RAxML for a ML tree using the GRTCAT model. The ML trees were visualized with FigTree (Rambaut, 2007) 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 (Pedersen, 2017).

2.4.7 TRANSPOSABLE ELEMENT CONTENT

Transposable elements (TEs) were identified and categorized using EDTA with the higher sensitivity setting (–sensitive 1) (Ou et al., 2019), which utilizes several programs: LTRharvest (Ellinghaus, et al., 2008), LTR_FINDER (Xu and Wang, 2007), LTR_FINDER_parrallel (Ou and Jiang, 2019), LTR_retriever (Ou and Jiang, 2018), TIR-Learner (Su et al., 2019), Generic Repeat Finder (Shi and Liang, 2019), HelitronScanner (Xiong et al., 2014), and TEsorter (Zhang et al., 2019). Output was aggregated and the total TE content per genome (stacked histogram) and TE content as a function of genome size (Mb) visualized.

2.4.8 GENOME ORGANIZATION

Pair-wise full genome alignments of long-read assemblies: I-73–1, D308, BFP, M4 (1.1), and DW-5 were performed with minimap2 (Li, 2018) and visualized with dotPlotly (Porten, 2018). For chromosomes with *ToxA* and *ToxB* putative transposons, syntenic blocks with a minimum length of 6500 bp were generated using Sibelia (v3.0.7) (Minkin et al., 2013) then visualized using Circos (v0.69–8) (Krzywinski et al., 2009). Selected proteins within the putative transposons were modelled with Phyre2 (Kelley et al., 2015). Mauve (Darling et al., 2010) alignments were used to identify target sites and target site duplications for transposons associated with *ToxA* and *ToxB/toxb*. Gene annotations (gff3), functional annotations (pfam), and BLAST outputs were used to manually construct the transposon schematics. Intergenic distances were calculated from gff3 files using an adapted R-script (Frantzeskakis et al., 2018) and plotted using a hexagonal heatmap of 2d bin counts geom_hex with 50 bins. Potential GDR and GSR genes were selected based on the 90th percentiles of ITL distances, with potential GDR genes having 5′ and 3′ ITLs < 2070 bp and GSR genes having 5′ and 3′ ITLs > 7394 bp. GDR and GSR genes were codon aligned to homologs of BFP using PRANK (Loytynoja and Goldman, 2008;

Jeffares et al., 2015) with some sequences being reverse complimented prior to alignment. Alignments were assessed for evolution rates via dN/dS ratios using codeml from the PAML package (Yang, 2007). dN/dS ratios were subject to unpaired Welch two-sample *t*-test assuming unequal variance to test significant differences. Long-read genomes were uploaded to RIPper (Van Wyk et al., 2019) 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.

2.5 REFERENCES

Aboukhaddour R, Cloutier S, Ballance G, Lamari L. Genome characterization of *Pyrenophora tritici-repentis* isolates reveals high plasticity and independent chromosomal location of *ToxA* and *ToxB*. *Mol Plant Pathol*. 2009;10(2):201-12.

Aboukhaddour R, Cloutier S, Lamari L, Strelkov S. Simple sequence repeats and diversity of globally distributed populations of *Pyrenophora tritici-repentis*. *Can J Plant Pathol*. 2011;33(3):389-99.

Aboukhaddour R, Hafez M, Strelkov S, Fernandez MR. Tan spot under the lenses of plant pathologists. In: Oliver RP, editor. *Achieving durable resistance in cereals*. Cambridge, United Kingdom: Burliegh Dodds; 2021.

Aboukhaddour R, Strelkov SE. Exploring de novo specificity: the *Pyrenophora tritici-repentis*-barley interaction. *Plant Pathol*. 2016;65(8):1347-57.

Aboukhaddour R, Turkington TK, Strelkov SE. Race structure of *Pyrenophora triciti-repentis* (tan spot of wheat) in Alberta, Canada. *Can J Plant Pathol*. 2013;35(2):256-68.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of Molecular Biology*. 1990;215(3):403-10.

Andrews S. FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom; 2010.

Andrie RM, Schoch CL, Hedges R, Spatafora JW, Ciuffetti LM. Homologs of *ToxB*, a host-selective toxin gene from *Pyrenophora tritici-repentis*, are present in the genome of sister-species *Pyrenophora bromi* and other members of the Ascomycota. *Fungal Genet Biol*. 2008;45(3):363-77.

Aylward J, Steenkamp ET, Dreyer LL, Roets F, Wingfield BD, Wingfield MJ. A plant pathology perspective of fungal genome sequencing. *IMA Fungus*. 2017;8(1):1-15.

Badet T, Croll D. The rise and fall of genes: origins and functions of plant pathogen pangenomes. *Curr Opin Plant Biol.* 2020;56:65-73.

Badet T, Oggenfuss U, Abraham L, McDonald BA, Croll D. A 19-isolate reference-quality global pangenome for the fungal wheat pathogen *Zymoseptoria tritici*. *BMC Biology*. 2020;18(1):1-18.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*. 2012;19(5):455-77.

Barrett K, Jensen K, Meyer AS, Frisvad JC, Lange L. Fungal secretome profile categorization of CAZymes by function and family corresponds to fungal phylogeny and taxonomy: Example *Aspergillus* and *Penicillium*. *Scientific Reports*. 2020;10(1):1-12.

Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, et al. The Pfam protein families database. *Nucleic Acids Research*. 2004;32(suppl_1):D138-D41.

Bates S, Hughes HB, Munro CA, Thomas WP, MacCallum DM, Bertram G, et al. Outer chain N-glycans are required for cell wall integrity and virulence of *Candida albicans. J Biol Chem*. 2006;281(1):90-8.

Cantarel BL, Korf I, Robb SM, Parra G, Ross E, Moore B, et al. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. 2008;18(1):188-96.

Ciuffetti LM, Manning VA, Pandelova I, Faris JD, Friesen TL, Strelkov SE, et al. *Pyrenophora tritici-repentis*: a plant pathogenic fungus with global impact. *Genomics of plant-associated fungi: monocot pathogens*: Springer; 2014. p. 1-39.

Croll D, McDonald BA. The accessory genome as a cradle for adaptive evolution in pathogens. *PLoS Pathogens*. 2012;8(4):e1002608.

Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PloS one*. 2010;5(6):e11147.

Davies K, De Lorono I, Foster S, Li D, Johnstone K, Ashby A. Evidence for a role of cutinase in pathogenicity of *Pyrenopeziza brassicae* on brassicas. *Physiol Mol Plant Pathol*. 2000;57(2):63-75.

de Man TJ, Stajich JE, Kubicek CP, Teiling C, Chenthamara K, Atanasova L, et al. Small genome of the fungus *Escovopsis weberi*, a specialized disease agent of ant agriculture. *Proc Natl Acad Sci*. 2016;113(13):3567-72.

Dhillon B, Gill N, Hamelin RC, Goodwin SB. The landscape of transposable elements in the finished genome of the fungal wheat pathogen *Mycosphaerella graminicola*. BMC Genomics. 2014;15(1):1-17.

Dong S, Raffaele S, Kamoun S. The two-speed genomes of filamentous pathogens: waltz with plants. *Curr Opin Genet Dev*. 2015;35:57-65.

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*. 2004;32(5):1792-7.

Effertz R, Meinhardt SW, Anderson J, Jordahl J, Francl L. Identification of a chlorosis-inducing toxin from *Pyrenophora tritici-repentis* and the chromosomal location of an insensitivity locus in wheat. *Phytopathology*. 2002;92(5):527-33.

Ellinghaus D, Kurtz S, Willhoeft UJBb. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. 2008;9(1):1-14.

Fouché S, Plissonneau C, Croll D. The birth and death of effectors in rapidly evolving filamentous pathogen genomes. *Curr Opin Microbiol*. 2018;46:34-42.

Fouts DE, Brinkac L, Beck E, Inman J, Sutton G. PanOCT: automated clustering of orthologs using conserved gene neighborhood for pan-genomic analysis of bacterial strains and closely related species. *Nucleic Acids Research*. 2012;40(22):e172-e.

Frantzeskakis L, Kracher B, Kusch S, Yoshikawa-Maekawa M, Bauer S, Pedersen C, et al. 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*. 2018;19(1):1-23.

Friesen TL, Ali S, Klein KK, Rasmussen JB. Population Genetic Analysis of a global collection of *Pyrenophora tritici-repentis*, causal agent of tan spot of wheat. *Phytopathology*. 2005;95(10):1144-50.

Friesen TL, Stukenbrock EH, Liu Z, Meinhardt S, Ling H, Faris JD, et al. Emergence of a new disease as a result of interspecific virulence gene transfer. *Nature Genetics*. 2006;38(8):953-6.

Gluck-Thaler E, Ralston T, Konkel Z, Ocampos CG, Ganeshan VD, Dorrance AE, et al. Giant Starship elements mobilize accessory genes in fungal genomes. *Mol Biol Evol*. 2021;39(5):msac109.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. 2011;29(7):644.

Guo J, Shi G, Kalil A, Friskop A, Elias E, Xu SS, et al. *Pyrenophora tritici-repentis* race 4 isolates cause disease on tetraploid wheat. *Phytopathology*. 2020;110(11):1781-90.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics.* 2013;29(8):1072-5.

Habig M, Schotanus K, Hufnagel K, Happel P, Stukenbrock EH. *Ago1* Affects the Virulence of the Fungal Plant Pathogen *Zymoseptoria tritici*. *Genes*. 2021;12(7):1011.

Hafez M, Despins T, Nakajima K, Aboukhaddour R. Identification of a novel *ToxA* haplotype of *Pyrenophora tritici-repentis* from Japan. *Phytopathology.* 2022(ja).

Haridas S, Albert R, Binder M, Bloem J, LaButti K, Salamov A, et al. 101 Dothideomycetes genomes: a test case for predicting lifestyles and emergence of pathogens. *Stud Mycol*. 2020;96:141-53.

Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. Whole-genome annotation with BRAKER. *Gene prediction*: Springer; 2019. p. 65-95.

Jagadeeswaran G, Veale L, Mort AJ. Do lytic polysaccharide monooxygenases aid in plant pathogenesis and herbivory? *Trends Plant Sci*. 2021;26(2):142-55.

Jeffares DC, Tomiczek B, Sojo V, Reis Md. A beginners guide to estimating the non-synonymous to synonymous rate ratio of all protein-coding genes in a genome. *Parasite genomics protocols*. New York, NY: Humana Press; 2015. p. 65-90.

Käll L, Krogh A, Sonnhammer EL. An HMM posterior decoder for sequence feature prediction that includes homology information. *Bioinformatics*. 2005;21(suppl_1):i251-i7.

Kamel S, Cherif M, Hafez M, Despins T, Aboukhaddour R. *Pyrenophora tritici-repentis* in Tunisia: race structure and effector genes. *Front Plant Sci*. 2019;10:1562.

Kang Z, Buchenauer H. Ultrastructural and cytochemical studies on cellulose, xylan and pectin degradation in wheat spikes infected by *Fusarium culmorum*. *Journal of Phytopathology*. 2000;148(5):263-75.

Kelkar YD, Ochman H, Evolution. Causes and consequences of genome expansion in fungi. *Genome Biology*. 2012;4(1):13-23.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Prot*. 2015;10(6):845-58.

Kim D, Langmead B, Salzberg SLJNm. HISAT: a fast spliced aligner with low memory requirements. 2015;12(4):357-60.

Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol*. 2019;37(5):540-6.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. *Genome Res*. 2009;19(9):1639-45.

Lamari L, Strelkov S, Yahyaoui A, Orabi J, Smith R. The identification of two new races of *Pyrenophora tritici-repentis* from the host center of diversity confirms a one-to-one relationship in tan spot of wheat. *Phytopathology*. 2003;93(4):391-6.

Lawrence JG. Common themes in the genome strategies of pathogens. *Curr Opin Genet Dev*. 2005;15(6):584-8.

Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics.* 2015;31(10):1674-6.

Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34(18):3094-100.

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*. 2014;42(D1):D490-D5.

Lorrain C, Feurtey A, Möller M, Haueisen J, Stukenbrock E. Dynamics of transposable elements in recently diverged fungal pathogens: lineage-specific transposable element content and efficiency of genome defenses. G3. 2021;11(4):jkab068.

Loytynoja A, Goldman N. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science*. 2008;320(5883):1632-5.

Lu S, Turgeon BG, Edwards MC. A ToxA-like protein from *Cochliobolus heterostrophus* induces light-dependent leaf necrosis and acts as a virulence factor with host selectivity on maize. Fungal *Genet Biol*. 2015;81:12-24.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience*. 2012;1(1):2047-217X-1-18.

Ma L-J, Van Der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, Di Pietro A, et al. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature*. 2010;464(7287):367-73.

Manning VA, Pandelova I, Dhillon B, Wilhelm LJ, Goodwin SB, Berlin AM, et al. Comparative genomics of a plant-pathogenic fungus, *Pyrenophora tritici-repentis*, reveals transduplication and the impact of repeat elements on pathogenicity and population divergence. *G3*. 2013;3(1):41-63.

Mat Razali N, Cheah BH, Nadarajah K. Transposable elements adaptive role in genome plasticity, pathogenicity and evolution in fungal phytopathogens. *International Journal of Molecular Sciences*. 2019;20(14):3597.

McCarthy CG, Fitzpatrick DA. Pangloss: A Tool for Pan-Genome Analysis of Microbial Eukaryotes. *Genes*. 2019;10(7):521.

McDonald MC, Ahren D, Simpfendorfer S, Milgate A, Solomon PS. The discovery of the virulence gene *ToxA* in the wheat and barley pathogen *Bipolaris sorokiniana. Mol Plant Pathol*. 2018;19(2):432-9.

McDonald MC, Taranto AP, Hill E, Schwessinger B, Liu Z, Simpfendorfer S, et al. Transposon-mediated horizontal transfer of the host-specific virulence protein *ToxA* between three fungal wheat pathogens. *MBio*. 2019;10(5).

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297-303.

Min B, Grigoriev IV, Choi IG. FunGAP: Fungal Genome Annotation Pipeline using evidence-based gene model evaluation. *Bioinformatics*. 2017;33(18):2936-7.

Minkin I, Patel A, Kolmogorov M, Vyahhi N, Pham S, editors. Sibelia: a scalable and comprehensive synteny block generation tool for closely related microbial genomes. *International Workshop on Algorithms in Bioinformatics*; 2013: Springer.

Moolhuijzen P, See PT, Hane JK, Shi G, Liu Z, Oliver RP, et al. Comparative genomics of the wheat fungal pathogen *Pyrenophora tritici-repentis* reveals chromosomal variations and genome plasticity. *BMC Genomics*. 2018;19(1):279.

Moolhuijzen P, See PT, Moffat CS. A new PacBio genome sequence of an Australian *Pyrenophora tritici-repentis* race 1 isolate. *BMC Res Notes*. 2019;12(1):1-5.

Moolhuijzen P, See PT, Moffat CS. PacBio genome sequencing reveals new insights into the genomic organisation of the multi-copy *ToxB* gene of the wheat fungal pathogen *Pyrenophora tritici-repentis*. BMC Genomics. 2020;21(1):1-12.

Nagendran S, Hallen-Adams HE, Paper JM, Aslam N, Walton JD. Reduced genomic potential for secreted plant cell-wall-degrading enzymes in the ectomycorrhizal fungus *Amanita bisporigera*, based on the secretome of *Trichoderma reesei*. *Fungal Genet Biol*. 2009;46(5):427-35.

Nakamura AM, Nascimento AS, Polikarpov I. Structural diversity of carbohydrate esterases. *Biotechnology Research and Innovation*. 2017;1(1):35-51.

Nesmelova IV, Hackett PB. DDE transposases: Structural similarity and diversity. *Adv Drug Delivery Rev*. 2010;62(12):1187-95.

Ochman H, Moran NA. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science*. 2001;292(5519):1096-9.

Ortiz E. vcf2phylip v2. 0: convert a VCF matrix into several matrix formats for phylogenetic analysis. https://doiorg/105281/zenodo. 2019;2540861.

Ou S, Jiang NJMD. LTR_FINDER_parallel: parallelization of LTR_FINDER enabling rapid identification of long terminal repeat retrotransposons. 2019;10(1):1-3.

Ou S, Jiang NJPp. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. 2018;176(2):1410-22.

Ou S, Su W, Liao Y, Chougule K, Agda JR, Hellinga AJ, et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome biology*. 2019;20(1):1-18.

Pedersen TL. Hierarchical sets: analyzing pangenome structure through scalable set visualizations. *Bioinformatics*. 2017;33(11):1604-12.

Porten T. dotPlotly. https://githubcom/tpoorten/dotPlotly. 2018.

Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, et al. InterProScan: protein domains identifier. 2005;33(suppl_2):W116-W20.

Raffaele S, Farrer RA, Cano LM, Studholme DJ, MacLean D, Thines M, et al. Genome evolution following host jumps in the Irish potato famine pathogen lineage. *Science*. 2010;330(6010):1540-3.

Rafiei V, Vélëz H, Tzelepis G. The Role of Glycoside Hydrolases in Phytopathogenic Fungi and Oomycetes Virulence. *International Journal of Molecular Sciences*. 2021;22(17):9359.

Rambaut A. FigTree, a graphical viewer of phylogenetic trees. https://githubcom/rambaut/figtree/. 2007.

Richard G-F. Eukaryotic pangenomes. In: Tettelin H, Medini D, editors. *The Pangenome*. Gewerbestrasse, Switzerland: Springer; 2020. p. 253-91.

Savary S, Willocquet L, Pethybridge SJ, Esker P, McRoberts N, Nelson A. The global burden of pathogens and pests on major food crops. *Nature Ecology and Evolution*. 2019;3(3):430-9.

Seemann T. Shovill. https://githubcom/tseemann/shovill.

Seppey M, Manni M, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness. *Gene Prediction*: Springer; 2019. p. 227-45.

Shi GS, Kariyawasam G, Liu S, Leng Y, Zhong S, Ali S, et al. A conserved hypothetical gene is required but not sufficient for Ptr ToxC production in *Pyrenophora tritici-repentis*. *Mol Plant-Microbe Interact*. 2022(ja).

Shi J, Liang CJPp. Generic Repeat Finder: a high-sensitivity tool for genome-wide de novo repeat detection. 2019;180(4):1803-15.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31(19):3210-2.

Smit AF, Hubley R. RepeatModeler Open-1.0. 2008.

Smit AFJhwro. Repeat-Masker Open-3.0. 2004.

Sperschneider J, Dodds PN, Gardiner DM, Singh KB, Taylor JM. Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. *Mol Plant Pathol*. 2018;19(9):2094-110.

Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30(9):1312-3.

Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern BJNar. AUGUSTUS: ab initio prediction of alternative transcripts. 2006;34(suppl_2):W435-W9.

Strelkov S, Kowatsch R, Ballance G, Lamari L. Characterization of the *ToxB* gene from North African and Canadian isolates of *Pyrenophora tritici-repentis*. *Physiol Mol Plant Pathol*. 2005;67(3-5):164-70.

Strelkov SE, Lamari L, Sayoud R, Smith RB. Comparative virulence of chlorosis-inducing races of *Pyrenophora tritici-repentis. Can J Plant Pathol*. 2002;24(1):29-35.

Su W, Gu X, Peterson TJMp. TIR-Learner, a new ensemble method for TIR transposable element annotation, provides evidence for abundant new transposable elements in the maize genome. 2019;12(3):447-60.

Torres DE, Oggenfuss U, Croll D, Seidl MF. Genome evolution in fungal plant pathogens: looking beyond the two-speed genome model. *Fungal Biology Reviews*. 2020.

Van Wyk S, Harrison CH, Wingfield BD, De Vos L, van Der Merwe NA, Steenkamp ET. The RIPper, a web-based tool for genome-wide quantification of Repeat-Induced Point (RIP) mutations. *PeerJ*. 2019;7:e7447.

Vanheule A, Audenaert K, Warris S, van de Geest H, Schijlen E, Höfte M, et al. Living apart together: crosstalk between the core and supernumerary genomes in a fungal plant pathogen. *BMC Genomics*. 2016;17(1):1-18.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*. 2014;9(11):e112963.

Wei B, Despins T, Fernandez MR, Strelkov SE, Ruan Y, Graf R, et al. Race distribution of *Pyrenophora tritici-repentis* in relation to ploidy level and susceptibility of durum and winter bread wheat. *Canadian Journal of Plant Pathology*. 2021;43(4):582-98.

Wei B, Moscou MJ, Sato K, Gourlie R, Strelkov S, Aboukhaddour R. Identification of a locus conferring dominant susceptibility to *Pyrenophora tritici-repentis* in barley. *Front Plant Sci*. 2020;11:158.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*. 2007;8(12):973-82.

Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biology*. 2019;20(1):257.

Xiong W, He L, Lai J, Dooner HK, Du CJPotNAoS. HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. 2014;111(28):10263-8.

Xu Z, Wang HJNar. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. 2007;35(suppl_2):W265-W8.

Yamamoto N, Wang Y, Lin R, Liang Y, Liu Y, Zhu J, et al. Integrative transcriptome analysis discloses the molecular basis of a heterogeneous fungal phytopathogen complex, *Rhizoctonia solani* AG-1 subgroups. *Scientific Reports*. 2019;9(1):1-14.

Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007;24(8):1586-91.

Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research*. 2018;46(W1):W95-W101.

Zhang R-G, Wang Z-X, Ou S, Li G-YJb. TEsorter: lineage-level classification of transposable elements using conserved protein domains. 2019:800177.

Zhao Z, Liu H, Wang C, Xu J-R. Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics*. 2013;14(1):1-15.

CHAPTER 3: THE VIRULENCE GENE *TOXB* IS BOTH AMPLIFIED AND DISRUPTED BY TRANSPOSONS IN THE WHEAT PATHOGEN *PYRENOPHORA TRITICI-REPENTIS*

3.0 SUMMARY

• Mechanisms that drive virulence gene duplication in plant pathogenic fungi remain poorly understood. In *Pyrenophora tritici-repentis* (*Ptr*), responsible for tan spot of wheat, *ToxB* is a multicopy virulence gene encoding a proteinaceous necrotrophic effector. *ToxB* exhibits a virulence dosage effect, where higher copy numbers are associated with increased disease severity. In this work, we sought to resolve a 25-year old question as to what drove the proliferation of *ToxB* within *Ptr.*

• To investigate this, 23 long-read assemblies were generated and analyzed from a collection of globally distributed isolates with various *ToxB* copy numbers, with a specific focus on regions containing *ToxB.*

• Extensive comparative alignments identified a Helitron-like element, *ToxB-HLE*, that appears to be driving the duplication of *ToxB* in an accessory region of chromosome 4. This region is entirely absent in isolates lacking *ToxB* or its nonfunctional homolog *toxb*. In addition to gene amplification by transposons, multiple independent transposon insertion events were identified in several isolates that disrupted the *ToxB* open reading frame creating inactive *toxb* haplotypes.

• This study provides strong evidence supporting the hypothesis that transposons play dual roles in the rapid evolution of fungal pathogenicity by both amplifying and disrupting a key virulence gene in a globally distributed plant pathogen.

3.1 INTRODUCTION

Gene duplications are fundamental drivers of functional diversification and play a key role in shaping adaptation in eukaryotes by generating genetic redundancy that facilitates the emergence of novel functions (Nei et al., 1969; Holland et al., 1999; Hastings et al., 2009; Ohno et al., 2013; Sacristán et al., 2021). When duplication occurs, this creates an opportunity for natural selection to modify gene functions. The happens because purifying selection is relaxed on the additional gene copies, allowing them to evolve new functions, while positive pressure maintains the original gene activity (Kimura, 1983; Long et al., 2003). The impact of a duplication event is dependent on not only the mechanism (e.g. retrotransposition, break-fusion-bridge, unequal cross over, etc.) and size (i.e. whole genome, local region, single genes) of a duplication but on the evolutionary pressures the organism faces.

For example, whole genome duplication in the ancestors of *Arabidopsis* is thought to have occurred three times (Simillion et al. 2002; Bowers et al. 2003). Ancient whole genome duplications have been identified in many other plant species, including crops such as barley, rice, and grape (Tian et al., 2005; International Rice Genome Sequencing Project, 2005; Jaillon et al., 2007; Thiel et al., 2009). These duplications are distinct from alloploids such as wheat where the additional genetic material is gained through species hybridization. Whole genome duplication is also thought to have occurred in other higher eukaryotes such as yeast (*Saccharomyces cerevisiae*) and vertebrates which shows the phenomenon is widespread and extremely influential (Kellis et al., 2004; Berthelot et al., 2014). Whole genome duplications are often followed by a large reduction in duplicated genes as redundant copies are removed through purifying selection (Kellis et al., 2004; Bloome et al., 2006; Wolf and Koonin, 2013).

As with other organisms, gene duplications in fungi are known to have led to functional diversification of many orthogroups, especially in genes that aid fungi in responding to stress or environments that confer strong selective pressure (Wapinski et al., 2007). For example, it has been shown that duplication of biosynthetic gene clusters have played an important role in the evolution of trichothecene production in *Fusarium* sp. (Ward et al., 2002; Brown et al., 2004; Proctor et al., 2018; Rokas et al., 2018). Trichothecenes provide pathogens with a powerful chemical weapon during host infection and confers a strong advantage compared to isolates which lack toxin production. Duplication of single genes rather than whole genomes or gene clusters can also have significant impact. For example, an ancient duplication of the *Sir5* gene, creating *Sir2* in *Zymoseptoria tritici* (a major leaf-spot pathogen), was recently reported to be significantly associated with thermal adaptation (Moser Tralamazza et al., 2023). It is likely that *Sir2* significantly contributed to, and shaped, *Z. tritici*'s ability to survive in highly variable environments (Moser Tralamazza et al., 2023). In *Z. tritici*, the *Sir2* gene remains closely associated with transposable elements that may have driven the duplication event, indicating that these elements can play an important role in initiating gene duplication in fungi.

There are a growing number of examples that demonstrate that gene duplications can occur very rapidly under strong selective pressure and remain polymorphic within a fungal population. For example, the duplication of the *CYP51* gene (syn. *ERG11*) decreases sensitivity to azole fungicides (Liu et al,. 2011; Stalder et al., 2022; Fan et al., 2023; Arnold et al., 2024). Purifying selection of resistance gene duplications may occur when the high selective pressure of fungicides are removed (Hawkins and Fraaije, 2018). Recently, a virulence gene duplication was explored in the genus *Rhynchosporium* which contains several pathogens of grasses such as bromegrass, barley, triticale, and rye. The necrosis inducing effector *NIP1* was found to have

substantial copy-number variations between the related pathogen species with higher copy-numbers linked to increased virulence (Mohd-Assaad et al., 2019). In these examples, long-read genome assemblies or large-scale population genomics data were key to uncovering the gene duplication events, because the gene duplications were still highly identical at the nucleotide level. These high-identity copies collapse in short-read assemblies into a single sequence and can be difficult to detect. This makes determining the exact sequence of events that led to duplications difficult to describe in detail, as these regions often fall into unassembled regions of the genome.

In plant pathogens, transposable elements are drivers of adaptation in regions encoding effectors, small secreted proteins or metabolites that interact and modulate host defense systems (Raffaele et al., 2010; Dong et al., 2015; Torres et al., 2020). In addition transposons can also alter expression profiles of nearby genes and facilitate genome instability by increasing the likelihood of deleterious rearrangements (Eichler and Sankoff, 2003; Mita and Boeke, 2016; Bhat et al., 2022). The reportedly rapid evolution facilitated by transposons also means that mutations driven by transposons are highly variable within the larger pathogen populations and thereby transient in nature (Fouché et al., 2018; Sigh et al., 2021). This variability can have both positive and negative impacts on pathogen evolution. For example, in the tomato pathogen *Fusarium oxysporum* f.sp *lycopersici* the insertion of a hAT transposon within the effector gene *AVR1* resulted in a loss of function mutation (Inami et al., 2012). While seemingly a negative event, due to the loss of a functional virulence gene, this transposon insertion under the right host environment leads to a gain of virulence on host genotypes carrying the *I* resistance gene (Inami et al., 2012). The tension or conflict between transposon activity and adaptive evolution in fungal pathogen genomes has recently been termed “the devil’s bargain” (Fouché et al., 2022).

However, there remains a limited number of examples where transposons have been shown to behave in antagonistic ways on the same fungal protein with a characterised function in virulence.

Differential gene gains and losses within a species are also known to impact pathogen virulence and host ranges (Lawrence et al., 2005; Friesen et al., 2006; Powell et al.,2008). In the leaf-spot pathogen *Z. tritici* for example, large numbers of genes with copy-number variations associated with virulence factors and secondary metabolite gene clusters were found near to transposable elements (Hartmann and Croll, 2017). Strong divergent selection on the presence-absence polymorphisms were also linked to local adaptation. Previous work found high numbers of differential gene gains and losses *Pyrenophora tritici-repentis* as well, however their links to adaptation remain to be explored (Gourlie et al., 2022).

The *ToxB* gene in *P. tritici-repentis* (*Ptr*), causing tan spot of wheat, is one rare example where a known multi-copy virulence gene, that is critically important for disease development, has also been impacted by transposon insertions to create inactive copies (Strelkov et al., 1999; Hafez et al., 2024). ToxB is the second proteinaceous effector identified in the tan spot pathosystem, after the necrosis-inducing effector, ToxA (Orolaza et al., 1995; Lamari et al., 1995; Lamari et al., 2003). Both effectors function in an inverse gene-for-gene manner, whereby necrosis/chlorosis is only observed in wheat genotypes carrying the corresponding susceptibility genes (Lamari and Bernier, 1989; Lamari et al., 1995; Lamari et al., 2003). When purified ToxB protein is infiltrated into wheat leaves, it induces chlorosis in sensitive wheat genotypes that carry the corresponding sensitivity gene, designated *Tsc2* (Faris et al., 2013). *ToxB* was found to be present in field isolates in as many as 10 copies through both PCR techniques and southern hybridization experiments (Martinez et al, 2001; 2004; Strelkov et al., 2005). Through a series of

cloning experiments these *ToxB* duplications were determined to be identical to each other, but the surrounding sequences were different (Martinez et al., 2004). This left a question as to how *ToxB* was duplicated in these genomes.

Importantly the number of *ToxB* copies has been shown to contribute additively to virulence leading to quantitative (dosage) effects (Amaike et al. 2008; Aboukhaddour et al., 2012). While these data suggest that the evolution of virulence in this pathogen is driven by duplication of the *ToxB* effector, there also exists a non-functional homolog, designated *toxb,* that is unable to induce chlorosis found in some *Ptr* isolates and other Ascomycetes (Hafez et al., 2024). Recent work found *ToxB* present in 21% of a worldwide *Ptr* population (422 isolates) while *toxb* was identified in 6%, indicating that both active and inactive forms persist in the global pathogen population (Hafez et al., 2024). The majority of chlorosis-inducing *Ptr* isolates carry the same *ToxB* haplotype (allele), *ToxB1*, though at least three functional *ToxB* (*ToxB1*, *B4*, and *B5*) and six non-functional *toxb* haplotypes (*toxb2*, *b3*, *b12*, *b13*, *b14*, and *b15*) have been identified in *Ptr* (reviewed in Hafez et al., 2024).

Understanding how *ToxB* was duplicated in the tan spot pathogen has been complicated by its multi-copy nature combined with the limited access to *ToxB* coding isolates (Aboukhaddour et al., 2023; Hafez et al., 2024). *ToxB* is absent from Australian isolates and very rare in isolate collections from North America, and is mainly found in isolates from the Fertile Crescent and the encompassing regions. As such, the *ToxB* containing isolates we have obtained from these regions offered a unique opportunity to further explore *ToxB* evolution. Our previous work using long read assemblies of isolates I-73-1 and D308 showed that *ToxB* (and its homolog *toxb*) were carried on chromosome 4 (Chr04) within what appeared to be an accessory region that shared several common features with a Starship transposon, but was missing the tyrosine

recombinase captain, and hence it was thought it may be an ancient/derelict Starship transposon (Gourlie et al., 2022). Earlier research using southern hybridisation approaches indicated that *ToxB* can be carried on two different chromosomes in the same isolate (Aboukhaddour et al., 2009). Recently, a Copia-like retrotransposon (*Copia-1_Ptr*) was identified in some races of *Ptr* which has disrupted the open-reading frame of *ToxB*, resulting in an inactive form of *toxb* (Hafez et al., 2024). Together, these studies suggest that transposon activity may be a key driver in both the expansion and inactivation of the *ToxB* gene in different populations. To address these questions, 23 new long-read assemblies of *Ptr* from 7 countries were generated including 8 isolates with *ToxB,* as well as 9 isolates with the inactive *toxb* allele and 6 isolates completely lacking either *ToxB* or *toxb* based on previous PCR and southern blot hybridization analyses (Aboukhaddour et al., 2009; 2013; Kamel et al., 2019; Wei et al., 2021; Hafez et al., 2024). This comprehensive collection of diverse isolates provides a framework to dissect the genomic mechanisms that led to *ToxB* duplication and pseudogenization.

3.2 METHODS

3.2.1 ISOLATES, DNA EXTRACTION, AND SEQUENCING

Twenty four isolates in total were included in this study and isolate details presented in Table 1. DNA was extracted as described in Gourlie et al., (2022). Briefly, isolates were grown in 100 mL PDB flasks @ 25°C in the dark. Mycelia mats were harvested after ~2 weeks then washed several times with milli-Q water and freeze-dried. Genomic DNA extracted using Genomic Tip 100/G kits from Qiagen. A subset of isolates were sequenced in 2022 by Genome Quebec using PacBio Sequel, and the remaining isolates were sequenced at Mayo Clinic in 2023 using PacBio RS II.

3.2.2 GENOME ASSEMBLY

The subset of isolates sequenced at Genome Quebec were assembled using CANU (v2.2) after converting subread.bam files to fastq using SAMtools (v1.9) (Li et al., 2009; Koren et al., 2017). The set sequenced at Mayo Clinic were assembled using HiCanu (v2.2) (Nurk et al., 2020). For both CANU and HiCanu the default settings were used. QUAST (Gurevich et al., 2013) and BUSCO scores were used to assess assembly quality using the Pleosporales db10 dataset (Simão et al., 2015). Completeness and *ToxB* regional confidence was also assessed by aligning raw reads to the assemblies.

3.2.3 ALIGNMENTS

Alignment of reads to assemblies was done with Bowtie2 (Langmead and Salzberg, 2012) and SAMtools (v1.9) (Li et al., 2009) then visualized with either IGV (Robinson et al., 2011) or dotplotly (Porten, 2018). Whole genome alignments were done with minimap2 (v2.18-r1015) (Li, 2018) and visualized with dotplotly (Porten, 2018). Genome alignment also done with progressiveMauve (Darling et al., 2010). Regional/local alignments were done within Geneious built-in dotplotter (Biomatters Ltd.), MAFFT (v1.5.0) (Katoh and Standley, 2013), LASTZ (v7.0.3) (Haris, 2007), or gggenomes R package (Hackl et al., 2024). Circular alignments were done with Sibelia (v3.0.7) (Minkin et al., 2013) and visualized with Circos (v0.69-8) (Krzywinski et al., 2009). Reference genomes Pt-1C-BFP and M4_2.2 were retrieved from NCBI for alignments (Manning et al., 2013; Moolhuijzen et al., 2018).

3.2.4 IDENTIFICATION OF THE *TOXB* HELIRON-LIKE-ELEMENT (*TOXB-HLE*) AND OTHER TRANSPOSONS

Total transposon content of each isolate was assessed with EDTA (v2.2) (Ou et al., 2019). Each copy of *ToxB* from each genome including various amounts of upstream and/or

downstream sequence. The up/down stream sequences ranged between 50 bp to 15 kbp , stopping at the next *ToxB* copy when in tandem. Multiple iterations of alignments (described above) were used to identify where sequence similarity ended between copies to establish putative edges/edge sequences. Putative edge sequences were used to extract copies of the repetitive region containing *ToxB*. Open-reading frames were identified within and just beyond the putative edges using Geneious ORF finder with default settings. ORFs were used to BLAST the UniProt database (Consortium, 2015) to provide functional information or identify putative domains. The shortest edge-to-edge sequence was used as a reference to compare copies. Variations were manually catalogued. Hairpins were identified using VectorBuilder (v2.1.813). Insertions >50 bp were further explored using similar methods: identifying edges, determining edge type (i.e. LTR or TIR), search for target site duplications, ORF prediction and annotation, BLASTN of RepBase (vSep2020), and literature comparison. Transposon schematics were created manually.

3.3 RESULTS

3.3.1 LONG-READ ASSEMBLIES REVEAL GENOME EXPANSION IS DRIVEN BY TRANSPSONS

The isolates selected for sequencing represent diverse geographical regions from North America, Japan, North Africa, and Caucasus region with sampling times that spanned from 1955 to 2020 (Table 3.1). Isolates were selected based on their estimated number of *ToxB* copies, along with a subset known to carry the inactive homolog, *toxb.* Six isolates that completely lack *ToxB* or *toxb* as determined previously were also selected (Kamel et al., 2019; Hafez et al., 2022). All isolates were *de novo* assembled with CANU or Hi-CANU (depending on sequencing technology used) with an expected genome size of ~40 Mb based on the reference M4_2.2. Assembled genome sizes varied from 34.5 to 44.7 Mb (Table 3.1). Genomes assembled from

HiFi data (circular consensus) contained more contiguous sequences than genomes assembled from older continuous long-read sequencing data. All assemblies had greater than 96% BUSCO scores based on the Pleosporales database (pleo_db10). The transposon annotation pipeline EDTA identified variable transposon content across isolates ranging from 12 to 26% of total genome content (Table 3.1). Genome size increased linearly with total predicted transposon content ($R^2 = 0.58$; $p<0.001$) (Figure 1). The reported R^2 value suggests transposons are responsible for 58% of the observed genome expansion with some other unknown variable(s) contributing the remaining ~42%. The main drivers of transposon expansion are LTR retrotransposons (i.e. Copia, Ty-3, and unknown elements with LTRs) as well as DNA hAT transposons.

Table 3.1 Assembly statistics and *ToxB* information of *Pyrenophora tritici-repentis* genomes used in this study. Table is ordered by *ToxB* copy number, followed by isolates with *toxb*, then isolates without *ToxB* or *toxb* (alphabetical by isolate name within groupings).

Isolate	Location	Year	Assembler	Size (Mb)	Contigs		N50 (kbp)	TE content (%)	*ToxB* copies	*ToxB* contig (size)
					Total	>1Mb				
Alg3-24	Algeria	1993	Hi-Canu	42.4	119	11	3,402	21.1	13	35 (3.4 Mb) 112 (2.2 Mb) 113 (152 kbp) 127 (84 kbp) 133 (39 kbp) 131 (38 kbp) 63 (29 kbp) 136 (28 kbp) 128 (24 kbp) 129 (23 kbp) 130 (23 kbp)
AlgH1	Algeria	1993	Hi-Canu	42.5	103	12	3,698	20.5	5	21 (3.8 Mb)
I-73-1*	Syria	2001	Flye	39.9	39	11	3,647	19.9	4	12 (3.6 Mb) 13 (6 kbp)
T103-1	Tunisia	2017	CANU	39.3	236	5	570	21.2	4	182 (229 kbp) 183 (209 kbp)
T128-1	Tunisia	2017	CANU	39.8	553	11	3,243	20.3	4	14 (3.7 Mb) 21 (28 kbp)
I-34-1	Azerbaijan	2001	Hi-Canu	43.3	150	12	3,253	21.7	3	151 (3.6 Mb)
92-171-R5	Canada	1992	Hi-Canu	44.7	36	14	2,880	26.4	2	4 (4.9 Mb)
Tptr3-1	Tunisia	2017	Hi-Canu	42.8	126	13	3,604	21.9	1	129 (3.7 Mb)
107224	Canada	1955	Hi-Canu	37.5	35	18	2,075	13.6	*toxb*	26 (911 kbp)

331-2	Canada	1985	Hi-Canu	43.7	123	12	3,470	23.4	*toxb*	43 (3.7 Mb)
90-2	Canada	1990	CANU	36.2	310	12	2,960	12.0	*toxb*	344 (3.1 Mb)
D308*	Canada	1985	Flye	39.6	70	11	3,667	20.2	*toxb*	5 (3.7 Mb)
SC22-2	Canada	1999	Hi-Canu	44.6	160	12	3,161	22.7	*toxb*	160 (3.7 Mb)
SC29-1	Canada	1999	CANU	41.5	490	6	468	22.8	*toxb*	457 (33 kbp)
SC29-8	Canada	1999	Hi-Canu	43.2	175	13	2,933	21.1	*toxb*	184 (3.9 Mb)
SW21-5	Canada	2016	Hi-Canu	42.1	96	13	2,768	21.9	*toxb*	39 (3.7 Mb)
G9-7	Canada	2016	Hi-Canu	43.3	135	11	3,552	17.1	*toxb*	69 (2.6 Mb)
Den17	Denmark	2009	Hi-Canu	42.7	97	12	3,465	20.1	0	51 (3.3 Mb)
I-33-16	Azerbaijan	2001	Hi-Canu	41.3	75	13	2,705	20.4	0	51 (2.7 Mb)
K11	Japan	2020	Hi-Canu	41.2	150	12	3,271	17.6	0	47 (3.2 Mb)
K6	Japan	2020	Hi-Canu	40.0	135	11	3,071	15.2	0	61 (3.0 Mb)
K9	Japan	2020	Hi-Canu	40.5	130	12	2,812	16.5	0	146 (2.7 Mb)
T126-1	Tunisia	2017	Hi-Canu	42.5	175	12	3,241	21.1	0	206 (3.6 Mb)

*Assembled previously in Gourlie et al., 2022

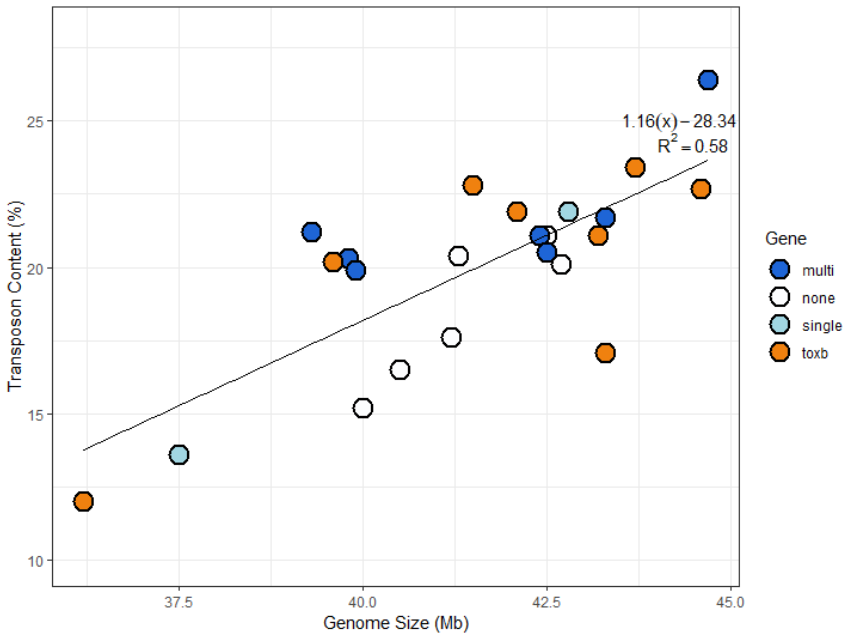


Figure 3.1 Comparison of genome size to total transposon content (predicted by EDTA) as a percentage of genome size. Each dot represents an individual isolate and the dot color indicates the *ToxB/toxb* genotype as indicated by the legend.

3.3.2 WHOLE CHROMOSOME ALIGNMENT SHOWS ACCESSORY CHROMOSOME ARM ASSOCIATED WITH TOXB PRESENCE/ABSENCE

BLASTN of the *ToxB* open reading frame (ORF) revealed *ToxB* copy-numbers in each isolate, their chromosomal location and their relative positions to other copies (Figure 3.2). Isolates found completely lacking *ToxB*/*toxb* were Den17, T126-1, K11, K9, K6, and I-33-16, which matches previous PCR genotyping and phenotypic screens (Kamel et al., 2019; Hafez et al., 2022; unpublished data). *ToxB* in its functional form was found as multiple copies in isolates: Alg3-24, AlgH1, I-34-1, I-73-1, T103-1, 92-171, and T128-1. Isolate Tptr3-1 contained only a single functional copy of *ToxB*. Isolates with a single non-functional *toxb* were 90-2, 331-2, D308, G9-7, SC22-2, SC29-1, SC29-8, SW21-5, and 107224. No isolates contained multiple *toxb* copies, and no isolate contained a combination of *ToxB* and *toxb* together. Of the 17 *ToxB/toxb* containing

isolates, 13 assemblies harbour the gene on chromosome sized contigs (defined here as ≥1 Mb). The remaining 4 assemblies contained *ToxB/toxb* on contigs ranging from 300 kbp down to only 27 kbp (Table 3.1). Based on alignments with the reference isolates Pt-1C-BFP (non-carrier; USA) and M4_2.2 (non-carrier; Australia) *ToxB/toxb* is located on Chr04 in 11 isolates. The exceptions being isolate Alg3-24, 92-171R5, and G97. In this Algerian isolate, Alg3-24, one *ToxB* copy was found on Chr05 and an additional two copies were found on Chr10 (Figure 3.2). Alg3-24 had 10 additional copies of *ToxB* found on contigs ≥ 150 kbp. It is not clear if these smaller contigs represent additional copies in other chromosomes and attempts to manually align these smaller contigs yielded inconclusive results. As such they were omitted from further analysis. Isolates 92-171R5 and G9-7 are phylogenetically distant from the other *Ptr* isolates (Gourlie et al., 2022) so it is not surprising *ToxB/toxb* was found on Chr03 and Chr07 respectively, rather than chromosome 4 as with most isolates. For three isolates (107224, SC29-1, T103-1) *ToxB/toxb* assembled into small contigs where we could not determine its precise location in the genome. Chromosome alignment of Chr04 *ToxB/toxb* isolates with isolates that do not contain either *ToxB/toxb* revealed a large accessory region (Figure 3.3). This *ToxB/toxb* containing region was variable in size depending on which isolates were aligned with sizes ranging from ~50 kbp to ~420 kbp suggesting an expansion of Chr04 (Additional File 3.1).

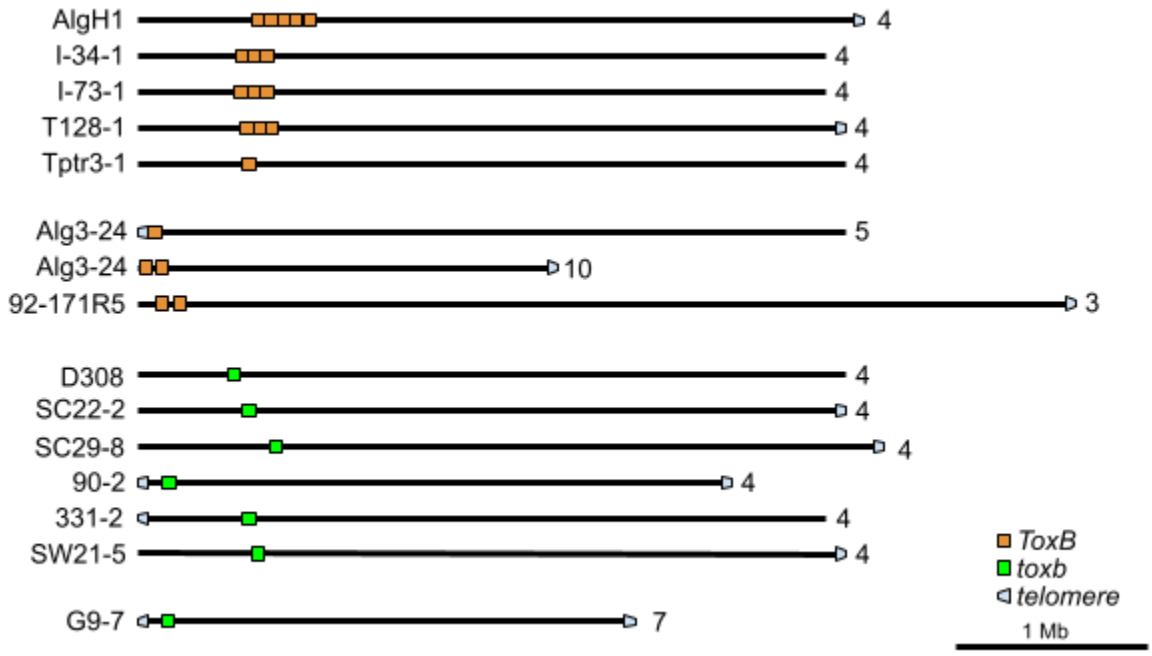


Figure 3.2 Relative positions of *ToxB/toxb* copies in isolates with chromosome sized contigs. Orange boxes indicate active *ToxB* isoforms and green boxes indicate inactive *toxb* haplotypes. Grey caps indicate telomeric repeats (TTAGGG/CCCTAA) were found. The number on the far right represents the corresponding chromosome number when aligned with M4_2.2.

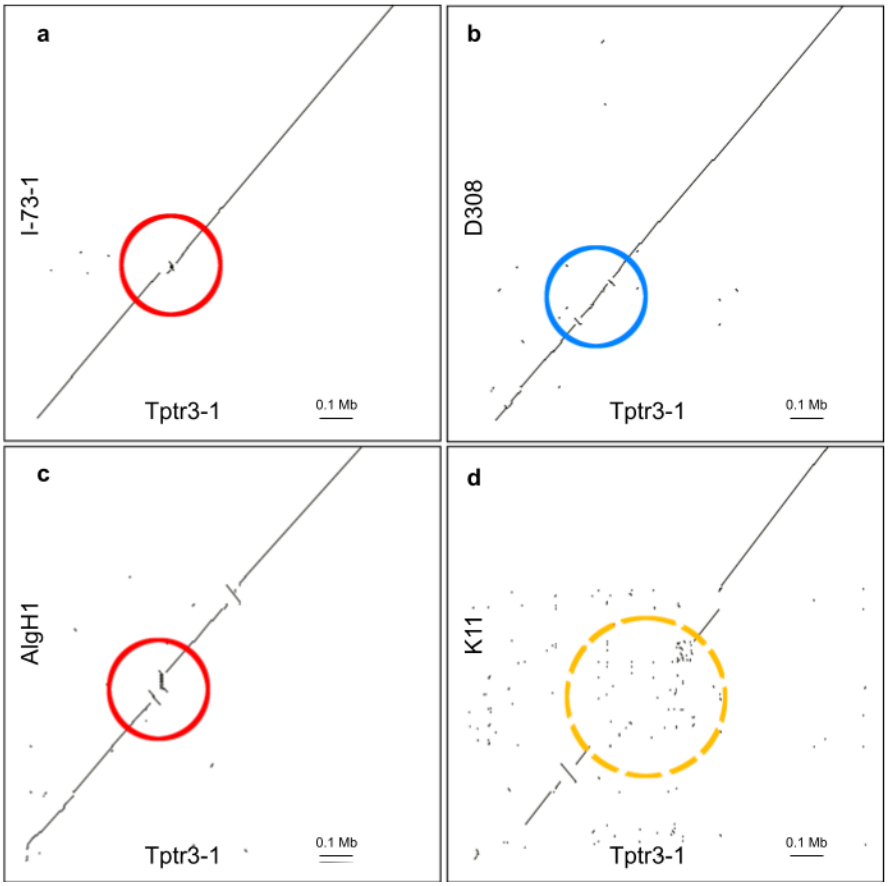


Figure 3.3 Representative dotpot alignments of *ToxB/toxb* containing and non-containing contigs. The x-axis in all alignments is a section of Chr04 from isolate Tptr3-1 which contains a single copy of *ToxB*. The y-axis in plots A-D are the same section of Chr04 from different isolates: a alignment with I-73-1 which contains three *ToxB* copies. Area containing *ToxB* is highlighted in the red circle; b alignment with toxb isolate D308, region containing *toxb* is highlighted in a blue circle; c alignment with isolate AlgH1, which contains five copies of *ToxB* marked by a red circle; and d alignment of K11 that does not contain *ToxB* or *toxb*. The accessory region is marked in an orange dashed circle.

3.3.3 TOXB COPIES ARE PRESENT AS TANDEM DUPLICATIONS

The BLASTN results reported above identified five isolates where multiple copies of *ToxB* were localized to the accessory region of Chr04 in close proximity to each other. Closer inspection of this region showed that *ToxB* and some flanking DNA were arranged in tandem copies (or direct repeats). In AlgH1, we identified five tandem copies and in each of I-34-1, I-73-1, T103-1, and T128-1 we identified three tandem copies (Figure 3.2

and 3.4). In order to assess whether this tandem array of *ToxB* was not due to a mis-assembly of this region, the raw PacBio reads were aligned against the *de novo* assembly for each isolate. Isolates I-73-1, T103-1, I-34-1, and T128-1 all have single reads spanning all three copies along with high coverage (30x to 50x), providing high confidence that the tandem array of *ToxB* was assembled correctly (Additional File 3.2). For AlgH1, with five tandem copies, we identified five reads with two of the five tandem copies present and no single read that spanned the full 56 kbp region due to its large size. Coverage of the region is high however (30x to 50x). In Alg3-24, we did not observe the same tandem copy structure seen in other isolates, rather two copies were found on a chromosome sized contig with inverted orientation and a single copy on a separate chromosome sized contig (Figure 3.5). In this assembly, there were several other *ToxB* copies on smaller contigs, all with very high transposon content likely explaining their inability to be assembled more contiguously. All *ToxB* containing reads from Alg3-24 contained only single copies of *ToxB*, therefore it is difficult to conclude if the chromosome assemblies with two copies is correct (Additional File 3.3).

All isolates carrying *toxb*, or lacking both *ToxB* and *toxb*, had high quality assemblies with high coverage (>30x) around *toxb* (where applicable), except for isolates SC29-1 and 107224 (Table 3.1). For isolate 107224, the contig containing *ToxB* is relatively small (0.9 Mb), and likely does not represent a full chromosome. Overall the assembly accuracy is high; however, for isolates Alg3-24, SC29-1, T103-1, 92-171R5, and 107224, coverage and contiguity near some *ToxB/toxb* sites are variable with coverage dropping to 10x in some regions, leading to possible inaccuracies in these assemblies. In summary, we have high confidence with high coverage in one assembly

with a single *ToxB* copy, four assemblies with multiple *ToxB* copies, seven assemblies with *toxb*, and six assemblies which lack *ToxB* or *toxb* entirely.

While the BLASTN results above indicated that each *ToxB* copy was identical and closely spaced along the chromosome. However it did not clarify whether the DNA in between the copies was also included in the duplication event. Leveraging the variation observed in copy number (i.e. isolates with one, three or five tandem copies), the genomic region carrying *ToxB/toxb* was extracted and aligned to all other isolates, including self-alignments. These alignments revealed distinct edges for the start and end of the replicated sequence (Additional File 3.4). In Tptr3-1, the only isolate carrying a single copy of *ToxB* this region was 8,416 bp, but ranged from 5,798 bp (I-73-1) to 11,434 (AlgH1) in multicopy isolates. When Tptr3-1 was compared to isolates with tandem arrays of this repeat (three copies or five copies) this region was nearly identical but each copy was separated by a conserved 251 bp "sequence junction", which is not present in isolates with a single copy of *ToxB* or *toxb* (Figure 3.5). An alignment of representative *ToxB/toxb* regions is shown below in Figure 3.4. These alignments also show that *ToxB* was not the only gene duplicated within this region. In addition, two small open reading frames (ORFs) were identified, as well as a conserved hAT transposon (Figure 3.4). The ORF near the 5' end of the duplicated sequence shared homology to domains in DUF3505 and C2H2-type zinc-fingers, while the second ORF near the 3' shares homology with reverse transcriptase (RVT_1; PF00078) and RNases (RNase_H; PF00075). Both ORFs are quite small, 333 and 444 bp respectively, suggesting they may be gene fragments rather than fully functional genes.

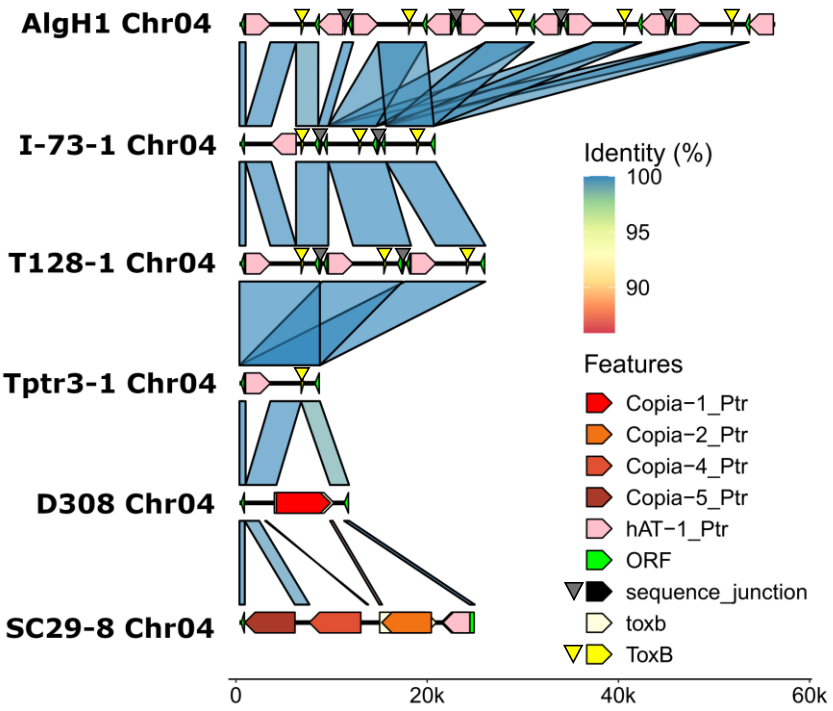


Figure 3.4 Linear alignments of the *ToxB/toxb* regions of Chr04 in *Pyrenophora tritici-repentis*. Colored arrows along the chromosome (black line) show *ToxB* (yellow), *toxb* (white), ORFs (green), transposons (reds and oranges), and the sequence junction between duplicated regions (black). Due to the alignment scale, vertical yellow and black arrows have been added to aid in locating the positions of *ToxB* and the sequence junction respectively. The grey blocks drawn between isolates show aligned DNA with >85% identity. Region duplications in isolates I-73-1 and AlgH1 are apparent when compared to the single *ToxB* containing isolate Tptr3-1. The position and number of *hAT-1_Ptr* transposons in the multi-copy isolates T128-1, I-73-1, and AlgH1 combined with the conserved edges suggest independent duplication events via the same mechanism. *ToxB* disruption via Copia retroelements is also visible in isolates D308 (*Copia-1_Ptr*) and SC29-8 (*Copia-2_Ptr*).

3.3.4 *TOXB* IS PRESENT IN A DYNAMIC LANDSCAPE OF REPETITVE ELEMENTS

In annotating the features surrounding *ToxB* and performing detailed pairwise alignments between all isolates it was clear that there were many unique sequence differences between different isolates with variable copies of *ToxB/toxb*. However, these differences were primarily driven by the nesting of other transposons near, or even within, the *ToxB* open-reading frame (Figure 3.5; Additional File 3.5). After extensive

manual curation of each sequence, we have classified 14 unique transposon insertions (5 DNA and 9 retrotransposons) near *ToxB* with an additional two insertions which appear to be transposons that we could not confidently classify (likely to be retrotransposons) (Table S1). We also identified three fragmented LTRs, two of which match the LTRs of the classified retrotransposons. Numerous copies of each were found throughout the *Ptr* genomes (Additional File 3.6). Independent annotation of these contigs after manual curation with the software EDTA confirmed the presence of these transposons and allowed quantification of the transposon density within the accessory region (Additional File 3.7).

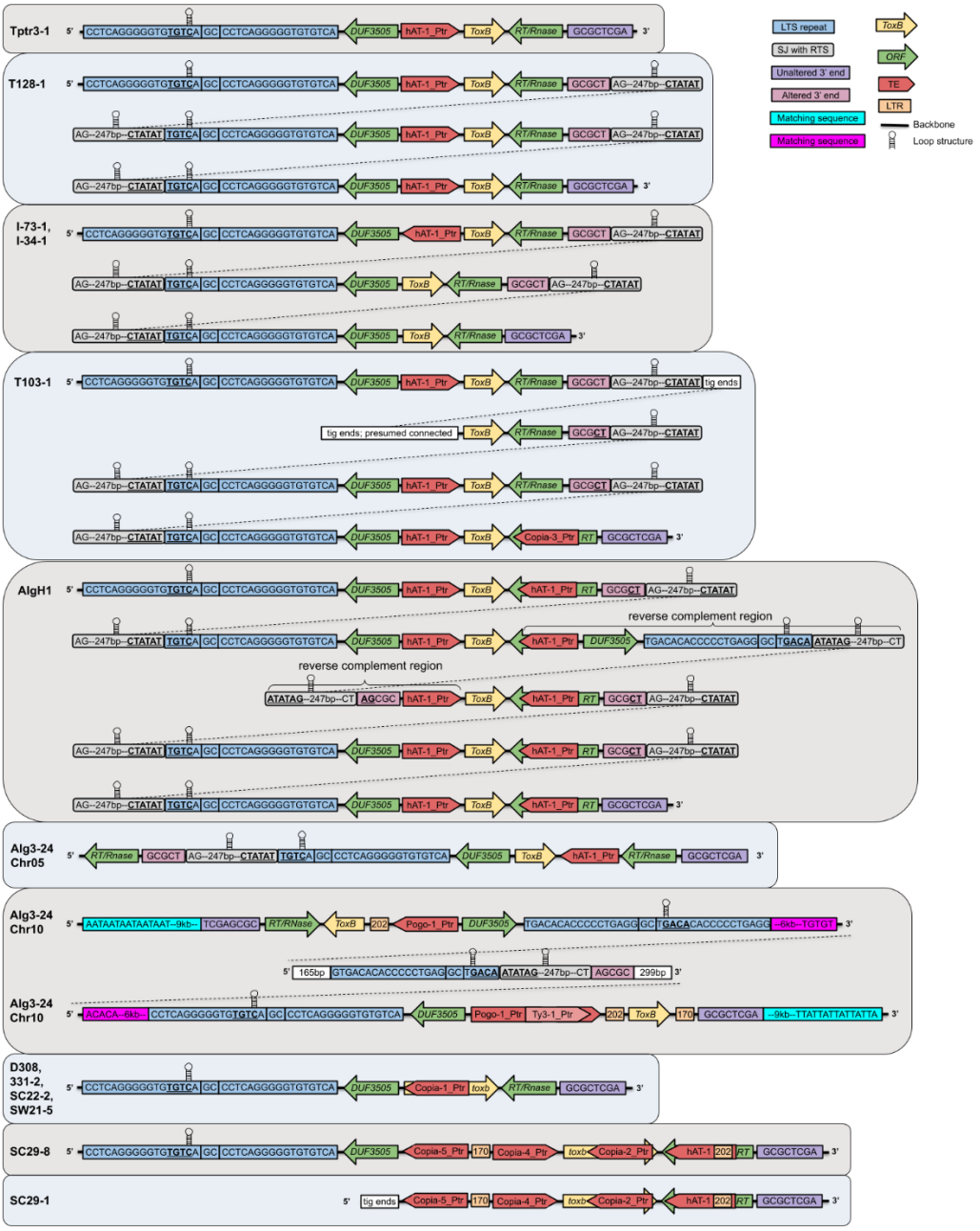


Figure 3.5 Schematics of features surrounding *ToxB* (yellow arrow) and the duplicated regions. Features which may explain region replication are noted: blue = 5' left-terminal-sequence repeat; purple = original 3' end; light pink = 3' end in duplicated regions; grey = 251 bp sequence junction with right-terminal-sequence; green = open reading frames; red = transposons; orange = abandoned LTRs; loop icon = loop structure. The sequence junction is only present in the duplicated copies. The sequence junctions holds the 3' right-terminal-sequence need for *ToxB-HLE* to replicate. Figure is not to scale only reflects orientation and relative positions.

3.3.4.1 INDEPENDENT COPIA TRANSPOSON INSERTIONS CREATE INACTIVE *TOXB* HAPLOTYPES

In an attempt to explain the inactivation of this effector, we examined each *toxb* containing contig by comparing *toxb* and its flanking sequences with copies of *ToxB* from isolates I-73-1 and Tptr3-1. Recently, we showed that a single insertion of a 5.6 kbp Copia-like element, named *Copia-1_Ptr,* disrupts the 5' end of the *ToxB* open-reading frame in isolate D308 (Hafez et al., 2024). The two isolates, SW21-5 and SC22-2, shared the same *toxb* haplotype (Hafez et al., 2024). Sequencing in this study revealed that both contained the same *Copia-1_Ptr* insertion (Figure 3.4 and 3.5). In this study, we sequenced another isolate with a different *toxb* haplotype, SC29-1, which had high variability in the 3' of the gene (Hafez et al., 2024). In this isolate we identified a novel 5.1 kbp insertion which had disrupted the *ToxB* open-reading frame (Figure 3.6). Within this 5.1 kbp insertion we identified three large ORFs which contained capsid protein (GAG), integrase (INT), retrovirus-related polyprotein (Pol), and reverse transcriptase (RT) domains. At the edges of the insertion, a four bp target site duplication was identified and the insertion site (5'-AAGT-3') was found using isolate I-73-1 that carries an uninterrupted copy of *ToxB*. Adjacent to both TSDs were 165 bp LTRs. A search of RepBase revealed relatively high homology to the retrotransposon *Copia-3_PaNo-I* reported in *Parastagonospora nodorum* (67% identity over 46% of query) (Kojima, 2023). A pairwise alignment of this region with *Copia-1_Ptr* (which has disrupted the *ToxB* ORF in D308) was performed to determined similarity with the previously reported retrotransposon. Based on the alignment of the two retrotransposons and the large difference in LTR sequences the two Copia elements are unrelated (51% pairwise identity) (Additional File 3.8) and following RepBase convention this transposon was

named *Copia-2_Ptr*. *Copia-2_Ptr* is inserted at base 222 within the *ToxB* ORF, which places it further towards the 3' end of the gene relative to the *Copia-1_Ptr* insertion in isolate D308 which was inserted at base 11. A second isolate SC29-8, that shares the same haplotype, has and identical copy of *Copia-2_Ptr* inserted in the same location. These data indicate that the inactive *toxb* haplotype has arisen independently by the insertion of two different *Copia* transposable elements.

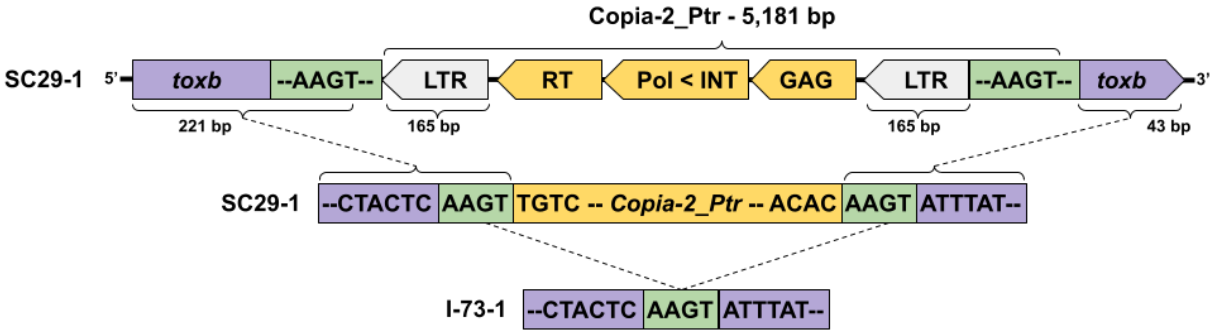


Figure 3.6 Schematic of a Copia-like transposon, *Copia-2_Ptr*, inserted within the 3' end of the *ToxB* (purple) open-reading in isolate SC29-1 creating the inactive toxb13 haplotype. A target site duplication (green) flanking long-terminal repeats (white) mark the transposon edges when compared with the non-transposon containing sequence of *ToxB* from isolate I-73-1. Open-reading frames within the transposon are marked by yellow arrows which show orientation (RT = reverse transcriptase; Pol = polyprotein; INT = integrase; GAG = capsid protein). Figure is not to scale.

3.3.4.2 HISTONE REPLICATING HAT TRANSPOSON CONFIRMS MULTIPLE INDEPENDENT TOXB DUPLICATIONS

The most commonly observed insertion within the putative *ToxB* replicative unit is 2,603 bp in length. This insertion appears near *ToxB* in the majority of the sequenced isolates, including all the multi-copy isolates and four of the single-copy (*ToxB/toxb*) isolates (Figure 3.5). At the edges of the insertions are 6 bp TSDs with the target site observable in non-insert containing versions and 17 bp TIRs adjacent to the TSDs. Four ORFs were identified within these edges with similarity to: histone H3, DNA polymerase

phi, HAT C-terminal dimerization, and topoisomerase domains (Figure 3.7). A search of RepBase returned no meaningful results. Copy numbers of this putative TE within entire genomes ranged from 0 to 35 copies, with more likely to be found if the stringent search (90% ID over 90% query) was loosened. The presence of multiple copies in close proximity likely created a composite transposon in AlgH1 as evidenced by an inversion event (Figure 3.5).

A similar, but significantly larger element (5.6 kbp) was describe by Manning et al., (2013) who previously reported the replication of histone by a putative hAT transposon near *ToxB*. Upstream of *ToxB* a larger version 5,873 bp in length was identified in isolate I-73-1. However this larger element is in fact the 2,603 bp hAT transposon with a different 3,278 bp hAT transposon nested inside (6 bp TSD [5'-ATAGTC-3'] identifiable). The nested hATs is likely what was described by Manning et al. (2013), however, the authors did not have access to other high quality assemblies for comparison. We believe the true replicative unit for histone H3 in *Ptr* is this smaller hAT which we are naming *hAT-1_Ptr*.

The activity of *hAT-1_Ptr* offers a key insight into the replication of *ToxB*, in particular when comparing the Tptr3-1, T128-1, I-73-1, I-34-1, and AlgH1. The single *ToxB* isolate Tptr3-1 and the triple-copy isolate T128-1 both contain *hAT-1_Ptr* in the exact same relative position (no matter which copy is considered) (Figure 3.4 and 3.5). In the case of T128-1 the replication of *ToxB* must have occurred after the insertion of *hAT-1_Ptr*, as three insertion events in the same relative position is highly improbable. However, in multi-*ToxB* AlgH1, there are two copies of *hAT-1_Ptr* present near *ToxB*, one upstream and one downstream. Each duplication event contains the same two copies of

hAT-1_Ptr indicating two insertions of *hAT-1_Ptr* occurred prior to the duplications. In isolates I-73-1 and I-34-1, only a single copy of *hAT-1_Ptr* is present and it is located near only one of three *ToxB* copies. This indicates that duplication of *ToxB* must have occurred prior to the insertion of *hAT-1_Ptr* in these isolates. This is clear evidence that duplication of *ToxB* and the surrounding regions has occurred multiple independent times in *Ptr* via the same mechanism.

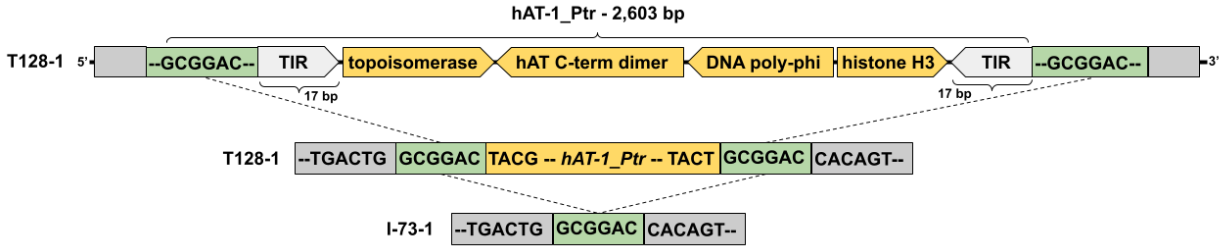


Figure 3.7 Schematic of a histone replicating hAT transposon present near *ToxB* in *Pyrenophora tritici-repentis* isolate T128-1. Target site duplications (green) flanking terminal inverted repeats (white) mark the transposon edges when compared with the non-transposon containing intergenic sequence (grey) of isolate I-73-1. Open-reading frames within the transposon are marked by yellow arrows which show orientation. Figure is not to scale.

3.3.4.3 A SMALL SEQUENCE WITH BIG IMPLICATIONS

In between each tandem repeat of the *ToxB* region, there is a small sequence of 251 bp which we refer to as the sequence junction (SJ) (Additional File 3.4). The SJ is adjacent to the *ToxB* region only in duplicated versions; in single *ToxB/toxb* carriers the SJ is not adjacent suggesting this sequence may be the lynch-pin for the duplication (Figure 3.5). The SJ is present many times (28 to 85 copies) in all *Ptr* genomes with >98% similarity over 90% of the query (Additional File 3.5; Additional File 3.6). Due to the high abundance of the element, we selected random instances for closer inspection. In some cases the SJ appears to be a conserved domain within reverse

transcriptases or LINE transposable elements (a type of retrotransposon). However, many instances of the SJ were not part of any ORF, annotated gene, or transposon. BLAST searches of NCBI and UniProt yielded limited similarity to reverse transcriptase, RNaseH1, hAT-Tnp-dimer, and DUF3824 sequences. This could indicate the SJ may be a truncated transposon which has somehow retained mobility, or is in some other manner being propagated throughout the *Ptr* genome. The potential mobility of the SJ makes untangling the observed genotypes a simpler task. In our proposed model, we assume a random insertion of the SJ near to the *ToxB* region provided the right-terminal sequence necessary to create a Helitron with the other structural features being present already (Figure 3.8).

3.3.5 DIAGNOSTIC FEATURES SUPPORT A HELITRON-LIKE-ELEMENT ASSOCIATED WITH *TOXB*

Transposons are a common mechanism by which duplications can occur. Other mechanisms of duplication are possible, such as microhomology break induced replication, however, given the unique sequence junction found between the tandemly arrayed copies that was absent in the single copies, micro-homology repair mechanisms did not seem to fit with the observed sequence data. Other mechanisms like break-fusion-bridge were also considered due to the lack of telomeric sequences on the ends of many isolates Chr04. However this was also rejected as BFB cycles produce inverted repeats in addition to the large distance that *ToxB* resides from the chromosomal end. Furthermore, the *hAT-1_Ptr* transposon indicates that whatever mechanism is driving duplication, it is not a ‘one-off’ event, with duplication occurring independently at least three times based on the observed genotypes. Given the high density of transposons in the area, we

explored the role that transposons might have played in duplicating the *ToxB* region. In isolates with tandem repeats, the edges identified via self and non-self alignments were unmatching sequences, meaning they were neither long-terminal repeats (LTR) nor terminal inverted repeats (TIR) common to most transposon families. Additionally, no discernable target-site duplication (TSD) was identifiable when compared to isolates without *ToxB/toxb* due to the genes presence in an accessory region. This excluded the *hAT* transposase located within the duplicated sequence as the driver of the duplication, as well as most described Class I or Class II transposons. Additionally, the presence of a conserved UTR 5' UTR intron (Hafez et al., 2024) would seem to rule out retropositional duplication (Magadum et al., 2013). Not all transposons have repeated edges however, nor do all create TSDs upon insertion.

We therefore expanded our search to less common transposon families. One such family of transposons are Helitrons. Helitrons are DNA transposons that are hypothesized to use a rolling-circle mode of replication. These transposons lack TSDs, are structurally asymmetrical (i.e. non-matching edges), and can produce copy-paste style tandem (or near tandem) duplications (Kaptinov and Jurka, 2001; 2007; Thomas and Pritham, 2015). Most Helitrons are non-autonomous, utilizing existing molecular machinery to mobilize themselves, however some do carry Replication initiator/helicase (RepHel) and/or replication protein A (RPA) domains (Boa and Jurka, 2013; Thomas and Pritham, 2015; Grabundzija, et al., 2016). All of these characteristics, with the exception of the RepHel/RPA domain, are observed within the repeating *ToxB* region. Furthermore, Helitrons are known to capture genes and/or gene fragments, which fits with the duplication of *ToxB* and other observed ORFs within the replicon.

To initiate replication in canonical Helitrons, the 5' end of a Helitron contains a 'TC' dinucleotide, where the Rep/Hel protein binds, known as the left-terminal sequence (LTS). This replication continues along the sequence until it reaches a 20-40 bp hairpin loop followed by a nearby 'CTRR' (R = A or G) sequence, known together as the right-terminal sequence (RTS) (Kaptinov and Jurka, 2007; Thomas and Pritham, 2015; Grabundzija, et al., 2016). The 5' end of the *ToxB* region in both single and multi-copy isolates contains two 17 bp repeats separated by a 'GC' dinucleotide (i.e. 36 bp total) (Figure 5). This 36 bp contains several potential 'TC' dinucleotide initiation sites. However in isolates with tandemly duplicated copies, this region has been truncated to 24 bp for each subsequent *ToxB* copy (Figure 3.5). In all cases, the 'TC' dinucleotide is not precisely at the defined edge which contains an additional 'TG' dinucleotide (i.e. sequence is 'TGTC') which has been observed in other fungal Helitrons (Chellapan et al., 2016). The 3' 'CTRR' sequence for canonical Helitrons is created by the 5' end of the SJ at the 3' end of the *ToxB* region which may be a potential RTS and was considered. However, it is more likely the 3' end of the SJ, whose sequence is 'CTATAT' (i.e. 'CTR TAT') is the RTS as it creates a simpler replication model (i.e. fewer steps) (Figure 3.8). The terminal 3' region copy in multi-copy isolates and in the single copy sequences is consistently 'CTCGA' (purple rectangles in Figure 5) suggests replication is occurring within the bounds of the conserved 5' 36 bp repeat and the conserved terminal 3' 'CTCGA' end (Figure 3.5).

As described, the RTS of Helitrons contains a ~20 to 40 bp GC dominant stem-loop upstream of the 3' 'CTRR' sequence. We used VectorBuilder to establish whether the sequences upstream of the putative 3' end formed such a loop structure (Additional

File 3.9). We found a multiple stem loops upstream of the SJ 3' 'CTATAT' sequence, however the most likely candidates begin 32 bp and 72 bp upstream of 'CTATAT'. The ΔG for the nearest stem loop was -4.20 and with a 57% GC content (Additional File 3.9). A ΔG of -4.20 indicates stem loop formation is spontaneous and energetically favorable but only moderately stable. The stem loop further upstream (72 bp) has a ΔG of -5.90. The second stem loop is considerably further upstream than stem loops reported in other Helitrons. The 5' end also contains a stem loop structure with a ΔG of -6.20.

Some Helitrons contain RepHel and/or RPA domains which facilitate their mobility, however we did not observe any within the putative replicon edges, indicating a non-autonomous nature. Functional gene annotations revealed at least ~80 Helicase genes are present within the *Ptr* genome, with at least three associated with Helitrons (i.e. Pif1-like helicases) (Heringer and Kuhn, 2021; Castanera et al., 2014; Thomas et al., 2014). Several DNA2-NAM7 helicases were also identified which have been linked to rolling-circle replications in viral plant pathogens but not Helitron activity specifically (Sahu et al., 2010; Maredza et al., 2016). Most annotated helicases were of the DEADbox class which have limited support for Helitron involvement in plants (Tuteja, 2003). Gene annotations also identified at least two RPA-like genes. Based on the presented features and the confirmation that replication has occurred independently multiple times (see histone replication hAT transposon above) we conclude that a **H**elitron-**L**ike **E**lement, which we've named *ToxB-HLE*, is a likely candidate to explain duplication of *ToxB* in the genome of *P. tritici-repentis*. A structural comparison of established Helitrons along with *ToxB-HLE* is provided in Figure 3.9.

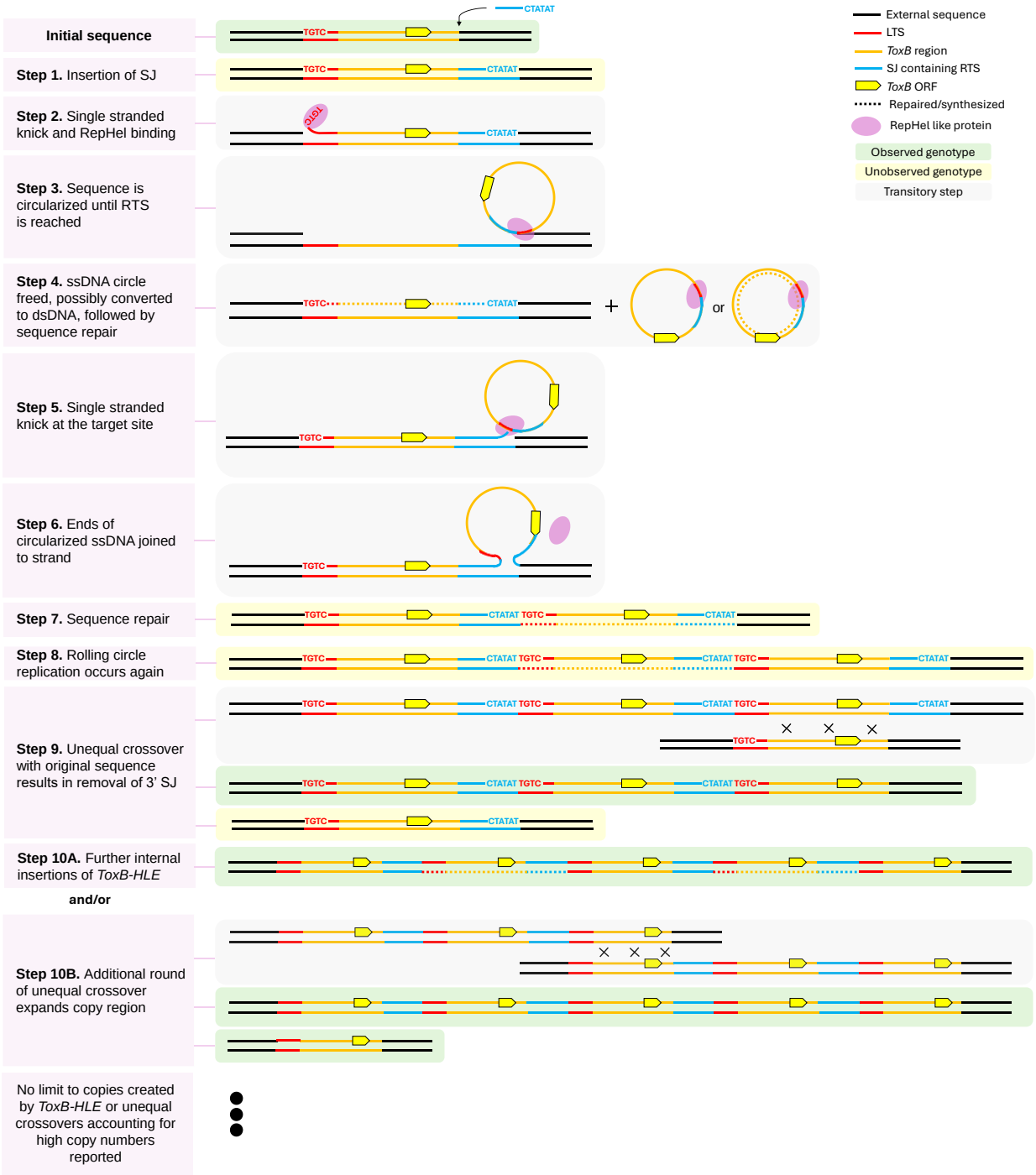


Figure 3.8 Steps which could lead to the observed genotypes through insertion, rolling-circle replication, and unequal crossing over. In this model, *ToxB-HLE* is composed of the *ToxB* region (orange) containing the left-terminal sequence 'TGTC' (red), and the sequence junction (blue) which contains the right-terminal sequence 'CTATAT' as well as the requisite hairpin loop (not pictured here). Observed genotypes are on a green background. Figure is partially adapted from Thomas and Pritham (2015), Chellepan et al., (2016) Grabundzija et al., (2016), and Harmer et al., (2022).

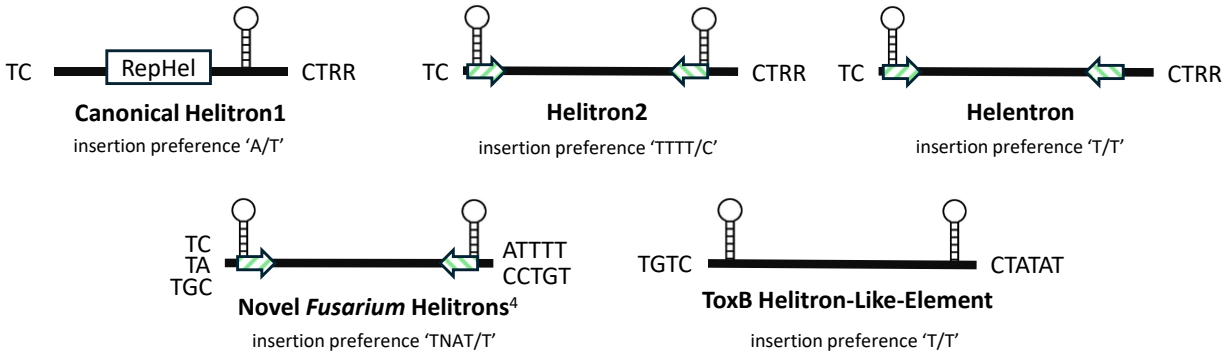


Figure 3.9 Generalized schematics of known Helitrons elements. Canonical Helitron1 adapted from Kaptinov and Jurka (2007); Helitron2 adapted from Boa and Jurka (2013); Helentron adapted from Thomas and Pritham (2015); and *Fusarium* Helitrons adapted from Chellapan et al. (2016). RepHel element may be present in all Helitrons however the majority of documented Helitrons lack these domains so are omitted from all but the canonical Helitron1. Stem loops and asymmetrical terminal inverted repeats (arrows) are shown along with the left-terminal sequences and right-terminal sequences where R = A or G.

3.4 DISCUSSION

The multi-copy nature of *ToxB* has remained largely unexplored since it was first identified (Strelkov et al. 1999; Martinez et al., 2001; Lamari et al., 2003). In an attempt to analyze *ToxB* replication, we've utilized the genomes of 23 isolates representing a time scale spanning 65 years with diverse geographical origins (Table 3.1). Many of these isolates have been used in the past 30 years to explore virulence on host plants, infection cytology, gene content, variation in effector copy numbers and *ToxB* gene silencing experiments (Strelkov et al., 2005; Amaike et al., 2008; Kim et al, 2007; 2010; Lamari et al., 2003; Aboukhaddour et al., 2009, 2012; 2013; Gourlie et al. 2022; Hafez et al., 2024). Here we provide the most comprehensive insight into the evolution of *ToxB* providing evidence that a non-autonomous Helitron-like-element, *ToxB-HLE*, is responsible for the replication of this important necrotrophic effector within an accessory region. This is an important finding as *ToxB* duplication has been directly linked to increased virulence on

wheat and is one of the few known examples of recent necrotrophic effector duplication (Strelkov et al., 2002; Martinez et al., 2004; Aboukhaddour et al., 2012). This is also an important example, in a growing number of studies, of how transposon activity appears to be driving adaptation in fungal plant pathogens, not only in virulence, but climate adaptation and spore production as well (Wang et al., 2021; Feurtey et al., 2023).

3.4.1 *TOXB* IS PRESENT IN AN ACCESSORY REGION RICH WITH TRANSPOSONS

In *Ptr*, the *ToxB* carrying chromosome is homologous to Chr04 as defined by the reference isolates Pt-1C-BFP and M4_2.2 (accession GCF_000149985 and GCF_003171515) (Manning et al., 2013; Moolhuijzen et al., 2018). In *ToxB/toxb* carrying isolates, both the active and inactive homologs were located near the ends of the carrying chromosome, generally ~400 to 600 kbp from the contig end. In isolates completely lacking *ToxB/toxb*, Chr04 appears to be intact but a large variable region (~50 to 420 kbp) harbouring *ToxB/toxb* is absent suggesting the region may be accessory in nature. The accessory genome of fungal pathogens is known to be enriched in virulence genes and contain many repetitive elements, which makes these regions subject to higher rates of rearrangement and mutations (Bertazzoni et al., 2018). Within fungal genomes, some accessory regions can be easily identified by either their high ‘AT’ or transposon content (Raffaele et al., 2010; Croll and McDonald, 2012; Dong et al., 2015; Torres et al., 2020). This led to the development of the “two-speed” or “multi-speed” genome hypotheses for fungal genomes, that these repeat-rich regions were hotbeds for the accumulation of mutations that could then be selected for under environmental or host stresses (Raffaele et al., 2010; Croll and McDonald, 2012; Dong et al., 2015; Torres et al.,

2020). Accessory regions commonly possess high numbers of transposons and repetitive elements and also tend to be subject to higher rates of rearrangements and mutations and are associated with the two-speed genome model (Raffaele et al., 2010; Croll and McDonald, 2012; Dong et al., 2015; Torres et al., 2020). Interestingly, *Ptr* does not have these signature genomic compartments based on recent work (Gourlie et al., 2022). It is possible however, that the genome of *Ptr* is in the process of compartmentalization, with the creation or acquisition of the *ToxB* accessory region being a potential early step. Alternatively, based on two isolates previously sequenced (I-73-1 and D308) it was suggested the region may be an ancient and derelict Starship due to the large size and the presence of several known Starship cargo genes. The region lacks the signature tyrosine recombinase captain however, so this hypothesis remains to be explored. Unfortunately it was beyond the scope of this paper to confirm if this accessory region is or is not an ancient Starship.

There are a number of previous examples in which virulence genes are embedded within accessory regions in the fungal clade Pleosporales, to which *Ptr* belongs. For instance, an accessory region in *Cochliobolus heterostrophus* (southern corn leaf blight) was found containing the Tox1 biosynthetic cluster, responsible for T-toxin production (Condon et al., 2013; Haridas et al., 2023). T-toxin is a race defining effector in this pathogen which significantly increases the virulence of *C. heterostrophus* (Haridas et al., 2023). Similarly in *Alternaria alternata*, various biosynthetic toxin clusters have been found within accessory regions, including ACT-toxin, ACR-toxin, AF-toxin, AM-toxin and AAL-toxin (Akamatsu et al., 1999; Hatta et al., 2002; Masunaka et al., 2005; Miyamoto et al., 2008; Akagi et al., 2009). A more comprehensive list, including fungi

outside the Pleosporales, is reviewed in Bertazzoni et al. (2018). In this context, it is not surprising that another necrotrophic effector, *ToxB*, is present in an accessory region within *Ptr*. Several of the toxins found in other species are also present as multiple copies where copy-number is directly correlated to increased virulence similar to *ToxB* (Amaike et al. 2008; Strelkov et al. 2005; Aboukhaddour et al. 2012).

Significant transposon activity was apparent within the accessory region and very near to *ToxB*. We observed independent inactivation of the effector gene through the insertion of two different Copia retrotransposons within the *ToxB* open-reading frame resulting in *toxb* haplotypes. Isolates with these *toxb* haplotypes were confirmed to have no expression (Aboukhaddour personal communication)Virulence gene disruption by transposons have been reported in other species including *Fusarium oxysporum* where the Avr gene *SIX1-H* was shown to be deactivated by a hAT transposon (Rep et al., 2005). A hAT transposon was also found to be replicating a histone H3-like protein, a component of nucleosomes and chromatin. H3 histone proteins are known to have been anciently replicated and, though they are generally conserved, can have functionally distinct variants (Waterborg, 2012). Manning et al., (2013) explored H3 replication in *Ptr* previously, and found distinct variants duplicating within the genome. Post-translational modifications to H3 variants could lead to alterations in transcription, DNA repair, and meiosis (or any process involving chromatin) (Kouzarides, 2007; Waterborg, 2012). This is also not the first case of histone proteins being replicated within an HLE, with histone gene capture observed in a Helentron of *Drosophila melanogaster* (Thomas et al., 2014). Within this dynamic genomic region *ToxB* has replicated, and we have provided evidence that a Helitron-like element (HLE) may be responsible.

3.4.2 HELITRON-LIKE ELEMENT IS LIKELY RESPONSIBLE FOR TOXB DUPLICATION

The mechanisms of gene duplication and copy number variation are relatively diverse, some of which are difficult to tease apart ex post facto (e.g. template-switching, break-fusion-bridge, illegitimate cross overs, etc.) (Reams and Roth, 2015). Through careful pair-wise alignments and sequence comparisons, we have examined and identified a number of structural features surrounding *ToxB* copies which provide strong evidence for Helitron activity.

Relative to other transposable elements, Helitrons are poorly understood and difficult to identify due to their odd characteristics (Kaptinov and Jurka, 2001; 2007; Thomas and Pritham, 2015; Grabundzija et al., 2016; Chellapan et al., 2016). As such they are likely to be under represented in transposon annotation pipelines and databases, particularly if they do not conform to the standard Helitron model. Despite this, examples of Helitrons with captured genes have been identified and documented in plant species, in for example *Zea mays*, *Oryza sativa*, *Brassica rapa,* and *Arabidposis thaliana* (Kapitonov and Jurka 2001; 2007; Sweredoski et al., 2008; Fu et al., 2013). In fungi specifically however, fewer examples have been identified. In the oyster mushroom *Pleurotus ostreatus*, HELPO1 and HELPO2 Helitron families were identified, with several HELPO1 Helitrons capturing genes with unknown functions (Castanera et al., 2014). HELPO1 also experienced incursion by retrotransposons within the Helitron boundaries, creating chimeric transposons similar to what we have observed within *Ptr ToxB-HLE* (Castanera et al., 2014). In the pathogen *Fusarium oxysporum*, Helitrons have captured at least 27 genes including a reverse transcriptase (Chellapan et al., 2016). Many novel Helitrons in *F. oxysporum* were very near to pathogenicity-related genes (e.g. SIX9a,

SIX9b, SIX6, ORX1-like, and ARG1) though none of these were captured (Chellapan et al., 2016). Other researchers have suggested that ectopic recombination or unequal crossing over caused by the presence of Helitrons may have played a role in the divergence of pathogenic races and the creation of accessory elements in *F. oxysporum* which is again reminiscent of what we observe with *Ptr* (Schmidt et al., 2013; Vlaardingerbroek et al., 2016; Biju et al., 2017). A genome-wide study of *R. commune*, found only a small number of Helitrons associated with increased copy-number (Stalder et al., 2022), however this may be due to the limited ability of pipelines to annotate these odd elements. In *P. tritici-repentis*, Helitrons appear to comprise only ~0.5% of the genome though this is likely a low estimation (Gourlie et al., 2022).

The first notable feature of the *ToxB* replication is the tandem nature of the duplication. Helitrons transpose via rolling-circle replication which is capable of generating such tandem repeats and is what allows the transposon to capture neighboring genes (Thomas and Pritham, 2015; Grabundzija et al., 2016; Xiong et al., 2016). While Helitrons do not possess repetitive edges as often seen in other transposons (i.e. LTRs in retrotransposon or TIRs in DNA transposons) they do possess key base pair sequences at their 5' and 3' ends, though both ends have been shown they can be variable in other fungal species (Chellapan et al., 2016). Standard Helitrons contain at their 5' end a 'TC' dinucleotide, while *ToxB-HLE* contains a conserved 'TGTC' sequence at the 5' end. Some Helitrons from *Fusarium oxysporum* however were found to have 'TA' or 'TGCCT' 5' sequences (Chellapan et al., 2016). Similarly, we observed the sequence 'CTATAT' at the 3' end of the *ToxB-HLE*. Again, Helitrons in *F. oxysporum* had 3' sequences as diverse as 'CTCCTGT' (Chellapan et al., 2016). The presence of a stem

loop near the 3' end and, in some cases, the 5' end as well, is also diagnostic feature and we have found both are present in *ToxB-HLE* (Thomas and Pritham, 2015). The 5' loop could indicate the *ToxB-HLE* belongs to the subgroup Helitron2 (or possibly Helentrons), however the characteristic asymmetrical inverted repeats (ATIR) of Helitron2 were not found (Boa and Jurka, 2013). The 5' loop may be coincidental, or *ToxB-HLE* may share only partial features of Helitron2, perhaps being an intermediate type between Helitron1 and Helitron2. Helitron1 elements have shown a preference for insertion between an 'A/T' dinucleotides while Helitron2 and Helentrons have a preference for 'T/T' or 'T/C' insertions (Thomas et al., 2014; Thomas and Pritham, 2015; Grabundzija, et al., 2016; Chellapan et al., 2016). Grabundzija et al., (2016) found *Helraiser* insertions between almost all possible dinucleotide pairs. In our model we expect *ToxB-HLE* to have a preference for 'T/T' dinucleotide insertion sites. The lack of the RepHel/RPA domains is noteworthy but it has been established that the majority of Helitrons are non-autonomous and therefore most will lack these domains, and we were able to find helicases and RPA genes within the *Ptr* genome which may support *ToxB-HLE* replication (Boa and Jurka, 2013). The activity of *ToxB-HLE* may not be the only mechanism implicated in *ToxB* copy-number variation. As the copy-number of *ToxB-HLE* increases it is possible other mechanisms may also begin to replicate *ToxB* such as compound transposons or unequal crossing over (Figure 8). Furthermore, as the region expanded with higher numbers of nested transposons in some isolates, other unexplained mechanisms may also replicate *ToxB* as seen in Alg3-24 where features of *ToxB-HLE* present but scattered within a large duplicated region which is also inverted (Figure 5). The structure surrounding copies in Alg3-24 could also be explained by genomic rearrangements which have broken up the

previously tandem copies. Additionally, in isolate 92-171R5 we observed two copies of *ToxB* but did not find any signature of *ToxB-HLE* in the region, rather the duplicated regions are flanked by undefined LTR retrotransposons (as annotated by EDTA) (Additional File 3.10). The poor coverage of this region in 92-171R5 combined with the abundance of transposons makes it difficult to draw confident conclusions for this isolate. However if the assembly is accurate, its possible the duplication of *ToxB* was driven by retroposition in 92-171R5.

3.4.3 UNEQUAL CORSSING OVER MAY PLAY A ROLE IN FURTHER REPLICATION *TOXB*

During meiosis, homologous chromosomes pair and undergo recombination, exchanging large sections of genetic material by physical crossing over at homologous regions. This process relies on the accurate pairing of homologous sequences; however, in regions with high repeat content misalignment can occur, leading to unequal crossover events in which genetic material is unevenly exchanged between chromosomes. Uneven transfer leads to a duplication in one chromosome and a deletion in the other. Tandemly duplicated regions become more likely to undergo further unequal crossing events due to the increased homology in the region, expanding the number of tandem duplications (Anderson and Roth, 1977; Silver, 2001; Taylor and Raes, 2004). These duplicated regions can vary greatly in size, from small sections of a few hundred base pairs to tens of thousands. We know that *ToxB* is located in a region with a high amount transposon activity, and it has been shown that transposons are able to provide the repetitive sequences necessary for unequal crossing over or ectopic recombination to occur. As an early example, Roeder (1983) showed unequal crossing over was possible between Ty

elements in yeast (Ty1 are Copia LTR retrotransposons). It has also been suggested that homothallics may themselves arise from unequal crossing over of heterothallics (Yun et al., 1999). For example, in some *Neurospora* species, unequal crossing over of mating type genes facilitated by LTR retrotransposons appear to have prompted the transition from heterothallism to homothallism (Gioti et al., 2012). Furthermore, the genome of *F. oxysporum* is hypothesized to have developed accessory regions and pathogenicity gene gains/losses (race changes) linked to unequal crossing over caused by repetitive elements (Biju et al., 2017). It bears repeating here that the variable positions and numbers of the transposon *hAT-1_Ptr* indicate the duplication event is a conserved mechanism which has occurred multiple times independently.

Unequal crossing over seems likely to have played a role in *ToxB* replication due to the number of transposons and repetitive elements which creates homology in the *ToxB* region on which the process could act. This process is able to create tandem duplications such as those observed with *ToxB* (and its surrounding region) (Figure 8). The mechanism also allows for large changes in copy-number variation which is as seen with *ToxB*, with the copy-number able to double with each successive event. It is also difficult to explain why the terminal copies do not have the sequence junction unless has been removed by unequal crossing over or perhaps through remodelling processes (Reams and Roth, 2015). Additionally, unequal crossover may account for some isolates of *Ptr* which completely lack *ToxB* or its inactive homolog, *toxb*, by creating a duplicate in one genome and a deletion in the other. We know, as a homothallic (selfing) species, *Ptr* can easily undergo sexual reproduction via the creation of ascospore containing pseudothecia

which are abundant on crop stubble so there is ample opportunity for such events to occur.

3.4.4 THE ROAD AHEAD

Over 25 years after the identification of *ToxB* we have provided the most detailed examination of the duplicated DNA region and provided a possible model to explain *ToxB* duplication and pseduogenisation in *Ptr*. This work highlights the double-edged nature of transposable elements, providing gene duplication and gene disruption (Fouché et al., 2022). The evolutionary history of *ToxB* and its homolog *toxb* in Ptr is still presenting an intriguing puzzle.

Unlike *ToxB*, which is unique to Ptr and primarily exists as a multicopy gene, *toxb* exists in numerous species within the Pleosporales order, with some species even carrying multiple copies (Hafez et al., 2024). In North America, *toxb* is abundant in Ptr and in other related species collected from grasses, whereas the *ToxB* gene is almost entirely absent in this region. In contrast, *ToxB* is prevalent in isolates from the Fertile Crescent, North Africa, and surrounding regions, suggesting a geographical adaptation and likely these two forms of the gene are following separate or independent evolutionary trajectories.. Interestingly, some of the explored Canadian Ptr isolates in this study, were collected mainly from durum and were shown to carry the non-functional *toxb* alongside other effectors, such as ToxC.. This raises the question of whether *ToxB* in the wheat infecting isolates is undergoing purfiying selection in North America, or if the lack of active *ToxB* on the continent is the result of a genetic founder effect. Cultivars sensitive to *ToxB* are widely grown in Canada, suggesting that retaining a functional *ToxB* would provide a fitness advantage which is support for a founder effect rather than selection against *ToxB*. To resolve this evolutionary mystery, further research on the fitness cost

related to *ToxB* in Ptr may be needed however. Additionally, studies on grass-infecting isolates are essential to better understand the diversity and distribution of *toxb*. Expanding sampling efforts beyond North America to underexplored regions may offer critical insights into the origins and evolutionary pressures shaping *ToxB* and its homologs.

Despite the extensive analysis, the exact replication mechanism of *ToxB* has yet to be determined with complete certainty. Long-read sequencing improved our confidence in determining the number of *ToxB* copies carried in a given *Ptr* genome. However, challenges persist with some isolates due to the high transposon and repeat content surrounding *ToxB*. To address these issues we suggest that future sequencing efforts utilize higher than average coverage to help ensure these regions can be assembled. Additionally, incorporating Hi-C data would also be advantageous for resolving existing assembly challenges, such as seen with Alg3-24, which have high copy numbers that remain unresolved.

The development of a qPCR protocol for determining *ToxB* copy number is a logical step forward in understanding the *ToxB* effector and it's evolution. Using isolates from our collection with high confidence copy numbers should allow for qPCR curve comparison and rapid copy number determination in new isolates. Such a protocol would also help in confirming that future genome assemblies reflect the correct number of *ToxB* copies. Combining a qPCR protocol with a crossing-population, it should be possible to identify isolates with novel *ToxB* duplications by comparing parents with progeny. Such an approach may help solidify our proposed model as we catch ToxB duplication in-the-

act. A qPCR protocol would also facilitate more comprehensive studies on the relationship between *ToxB* copy-number and virulence.

In fungal pathogens in particular, duplications have been reported in various metabolic systems, however rarely is a specific of mechanism identified (Powell et al., 2008; Stalder et al., 2022). Here we have provided a large body of evidence to support the hypothesis that multiple types of transposons act to both amplify and disrupt an important necrotrophic effector in a global plant pathogen. In *Ptr*, it is unclear how long the these duplicated versions of *ToxB* will be maintained before they are either lost, pseudogenized, or retained. However, the sixteen year time span between the collection of multi-copy isolates I-73-1 (2001; Syria) and T128-1 (2017; Tunisia) hints that additional copies are more likely to be maintained rather than lost or pseudogenized. Copy retention and conservation is likely linked to the increased fitness the extra copies provide on susceptible hosts. However, should additional copies no longer provide an advantage, such redundancy is unlikely to be preserved and additional copies would be free to mutate and diverge (Silver, 2001; Taylor and Raes, 2004). Therefore, it is possible that in the future some *ToxB* copies may evolve to become novel effectors with altered functions or new host target sites.

3.5 REFERENCES

Aboukhaddour, R., Cloutier, S., Ballance, G.M. and Lamari, L., 2009. Genome characterization of *Pyrenophora tritici-repentis* isolates reveals high plasticity and independent chromosomal location of *ToxA* and *ToxB*. *Molecular Plant Pathology*, **10**: 201-212.

Aboukhaddour, R., Kim, Y.M. and Strelkov, S.E., 2012. RNA-mediated gene silencing of *ToxB* in *Pyrenophora tritici-repentis*. *Molecular Plant Pathology*, **13**: 318-326.

Aboukhaddour, R., Turkington, T.K. and Strelkov, S.E., 2013. Race structure of *Pyrenophora triciti-repentis* (tan spot of wheat) in Alberta, Canada. *Canadian Journal of Plant Pathology*, **35**: 256-268.

Aboukhaddour, R., Hafez, M., McDonald, M., Moffat, C.S., Navathe, S., Friesen, T.L., Strelkov, S.E., Oliver, R.P., Tan, K.C., Liu, Z. and Moolhuijzen, P.M., 2023. A revised nomenclature for *ToxA* haplotypes across multiple fungal species. *Phytopathology*, **113**: 1180-1184.

Akamatsu, H., Taga, M., Kodama, M., Johnson, R., Otani, H. and Kohmoto, K., 1999. Molecular karyotypes for *Alternaria* plant pathogens known to produce host-specific toxins. *Current Genetics*, **35:** 647-656.

Akagi, Y., Akamatsu, H., Otani, H. and Kodama, M., 2009. Horizontal chromosome transfer, a mechanism for the evolution and differentiation of a plant-pathogenic fungus. *Eukaryotic Cell*, **8**: 1732-1738.

Amaike, S., Ozga, J.A., Basu, U. and Strelkov, S.E., 2008. Quantification of *ToxB* gene expression and formation of appressoria by isolates of *Pyrenophora tritici-repentis* differing in pathogenicity. *Plant Pathology*, ***57***: 623-633.

Anderson, R.P. and Roth, J.R., 1977. Tandem genetic duplications in phage and bacteria. *Annual Review of Microbiology*, **31**: 473-505.

Arnold, C.J., Hahn, E.A., Whetten, R., Chartrain, L., Cheema, J., Brown, J.K. and Cowger, C., 2024. Multiple routes to fungicide resistance: Interaction of Cyp51 gene sequences, copy number and expression. *Molecular Plant Pathology*, **25**: e13498.

Bao, W. and Jurka, J., 2013. Homologues of bacterial TnpB_IS605 are widespread in diverse eukaryotic transposable elements. *Mobile DNA*, **4**: 1-16.

Bertazzoni, S., Williams, A.H., Jones, D.A., Syme, R.A., Tan, K.C. and Hane, J.K., 2018. Accessories make the outfit: accessory chromosomes and other dispensable DNA regions in plant-pathogenic fungi. *Molecular Plant-Microbe Interactions*, *31*(8: 779-788.

Berthelot, C., Brunet, F., Chalopin, D., Juanchich, A., Bernard, M., Noël, B., Bento, P., Da Silva, C., Labadie, K., Alberti, A. and Aury, J.M., 2014. The rainbow trout genome

provides novel insights into evolution after whole-genome duplication in vertebrates. *Nature Communications*, **5**: 1-10.

Bhat A, Ghatage T, Bhan S, Lahane GP, Dhar A, Kumar R, Pandita RK, Bhat KM, Ramos KS, Pandita TK. 2022. Role of Transposable Elements in Genome Stability: Implications for Health and Disease. *International Journal of Molecular Science*, **23**: 7802.

Biju, V.C., Fokkens, L., Houterman, P.M., Rep, M. and Cornelissen, B.J., 2017. Multiple evolutionary trajectories have led to the emergence of races in *Fusarium oxysporum* f. sp. *lycopersici*. *Applied and Environmental Microbiology*, **83**: e02548-16.

Blomme, T., Vandepoele, K., De Bodt, S., Simillion, C., Maere, S. and Van de Peer, Y., 2006. The gain and loss of genes during 600 million years of vertebrate evolution. *Genome Biology*, **7**: 1-12.

Bowers J. E., Chapman B. A., Rong J. and Paterson A. H. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* **422**: 433–438.

Carbone, I., Ramirez-Prado, J.H., Jakobek, J.L. and Horn, B.W., 2007. Gene duplication, modularity and adaptation in the evolution of the aflatoxin gene cluster. *BMC Evolutionary Biology*, **7**: 1-12.

Castanera, R., Pérez, G., López, L., Sancho, R., Santoyo, F., Alfaro, M., Gabaldón, T., Pisabarro, A.G., Oguiza, J.A. and Ramírez, L., 2014. Highly expressed captured genes and cross-kingdom domains present in Helitrons create novel diversity in *Pleurotus ostreatus* and other fungi. *BMC Genomics*, **15**: 1-17.

Chellapan, B.V., van Dam, P., Rep, M., Cornelissen, B.J. and Fokkens, L., 2016. Non-canonical Helitrons in *Fusarium oxysporum*. *Mobile DNA*, **7**: 1-16.

Condon, B.J., Leng, Y., Wu, D., Bushley, K.E., Ohm, R.A., Otillar, R., Martin, J., Schackwitz, W., Grimwood, J., MohdZainudin, N. and Xue, C., 2013. Comparative genome structure, secondary metabolite, and effector coding capacity across *Cochliobolus* pathogens. *PLoS Genetics*, **9***:* e1003233.

Consortium, U., 2015. UniProt: a hub for protein information. *Nucleic Acids Research*, **43**: D204-D212.

Croll, D. and McDonald, B.A., 2012. The accessory genome as a cradle for adaptive evolution in pathogens. *PLoS Pathogens*, **8**: e1002608.

Darling, A.E., Mau, B. and Perna, N.T., 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PloS One*, **5**: e11147.

Dong, S., Raffaele, S. and Kamoun, S., 2015. The two-speed genomes of filamentous pathogens: waltz with plants. *Current Opinion in Genetics & Development*, **35**: 57-65.

Eichler, E.E. and Sankoff, D., 2003. Structural dynamics of eukaryotic chromosome evolution. *Science*, **301**: 793-797.

Fan, X., Dai, R.C., Zhang, S., Geng, Y.Y., Kang, M., Guo, D.W., Mei, Y.N., Pan, Y.H., Sun, Z.Y., Xu, Y.C. and Gong, J., 2023. Tandem gene duplications contributed to high-level azole resistance in a rapidly expanding *Candida tropicalis* population. *Nature Communications*, **14:** 8369.

Faris, J.D., Liu, Z. and Xu, S.S., 2013. Genetics of tan spot resistance in wheat. *Theoretical and Applied Genetics*, **126**: 2197-2217.

Feurtey, A., Lorrain, C., McDonald, M.C., Milgate, A., Solomon, P.S., Warren, R., Puccetti, G., Scalliet, G., Torriani, S.F., Gout, L. and Marcel, T.C., 2023. A thousand-genome panel retraces the global spread and adaptation of a major fungal crop pathogen. *Nature Communications*, **14**: 1059.

Fouché, S., Plissonneau, C. and Croll, D., 2018. The birth and death of effectors in rapidly evolving filamentous pathogen genomes. *Current Opinion in Microbiology*, **46**: 34-42.

Fouché, S., Oggenfuss, U., Chanclud, E. and Croll, D., 2022. A devil's bargain with transposable elements in plant pathogens. *Trends in Genetics*, **38**: 222-230.

Fu, D., Wei, L., Xiao, M. and Hayward, A., 2013. New insights into Helitron transposable elements in the mesopolyploid species *Brassica rapa*. *Gene*, **532**: 236-245.

Gioti, A., Mushegian, A.A., Strandberg, R., Stajich, J.E. and Johannesson, H., 2012. Unidirectional evolutionary transitions in fungal mating systems and the role of transposable elements. *Molecular Biology and Evolution*, **29**: 3215-3226.

Gourlie, R., McDonald, M., Hafez, M., Ortega-Polo, R., Low, K.E., Abbott, D.W., Strelkov, S.E., Daayf, F. and Aboukhaddour, R., 2022. The pangenome of the wheat pathogen *Pyrenophora tritici-repentis* reveals novel transposons associated with necrotrophic effectors *ToxA* and *ToxB*. *BMC Biology*, **20**: 239.

Grabundzija, I., Messing, S.A., Thomas, J., Cosby, R.L., Bilic, I., Miskey, C., Gogol-Döring, A., Kapitonov, V., Diem, T., Dalda, A. and Jurka, J., 2016. A Helitron transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. *Nature Communications*, **7**: 10716.

Gurevich A, Saveliev V, Vyahhi N, and Tesler G., 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, **29**: 1072–1075.

Hafez, M., Despins, T., Nakajima, K. and Aboukhaddour, R., 2022. Identification of a novel *ToxA* haplotype of *Pyrenophora* tritici-repentis from Japan. *Phytopathology*, **112**: 1597-1602.

Hackl, T., Ankenbrand, M., van Adrichem, B., Wilkins, D. and Haslinger, K., 2024. gggenomes: effective and versatile visualizations for comparative genomics. *arXiv preprint arXiv:2411.13556*.

Hafez, M., Gourlie, R., McDonald, M., Telfer, M., Carmona, M.A., Sautua, F.J., Moffat, C.S., Moolhuijzen, P.M., See, P.T. and Aboukhaddour, R., 2024. Evolution of the *ToxB* gene in *Pyrenophora tritici-repentis* and related species. *Molecular Plant-Microbe Interactions*, **37**: 327-337.

Haridas, S., González, J.B., Riley, R., Koriabine, M., Yan, M., Ng, V., Rightmyer, A., Grigoriev, I.V., Baker, S.E. and Turgeon, B.G., 2023. T-toxin virulence genes: unconnected dots in a sea of repeats. *MBio*, **14**: e00261-23.

Harmer, C.J., Lebreton, F., Stam, J., McGann, P.T. and Hall, R.M., 2022. Mechanisms of IS 26-mediated amplification of the *aphA1* gene leading to tobramycin resistance in an *Acinetobacter baumannii* isolate. *Microbiology Spectrum*, **10**: e02287-22.

Harris, R.S., 2007. *Improved pairwise alignment of genomic DNA*. The Pennsylvania State University.

Hastings, P. J., Lupski, J. R., Rosenberg, S. M., & Ira, G. 2009. Mechanisms of change in gene copy number. *Nature Reviews Genetics*, **10**: 551– 564.

Hatta, R., Ito, K., Hosaki, Y., Tanaka, T., Tanaka, A., Yamamoto, M., Akimitsu, K., and Tsuge, T. 2002. A conditionally dispensable chromosome controls host-specific pathogenicity in the fungal plant pathogen *Alternaria alternata*. *Genetics,* **161**: 59-70

Hawkins, N.J. and Fraaije, B.A., 2018. Fitness penalties in the evolution of fungicide resistance. *Annual Review of Phytopathology*, **56**: 339-360.

Heringer, P. and Kuhn, G.C., 2022. Pif1 helicases and the evidence for a prokaryotic origin of Helitrons. *Molecular Biology and Evolution*, **39**: msab334.

Holland, P.W., 1999, October. Gene duplication: past, present and future. In *Seminars in Cell & Developmental Biology*, **10**: 41-547. Academic Press.

International Rice Genome Sequencing Project, 2005. The map-based sequence of the rice genome. *Nature*, **436**: 793-800.

Inami, K., Yoshioka-Akiyama, C., Morita, Y., Yamasaki, M., Teraoka, T. and Arie, T., 2012. A genetic mechanism for emergence of races in *Fusarium oxysporum* f. sp.

lycopersici: inactivation of avirulence gene *AVR1* by transposon insertion. *PLoS One*, **7**: e44101.

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Hugueney P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyère C, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Del Fabbro C, Alaux M, Di Gaspero G, Dumas V, et al., 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, **449**: 463-467.

Kamel, S., Cherif, M., Hafez, M., Despins, T. and Aboukhaddour, R., 2019. *Pyrenophora tritici–repentis* in Tunisia: race structure and effector genes. *Frontiers in Plant Science*, **10**: 1562.

Kapitonov, V.V. and Jurka, J., 2001. Rolling-circle transposons in eukaryotes. *Proceedings of the National Academy of Sciences*, **98**: 8714-8719.

Kapitonov, V.V. and Jurka, J., 2007. Helitrons on a roll: eukaryotic rolling-circle transposons. *Trends in Genetics*, **23**: 521-529.

Katoh, K. and Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, **30**: 772-780.

Kellis M., Birren B. W. and Lander E. S. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cereviseae*. *Nature*, 428, 617–624.

Kim, Y.M. and Strelkov, S.E., 2007. Heterologous expression and activity of Ptr *ToxB* from virulent and avirulent isolates of *Pyrenophora tritici-repentis*. *Canadian Journal of Plant Pathology*, **29**: 232-242.

Kim, Y.M., Bouras, N., Kav, N.N. and Strelkov, S.E., 2010. Inhibition of photosynthesis and modification of the wheat leaf proteome by Ptr *ToxB*: a host-specific toxin from the fungal pathogen *Pyrenophora tritici-repentis*. *Proteomics*, **10**: 2911-2926.

Kimura, M. 1983. The Neutral Theory of Molecular Evolution. Cambridge Univ. Press, Cambridge

Kojima, K.K., 2023. Copia-3_PaNo-I LTR retrotransposon from the *Parastagonospora nodorum* genome - internal portion consensus. *Repbase Reports*, **20**: 829

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H. and Phillippy, A.M., 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*, **27**: 722-736.

Kouzarides, T., 2007. Chromatin modifications and their function. *Cell*, **128**: 693-705.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J. and Marra, M.A., 2009. Circos: an information aesthetic for comparative genomics. *Genome Research*, **9**: 1639-1645.

Lamari, L. and Bernier, C.C., 1989. Toxin of *Pyrenophora tritici-repentis*: host-specificity, significance in disease. *Phytopathology*, **79**: 740-744.

Lamari, L., Sayoud, R., Boulif, M. and Bernier, C.C., 1995. Identification of a new race in *Pyrenophora tritici-repentis*: implications for the current pathotype classification system. *Canadian Journal of Plant Pathology*, **17**: 312-318.

Lamari, L., Strelkov, S.E., Yahyaoui, A., Orabi, J. and Smith, R.B., 2003. The identification of two new races of *Pyrenophora tritici-repentis* from the host center of diversity confirms a one-to-one relationship in tan spot of wheat. *Phytopathology*, **93**: 391-396.

Langmead, B. and Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods*, **9**: 357-359.

Lawrence, J.G., 2005. Common themes in the genome strategies of pathogens. *Current Opinion in Genetics & Development*, **15**: 584-588.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup, 2009. The sequence alignment/map format and SAMtools. B*ioinformatics*, **25**: 2078-2079.

Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, **34**: 3094-3100.

Liu, X., Yu, F., Schnabel, G., Wu, J., Wang, Z. and Ma, Z., 2011. Paralogous *cyp51* genes in *Fusarium graminearum* mediate differential sensitivity to sterol demethylation inhibitors. *Fungal Genetics and Biology*, **48**: 113-123.

Long, M., Betrán, E., Thornton, K. and Wang, W., 2003. The origin of new genes: glimpses from the young and old. *Nature Reviews Genetics*, **4**: 865-875.

Magadum, S., Banerjee, U., Murugan, P., Gangapur, D. and Ravikesavan, R., 2013. Gene duplication as a major force in evolution. *Journal of Genetics*, ***92***: 155-161.

Manning, V.A., Pandelova, I., Dhillon, B., Wilhelm, L.J., Goodwin, S.B., Berlin, A.M., Figueroa, M., Freitag, M., Hane, J.K., Henrissat, B. and Holman, W.H., 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: Genes, Genomes, Genetics*, **3**: 41-63.

Maredza, A.T., Allie, F., Plata, G. and Rey, M.E.C., 2016. Sequences enhancing cassava mosaic disease symptoms occur in the cassava genome and are associated with South African cassava mosaic virus infection. *Molecular Genetics and Genomics*, **291**: 1467-1485.

Martinez, J.P., Ottum, S.A., Ali, S., Francl, L.J. and Ciuffetti, L.M., 2001. Characterization of the *ToxB* gene from *Pyrenophora tritici-repentis*. *Molecular Plant-Microbe Interactions*, **14**: 675-677.

Martinez, J.P., Oesch, N.W. and Ciuffetti, L.M., 2004. Characterization of the multiple-copy host-selective toxin gene, *ToxB*, in pathogenic and nonpathogenic isolates of *Pyrenophora tritici-repentis*. *Molecular Plant-Microbe Interactions*, **17**: 467-474.

Masunaka, A., Ohtani, K., Peever, T. L., Timmer, L. W., Tsuge, T., Yamamoto, M., Yamamoto, H., and Akimitsu, K. 2005. An isolate of *Alternaria alternata* that is pathogenic to both tangerines and rough lemon and produces two host-selective toxins, ACT- and ACR-toxins. *Phytopathology*, **95**: 241-247.

Mohd-Assaad, N., McDonald, B.A. and Croll, D., 2019. The emergence of the multi-species NIP1 effector in *Rhynchosporium* was accompanied by high rates of gene duplications and losses. *Environmental Microbiology*, **21**: 2677-2695.

Moolhuijzen, P., See, P.T., Hane, J.K., Shi, G., Liu, Z., Oliver, R.P. and Moffat, C.S., 2018. Comparative genomics of the wheat fungal pathogen *Pyrenophora tritici-repentis* reveals chromosomal variations and genome plasticity. *BMC Genomics*, **19**: 1-23.

Minkin, I., Patel, A., Kolmogorov, M., Vyahhi, N. and Pham, S., 2013. Sibelia: a scalable and comprehensive synteny block generation tool for closely related microbial genomes. In *Algorithms in Bioinformatics: 13th International Workshop, WABI 2013, Sophia Antipolis, France, September 2-4, 2013. Proceedings,* **13**: 215-229. Springer Berlin Heidelberg.

Mita, P. and Boeke, J.D., 2016. How retrotransposons shape genome regulation. *Current Opinion in Genetics & Development*, **37**: 90-100.

Miyamoto, Y., Masunaka, A., Tsuge, T., Yamamoto, M., Ohtani, K., Fukumoto, T., Gomi, K., Peever, T. L., and Akimitsu, K. 2008. Functional analysis of a multicopy host-selective ACT-toxin biosynthesis gene in the tangerine pathotype of *Alternaria alternata* using RNA silencing. *Molecular Plant-Microbe Interactions*, **21**: 1591-1599.

Nei, M., 1969. Gene duplication and nucleotide substitution in evolution. *Nature*, **221**: 40-42.

Nurk, S., Walenz, B.P., Rhie, A., Vollger, M.R., Logsdon, G.A., Grothe, R., Miga, K.H., Eichler, E.E., Phillippy, A.M. and Koren, S., 2020. HiCanu: accurate assembly of

segmental duplications, satellites, and allelic variants from high-fidelity long reads. *Genome Research*, **30**: 1291-1305.

Ohno, S., 2013. Evolution by gene duplication. Springer Science & Business Media

Orolaza, N.P., Lamari, L. and Ballance, G.M., 1995. Evidence of a host-specific chlorosis toxin from *Pyrenophora tritici-repentis*, the causal agent of tan spot of wheat. *Phytopathology*, **85**: 1282-1287.

Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J.R., Hellinga, A.J., Lugo, C.S.B., Elliott, T.A., Ware, D., Peterson, T. and Jiang, N., 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology*, **20**: 1-18.

Porten T., 2018. dotPlotly. GitHub, https://github.com/tpoorten/dotPlotly.

Powell, A.J., Conant, G.C., Brown, D.E., Carbone, I. and Dean, R.A., 2008. Altered patterns of gene duplication and differential gene gain and loss in fungal pathogens. *BMC Genomics*, **9**: 1-15.

Proctor, R.H., McCormick, S.P., Kim, H.S., Cardoza, R.E., Stanley, A.M., Lindo, L., Kelly, A., Brown, D.W., Lee, T., Vaughan, M.M. and Alexander, N.J., 2018. Evolution of structural diversity of trichothecenes, a family of toxins produced by plant pathogenic and entomopathogenic fungi. *PLoS Pathogens*, **14**: e1006946.

Raffaele, S., Farrer, R.A., Cano, L.M., Studholme, D.J., MacLean, D., Thines, M., Jiang, R.H., Zody, M.C., Kunjeti, S.G., Donofrio, N.M. and Meyers, B.C., 2010. Genome evolution following host jumps in the Irish potato famine pathogen lineage. *Science*, **330**: 1540-1543.

Reams, A.B. and Roth, J.R., 2015. Mechanisms of gene duplication and amplification. *Cold Spring Harbor Perspectives in Biology*, **7**: a016592.

Rep, M., van der Does, H.C. and Cornelissen, B.J., 2005. Drifter, a novel, low copy hAT-like transposon in *Fusarium oxysporum* is activated during starvation. *Fungal Genetics and Biology*, **42**: 546-553.

Roeder, G.S., 1983. Unequal crossing-over between yeast transposable elements. *Molecular and General Genetics MGG*, **190**: 117-121.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. and Mesirov, J.P., 2011. Integrative genomics viewer. *Nature Biotechnology*, **29**: 24-26.

Robens J, Cardwell KF: The costs of mycotoxin management in the United States. In *Aflatoxin and Food Safety* Edited by: Abbas HK. Boca Raton , CRC Taylor & Francis; 2005:1-12.

Sacristán, S., Goss, E.M. and Eves-van den Akker, S., 2021. How do pathogens evolve novel virulence activities?. *Molecular Plant-Microbe Interactions*, **34**: 576-586.

Sahu, P.P., Rai, N.K., Chakraborty, S., Singh, M., Chandrappa, P.H., Ramesh, B., Chattopadhyay, D. and Prasad, M., 2010. Tomato cultivar tolerant to Tomato leaf curl New Delhi virus infection induces virus-specific short interfering RNA accumulation and defence-associated host gene expression. *Molecular Plant Pathology*, **11**: 531-544.

Schmidt, S.M., Houterman, P.M., Schreiver, I., Ma, L., Amyotte, S., Chellappan, B., Boeren, S., Takken, F.L. and Rep, M., 2013. MITEs in the promoters of effector genes allow prediction of novel virulence genes in *Fusarium oxysporum*. *BMC Genomics*, **14**: 1-21.

Silver L., 2001. Evolution of Gene Families, Editor(s): Sydney Brenner, Jefferey H. Miller, *Encyclopedia of Genetics*, Academic Press, pages 666-669.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**: 3210-3212.

Simillion, C., Vandepoele, K., Van Montagu, M.C., Zabeau, M. and Van de Peer, Y., 2002. The hidden duplication past of Arabidopsis thaliana. *Proceedings of the National Academy of Sciences*, **99**: 13627-13632.

Singh, N.K., Badet, T., Abraham, L. and Croll, D., 2021. Rapid sequence evolution driven by transposable elements at a virulence locus in a fungal wheat pathogen. *BMC Genomics*, **22**: 393.

Stalder, L., Oggenfuss, U., Mohd-Assaad, N. and Croll, D., 2022. The population genetics of adaptation through copy number variation in a fungal plant pathogen. *Molecular Ecology*, **32**: 2443-2460.

Strelkov, S.E., Lamari, L. and Ballance, G.M., 1999. Characterization of a host-specific protein toxin (Ptr ToxB) from *Pyrenophora tritici-repentis*. *Molecular Plant-Microbe Interactions*, **12**: 728-732.

Strelkov, S.E., Lamari, L., Sayoud, R. and Smith, R.B., 2002. Comparative virulence of chlorosis-inducing races of *Pyrenophora tritici-repentis*. *Canadian Journal of Plant Pathology*, **24**: 29-35.

Strelkov, S.E., Kowatsch, R.F., Ballance, G.M. and Lamari, L., 2005. Characterization of the *ToxB* gene from North African and Canadian isolates of *Pyrenophora tritici-repentis*. *Physiological and Molecular Plant Pathology*, **67**: 164-170.

Sweredoski, M., DeRose-Wilson, L. and Gaut, B.S., 2008. A comparative computational analysis of nonautonomous Helitron elements between maize and rice. *BMC Genomics*, **9**: 1-13.

Taylor, J.S. and Raes, J., 2004. Duplication and divergence: the evolution of new genes and old ideas. *Annual Review Genetics*, **38**: 615-643.

Thiel, T., Graner, A., Waugh, R., Grosse, I., Close, T.J. and Stein, N., 2009. Evidence and evolutionary analysis of ancient whole-genome duplication in barley predating the divergence from rice. *BMC Evolutionary Biology*, **9**: 1-18.

Thomas, J. and Pritham, E.J., 2015. Helitrons, the eukaryotic rolling-circle transposable elements. *Mobile DNA* **3**: 891-924.

Thomas, J., Vadnagara,K. and Pritham,E.J., 2014. DINE-1, the highest copy number repeats in *Drosophila melanogaster* are non-autonomous endonuclease-encoding rolling-circle transposable elements (Helentrons). *Mobile DNA*, **5**: 18.

Tian, C.G., Xiong, Y.Q., Liu, T.Y., Sun, S.H., Chen, L.B. and Chen, M.S., 2005. Evidence for an ancient whole-genome duplication event in rice and other cereals. *Yi Chuan Xue Bao*, **32**: 519-527.

Torres, D.E., Oggenfuss, U., Croll, D. and Seidl, M.F., 2020. Genome evolution in fungal plant pathogens: looking beyond the two-speed genome model. *Fungal Biology Reviews*, **34**: 136-143.

Tralamazza, S.M., Gluck-Thaler, E., Feurtey, A. and Croll, D., 2024. Copy number variation introduced by a massive mobile element facilitates global thermal adaptation in a fungal wheat pathogen. Nature Communications, 15: 5728.

Tuteja, N., 2003. Plant DNA helicases: the long unwinding road. *Journal of Experimental Botany*, **54**: 2201-2214.

Vlaardingerbroek, I., Beerens, B., Rose, L., Fokkens, L., Cornelissen, B.J. and Rep, M., 2016. Exchange of core chromosomes and horizontal transfer of lineage-specific chromosomes in *Fusarium oxysporum*. *Environmental Microbiology*, **18**: 3702-3713.

Xiong, W., Dooner, H.K. and Du, C., 2016. Rolling-circle amplification of centromeric Helitrons in plant genomes. *The Plant Journal*, **88**: 1038-1045.

Wang, C., Milgate, A.W., Solomon, P.S. and McDonald, M.C., 2021. The identification of a transposon affecting the asexual reproduction of the wheat pathogen *Zymoseptoria tritici*. *Molecular Plant Pathology*, **22**: 800-816.

Wapinski, I., Pfeffer, A., Friedman, N. and Regev, A., 2007. Natural history and evolutionary principles of gene duplication in fungi. *Nature*, **449**: 54-61.

Ward, T.J., Bielawski, J.P., Kistler, H.C., Sullivan, E. and O'Donnell, K., 2002. Ancestral polymorphism and adaptive evolution in the trichothecene mycotoxin gene cluster of phytopathogenic *Fusarium*. *Proceedings of the National Academy of Sciences*, **99**: 9278-9283

Waterborg, J.H., 2012. Evolution of histone H3: emergence of variants and conservation of post-translational modification sites. *Biochemistry and Cell Biology*, **90**: 79-95.

Wei, B., Despins, T., Fernandez, M.R., Strelkov, S.E., Ruan, Y., Graf, R. and Aboukhaddour, R., 2021. Race distribution of Pyrenophora tritici-repentis in relation to ploidy level and susceptibility of durum and winter bread wheat. *Canadian Journal of Plant Pathology*, **43**: 582-598.

Wolf, Y.I. and Koonin, E.V., 2013. Genome reduction as the dominant mode of evolution. *Bioessays*, **35**: 829-837

Yun, S.H., Berbee, M.L., Yoder, O.C. and Turgeon, B.G., 1999. Evolution of the fungal self-fertile reproductive life style from self-sterile ancestors. *Proceedings of the National Academy of Sciences*, **96**: 5592-5597.

Zhaxybayeva, O., Lapierre, P. and Gogarten, J.P., 2005. Ancient gene duplications and the root(s) of the tree of life. *Protoplasma*, **227**: 53-64.

CHAPTER 4: GENOME WIDE ASSOCIATION STUDY REVEALS SNPS ASSOCIATED WITH THE TOXC PHENOTYPE IN *PYRENOPHORA TRITICI-REPENTIS*

4.0 ABSTRACT

The nature of ToxC, a race defining effector in the wheat pathogen *Pyrenophora tritici-repentis*, has remained elusive for over 20 years. We have utilized a set of 36 genomes previous used to create a pangenome to perform a genome wide association study linking single nucleotide polymorphisms to the ToxC phenotype. The data set represents a diverse set of isolates and contained roughly equal numbers of ToxC producers and non-producers. The highest -log(p-value) observed in the GWAS analysis was 5.49 which passes the generally accepted significance threshold of 5. As a relatively small number of genomes was used, this is a promising result. Located ~216 kb from the end of chromosome 6 a number of interesting genes were found in the vicinity including: eight with unknown functions, a CAzyme associated with plant cell wall degradation, senescence and autophagy related proteins, and a transmembrane transport protein. This work provides ample new information and avenues of research by which the identity of ToxC may be uncovered.

4.1 INTRODUCTION

Pyrenophora tritici-repentis (Ptr) is a fungal pathogen causing tan spot of wheat which produces three primary necrotrophic effectors: ToxA, ToxB, and ToxC. Both ToxA and ToxB are proteinaceous and interact with the host in an inverse gene-for-gene manner (Tomas and Bockus, 1987; Lamari and Bernier, 1989; Strelkov et al., 1999; Lamari et al., 2003). The specific genes which encode ToxA and ToxB are both known as well as the host genes with which they interact, *Tsn1* and *Tsc2* respectively (Ballance et al., 1996;

Faris et al., 1996; Ciuffetti et al., 1997; Anderson et al., 1999; Strelkov, et al. 1999; Haen et al., 2004; Kim et al., 2010). However, the identity of the third effector, ToxC, has remained elusive for several decades. Early work with culture filtrates concluded that ToxC was likely a polar, non-ionic metabolite with a low-molecular weight (Effertz et al., 2002). Recently, Shi et al., (2022) utilized two segregating populations of Ptr to identify the gene *ToxC1* which appears to plays a role in the synthesis of ToxC which was confirmed via knock-out experiments. This was complimented by other recent work which showed that isolate Biotrigo9-1 had a large insertion within *ToxC1*, which may account for this isolates lack of ToxC production (Moolhuijzen et al., 2022). The wheat host gene with which ToxC interacts is known as *Tsc1* and the presence of this susceptibility gene plays an important role in disease development in both durum and bread wheat (Effertz et al., 2002; Liu Y et al., 2020; Running et al., 2022). Beyond these works little else has been done to identify the nature of ToxC or which gene(s) are part of its synthesis or secretion.

As we recently made significant contributions to the work surrounding the *ToxA* and *ToxB* genes, we were spurred to leverage our existing data set to shine new light on ToxC as well (Gourlie et al., 2022; Hafez et al. 2024). Here we utilized the same set of Ptr genomes used to create a pangenome as this set contains roughly equal numbers of ToxC producers and non-producers from diverse regions and time points. The majority of these isolates have also been utilized in many others works with confident phenotype information and it is a suitable population for attempting a genome-wide association study (GWAS) (Lamari et al., 2003; Strelkov et al., 2005; Amaike et al., 2008; Kim et al., 2010; Aboukhaddour et al., 2012; Aboukhaddour et al., 2013; Kamel et al., 2019; Gourlie

et al., 2022; Hafez et al., 2024). The aim of this work was to provide a set of candidate single nucleotide polymorphisms (SNPs) and/or genes which are linked to or involved with ToxC biosynthesis, transport, secretion, or regulation for future work to investigate.

4.2 METHODS AND MATERIALS

4.2.1 ISOLATES, DNA EXTRACION, SEQUENCING, AND ASSEMBLY

The isolates used in this work are from a previous pangenome project therefore information on isolate collection, DNA extraction, genome sequencing, *de novo* assembly, and annotation details are described in Gourlie et al., (2022). Here in brief, isolates were grown on ¼ PDB broth until fungal mats were sufficiently large to extract DNA. Fungal mats were washed with sterile water, freeze dried and stored at -20°C until extraction. Genomic DNA was extracted using Qiagen 'Genomic-tip 20/G' kits. Short-read sequencing of isolates was performed with with 150-bp paired-end, 400-bp inserts at 100 × coverage with Illumina HiSeq X and long-read sequencing of the reference isolate I-73-1 was done with PacBio RS II at Genome Quebec (Montreal, Canada). Short-read assemblies were created using Shovill and SPAdes (Bankevich et al., 2012) and the long-read assembly with Flye (vAPRIL-2020) (Kolmogorov et al., 2019) and Pilon (Walker et al., 2014). Genomes were annotated with the FunGAP pipeline (v1.0.1) (Min et al., 2017). See Gourlie et al., (2022) (thesis Table 2.1) for more isolate details.

4.2.2 VARIANT CALLING WITH GENOME ANALYSIS TOOLKIT AND SNP ASSOCIATION WITH GAPIT

Variant calls were generated with Genome Analysis Toolkit (v4.4.0.0) (GATK) (McKenna et al., 2010) for 36 isolates (19 ToxC producers; 17 non-producers) using the ToxC producer I-73-1 as the reference. The reference genome was indexed using bowtie2 (v2.5.1) (Langmead and Salzberg, 2012) (-build) and SAMtools (1.17) (Li et al., 2009) (-

faidx), and a sequence dictionary made with GATK (-CreateSequenceDictionary). Raw reads of the 36 isolates were aligned to the reference with bowtie2. Alignments were converted from SAM to BAM format and then sorted via SAMtools (-view -S -b -sort). Duplicates were removed from the BAM alignments with Picard (v3.0.0) (-MarkDuplicates --REMOVE_DUPLICATES) and read groups added (-AddOrReplaceReadGroups). Read groups were indexed with SAMtools (-index) and variant calls for each isolate were done with GATK HaplotypeCaller. Variant call files were then merged (-GenomicsDBImport --interval_padding 100) and genotyped (-GenotypeGVCFs). Variant calls were then assessed using various quality scores (QUAL=250; QD=20; FS=0.1; SOR=3; MQ=30; ReadPosRankSum>2; ReadPosRankSum<-2; MQRankSum>2.5; MQRankSum<-2.5) and filter calls added (-VariantsToTable). SNPs which did not pass the filters were removed using vcftools (v0.1.16) (Danecek et al., 2011) as well as SNPs which did not meet a minimum allele frequency of 0.05 (--remove-indels --remove-filtered-all --maf --max-missing 0.7 --recode). The filtered vcf file was converted to hapmap format with Tassel (v5.2.58) (Bradbury et al., 2007). The GAPIT (v3) package (Wang and Zhang, 2021) was used within RStudio to conduct GWAS. The variant call hapmap file and a binary phenotype matrix were imported. GWAS was run using the FarmCPU model (Liu X et al., 2016), the ‘Zhang’ kinship algorithm and four principal components. The GWAS output file was reimported and the qqman package (Turner, 2014) was used to create a Manhattan plot.

4.2.3 METABOLIC CLUSTER IDENTIFICATION WITH ANTISMASH

As it is assumed ToxC is the product of a metabolic pathway we used antiSMASH (v7.1.0) (Blin et al., 2023) to search for metabolite gene clusters which may be near to

associated SNPs. antiSMASH was run on the reference isolate I-73-1 using the following arguments: --taxon fungi --cb-general --cb-knownclusters --cb-subclusters --asf --pfam2go --smcog-trees --genefinding-gff3. Output was manually assessed.

4.3 RESULTS

HaplotypeCaller identified a total of 66,574 variant sites across the 36 isolates. Post-filtering, 18,543 high-quality variant calls remained for GWAS analysis. The Q-Q plot generated by the FarmCPU algorithm indicated the model was suitable for association analysis (Additional File 4.1). The highest -log(pvalue) was 5.49, located at the 5' end of chromosome 6 at position 216,200 in the I-73-1 genome. Several other SNPs with relatively high associations were nearby which creates a clear visual peak in the Manhattan plot (Figure 4.1).

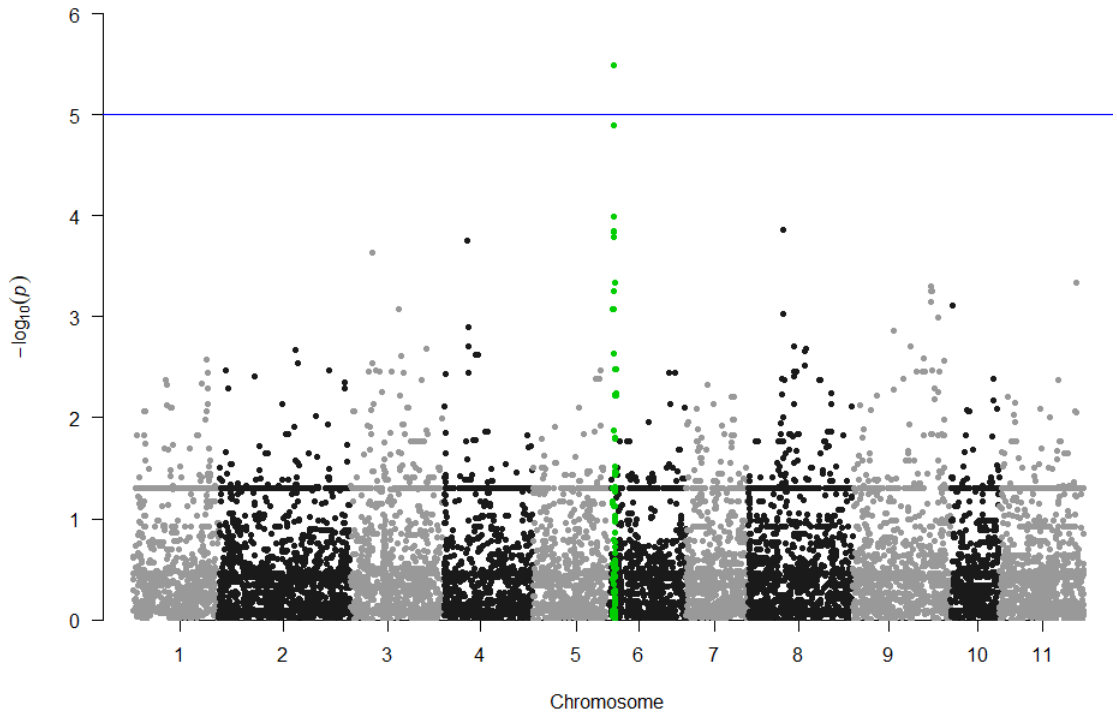


Figure 4.1 Manhattan plot visualizing the association of SNPs across the genome of *P. tritici-repentis* with the ToxC phenotype. SNP peak on the 5' of chromosome 6 are highlighted in green. A widely used significance cutoff of p = 1E-05 is indicated with a horizontal line.

The SNP peak was located within gene 04169 which was not annotated by previously utilized Pfam or Interproscan databases. However, manual searching of the UniProt database did yield some amino acid similarity to DUF1479 (domain of unknown function) (39.5%) and F-Box domain containing proteins (41.5%). Surrounding genes include several other unannotated genes, a second DUF1479 gene, glycosyl hydrolase family 1, senescence-associated protein, autophagy-related protein, major facilitator superfamily (MFS), and several other interesting genes and protein domains (Figure 4.2). A list of nearby genes and their annotations can be found in Additional File 2. AntiSMASH identified 28 secondary metabolite clusters within I-73-1. However only two metabolite synthesis clusters were found on chromosome 6: a type III polyketide

synthase cluster (T3PKS) and a fungal-RiPP-like cluster. Neither of these biosynthetic clusters were near the GWAS peak unfortunately (>500 kb distant). There are hints of peaks on chromosomes 4 and 8 whose SNPs also reside within genes. The highest SNP on chromosome 4 is located within a glutathione s-transferase and the SNP on chromosome 8 resides in a gene with unknown function.

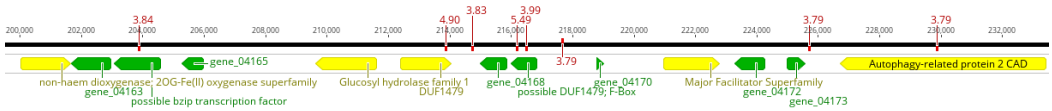


Figure 4.2 Gene annotations on Chr06 of I-73-1 in the region near the GWAS peak. Yellow genes were annotated with the Pfam database, while green genes did not return results from Pfam. Uniprot BLAST searches provided some translated protein similarities for genes not annotated by Pfam. Red marks and numbers indicate the position and -log(p-value) of SNPs identified via GWAS.

To identify potential structural variations within the GWA region assembly alignment is required. The SNP calling was performed using short-read data and alignments of the GWAS region with the short-read assemblies did not yield useful information. Utilizing long-read assemblies generated in Chapter 3, it was possible to see some structural variation between ToxC producers and ToxC non-producers for isolates that were used in the GWAS analysis, namely isolates T126-1, T128-1, and Alg3-24. In particular, the presence of retrotransposon insertion near the GWAS peak in isolate T128-1 and T126-1 (Figure 4.3).

In isolate T126-1, a 5400 bp retrotransposon is inserted between genes 04173 (unknown function) and 04174 (autophagy-related protein). In isolate T128-1, a 1468 bp retrotransposon is inserted in the promoter region of gene 04169 (unknown function,

possibly DUF1479) which is the gene containing the highest GWAS peak. In isolate Alg3-24, a 5922 bp hAT transposon is inserted upstream of gene 04161 (senescence-associated protein) which is followed by 5536 bp deletion which includes 7 bp of the terminal end of gene 04160 (unknown function). There is another 2107 bp deletion in Alg3-24 between genes 04159 (TRAP) and 04158 (PF00400 WD domain).

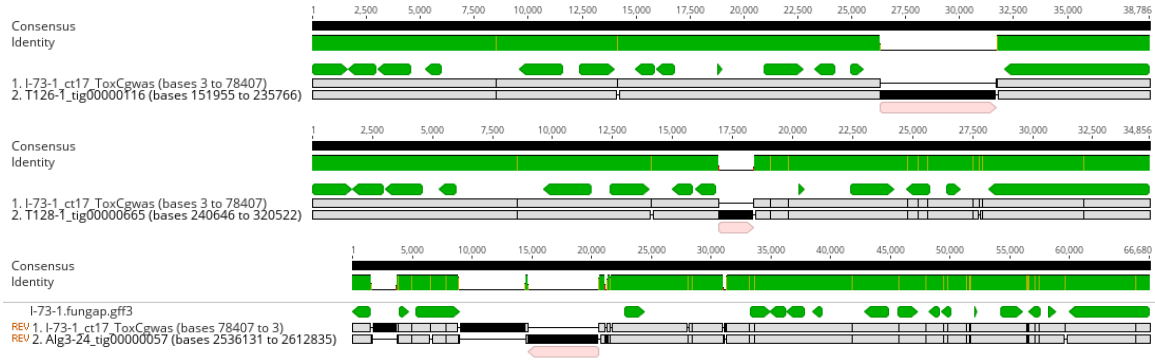


Figure 4.3 Alignment of GWAS peak region between ToxC production isolate I-73-1 and ToxC non-producers T126-1 (top alignment); T128-1 (middle alignment); and Alg3-24 (bottom alignment). Transposon insertions in the region are annotated with pink and genes annotated in green.

BLAST of the our reference I-73-1 for *ToxC1* revealed it was present on a 258 kb contig (contig 15) which did not have any significantly associated SNPs. Contig 15 failed to align to either Pt-1C-BFP or M4 (v2.2) reference genomes from NCBI. All the short-read assemblies were also searched for the presence of *ToxC1* and compared to their phenotype. Here we observed that some isolates with the ToxC phenotype appear to lack the *ToxC1* gene (AlgH1, SW20-7, SW21-1, and SW7-5). Additionally some isolates possessed *ToxC1* but lack the ToxC phenotype (L2-1, SW1-2, and T176-2) (Table 4.1).

Table 4.1 Number of isolates with or without the previously identified gene *ToxC1* based on BLASTN of short-read assemblies compared with the presence of the ToxC phenotype.

		***ToxC1* gene**		
		Present	**Absent**	**Total**
ToxC phenotype	**Observed**	16	4	20
	No reaction	3	14	17
	Total	19	18	37

4.4 DISCUSSION

The existence of ToxC has been known for over 20 years and is a defining effector for the Ptr race system (Effertz et al., 2002; Strelkov and Lamari, 2003). However the identity of ToxC has proven elusive. Recent work revealed one potential gene, dubbed *ToxC1*, in the ToxC synthesis pathway (Shi et al., 2022). This work ignited renewed interest in identifying ToxC and finding genes associated with its creation or secretion. Here, we've provided fresh clues via a genome wide association study.

Perhaps the most pertinent finding of this work is that, of the 12 genes nearest the GWAS peak, 8 are of unknown function. The unknown nature of these genes is tantalizing and it is possible they represent key steps in the synthesis of ToxC. Of the genes that had identifiable functional domains, one is the carbohydrate active enzyme (CAZyme) glycosyl hydrolase 1 (GH1) (gene_04166). Members of GH1 the gene family target glycosidic bonds between residues in disaccharide which are present in a number of the substrates (Hill and Reilly 2008). In fungal pathogens, CAZymes such as GH1 are broadly involved with the enzymatic activity of degrading cell walls, however their functions can be quite diverse and are not limited to cell walls degradation (Zhao et al., 2014; Garron and Henrissat, 2019). It is thought that ToxC is not a protein (enzyme)

however but rather a small metabolite (Effertz et al., 2002). It is possible that this GH1 CAZyme is in some way required for the delivery of ToxC into the host.

Another annotated gene is a transmembrane transporter from the Major Facilitator Superfamily (MFS; gene_04171). This group of transporters are capable of mobilizing diverse substrates including ions and other small molecules across membranes (Quistgaard et al. 2016). This gene is of particular interest as it may be involved with the secretion or and delivery of ToxC to host cells. Two DUF1479 (gene_04167) genes was also found which was recently suggested to be an oxidoreductase (catalyzes transfer of electrons) and may play a role in the virulence of *Candida albicans* (Rani et al., 2020). The fact DUF1479 could be facilitating virulence in another fungal pathogen is interesting and makes DUF1479 a strong candidate for further study.

A non-haem dioxygenase (gene_04162) is nearby which are known to be involved in diverse oxidation reactions, and in mammals, have been shown to act as cellular sensors or regulators in hypoxic response pathways (Ozer and Bruick, 2007). One possibility for this gene in relation to ToxC, is the product from this gene is utilized in some manner to sense the host (or host phenotype) or the contents of disrupted host cells which then elicits ToxC production. A senescence-associated protein (gene_04161) immediately stands out as these proteins are associated with cell death/degradation in plants. Another cell death associated protein domain, autophagy-related protein 2 CAD (gene_04174), is nearby as well. Autophagy genes are generally associated with self-degradation and homeostasis, but there are links to fungal pathogenicity (Tam et al., 2016; Ren et al., 2017).

For the three ToxC non-producing isolates which had long-read assemblies available, three different transposon insertions were noted near to the GWAS peak. It is not possible to say conclusively if these transposon insertions play any role in ToxC disruption. However, one possibility may be that genomic defenses against transposon proliferation such as DNA methylation have disrupted regulation of genes in this region. Such a mechanism would explain why the genes themselves share synteny and nearly identical sequences.

It was notable that *ToxC1* did not appear in our GWAS analysis. Additionally some isolates in our collection exhibit the ToxC phenotype but appear to lack *ToxC1* (Table 4.1). Although Shi et al., (2022) showed that *ToxC1* is necessary for the appearance of the ToxC phenotype, the above observations may suggest that role of *ToxC1* is more generalized and not specific to ToxC production. It is possible *ToxC1* may serve various functions in multiple pathways one of which is ToxC production or secretion. Another possibility is *ToxC1* is part of one synthesis pathway while some ToxC producers which lack *ToxC1* rely on a different pathway. Whichever case is true, the locus identified here showed a strong correlation to the ToxC phenotype using a diverse collection of isolates and should be investigated further.

Moving forward, a transcriptomic approach may be utilized to observe gene expression changes during host infection. Assuming ToxC production increases during the infection cycle, it should be possible to see which, if any, genes near the GWAS peak have altered expression during host colonization. Such an approach should narrow the candidate(s) for gene-knockout experiments to confirm what we have found in silico and

contribute to answering the long standing questions surrounding the nature of ToxC in *P. tritici-repentis*.

4.5 REFERENCES

Aboukhaddour R, Kim YM, Strelkov SE. 2012. RNA-mediated gene silencing of ToxB in *Pyrenophora tritici-repentis*. *Molecular Plant Pathology*. **13**: 318-326.

Aboukhaddour R, Turkington TK, Strelkov SE. 2013. Race structure of *Pyrenophora triciti-repentis* (tan spot of wheat) in Alberta, Canada. *Canadian journal of plant pathology*. **35**: 256-268.

Amaike S, Ozga J, Basu U, Strelkov S. 2008. Quantification of *ToxB* gene expression and formation of appressoria by isolates of *Pyrenophora tritici-repentis* differing in pathogenicity. *Plant Pathology*. **57**: 623-633.

Anderson J, Effertz R, Faris J, Francl L, Meinhardt S, Gill B. 1999. Genetic analysis of sensitivity to a *Pyrenophora tritici-repentis* necrosis-inducing toxin in durum and common wheat. *Phytopathology*. **89**:293-297.

Ballance G, Lamari L, Kowatsch R, Bernier C. 1996. Cloning, expression and occurrence of the gene encoding the Ptr necrosis toxin from *Pyrenophora tritici-repentis*.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology*. **19**: 455-477.

Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M, Fetter A, Terlouw BR, Metcalf WW, Helfrich EJ. 2023. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic acids research*. **51**: W46-W50.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*. **23**: 2633-2635.

Ciuffetti LM, Tuori RP, Gaventa JM. 1997. A single gene encodes a selective toxin causal to the development of tan spot of wheat. *The Plant Cell*. **9**: 135-144.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST. 2011. The variant call format and VCFtools. *Bioinformatics*. **27**: 2156-2158.

Effertz R, Meinhardt SW, Anderson J, Jordahl J, Francl L. 2002. Identification of a chlorosis-inducing toxin from *Pyrenophora tritici-repentis* and the chromosomal location of an insensitivity locus in wheat. *Phytopathology*. **92**: 527-533.

Faris J, Anderson JA, Francl L, Jordahl J. 1996. Chromosomal location of a gene conditioning insensitivity in wheat to a necrosis-inducing culture filtrate from *Pyrenophora tritici-repentis. Phytopathology*. **86**: 459-463.

Garron M-L, Henrissat B. 2019. The continuing expansion of CAZymes and their families. *Current opinion in chemical biology*. **53**: 82-87.

Gourlie R, McDonald M, Hafez M, Ortega-Polo R, Low KE, Abbott DW, Strelkov SE, Daayf F, Aboukhaddour R. 2022. The pangenome of the wheat pathogen *Pyrenophora tritici-repentis* reveals novel transposons associated with necrotrophic effectors *ToxA* and *ToxB*. *BMC biology*. **20**: 239.

Haen KM, Lu H, Friesen TL, Faris JD. 2004. Genomic targeting and high-resolution mapping of the *Tsn1* gene in wheat. *Crop Science*. **44**: 951-962.

Hafez M, Gourlie R, McDonald M, Telfer M, Carmona MA, Sautua FJ, Moffat CS, Moolhuijzen PM, See PT, Aboukhaddour R. 2024. Evolution of the *ToxB* Gene in *Pyrenophora tritici-repentis* and Related Species. *Molecular Plant-Microbe Interactions*. **37**: 327-337.

Hill AD, Reilly PJ. 2008. Computational analysis of glycoside hydrolase family 1 specificities. *Biopolymers*. **89**: 1021-1031.

Kamel S, Cherif M, Hafez M, Despins T, Aboukhaddour R. 2019. *Pyrenophora tritici–repentis* in Tunisia: Race structure and effector genes. *Frontiers in Plant Science*. **10**: 1562.

Kim YM, Bouras N, Kav NN, Strelkov SE. 2010. Inhibition of photosynthesis and modification of the wheat leaf proteome by Ptr ToxB: a host-specific toxin from the fungal pathogen *Pyrenophora tritici-repentis*. *Proteomics*. **10**: 2911-2926.

Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nature biotechnology*. **37**: 540-546.

Lamari L, Bernier C. 1989. Toxin of *Pyrenophora tritici-repentis*: Host-specificity, significance in disease. *Phytopathology*. **79**: 740-744.

Lamari L, Strelkov S, Yahyaoui A, Orabi J, Smith R. 2003. The identification of two new races of *Pyrenophora tritici-repentis* from the host center of diversity confirms a one-to-one relationship in tan spot of wheat. *Phytopathology*. **93**: 391-396.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods*. **9**: 357-359.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPDP. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*. **25**: 2078-2079.

Liu X, Huang M, Fan B, Buckler ES, Zhang Z. 2016. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS genetics*. **12**: e1005767.

Liu Y, Salsman E, Wang R, Galagedara N, Zhang Q, Fiedler JD, Liu Z, Xu S, Faris JD, Li X. 2020. Meta-QTL analysis of tan spot resistance in wheat. *Theoretical and Applied Genetics*. **133**: 2363-2375.

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.

Min B, Grigoriev IV, Choi I-G. 2017. FunGAP: Fungal Genome Annotation Pipeline using evidence-based gene model evaluation. *Bioinformatics*. **33**: 2936-2937.

Moolhuijzen PM, See PT, Shi G, Powell HR, Cockram J, Jørgensen LN, Benslimane H, Strelkov SE, Turner J, Liu Z. 2022. A global pangenome for the wheat fungal pathogen *Pyrenophora tritici-repentis* and prediction of effector protein structural homology. *Microbial Genomics*. **8**: 000872.

Ozer A, Bruick RK. 2007. Non-heme dioxygenases: cellular sensors and regulators jelly rolled into one? *Nature chemical biology*. **3**: 144-153.

Quistgaard EM, Löw C, Guettou F, Nordlund P. 2016. Understanding transport by the major facilitator superfamily (MFS): structures pave the way. *Nature Reviews Molecular Cell Biology*. **17**: 123-132.

Ren W, Zhang Z, Shao W, Yang Y, Zhou M, Chen C. 2017. The autophagy-related gene BcATG1 is involved in fungal development and pathogenesis in *Botrytis cinerea*. *Molecular plant pathology*. **18**: 238-248.

Running KL, Momotaz A, Kariyawasam GK, Zurn JD, Acevedo M, Carter AH, Liu Z, Faris JD. 2022. Genomic analysis and delineation of the tan spot susceptibility locus *Tsc1* in wheat. *Frontiers in Plant Science*. **13**: 793925.

Shi G, Kariyawasam G, Liu S, Leng Y, Zhong S, Ali S, Moolhuijzen P, Moffat CS, Rasmussen JB, Friesen TL. 2022. A conserved hypothetical gene is required but not sufficient for Ptr ToxC production in *Pyrenophora tritici-repentis*. *Molecular Plant-Microbe Interactions*. **35**: 336-348.

Strelkov S, Kowatsch R, Ballance G, Lamari L. 2005. Characterization of the *ToxB* gene from North African and Canadian isolates of *Pyrenophora tritici-repentis. Physiological and molecular plant pathology*. **67**: 164-170.

Strelkov S, Lamari L. 2003. Host–parasite interactions in tan spot [*Pyrenophora tritici-repentis*] of wheat. *Canadian Journal of Plant Pathology*. **25**: 339-349.

Strelkov S, Lamari L, Ballance GM. 1999. Characterization of a host-specific protein toxin (Ptr ToxB) from *Pyrenophora tritici-repentis*. *Molecular Plant-Microbe Interactions*. **12**: 728-732.

Tam JM, Mansour MK, Acharya M, Sokolovska A, Timmons AK, Lacy-Hulbert A, Vyas JM. 2016. The role of autophagy-related proteins in *Candida albicans* infections. *Pathogens*. **5**: 34.

Tomas A, Bockus W. 1987. Cultivar-specific toxicity of culture filtrates of *Pyrenophora tritici-repentis. Phytopathology*. **77**: 1337-1340.

Turner SD. 2014. qqman: an R package for visualizing GWAS results using QQ and manhattan plots. *Biorxiv*.005165.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS one*. **9**: e112963.

Wang J, Zhang Z. 2021. GAPIT version 3: boosting power and accuracy for genomic association and prediction. *Genomics, Proteomics and Bioinformatics*. **19**: 629-640.

Zhao Z, Liu H, Wang C, Xu J-R. 2014. Erratum to: comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC genomics.* **15**: 1-15.

CHAPTER 5: GENERAL DISCUSSION

In this body of work, the genome and race defining effectors of *Pyrenophora tritici-repentis* were examined in detail. Analysis of the Ptr genome was done by sequencing over 60 genomes, 40 via short-reads and 23 via long-reads. The assembled genomes were annotated for gene and transposon content then used to create a pangenome and assess genomic architecture. Tox gene mobility and replication were explored, and a genome wide association study conducted. As a whole, this thesis constitutes one of the largest explorations of the Ptr genome to date.

In summary, the primary findings of this thesis are as follows:

1. Ptr possesses an open pangenome with significant gene gains and losses across the compared isolates, particularly between pathogenic and non-pathogenic races;

2. In the Ptr pangenome, 43% of genes are core (shared by all isolates) and 57% of genes are accessory (shared by a subset of isolates);

3. Ptr isolates in the phylogenetic outgroup contain high numbers of singleton genes (i.e. unique genes present in a single genome);

4. Carbohydrate-active enzymes (CAZymes) are conserved while predicted putative effectors are mostly accessory in nature;

5. Based on two long-read assemblies, Ptr possesses a ‘one-compartment’ genome architecture (i.e. in general, genes appear to be spread evenly throughout the genome);

6. Significant chromosomal rearrangements were observed across isolates;

7. Expansion of the Ptr genome is driven, in part, by transposon incursions and replication, primarily retrotransposon Copia and Ty-3 elements as well as DNA hAT transposons;

8. The necrosis-inducing effector gene *ToxA* is contained within an actively mobile Starship transposon (*Horizon*) which has translocated *ToxA* to different chromosomes in different isolates;

9. The chlorosis-inducing effector gene *ToxB,* or its inactive homolog *toxb,* is present within a large accessory region of chromosome 4, which may be a defunct Starship transposon (*Icarus*), an accessory chromosomal arm, or the first stages of genome compartmentalization;

10. The *ToxB* gene has been duplicated by a Helitron-like element (*ToxB-HLE*), a DNA transposon which uses rolling-circle replication, with unequal crossing-over as a potential partner for further *ToxB* replication;

11. Multiple *Copia* transposon insertion events within the *ToxB* open reading frame have disrupted the gene, explaining two inactive *toxb* haplotypes;

12. Single-nucleotide polymorphisms are associated with the ToxC phenotype and provide candidate genes that may be involved in with biosynthesis or secretion of the ToxC metabolite.

5.1 GENOME PLASTICITY AND THE ADAPTABILITY OF *PYRENOPHORA TRITICI-REPENTIS*

The environmental adaptability of Ptr is readily apparent with its worldwide distribution, being found in most areas where wheat is cultivated, such as North and

South America, Europe, the Middle East, Asia, Africa, and Australia (reviewed in Kamel et al., 2019). Ptr has also been isolated from areas where no wheat is cultivated (Lamari et a1., 2005; Strelkov and Lamari, 2003). In addition to durum wheat and bread wheat, Ptr has been shown to cause disease on a variety of other grass hosts, such as brome grass, rye, barley, and *Agropyron* sp. (Hosford, 1971; Krupinsky, 1992a; 1992b). Such flexibility can be attributed to Ptr's genome plasticity, which has been previously explored using traditional molecular biology techniques and early genomics work. Among the findings of this research were: potential aneuploidies, large scale genomic rearrangements and transduplications, different chromosomal locations of *ToxA* and *ToxB*, and evidence of large numbers of transposable elements (Aboukhaddour et al., 2009; 2011; 2012; Manning et al., 2013; Moolhuijzen et al., 2018a). The genomics work to date has relied on a limited set of isolate genomes which have not represented all Ptr races nor reflected the species wide geographical distribution. Furthermore, a comparison of differential gene content across the available genomes has not been explored.

The idea that individuals of the same species may have differences in the presence and absence of genes is a relatively recent one, yet is already a cornerstone in our understanding of adaptative evolution (Tettelin et al., 2005). This idea, dubbed pangenomics, was initially limited to prokaryotes (e.g. *E. coli*) as their genomes were small enough to sequence in bulk back when sequencing costs were still relatively high (Tettelin et al., 2005). The number of pangenomic studies has been steadily increasing however, including those of plant pathogens. The addition of the Ptr pangenome created here joins the growing list of economically important pathogen species which have had their pangenomes analyzed: *Zymoseptoria tritici* (wheat; septoria leaf blotch) (Badet et

al., 2019; Chen et al., 2023), *Rhizoctonia solani* (various hosts) (Kaushik et al., 2022), *Cercospora zeina* (maize; gray leaf spot) (Welgemoed et al., 2023), *Magnaporthe oryzae* (grasses; blast diseases) (Joubert and Krasileva, 2024), *Claviceps purpurea* (cereals; ergot) (Wyka et al., 2022), *Pyricularia oryzae* (rice; blast disease) (Bao et al., 2022), *Fusarium oxysporum* (various hosts) (van Westerhoven et al., 2024), and *Parastagonospora* sp. (wheat; Septoria nodorum blotch) (Syme et al., 2018).

A common theme across all these studies, including Ptr, is a high accessory gene content. Accessory genes are a key component of adaptability, as these genes are often related to environmental or host adaptations (Croll and McDonald, 2012). Environmental adaptations help pathogens become global problems as they enable the colonization of environments which were previously unsuitable. In our changing climate, this type of adaptability poses a major concern that pathogens may adapt more rapidly than we can alter their agronomic hosts. Gene gains and losses within a species can also influence pathogen virulence and host specificity (Lawrence et al., 2005; Powell et al., 2008). For example, in the leaf-spot pathogen *Z. tritici*, a significant number of genes exhibiting copy-number variations were linked to virulence factors and secondary metabolite gene clusters (Hartmann and Croll, 2017). Many of these CNVs were associated with transposons. In Ptr, it remains to be seen when or by what mechanisms the observed gene gains originated. It is likely that most are the result of gene duplications which have since diverged, but potentially some are from unknown horizontal transfer events. Other sources might include dispensable/accessory chromosomes, chromosomal regions, or differential conservation of gene-carrying transposable elements such as Starships.

Ptr was also found to have a significant number of transposons throughout its genome. Combined with the apparent lack of RIP (repeat-induced point mutations), a common defensive mechanism against transposons unique to fungal genomes, it's likely that Ptr is subject to genome instability. Genome instability can lead to large scale genomic rearrangements such as translocations, inversions, duplications, deletions, etc., such as those seen in Ptr (Wöstemeyer and Kreibich, 2002; Chan and Kolodner, 2011; Faino et al., 2016; Bourque et al., 2018). Such rearrangements can serve adaptability through gene and regulatory element reordering. Interestingly, it has been shown that asexual and homothallic fungal species such as Ptr tend to have high numbers of chromosomal rearrangements compared to their heterothallic counterparts (reviewed by Whittle et al., 2011).

While a large number of transposons were annotated in the Ptr genome, it did not have clear signatures of a 'two-compartment' genome. In such an architecture, there are gene-dense, transposon-sparse regions and gene-sparse, transposon-dense regions (Raffaele et al., 2010; Dong et al., 2015). In this model, rapidly evolving genes, such as those involved with virulence, are located in transposon-dense regions which are linked to higher mutation rates and rearrangements. Genes within Ptr appeared to be spaced relatively evenly, however, suggesting that no compartmentalization has occurred. Some researchers have started to look beyond the 'two-compartment' model to suggest that rapidly evolving pathogens derive their 'second speed' via high copy-number variability (Frantzeskakis et al., 2018; Torres et al., 2020). My work did not explore copy-number variation throughout the Ptr genome, however it is possible that Ptr adaptability is driven by copy-number variation and gene duplications rather than through

compartmentalization and high mutation rates. This seems likely, since we know the two main virulence genes, *ToxA* and *ToxB,* are highly conserved and that *ToxB* is subject to copy-number variation. It is also possible that the Ptr genome is in the process of compartmentalization, which is an exciting prospect that may allow researchers to better understand the process of compartmentalization as they continue to sequence Ptr genomes far into the future.

5.2 ROLE OF TRANSPOSONS IN THE EVOLUTION OF *PYRENOPHORA TRITIC-REPENTIS*

The proteinaceous necrotrophic effector *ToxA* is present not only in Ptr, but in two other leaf-spot pathogens as well: *Parastagonospora nodorum* and *Biploaris sorokiniana* (Friesen et al., 2006; McDonald et al., 2018; 2019). The presence of *ToxA* in these three species has been attributed to the mobility of a 14 kb DNA transposon, *ToxhAT* (McDonald et al., 2019). This thesis has shown that *ToxA* (including *ToxhAT*) is nested within a much larger Starship transposon, *Horizon*, which is actively mobile within the Ptr genome. *ToxA* was also recently found within another Starship (*Sanctuary*) within *B. sorokiniana* (Bucknell et al., 2024). The presence of *ToxA* on multiple types of transposons across multiple species is a clear indication that virulence genes benefit from a close association with mobile elements. The benefits are likely even greater for necrotrophic effectors due to their inverse gene-for-gene interaction which don't rely on circumventing host defense systems, meaning that other genes are not necessarily required to cause disease. The mobility of necrotrophic effectors which may confer pathogenicity to previously non-pathogenic organisms is of a major concern as it could lead to rapid emergence of new pathogens.

Virulence genes for other pathogens, such as *Argo* in *Z. tritici*, have also been identified within Starship elements (Habig et al., 2021; Gluck-Thaler et al., 2022). However, Starship cargo is reportedly quite diverse, containing a variety of accessory genes which are beneficial in certain situations like those that require heavy metal tolerance (Urquhart et al., 2022; Gluck-Thaler et al., 2022). A large number of genes were found within *Horizon*, the majority of which had no known functional annotations. It will be interesting in the future to determine if there is differential presence/absence of genes within *Horizon* between isolates from different regions or hosts, which may indicate that *Horizon* harbours genes contributing to localized Ptr adaptations. Searching for a genome containing both *Horizon* and *Sanctuary* (either in Ptr or *B. sorokiniana*) would also be of value, as it might help to explain how *ToxhAT* was acquired by two different Starships.

During analysis of the *ToxB* region, multiple Starship cargo genes were identified (NLR – HET; patatin-like phosphatase; and DUF3723), however, the characteristic tyrosine recombinase was absent in the two isolates examined. The lack of the captain, the differential presence/absence of cargo genes, and a number of transposon incursions suggest that the region may be an ancient, derelict Starship. After a larger number of genomes had been sequenced, it was found that the size of the accessory region (i.e. *Icarus*) containing *ToxB* is highly variable between isolates. The putative *Icarus* remains to be explored in detail in these newer assemblies, being beyond the scope of this thesis. If it is confirmed that the accessory region is indeed a Starship, by finding a non-functional captain for example, it opens the door to the possibility that *ToxB* was acquired horizontally at some point in the evolutionary history of Ptr. This HGT, if it occurred, likely would have preceded the splitting of Ptr from sister species such as *P. bromi* and *P.*

seminiperda, each of which also contain *ToxB*. Therefore, genomes of these species should be included in this type of analysis. Identifying two primary necrotrophic effectors on Starships in Ptr would be significant, indicating that Starships have played a critical role in transforming Ptr into a dangerous pathogen.

Irrespective of whether or not the *ToxB* accessory region is a Starship, transposons have significantly impacted the evolution of *ToxB*. The duplication of *ToxB* from a single-copy gene into a multi-copy gene has been a longstanding question in the Ptr community. This work has shown that *ToxB* is amplified through the Helitron-like-element, *ToxB-HLE*. Helitrons are unique transposable elements which undergo rolling-circle replication and which have limited genomic fingerprints to help identify their activity (Boa and Jurka, 2013; Thomas and Pritham, 2015; Grabundzija, et al., 2016). Helitrons also have the potential to be horizontally transferred (Thomas et al., 2010; Castanera et al., 2014; Coastes et al., 2015; Heringer and Kuhn, 2022), meaning *ToxB* may in the future be transferred to other leaf-spot pathogens such as *B. sorokiniana* or *P. nodorum* via *ToxB-HLE* or *Icarus*. Transposons have long been considered 'selfish' in the sense they appear to be parasitic in nature. However, when genes carried by a transposon confer beneficial or adaptive traits, transposon fitness becomes closely linked to the host organisms reproductive success (Billane et al., 2022; Gluck-Thaler et al., 2023).

The nature of transposons is such that in addition to conferring mobility to virulence genes, they also are capable of gene disruptions. In this work, it was shown that multiple independent retrotransposon insertions from different *Copia* elements disrupted the *ToxB* gene, creating inactive *toxb* haplotypes. It would not be surprising to find that transposon insertions have been responsible for the creation of other *toxb* haplotypes in

the future. The *ToxA* gene has also had insertions in some isolates of Ptr. The first identified was a 166 bp insertion element, *PtrHp1*, which had no effect on *ToxA* expression (Moolhuijzen et al., 2018b). Recently though, it was reported that a different insertion element, *PtrHp2* (170 bp), had disrupted the *ToxA* gene, and in this case prevented expression of the effector (Mironenko et al., 2024). The duality of transposons is on clear display within the Ptr genome. Transposons have benefitted Ptr with the acquisition of *ToxA* and the amplification of *ToxB;* meanwhile, other transposons or mobile elements have disrupted both genes. Figure 5.1 below shows how various transposons may have influenced the virulence of Ptr within the context of the currently described race system (see Figure 1.1 for a description of Ptr races).

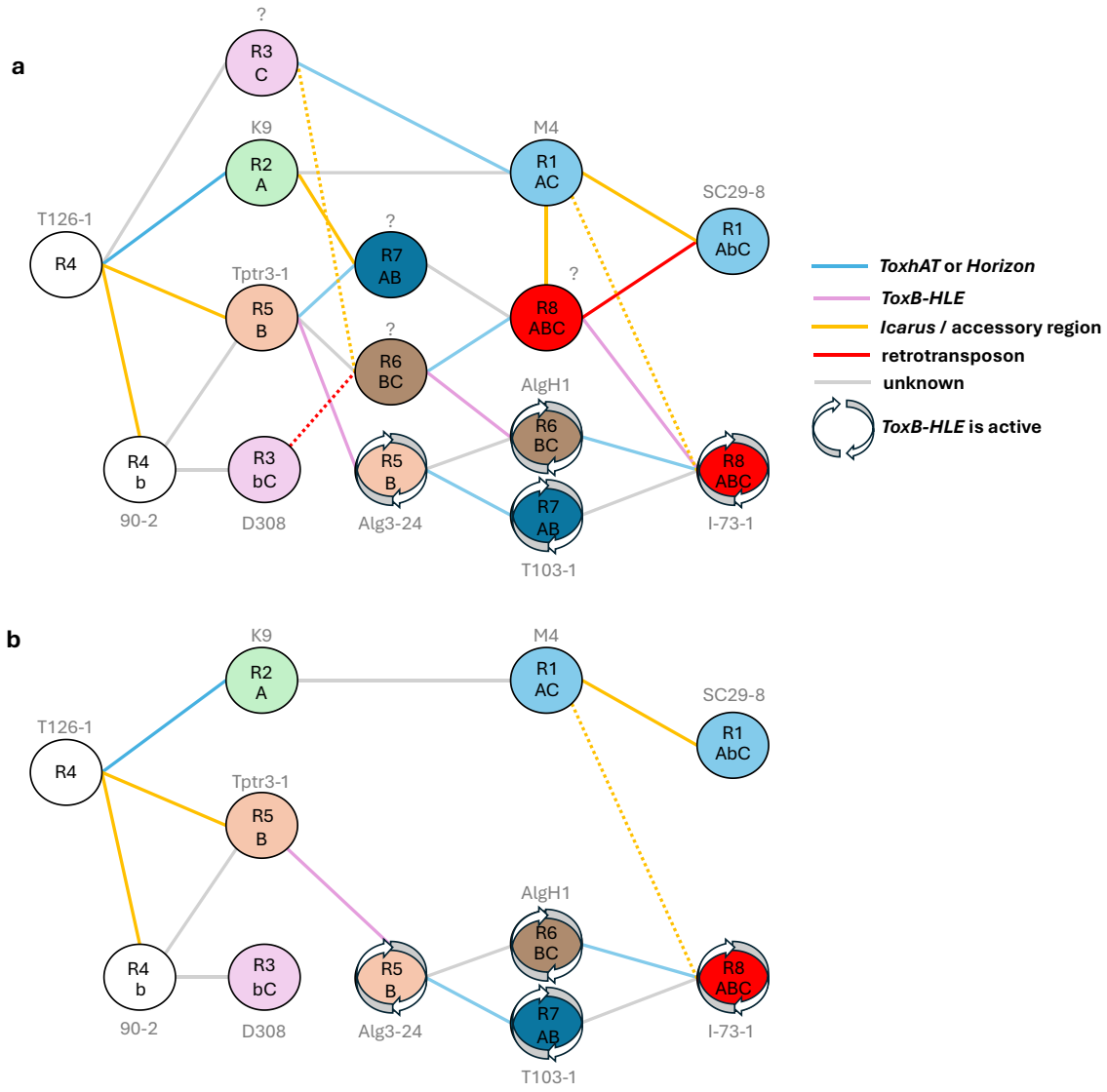


Figure 5.1 Model illustrating how transposons may link the virulence races in *Pyrenophora tritici-repentis*. Circles indicate specific genotypes where presence of a Tox gene is indicated with A=*ToxA*, B=*ToxB*, and b=*toxb*. In the case of ToxC the phenotype is indicated with C as the gene(s) are unknown. Blue lines indicate activity of *ToxhAT* or *Horizon*; pink lines indicate the formation of *ToxB-HLE*; orange lines indicate activity of *Icarus* or accessory chromosomal arm containing *ToxB* or *toxb*; red lines indicate activity of retrotransposons; grey lines indicate an unknown mechanism (generally the gain/disruption/loss of ToxC); circular arrows indicate *ToxB-HLE* is active; dashed lines are used for clarity when lines are crossing. Names of isolates beside a circle are representative examples of the genotype. **a** contains unobserved genotypes**; b** contains only observed genotypes.

5.3 FUTURE WORK IN THE STUDY OF THE *PYRENOPHORA TRITICI-REPENTIS* GENOME

As in depth as this work has been, there are significant extensions to each avenue of research. For example, continuing to explore the role of Starships in the evolution of Ptr is a high priority. Searching for more Starships and discovering what types of cargo are being mobilized in Ptr is a key concern. Additionally, extending analysis of the *ToxA* containing *Horizon* could yield insights into how Starships grow and acquire cargo genes or how they decrease in size by removing or deleting cargo. It is also still unclear whether the accessory region containing *ToxB* is a derelict Starship. With the large number of *ToxB* containing isolates already sequenced, extending analysis of this region is a logical step. Gene knock-out experiments and the use of transcriptomics can help determine which, if any, genes near the GWAS peak contribute to the production or secretion of the as yet unknown ToxC effector. Other avenues of research are also possible such as a further examination of predicted effectors (both core and accessory) and a systematic search for CNVs in the Ptr genome. Continued sequencing of Ptr genomes will undoubtedly reveal more insights and discoveries surrounding the evolution and emergence of *Pyrenophora tritici-repentis*.

5.4 REFERENCES

Aboukhaddour, R., Cloutier, S., Ballance, G.M. and Lamari, L., 2009. Genome characterization of *Pyrenophora tritici-repentis* isolates reveals high plasticity and independent chromosomal location of *ToxA* and *ToxB*. *Molecular Plant Pathology*, **10**: 201-212.

Aboukhaddour, R., Cloutier, S., Lamari, L. and Strelkov, S.E., 2011. Simple sequence repeats and diversity of globally distributed populations of *Pyrenophora tritici-repentis*. *Canadian Journal of Plant Pathology*, **33**: 389-399.

Aboukhaddour, R., Kim, Y.M. and Strelkov, S.E., 2012. RNA-mediated gene silencing of *ToxB* in *Pyrenophora tritici-repentis*. *Molecular Plant Pathology*, **13**: 318-326.

Badet, T., Oggenfuss, U., Abraham, L., McDonald, B.A. and Croll, D., 2020. A 19-isolate reference-quality global pangenome for the fungal wheat pathogen *Zymoseptoria tritici*. *BMC Biology*, **18**: 1-18.

Bao, W. and Jurka, J., 2013. Homologues of bacterial TnpB_IS605 are widespread in diverse eukaryotic transposable elements. *Mobile DNA*, **4**: 1-16.

Bao, J., Wang, Z., Chen, M., Chen, S., Chen, X., Xie, J., Tang, W., Zheng, H. and Wang, Z., 2022. Pan-genomics reveals a new variation pattern of secreted proteins in *Pyricularia oryzae*. *Journal of Fungi*, **8:** 1238.

Billane, K., Harrison, E., Cameron, D. and Brockhurst, M.A., 2022. Why do plasmids manipulate the expression of bacterial phenotypes?. *Philosophical Transactions of the Royal Society,* **377**: 20200461.

Bourque, G., Burns, K.H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H.L., Macfarlan, T.S. and Mager, D.L., 2018. Ten things you should know about transposable elements. *Genome Biology*, **19**: 1-12.

Bucknell, A., Wilson, H., Goncalves dos Santos, K.C., Simpfendorfer, S., Milgate, A., Germain, H., Solomon, P., Bentham, A. and McDonald, M., 2024. Sanctuary: A Starship transposon facilitating the movement of the virulence factor *ToxA* in fungal wheat pathogens. *bioRxiv*: 2024-03.

Castanera, R., Pérez, G., López, L., Sancho, R., Santoyo, F., Alfaro, M., Gabaldón, T., Pisabarro, A.G., Oguiza, J.A. and Ramírez, L., 2014. Highly expressed captured genes and cross-kingdom domains present in Helitrons create novel diversity in *Pleurotus ostreatus* and other fungi. *BMC Genomics*, **15**: 1-17.

Chan, J.E. and Kolodner, R.D., 2011. A genetic and structural study of genome rearrangements mediated by high copy repeat *Ty1* elements. *PLoS Genetics*, **7**: e1002089.

Chen, H., King, R., Smith, D., Bayon, C., Ashfield, T., Torriani, S., Kanyuka, K., Hammond-Kosack, K., Bieri, S. and Rudd, J., 2023. Combined pangenomics and transcriptomics reveals core and redundant virulence processes in a rapidly evolving fungal plant pathogen. *BMC Biology*, **21**: 24.4

Coates, B.S., 2015. Horizontal transfer of a non-autonomous Helitron among insect and viral genomes. *BMC Genomics*, **16**: 1-10.

Croll, D. and McDonald, B.A., 2012. The accessory genome as a cradle for adaptive evolution in pathogens. *PLoS pathogens*, **8**: e1002608.

Dong, S., Raffaele, S. and Kamoun, S., 2015. The two-speed genomes of filamentous pathogens: waltz with plants. *Current Opinion in Genetics & Development*, **35**: 57-65.

Faino, L., Seidl, M.F., Shi-Kunne, X., Pauper, M., van den Berg, G.C., Wittenberg, A.H. and Thomma, B.P., 2016. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. *Genome Research*, **26**: 1091-1100.

Frantzeskakis, L., Kusch, S. and Panstruga, R., 2018. The need for speed: compartmentalized genome evolution in filamentous phytopathogens. *Molecular Plant Pathology*, **20**: 3.

Friesen, T.L., Stukenbrock, E.H., Liu, Z., Meinhardt, S., Ling, H., Faris, J.D., Rasmussen, J.B., Solomon, P.S., McDonald, B.A. and Oliver, R.P., 2006. Emergence of a new disease as a result of interspecific virulence gene transfer. *Nature Genetics*, **38**: 953-956.

Gluck-Thaler, E., Ralston, T., Konkel, Z., Ocampos, C.G., Ganeshan, V.D., Dorrance, A.E., Niblack, T.L., Wood, C.W., Slot, J.C., Lopez-Nicora, H.D. and Vogan, A.A., 2022. Giant starship elements mobilize accessory genes in fungal genomes. *Molecular Biology and Evolution*, **39**: msac109.

Grabundzija, I., Messing, S.A., Thomas, J., Cosby, R.L., Bilic, I., Miskey, C., Gogol-Döring, A., Kapitonov, V., Diem, T., Dalda, A. and Jurka, J., 2016. A Helitron transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. *Nature Communications*, **7**: 10716.

Habig, M., Schotanus, K., Hufnagel, K., Happel, P. and Stukenbrock, E.H., 2021. *Ago1* affects the virulence of the fungal plant pathogen *Zymoseptoria tritici*. *Genes*, **12**: 1011.

Hartmann, F.E. and Croll, D., 2017. Distinct trajectories of massive recent gene gains and losses in populations of a microbial eukaryotic pathogen. *Molecular Biology and Evolution*, **34**: 2808-2822.

Heringer, P. and Kuhn, G.C., 2022. Pif1 Helicases and the evidence for a prokaryotic origin of helitrons. *Molecular Biology and Evolution*, **39**: msab334.

Hosford Jr, R.M., 1971. A form of *Pyrenophora trichostoma* pathogenic to wheat and other grasses. *Phytopathology*, **61**: 28-32.

Joubert, P.M. and Krasileva, K.V., 2024. Distinct genomic contexts predict gene presence–absence variation in different pathotypes of *Magnaporthe oryzae*. *Genetics*, **226**: iyae012.

Kamel, S., Cherif, M., Hafez, M., Despins, T. and Aboukhaddour, R., 2019. *Pyrenophora tritici–repentis* in Tunisia: Race structure and effector genes. *Frontiers in Plant Science*, **10**: 1562.

Kaushik, A., Roberts, D.P., Ramaprasad, A., Mfarrej, S., Nair, M., Lakshman, D.K. and Pain, A., 2022. Pangenome analysis of the soilborne fungal phytopathogen *Rhizoctonia solani* and development of a comprehensive web resource: RsolaniDB. *Frontiers in Microbiology*, **13**: 839524.

Krupinsky, J.M., 1992a. Aggressiveness of *Pyrenophora tritici-repentis* isolated from grass and barley hosts. *Plant Disease*, **78**: 783-789.

Krupinsky, J.M., 1992b. Grass hosts of *Pyrenophora tritici-repentis*. *Plant Disease*, **76**: 92-95.

Lamari, L., Strelkov, S.E., Yahyaoui, A., Amedov, M., Saidov, M., Djunusova, M. and Koichibayev, M., 2005. Virulence of *Pyrenophora tritici-repentis* in the countries of the Silk Road. *Canadian Journal of Plant Pathology*, **27**: 383-388.

Lawrence, J.G., 2005. Common themes in the genome strategies of pathogens. *Current Opinion in Genetics & Development*, **15**: 584-588.

Manning, V.A., Pandelova, I., Dhillon, B., Wilhelm, L.J., Goodwin, S.B., Berlin, A.M., Figueroa, M., Freitag, M., Hane, J.K., Henrissat, B. and Holman, W.H., 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: Genes, Genomes, Genetics*, **3**: 41-63.

McDonald, M.C., Ahren, D., Simpfendorfer, S., Milgate, A. and Solomon, P.S., 2018. The discovery of the virulence gene *ToxA* in the wheat and barley pathogen *Bipolaris sorokiniana*. *Molecular Plant Pathology*, **19**: 432-439.

McDonald, M.C., Taranto, A.P., Hill, E., Schwessinger, B., Liu, Z., Simpfendorfer, S., Milgate, A. and Solomon, P.S., 2019. Transposon-mediated horizontal transfer of the host-specific virulence protein *ToxA* between three fungal wheat pathogens. *MBio*, **10**: 10-1128.

Mironenko, N.V., Orina, A.S. and Kovalenko, N.M., 2024. Novel *ToxA* insertion element in *Pyrenophora tritici-repentis*. *Russian Journal of Genetics*, **60**: 1161-1167.

Moolhuijzen, P., See, P.T., Hane, J.K., Shi, G., Liu, Z., Oliver, R.P. and Moffat, C.S., 2018a. Comparative genomics of the wheat fungal pathogen *Pyrenophora tritici-repentis* reveals chromosomal variations and genome plasticity. *BMC Genomics*, **19**: 1-23.

Moolhuijzen, P.M., See, P.T., Oliver, R.P. and Moffat, C.S., 2018b. Genomic distribution of a novel *Pyrenophora tritici-repentis ToxA* insertion element. *PLoS One*, **13:** p.e0206586.

Powell, A.J., Conant, G.C., Brown, D.E., Carbone, I. and Dean, R.A., 2008. Altered patterns of gene duplication and differential gene gain and loss in fungal pathogens. *BMC Genomics*, **9**: 1-15.

Raffaele, S., Farrer, R.A., Cano, L.M., Studholme, D.J., MacLean, D., Thines, M., Jiang, R.H., Zody, M.C., Kunjeti, S.G., Donofrio, N.M. and Meyers, B.C., 2010. Genome evolution following host jumps in the Irish potato famine pathogen lineage. *Science*, **330**: 1540-1543.

Strelkov, S.E. and Lamari, L., 2003. Host–parasite interactions in tan spot [*Pyrenophora tritici-repentis*] of wheat. *Canadian Journal of Plant Pathology*, **25**: 339-349.

Syme, R.A., Tan, K.C., Rybak, K., Friesen, T.L., McDonald, B.A., Oliver, R.P. and Hane, J.K., 2018. Pan-*Parastagonospora* comparative genome analysis—effector prediction and genome evolution. *Genome Biology and Evolution*, **10**: 2443-2457.

Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S.V., Crabtree, J., Jones, A.L., Durkin, A.S. and DeBoy, R.T., 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proceedings of the National Academy of Sciences*, **102**: 13950-13955.

Thomas, J., Schaack, S. and Pritham, E.J., 2010. Pervasive horizontal transfer of rolling-circle transposons among animals. *Genome Biology and Evolution*, **2**: 656-664.

Thomas, J. and Pritham, E.J., 2015. Helitrons, the eukaryotic rolling-circle transposable elements. *Mobile DNA,* **3**: 891-924.

Torres, D.E., Oggenfuss, U., Croll, D. and Seidl, M.F., 2020. Genome evolution in fungal plant pathogens: looking beyond the two-speed genome model. *Fungal Biology Reviews*, **34**: 136-143.

Urquhart, A.S., Chong, N.F., Yang, Y. and Idnurm, A., 2022. A large transposable element mediates metal resistance in the fungus *Paecilomyces variotii*. *Current Biology*, **32**: 937-950.

van Westerhoven, A., Fokkens, L., Wissink, K., Kema, G.H., Rep, M. and Seidl, M.F., 2024. Reference-free identification and pangenome analysis of accessory chromosomes in a major fungal plant pathogen. *bioRxiv*: 2024-12.

Welgemoed, T., Duong, T.A., Barnes, I., Stukenbrock, E.H. and Berger, D.K., 2023. Population genomic analyses suggest recent dispersal events of the pathogen *Cercospora zeina* into East and Southern African maize cropping systems. *G3: Genes, Genomes, Genetics*, **13**: jkad214.

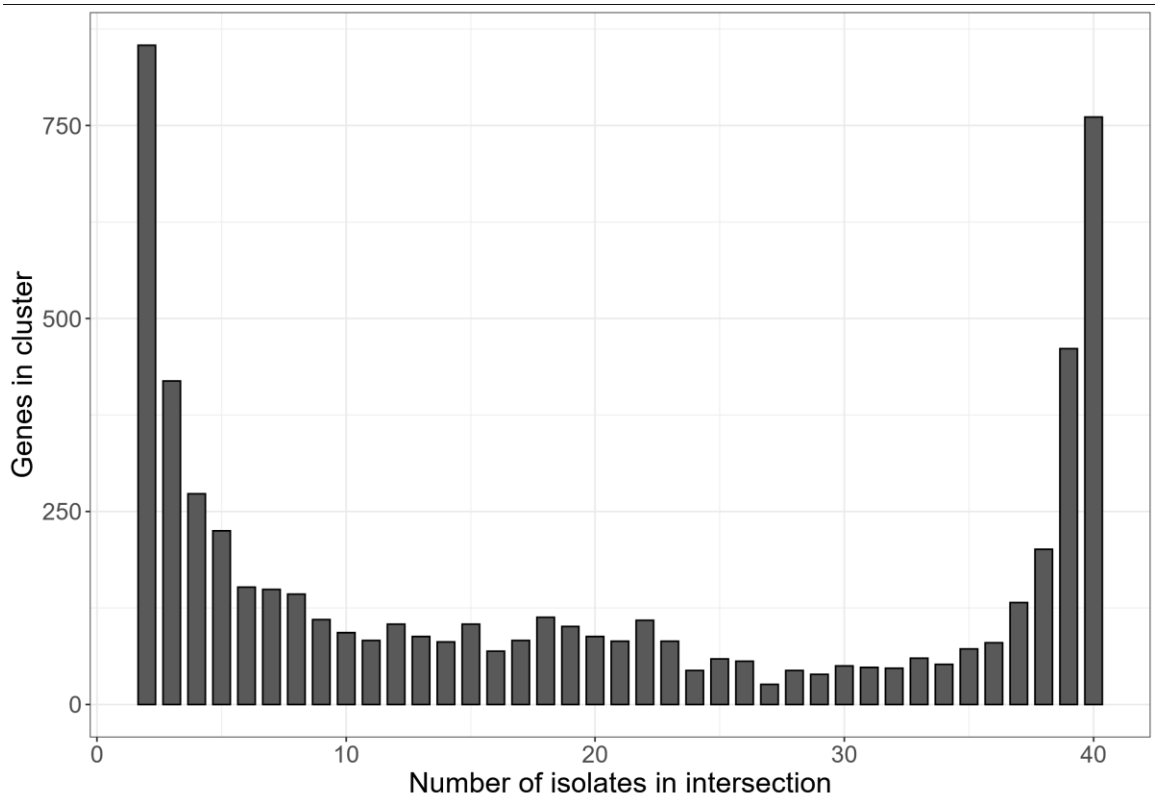
Wöstemeyer, J. and Kreibich, A., 2002. Repetitive DNA elements in fungi (Mycota): impact on genomic architecture and evolution. *Current Genetics*, **41**: 189-198.

Whittle, C.A., Nygren, K. and Johannesson, H., 2011. Consequences of reproductive mode on genome evolution in fungi. *Fungal Genetics and Biology*, **48**: 661-667.

Wyka, S., Mondo, S., Liu, M., Nalam, V. and Broders, K., 2022. A large accessory genome and high recombination rates may influence global distribution and broad host range of the fungal plant pathogen *Claviceps purpurea*. *PloS One*, **17**: e0263496.

Appendices

Appendix A: Supplementary material for Chapter 2



Additional File 2.1 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 isolates 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.

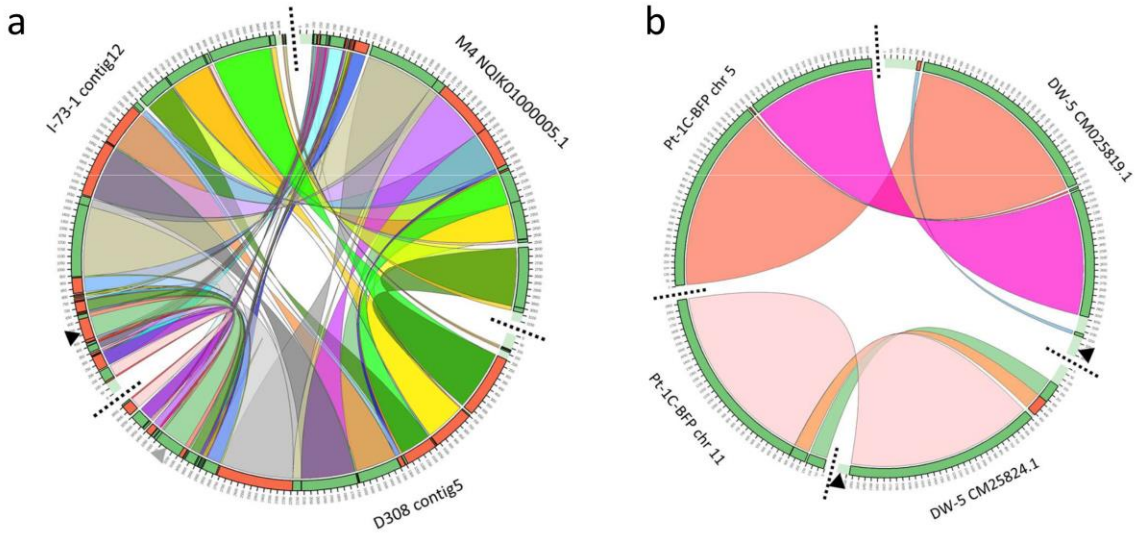
Additional File 2.2 Maximum likelihood phylogenies created by RAxML based on SNP data. **a** ML tree with all isolates; **b** ML tree with the divergent outgroup omitted to aid reading of the other branches

Additional File 2.3 Curated output from dbCAN for grouping and counting the number of genes in specific CAZyme families. Table too large to display; for table see https://github.com/fungal-spore/Ptr-pangenome-paper

Additional File 2.4 Annotation of genes present within the 143 kb Starship transposon 'Horizon' present in the race 8 isolate I-73-1.

Gene	PfamDB	Annotation
gene_09854	PF00400	WD domain, G-beta repeat
gene_09854	PF17111	Fungal N-terminal domain of STAND proteins
gene_09854	PF05729	NACHT domain
gene_09856	PF00249	Myb-like DNA-binding domain
gene_09859	PF00069	Protein kinase domain
gene_09868	PF02373	JmjC domain, hydroxylase
gene_09870	PF00704	Glycosyl hydrolases family 18
gene_09875	PF08030	Ferric reductase NAD binding domain
gene_09883	PF02714	Calcium-dependent channel, 7TM region, putative phosphate
gene_09883	PF13967	Late exocytosis, associated with Golgi transport
gene_09883	PF14703	Cytosolic domain of 10TM putative phosphate transporter
gene_09892	PF00172	Fungal Zn(2)-Cys(6) binuclear cluster domain
gene_09894	PF00856	SET domain
gene_09894	PF02373	JmjC domain, hydroxylase
gene_09899	PF07859	alpha/beta hydrolase fold
gene_09853	PF11917	Protein of unknown function (DUF3435)
gene_09881	PF12520	Protein of unknown function (DUF3723)
gene_09883	PF02714	Calcium-dependent channel, 7TM region, putative phosphate
gene_09883	PF13967	Late exocytosis, associated with Golgi transport
gene_09883	PF14703	Cytosolic domain of 10TM putative phosphate transporter
gene_09883	PF02714	Calcium-dependent channel, 7TM region, putative phosphate
gene_09883	PF13967	Late exocytosis, associated with Golgi transport
gene_09883	PF14703	Cytosolic domain of 10TM putative phosphate transporter
gene_09854	PF00400	WD domain, G-beta repeat
gene_09854	PF00400	WD domain, G-beta repeat
gene_09854	PF00400	WD domain, G-beta repeat
gene_09854	PF00400	WD domain, G-beta repeat
gene_09854	PF00400	WD domain, G-beta repeat
gene_09854	PF00400	WD domain, G-beta repeat
gene_09854	PF17111	Fungal N-terminal domain of STAND proteins
gene_09854	PF05729	NACHT domain
gene_09869	PF12796	Ankyrin repeats (3 copies)
gene_09886	PF11584	Proteinaceous host-selective toxin ToxA
gene_09893	PF06985	Heterokaryon incompatibility protein (HET)
gene_09895	PF05700	Breast carcinoma amplified sequence 2 (BCAS2)
gene_09895	PF12520	Protein of unknown function (DUF3723)
gene_09855	N/A	not in database
gene_09857	N/A	not in database
gene_09858	N/A	not in database
gene_09860	N/A	not in database
gene_09861	N/A	not in database
gene_09862	N/A	not in database
gene_09863	N/A	not in database
gene_09864	N/A	not in database
gene_09865	N/A	not in database
gene_09866	N/A	not in database

gene_09867		not in database
gene_09871	N/A	not in database
gene_09872	N/A	not in database
gene_09873	N/A	not in database
gene_09874	N/A	not in database
gene_09876	N/A	not in database
gene_09877	N/A	not in database
gene_09878	N/A	not in database
gene_09879	N/A	not in database
gene_09880	N/A	not in database
gene_09882	N/A	not in database
gene_09884	N/A	not in database
gene_09885	N/A	not in database
gene_09887	N/A	not in database
gene_09888	N/A	not in database
gene_09889	N/A	not in database
gene_09890	N/A	not in database
gene_09891	N/A	not in database
gene_09896	N/A	not in database
gene_09897	N/A	not in database
gene_09898	N/A	not in database
gene_09900	N/A	not in database

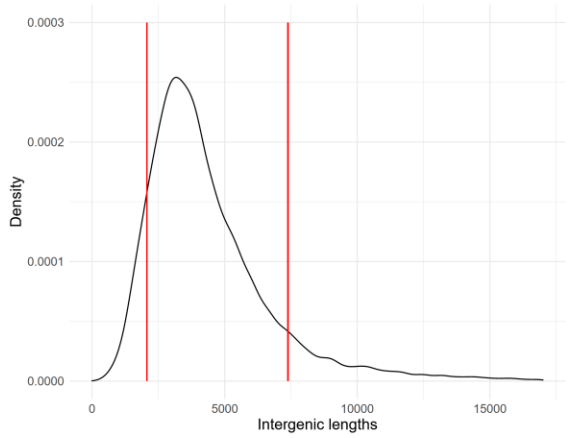


Additional File 2.5 Circular alignments of *ToxB* carrying contigs. **a** contig 5 from race 3 isolate D308, contig 12 from race 8 isolates I-73-1, and contig NQIK01000005 from race 1 isolates M4. A large 294 Kb region which contains three copies of the *ToxB* (black arrow) is visible which aligns with a section in D308 containing a single copy of the inactive *toxb* (grey arrow). **b** DW-5 contigs (CM025819.1 and CM025824.1) align to Pt-1C-BFP chromosomes (chr 5 and 11 respectively). Sections containing *ToxB* (black arrows) do not appear to be co-linear with each other or the reference chromosomes indicating possible transposon activity. These segments of DW-5 do not appear to align with 'Icarus' from I-73-1 or D308.

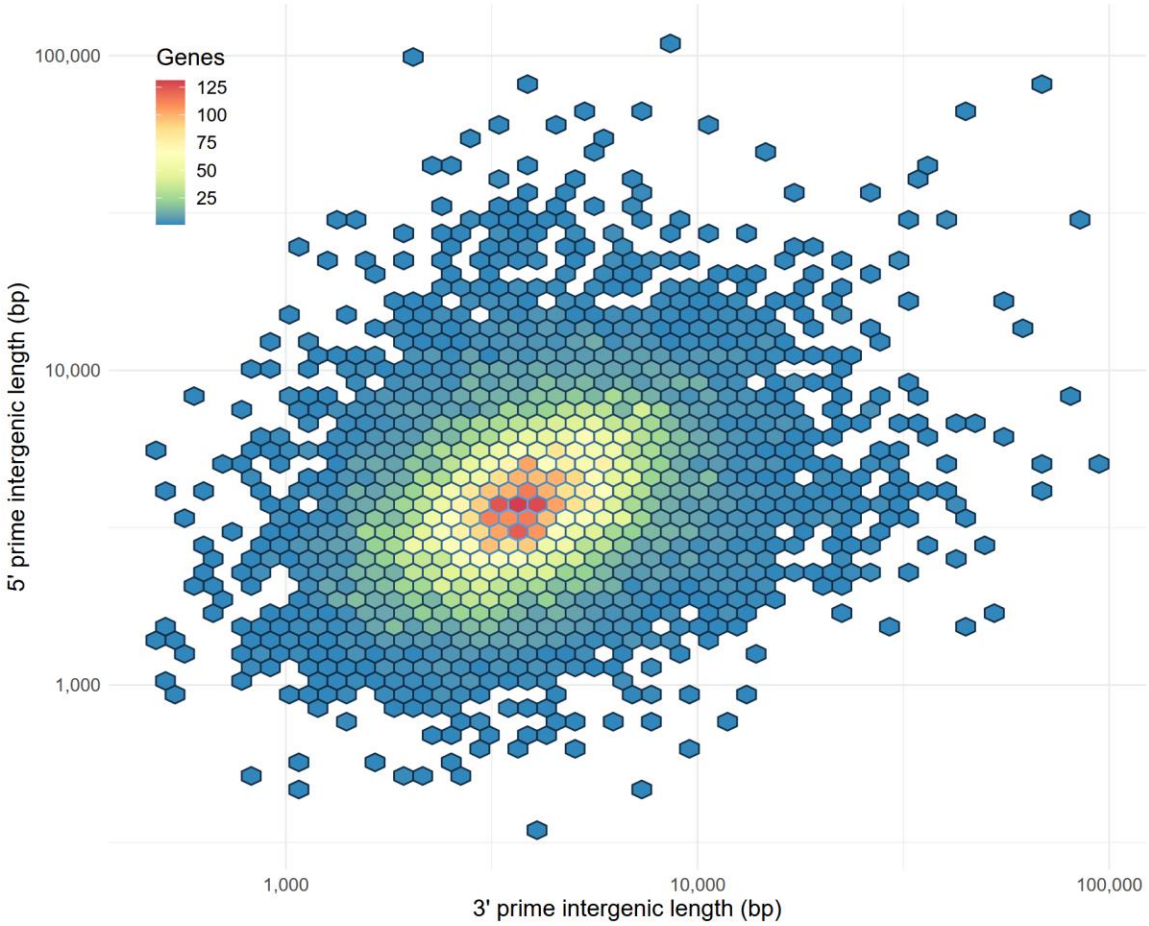
Additional File 2.6 Number of core, accessory, and singleton genes on contigs of the long-read assemblies of I-73-1 and D308.

Isolate	Contig	Size (Mb)	Number of genes			% of category		% of all genes			Gene per Mb			Core/Acc
			core	acc	total	core	acc	core	acc	total	core	acc	total	
D308	contig_1	6.85	1,976	323	2,299	19.45	13.80	15.81	2.58	18.39	289	47	336	6.12
D308	contig_2	5.69	1,596	282	1,878	15.71	12.05	12.77	2.26	15.02	233	41	274	5.66
D308**	contig_9	5.25	1,323	347	1,670	13.02	14.82	10.58	2.78	13.36	193	51	244	3.81
D308	contig_7	3.37	980	167	1,147	9.65	7.13	7.84	1.34	9.18	143	24	168	5.87
D308	contig_3	3.56	869	284	1,153	8.55	12.13	6.95	2.27	9.22	127	41	168	3.06
D308	contig_5	3.67	846	269	1,115	8.33	11.49	6.77	2.15	8.92	124	39	163	3.14
D308	contig_15	1.79	514	85	599	5.06	3.63	4.11	0.68	4.79	75	12	87	6.05
D308	contig_10	1.98	498	133	631	4.90	5.68	3.98	1.06	5.05	73	19	92	3.74
D308	contig_8	1.77	426	123	549	4.19	5.25	3.41	0.98	4.39	62	18	80	3.46
D308	contig_11	1.69	403	100	503	3.97	4.27	3.22	0.80	4.02	59	15	73	4.03
D308	contig_14	1.06	275	59	334	2.71	2.52	2.20	0.47	2.67	40	9	49	4.66
D308	contig_25	0.86	235	48	283	2.31	2.05	1.88	0.38	2.26	34	7	41	4.90
D308	contig_30	0.78	174	40	214	1.71	1.71	1.39	0.32	1.71	25	6	31	4.35
D308	contig_90	0.13	21	11	32	0.21	0.47	0.17	0.09	0.26	3	2	5	1.91
D308	contig_120	0.10	20	1	21	0.20	0.04	0.16	0.01	0.17	3	0	3	20.00
D308	contig_12	0.37	3	17	20	0.03	0.73	0.02	0.14	0.16	0	2	3	0.18
D308	contig_107	0.02	0	6	6	0.00	0.26	0.00	0.05	0.05	0	1	1	0.00
D308	contig_135	0.00	0	1	1	0.00	0.04	0.00	0.01	0.01	0	0	0	0.00
D308	contig_158	0.02	0	5	5	0.00	0.21	0.00	0.04	0.04	0	1	1	0.00
D308	contig_18	0.07	0	1	1	0.00	0.04	0.00	0.01	0.01	0	0	0	0.00
D308	contig_184	0.13	0	5	5	0.00	0.21	0.00	0.04	0.04	0	1	1	0.00
D308	contig_192	0.04	0	9	9	0.00	0.38	0.00	0.07	0.07	0	1	1	0.00
D308	contig_20	0.01	0	1	1	0.00	0.04	0.00	0.01	0.01	0	0	0	0.00
D308	contig_21	0.07	0	5	5	0.00	0.21	0.00	0.04	0.04	0	1	1	0.00
D308	contig_34	0.01	0	1	1	0.00	0.04	0.00	0.01	0.01	0	0	0	0.00
D308	contig_4	0.01	0	1	1	0.00	0.04	0.00	0.01	0.01	0	0	0	0.00

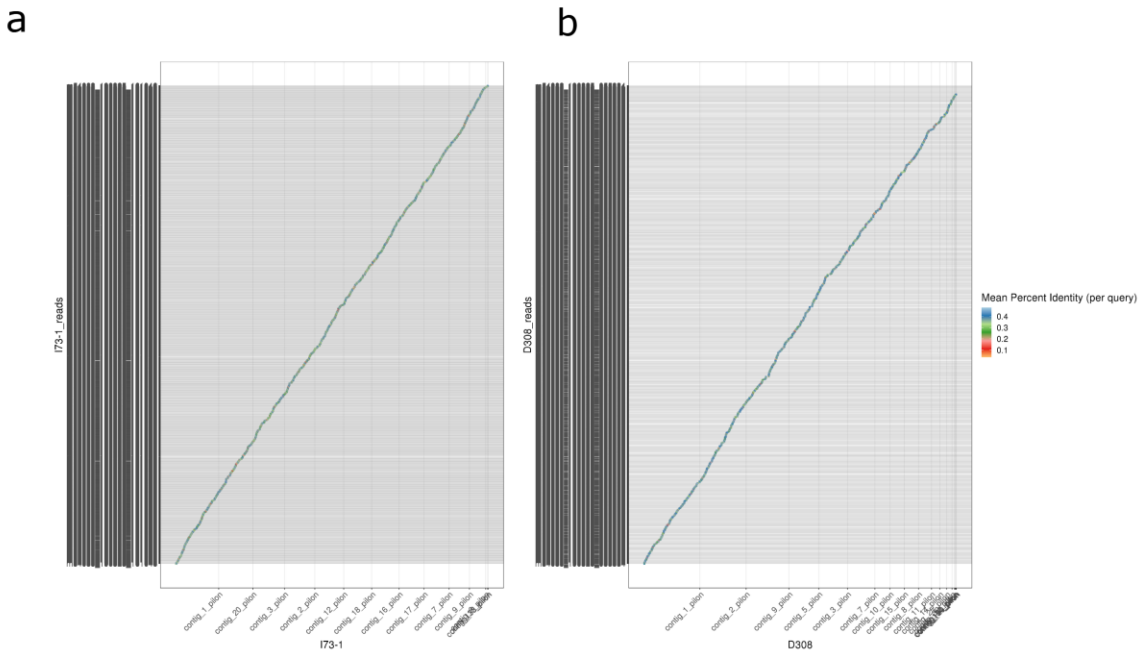
D308	contig_48	0.01	0	1	1	0.00	0.04	0.00	0.01	0.01	0	0	0	0.00
D308	contig_52	0.01	0	2	2	0.00	0.09	0.00	0.02	0.02	0	0	0	0.00
D308	contig_56	0.08	0	12	12	0.00	0.51	0.00	0.10	0.10	0	2	2	0.00
D308	contig_98	0.01	0	1	1	0.00	0.04	0.00	0.01	0.01	0	0	0	0.00
D308	contig_99	0.00	0	1	1	0.00	0.04	0.00	0.01	0.01	0	0	0	0.00
I73-1	contig_1	5.43	1,621	290	1,911	15.98	11.22	12.74	2.28	15.02	299	53	352	5.59
I73-1	contig_20	4.32	1,241	236	1,477	12.24	9.13	9.75	1.85	11.61	287	43	272	5.26
I73-1*	contig_3	4.00	1,050	277	1,327	10.35	10.72	8.25	2.18	10.43	262	51	245	3.79
I73-1	contig_18	3.55	1,039	173	1,212	10.24	6.69	8.16	1.36	9.52	293	32	223	6.01
I73-1	contig_2	3.83	1,031	325	1,356	10.17	12.57	8.10	2.55	10.65	269	60	250	3.17
I73-1**	contig_12	3.65	994	292	1,286	9.80	11.30	7.81	2.29	10.10	273	54	237	3.40
I73-1	contig_16	3.46	981	199	1,180	9.67	7.70	7.71	1.56	9.27	283	37	217	4.93
I73-1	contig_17	3.21	796	248	1,044	7.85	9.59	6.25	1.95	8.20	248	46	192	3.21
I73-1	contig_7	3.14	646	196	842	6.37	7.58	5.08	1.54	6.62	205	36	155	3.30
I73-1	contig_6	2.05	515	152	667	5.08	5.88	4.05	1.19	5.24	251	28	123	3.39
I73-1	contig_9	2.59	186	142	328	1.83	5.49	1.46	1.12	2.58	72	26	60	1.31
I73-1	contig_106	0.06	10	12	22	0.10	0.46	0.08	0.09	0.17	172	2	4	0.83
I73-1	contig_15	0.26	6	7	13	0.06	0.27	0.05	0.06	0.10	23	1	2	0.86
I73-1	contig_49	0.03	6	7	13	0.06	0.27	0.05	0.06	0.10	181	1	2	0.86
I73-1	contig_133	0.07	5	5	10	0.05	0.19	0.04	0.04	0.08	68	1	2	1.00
I73-1	contig_66	0.02	4	5	9	0.04	0.19	0.03	0.04	0.07	214	1	2	0.80
I73-1	contig_22	0.02	3	4	7	0.03	0.15	0.02	0.03	0.06	146	1	1	0.75
I73-1	contig_84	0.04	3	7	10	0.03	0.27	0.02	0.06	0.08	73	1	2	0.43
I73-1	contig_134	0.01	1	2	3	0.01	0.08	0.01	0.02	0.02	81	0	1	0.50
I73-1	contig_19	0.01	1	1	2	0.01	0.04	0.01	0.01	0.02	70	0	0	1.00
I73-1	contig_45	0.03	1	1	2	0.01	0.04	0.01	0.01	0.02	29	0	0	1.00
I73-1	contig_60	<0.01	1	1	2	0.01	0.04	0.01	0.01	0.02	0	0	0	1.00
I73-1	contig_89	0.05	1	3	4	0.01	0.12	0.01	0.02	0.03	22	1	1	0.33



Additional File 2.7 Density of ITL sizes for I-73-1, red bars indicate 90th percentile cut-off values.



Additional File 2.8 Intergenic distances of all genes in *Pyrenophora tritici-repentis* isolate D308. 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.



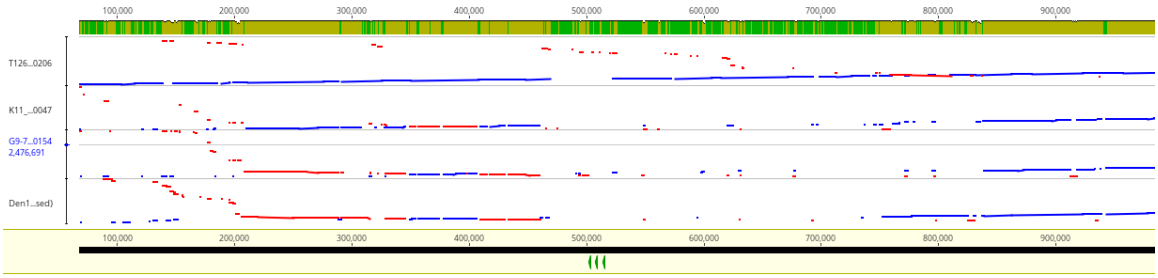
Additional File 2.9 Alignment of raw read data to long-read assemblies; **a** I-73-1; **b** D308.

Additional File 2.10 List of GenBank accession numbers associated with the genomes generated by this study and of the genomes downloaded and used for comparison.

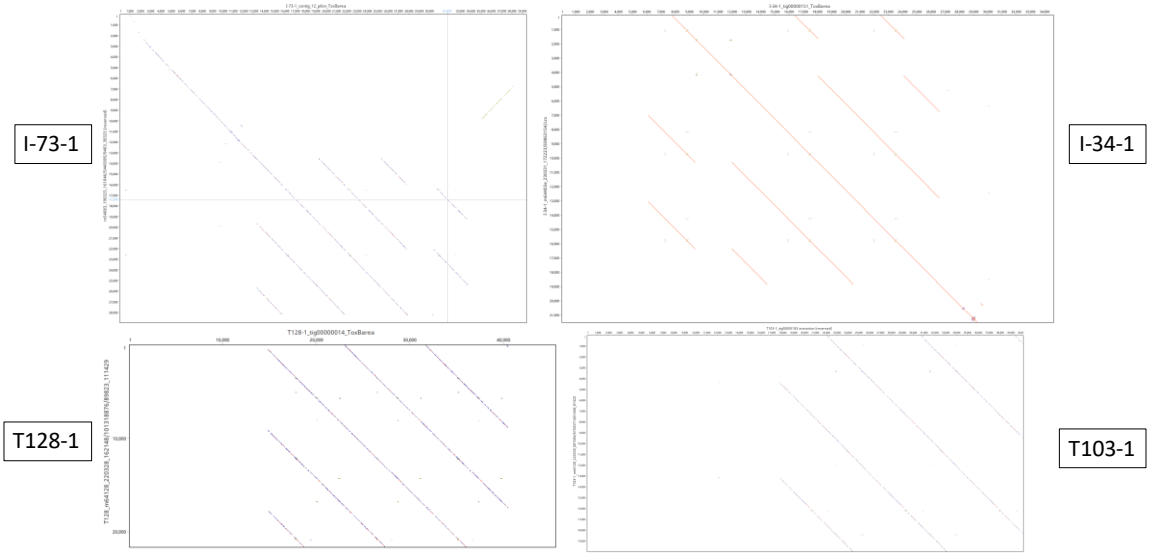
Isolate	BioProject ID	Genome Accession	BioSample Accession	Source
331-2	PRJNA803191	JAMBYI000000000	SAMN26027264	this study
86-124	PRJNA803191	JAMBYO000000000	SAMN26027258	this study
90-2	PRJNA803191	JAMBYD000000000	SAMN26027269	this study
92-171R5	PRJNA803191	JAMBYA000000000	SAMN26027272	this study
AB88-2	PRJNA803191	JAMBYN000000000	SAMN26027259	this study
ASC1	PRJNA803191	JAMBYY000000000	SAMN26027248	this study
AZ35-5	PRJNA803191	JAMBXS000000000	SAMN26027280	this study
Alg3-24	PRJNA803191	JAMBXZ000000000	SAMN26027273	this study
Alg4x-1	PRJNA803191	JAMBXY000000000	SAMN26027274	this study
AlgH1	PRJNA803191	JAMBXT000000000	SAMN26027279	this study
D308	PRJNA803191	JAKRES000000000	SAMN25871520	this study
G9-4	PRJNA803191	JAMBYC000000000	SAMN26027270	this study
I-17-2	PRJNA803191	JAMBXX000000000	SAMN26027275	this study
I-33-1	PRJNA803191	JAMBYX000000000	SAMN26027249	this study
I-34-1	PRJNA803191	JAMBXP000000000	SAMN26027283	this study
I-34-5	PRJNA803191	JAMBXW000000000	SAMN26027276	this study
I-35-18	PRJNA803191	JAMBXO000000000	SAMN26027284	this study

I-35-56	PRJNA803191	JAMBXV000000000	SAMN26027277	this study
I-36-1	PRJNA803191	JAMBXU000000000	SAMN26027278	this study
I-72-1	PRJNA803191	JAMBYH000000000	SAMN26027265	this study
I-72-7	PRJNA803191	JAMBYG000000000	SAMN26027266	this study
I-73-1	PRJNA803191	JAKRER000000000	SAMN25870380	this study
L2-1	PRJNA803191	JAMBYM000000000	SAMN26027260	this study
L3-1	PRJNA803191	JAMBYW000000000	SAMN26027250	this study
L4-1	PRJNA803191	JAMBYV000000000	SAMN26027251	this study
SC29-1	PRJNA803191	JAMBYF000000000	SAMN26027267	this study
SW1-2	PRJNA803191	JAMBYL000000000	SAMN26027261	this study
SW15-1	PRJNA803191	JAMBYK000000000	SAMN26027262	this study
SW2-1	PRJNA803191	JAMBYT000000000	SAMN26027253	this study
SW20-7	PRJNA803191	JAMBYU000000000	SAMN26027252	this study
SW21-1	PRJNA803191	JAMBYS000000000	SAMN26027254	this study
SW21-5	PRJNA803191	JAMBYE000000000	SAMN26027268	this study
SW21-7	PRJNA803191	JAMBYR000000000	SAMN26027255	this study
SW21-8	PRJNA803191	JAMBYQ000000000	SAMN26027256	this study
SW7-5	PRJNA803191	JAMBYP000000000	SAMN26027257	this study
T126-1	PRJNA803191	JAMBYB000000000	SAMN26027271	this study
T128-1	PRJNA803191	JAMBXN000000000	SAMN26027285	this study
T132-2	PRJNA803191	JAMBYJ000000000	SAMN26027263	this study
T176-2	PRJNA803191	JAMBXR000000000	SAMN26027281	this study
T181-1	PRJNA803191	JAMBXQ000000000	SAMN26027282	this study
Pt-1C-BFP	AAXI01	GCA_000149985	SAMN02953676	Manning et al., 2013
M4	NQIK02	GCA_003171515	SAMN07420926	Moolhuijzen et al., 2018
DW-5	MUXC02	GCA_003231415	SAMN04550943	Moolhuijzen et al., 2020

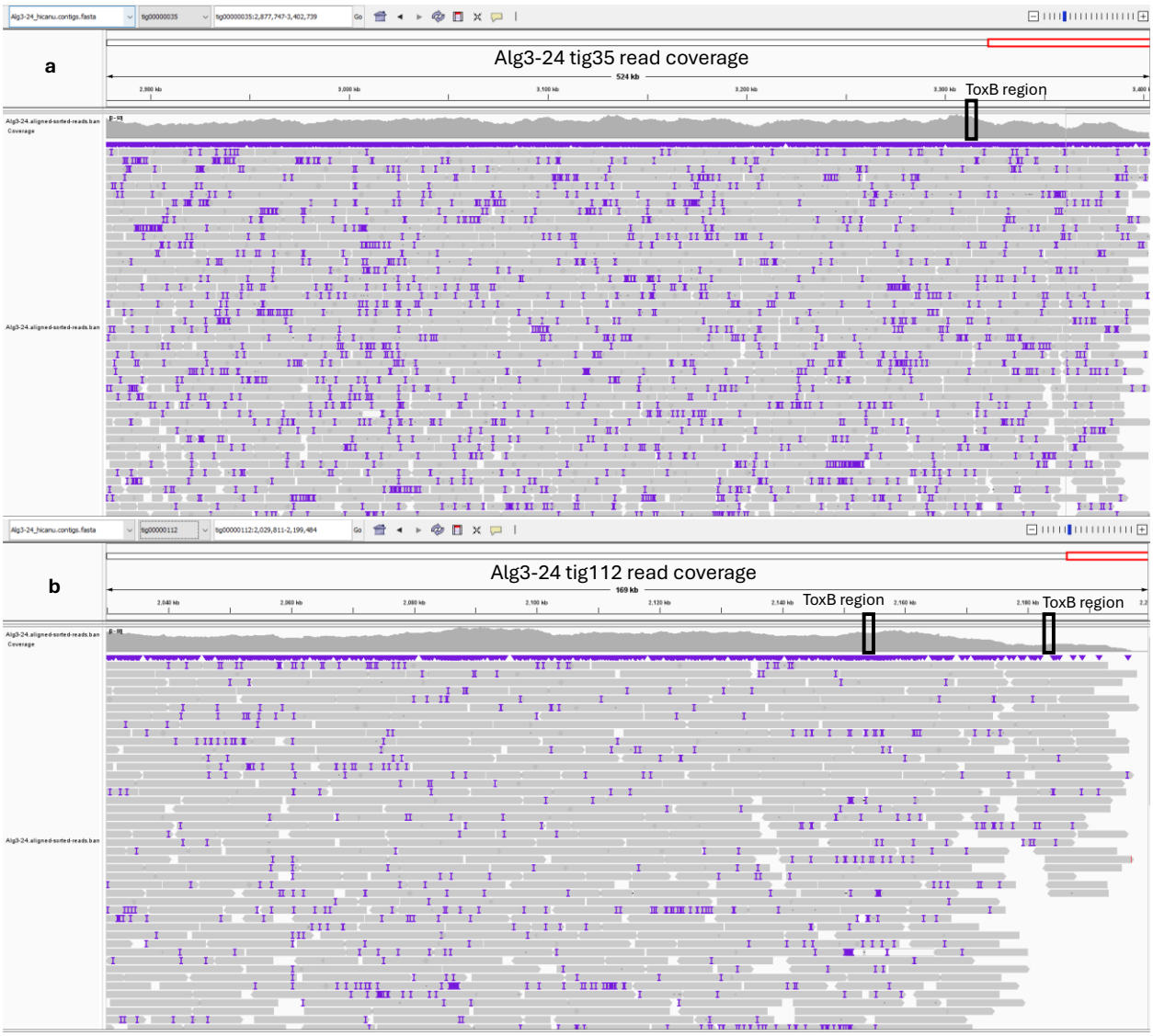
Appendix B: Supplementary material for Chapter 3



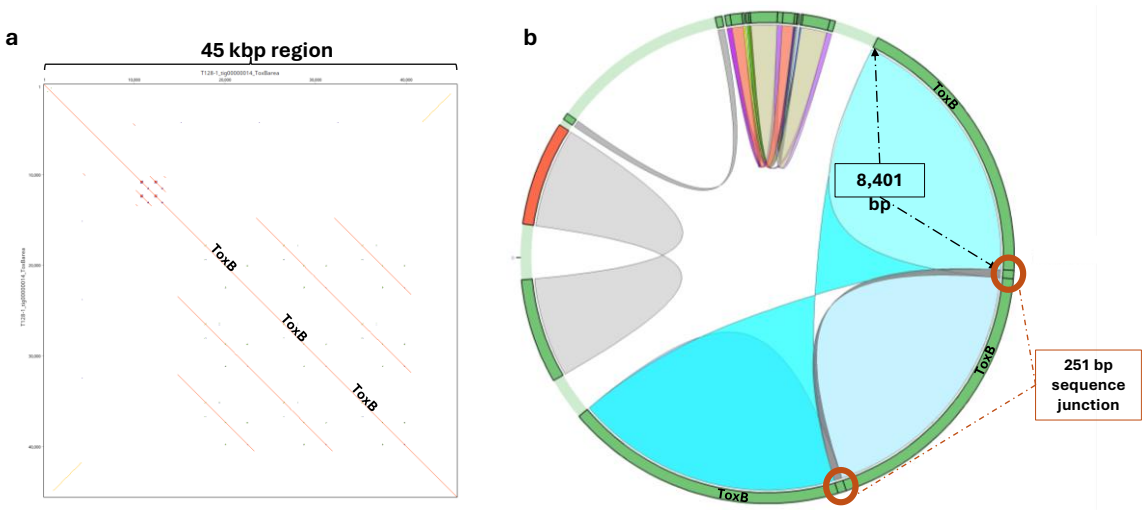
Additional File 3.1 Zoomed in contig alignment of select isolates T126-1 (track 1), K11 (track 2), G9-7 (track 3), and Den17 (track 4) which lack *ToxB/toxb* to isolate I-73-1 *ToxB* multi-copy contig12 (black bar). *ToxB* copies are marked with green triangles. Variable gap sizes are present in the *ToxB* lacking isolates, ranging from ~50 Kb to ~375 Kb. Gap sizes alter slightly when compared to single copy carrier Tptr3-1 or *toxb* carrier D308 (not shown).



Additional File 3.2 Dotplots showing single reads (y-axes) covering all copies of *ToxB* in the assemblies of four multicopy isolates (x-axes).



Additional File 3.3 Read coverage around *ToxB* in isolate Alg3-24 **a** Read coverage of *ToxB* region in tig35 (Chr05) in isolate Alg3-24; **b** Read coverage of multi *ToxB* copy tig112 (Chr10) in isolate Alg3-24. A drop in coverage is visible for the second copy.



Additional File 3.4 Example of self-alignments done with isolate T128-1 which were used to locate edges of the *ToxB* replicative unit. **a** linear dotplot showing the unidirectional nature of the duplications and helps locate the specific start and end locations; **b** circular alignment of the same 45 kbp shows the replication of the *ToxB* region more clearly, with duplications separated by a sequence junction. The alignments also revealed a pocket of nearby putative transposons and that the entire region is flanked by a hAT-like transposon.

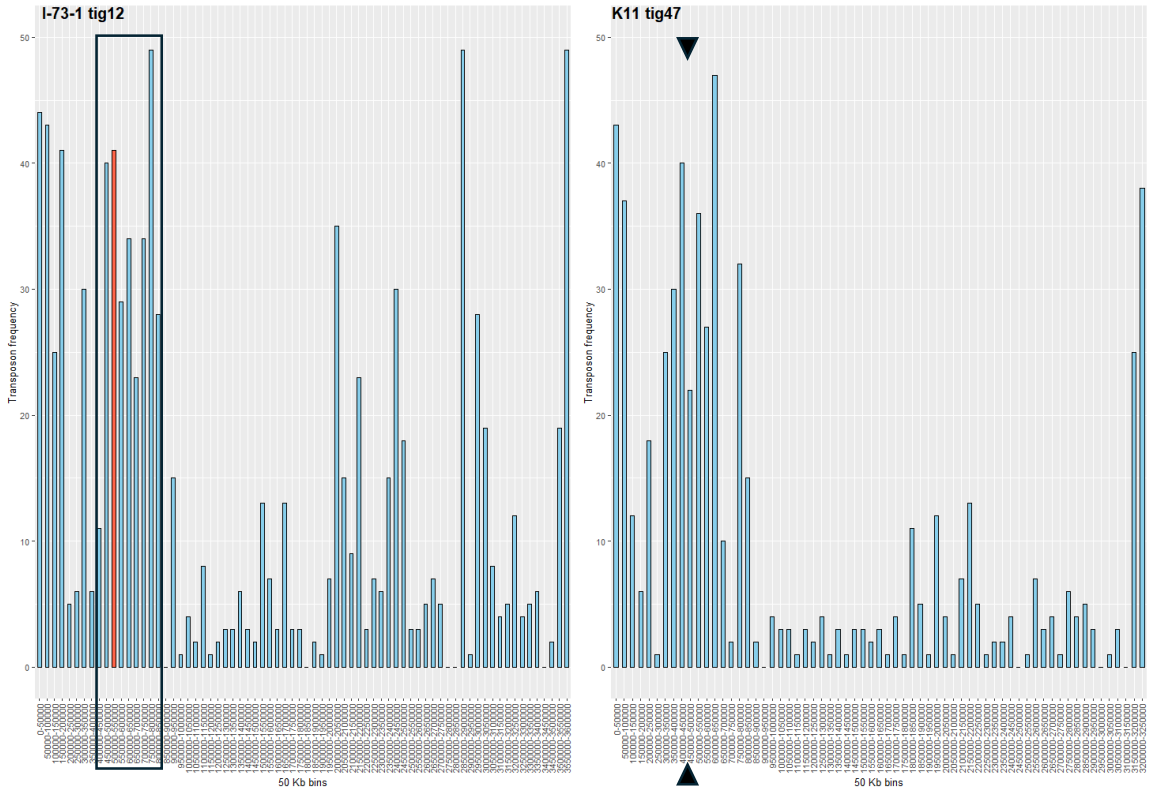
Additional File 3.5 Details of transposons and insertions near or within the *ToxB* ORF in Pyrenophora tritici-repentis.

Length	Name	Category	LTR/TIR length (bp)
5,661	*Copia-1_Ptr*	LTR retrotransposon	192 LTR
5,181	*Copia-2_Ptr*	LTR retrotransposon	165 LTR
5,549	*Copia-3_Ptr*	LTR retrotransposon	195 LTR
5,396	*Copia-4_Ptr*	LTR retrotransposon	240 LTR
5,251	*Copia-5_Ptr*	LTR retrotransposon	172 LTR
5,345	*Copia-6_Ptr*	LTR retrotransposon	102 LTR
6,070	*Copia-7_Ptr*	LTR retrotransposon	202 LTR
2,603	*hAT-1_Ptr**	DNA transposon	17 TIR
2,805	*hAT-1.2_Ptr***	DNA transposon	17 TIR
4,036	*hAT-2_Ptr*	hAT DNA transposon	257 LTR

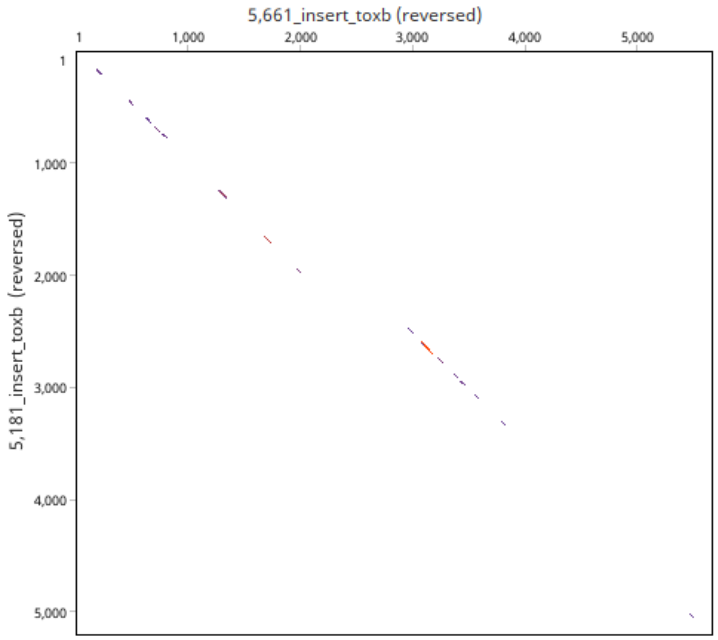
1,060	*hAT-3_Ptr*	DNA transposon	-
3,005	*hAT-4_Ptr*	DNA transposon	20 TIR
2,110	*Pogo-1_Ptr*	DNA transposon	25 TIR
5,620	*Ty3-1_Ptr*	LTR retrotransposon	256 LTR
6,770	*Ty3-2_Ptr*	LTR retrotransposon	454 LTR
5,601	-	Possible non-LTR retrotransposon	-
1,677	-	LTR transposon	147 LTR
240	-	Abandoned LTR from Copia-4	-
202	-	Abandoned LTR from Copia-7	-
170	-	Abandoned LTR from unknown	-

Additional File 3.6 Counts of transposons and the sequence junction throughout the genomes of *P. tritici-repentis*. Counts based on BLASTN 90% identity over 90% of the sequence.

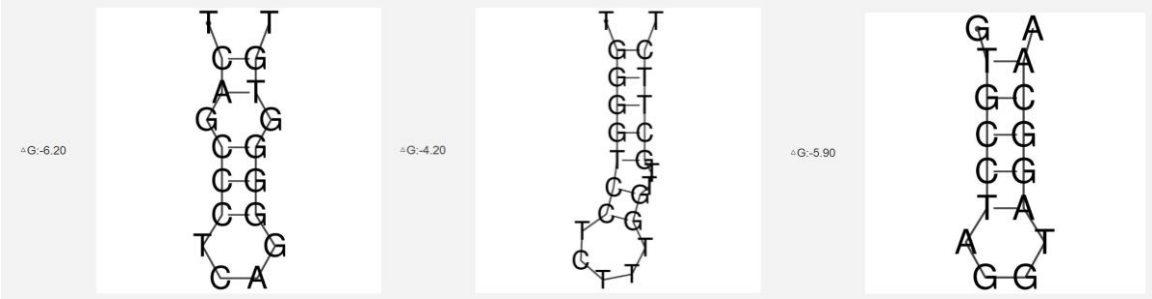
Isolate	Transposon copy count														
	Copia-1_Ptr	*Copia-2_Ptr*	*Copia-3_Ptr*	*Copia-4_Ptr*	*Copia-5_Ptr*	*Copia-6_Ptr*	*Copia-7_Ptr*	*hAT-1_Ptr*	*hAT-2_Ptr*	*hAT-3_Ptr*	*hAT-4_Ptr*	*Pogo-1_Ptr*	*Ty3-1_Ptr*	*Ty3-2_Ptr*	SJ
107224_b	3	0	4	0	0	1	6	0	0	1	0	8	0	1	53
331-2_b	12	4	5	15	19	0	25	26	0	44	54	24	0	7	53
90-2_b	0	0	2	0	0	0	4	0	0	0	0	7	0	0	58
92-171-R5_B	9	0	2	0	2	21	2	0	0	5	12	38	0	0	58
Alg3-24_B	1	5	0	19	9	1	24	20	1	21	34	28	6	6	63
AlgH1_B	10	2	5	18	13	0	24	33	0	58	62	26	0	8	48
D308*_b	9	3	5	14	15	0	25	16	0	45	63	23	0	6	30
Den17_b	3	2	1	21	9	0	28	20	0	26	48	25	1	5	44
G9-7_b	4	0	1	0	10	3	2	0	0	0	0	17	0	4	85
I-33-16_N	5	3	1	18	9	0	22	19	0	22	41	30	3	5	42
I-34-1_B	16	2	5	16	16	0	19	28	0	35	56	28	0	12	28
I-73-1_B	15	2	5	15	14	0	19	25	0	39	51	24	0	5	29
K11_N	0	0	0	0	0	0	0	0	0	1	31	14	1	11	51
K6_N	0	0	0	0	0	2	0	0	0	0	17	11	3	9	46
K9_N	0	0	0	0	0	3	0	0	0	1	30	13	2	10	47
SC22-2_b	12	4	5	14	19	0	25	23	0	42	53	26	0	7	64
SC29-1_b	7	4	3	18	17	0	22	25	0	51	41	28	0	7	31
SC29-8_b	7	4	3	18	15	0	22	25	0	53	44	27	0	7	31
SW21-5_b	8	2	1	13	13	0	21	25	0	57	53	30	0	4	40
T103-1_B	11	2	4	16	16	0	20	28	0	34	59	23	0	7	37
T126-1_N	11	2	6	15	13	0	17	24	0	42	57	25	0	8	30
T128-1_B	4	1	6	13	15	0	20	31	0	44	53	22	0	3	45
Tptr3-1_B	11	1	5	13	13	0	20	27	0	44	54	24	0	8	36



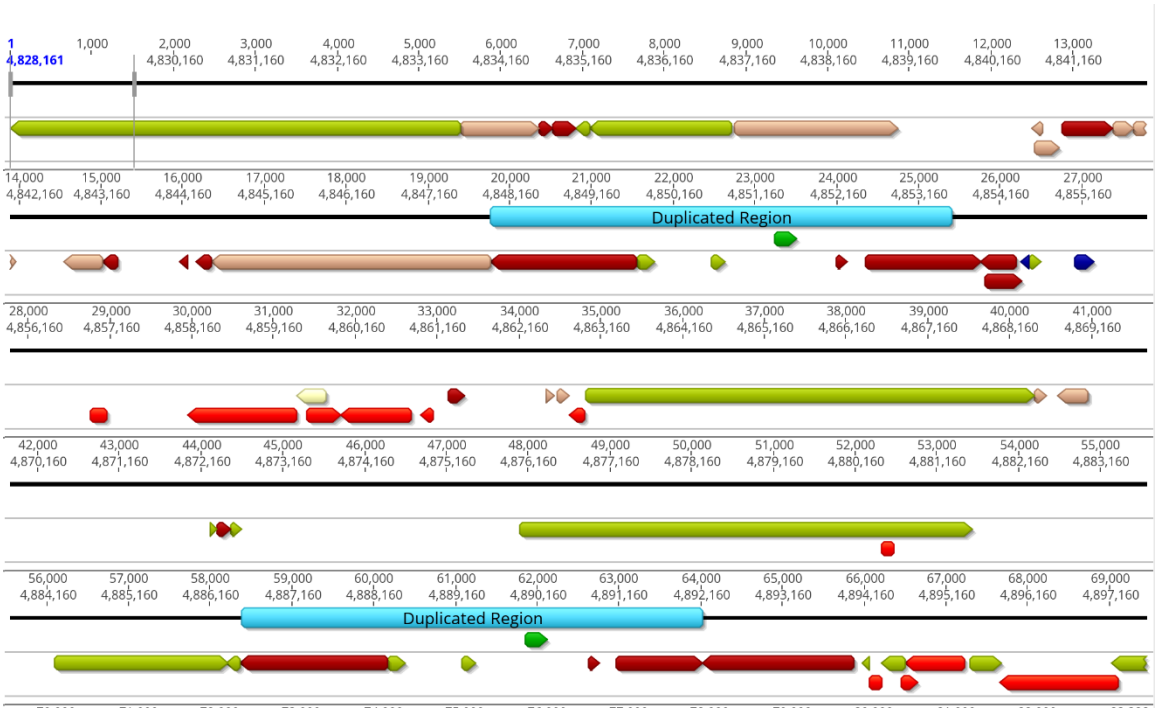
Additional File 3.7 Contig with *ToxB* in I-73-1 and homologous contig in K11 with no *ToxB/toxb*. Contigs were binned into 50 kb regions and transposon annotations counted. The ~450 kbp gap (Additional File 3.1) is not reflect in the chart of K11, a black triangle is placed in the location of the gap which correspond to the boxed bars in I-73-1. Red bar is the *ToxB* containing bin.



Additional File 3.8 Dotplot alignment (30 bp sliding window) of elements *Copia-1_Ptr* from D308 (x-axis) and *Copia-2_Ptr* from SC29-1 (y-axis) which were inserted in *ToxB* ORF of the respective isolates. Pairwise alignment with MUSCLE showed an identity of ~51%.

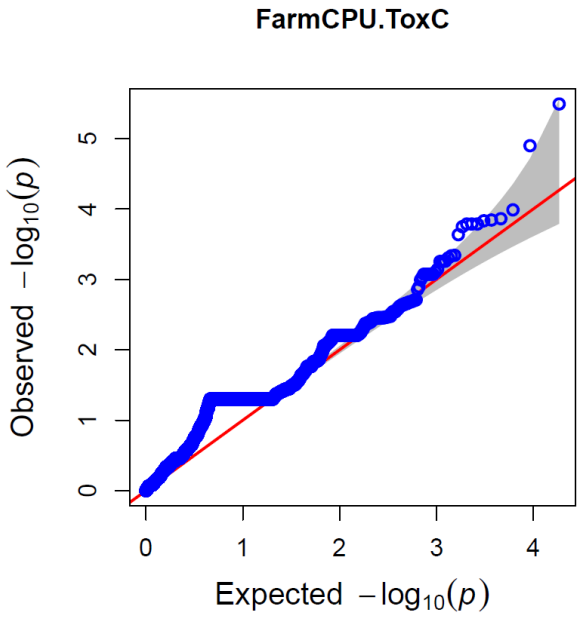


Additional File 3.9 Stemloop structures found near the terminal edges of *ToxB-HLE*. From left to right; the 5' LTS hairpin (red text = 'TGTC' start sequence); 3' hairpin beginning 32 bp from RTS; another 3' hairpin begining 72 bp from RTS with a stronger delta G value. 5' hairpins are seen in Helitron2 and Helentrons subfamilies but not in Heltitron1.



Additional File 3.10 Geneious annotations surrounding copies of *ToxB* (green arrows) in isolate 92-171R5. The duplication of *ToxB* and surrounding sequences are marked in light blue. All other annotations are transposons, or transposon remnants annotated by EDTA. The duplicated *ToxB* is flanked on either side by undefined retrotransposons.

Appendix C: Supplementary material for Chapter 4



Additional File 4.1 Q-Q plot of ToxC GWAS SNPs. The red lines is the expected -log(p) values assuming no association. Blue circles represent observed SNP -log(p) values.

Additional File 4.2 Table of gene annotations near to the GWAS peak. Negative log(p) values represent a SNP within the gene.

Gene	Annotation	-log(p)	General function
gene_04159	Transport protein particle (TRAPP) component		ER-to-Golgi transport
gene_04160	not in database		unknown
gene_04161	Senescence-associated protein		plant cell death
gene_04162	non-haem dioxygenase in morphine synthesis N-terminal 2OG-Fe(II) oxygenase superfamily		oxidation reactions cellular sensors/regulators
gene_04163	not in database	3.3	unknown
gene_04164	not in database	3.8	unknown
gene_04165	not in database		unknown
gene_04166	Glycosyl hydrolase family 1		CAZyme; break down of carbohydrates
gene_04167	Protein of unknown function (DUF1479)	4.9	possibly catalyzes transfer of electrons
gene_04168	not in database		unknown
gene_04169	not in database (manual search: some similarity to DUF1479)	5.5	unknown
gene_04170	not in database		unknown
gene_04171	Major Facilitator Superfamily		movement of diverse substrates including ions
gene_04172	not in database		unknown
gene_04173	not in database		unknown
gene_04174	Autophagy-related protein 2 CAD motif	3.8	cell death; known roles in pathogenicity