GENOMIC AND MOLECULAR CHARACTERIZATION *PYRENPHORA OF TERES* FORMA *TERES*

A Dissertation
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Nathan Andrew Wyatt

In Partial Fulfillment of the Requirements
for the Degree of
DOCTOR OF PHILOSPOHY

Major Program:
Genomics and Bioinformatics
Option:
Functional Genomics

November 2019

Fargo, North Dakota

North Dakota State University
Graduate School

Title

GENOMIC AND MOLECULAR CHARACTERIZATION OF *PYRENPHORA TERES* F. *TERES*

By

Nathan Andrew Wyatt

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

DOCTOR OF PHILOSOPHY

SUPERVISORY COMMITTEE:

Dr. Phil McClean
Chair

Dr. Timothy Friesen

Dr. Justin Faris

Dr. Robert Brueggeman

Approved:

November 8, 2019 — Date

Dr. Phil McClean — Department Chair

ABSTRACT

Pyrenophora teres f. *teres* is the causal agent of net form net blotch of barley. *P. teres* f. *teres* is prevalent globally across all barley growing regions and globally is the most devastating foliar disease of barley. Though economically important, the molecular mechanism whereby *P. teres* f. *teres* causes disease is poorly understood and investigations into these mechanisms have been hindered by a lack of genomic resources. To set a genomic foundation for *P. teres* f. *teres* the reference isolate 0-1 was sequenced and assembled using PacBio single molecule real-time (SMRT) sequencing and scaffolded into 12 chromosomes to provide the first finished genome of *P. teres* f. *teres.* High confidence gene models were generated for the reference genome of isolate 0-1 using a combination of pure culture and *in planta* RNA sequencing. An additional four *P. teres* f. *teres* isolates were sequenced and assembled to the same quality as the reference isolate 0-1 and used in a comparative genomic study. Comparisons of the five *P. teres* f. *teres* isolates showed a two-speed genome architecture with the genome being partitioned into core and accessory genomic compartments. Accessory genomic compartments clustered in sub-telomeric regions of the *P. teres* f. *teres* genome with a majority of previously identified quantitative trait loci (QTL) associated with avirulence/virulence being spanned by these accessory regions. Using these genomic resources, with a bi-parental mapping population and a natural population for QTL analysis and genome wide association study (GWAS), respectively, we identified a candidate gene for the previously mapped *AvrHar*. QTL analysis identified a locus extending off the end of *P. teres* f. *teres* chromosome 5 and GWAS analysis identified significant associations with a gene encoding a small secreted protein. The candidate *AvrHar* gene was validated using CRISPR-Cas9-RNP gene disruption in parental isolates 15A and 0-1. Disruption of *AvrHar* in isolate 15A did not result in a phenotypic change while disruption of the

0-1 allele resulted in a complete loss of pathogenicity. This is the first identification of an effector from *P. teres* f. *teres* validated using CRISPR-Cas9-RNP gene editing.

ACKNOWLEDGEMENTS

I would first and foremost like to acknowledge my advisor, Dr. Timothy Friesen, for guiding me and helping mature as a young scientist. Tim's passion for science is infectious and was a catalyst for my own scientific passion. I was given every opportunity to learn, think, explore new ideas, and take ownership of my projects.

I would also like to acknowledge Danielle Holmes, Lab manager of the Friesen Lab. Danielle's technical skills in molecular biology are second only to her ability to teach those skills to naïve students such as myself. Danielle's ability to calmly think through experiments and trouble shoot problems was of immense value during my Ph.D. and her passion for science and learning new techniques is inspiring.

I would like to acknowledge all present and former students of the Friesen lab for their help, camaraderie, and for putting up with my, sometimes large, personality would also like to specifically thank Dr. Jonathan Richards who I had the pleasure of sitting next to in the Friesen lab where I could pester him with endless questions regarding bioinformatics, his patience is legendary.

I would also like to thank the members of my Ph.D. committee; Dr. Robert Brueggeman, Dr. Phil McClean, and Dr. Justin Faris for their help and support during my time as a student. I would also like to thank Dr. Zhaohui Liu for his guidance in my work with different fungal pathogens and his previous work in the Friesen lab that set the foundation for my work.

Finally, I would like to thank my parents Bill and Julie, my brothers Zach and Sam, and my girlfriend Bri who have all been extremely encouraging and motivating in this pursuit.

DEDICATION

This dissertation is dedicated to my family; Bill, Julie, Zach and Sam. From very early on in my collegiate career my parents, Bill and Julie, encouraged me to take my education as far as I could. The values they instilled in me while raising me have been of immense value during my time as a student, and will serve me well as I begin my career. There are few people I look up to more than my brothers Zach and Sam, and their accomplishments have been inspiring. The encouragement and support they have given me in my lifetime is of immeasurable value. My brothers are my closest friends and their love and support have meant the world to me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDIX TABLES

LIST OF APPENDIX FIGURES

CHAPTER 1. LITERATURE REVIEW

Pyrenophora teres taxonomy and life cycle

The taxonomy of *Pyrenophora teres* classified the species in the Kingdom Fungi, Phylum Ascomycota, Class Dothideomycete, Order Pleosporales, Family Pleosporaceae, Genus *Pyrenophora*, and Species *teres*. *P. teres* was placed into the genus *Helminthosporium* and later changed due to the cylindrical morphology of its conidia and association with *Pyrenophora* species (Shoemaker 1959; Alcorn 1988). *P. teres* classification into the Pleosporales order was confirmed in a phylogenetic analysis using mating type gene sequences (Rau et al. 2005). *P. teres* has two forms, *P. teres* f. *teres* that causes net form net blotch, and *P. teres* f. *maculata* that causes spot form net blotch on barley (Smedegard-Petersen 1971). The two forms are indistinguishable morphologically though they may be distinguished based on disease symptom morphology (Smedegard-Petersen 1971). The two forms of *P. teres* are genetically distinct with rare identification of natural hybridization events in the field (Campbell et al. 1999; Poudel et al. 2017) though *P. teres* f. *teres* and *P. teres* f. *maculata* readily cross under laboratory conditions (Smedegard-Peterson 1978; Campbell 2003)

P. teres f. *teres* is heterothallic in nature, requiring two separate mating types to produce fertile sexual structures called pseudothecia (Mathre 1997). Pseudothecia consist of club shaped, bitunicate asci that typically hold eight ascospores (Mathre 1997). Ascospores are light brown in color and have three to four transverse septa and a single longitudinal septum specifically in the median cells (Mathre 1997). Production of fertile pseudothecia can take as long as six months and may depend on temperature and moisture content in barley stubble (Shipton et al. 1973) as formation time of pseudothecia under laboratory conditions can be as short as two months. *P. teres* f. *teres* overwinters on barley stubble where pseudothecia form and at the beginning of the

following growing season, ascospores are ejected from pseudothecia. Ascospores are wind-dispersed and become the primary inoculum for the growing season (Shipton et al. 1973; Jordan 1981). After primary infection, conidia are produced and act as secondary inoculum with multiple generations of conidia occurring throughout a growing season. Secondary inoculum, conidia, are dispersed by wind and rain splash. Senescent tissue is subsequently colonized by *P. teres* f. *teres* and saprotrophic growth begins until pseudothecia form and produce the primary inoculum for the next growing season, thus completing the cycle (Shipton et al. 1973; Mathre 1997).

Net form net blotch disease symptoms, progression, and management

P. teres f. *teres* is a necrotrophic fungal pathogen of barley that causes the disease net form net blotch. Annual yield losses are in the range of 10-40% with the potential for complete yield loss given optimal pathogen conditions, namely, cool and wet climate with susceptible barley cultivars planted (Mathre 1997; Murray and Brennan 2010). *P. teres* f. *teres* can infect barley leaves, leaf sheaths, stems, and kernels however the foliar symptoms are the most economically important (Liu et al. 2011). Malting quality of barley is the most important economic factor and malting characteristics such as kernel plumpness, size, and bulk density are reduced as a result of net form net blotch (Grewal et al. 2008).

Net form net blotch is characterized by longitudinal and transverse necrotic lesions that are produced on barley foliar tissue and form a net-like pattern that gives the disease its name (Atanasoff and Johnson 1920). The infection process begins when conidia or ascospores come into contact with leaf tissue and germinate. Germination begins within a few hours given suitable moisture and temperature (Shipton et al. 1973). Conidial germination usually begins with the production of a germ tube from a terminal cell, though as many as four simultaneous germ tubes

have been observed (Van Caeseele and Grumbles 1979). *P. teres* f. *teres* has been observed to form an appressorial like structure used for direct penetration into epidermal cells, resulting in pin-point lesions on the leaf surface (Jorgensen et al. 1998; Keon and Hargreaves 1983; Van Caeseele and Grumbles 1979; Liu et al. 2011). After penetration, a large primary intracellular vesicle is formed followed by a secondary intracellular vesicle within the epidermal cell layer from which intercellular hyphae emerge (Keon and Hargreaves 1983; Liu et al. 2011). Hyphae grow towards the mesophyll tissue where they continue to grow intercellularly, host cells adjacent to intercellular hyphae often show signs of death and disruption within 48 hours (Keon and Hargreaves 1983). Microscopic examination of the disease progression showed that chlorotic tissue surrounding necrotic lesions was free of hyphae, yet chloroplasts were disrupted providing evidence for the production of effectors secreted by *P. teres* f. *teres* to further disease progression (Keon and Hargreaves 1983; Liu et al. 2011). Disease lesions continue to grow near the leaf vein and progress to form the typical netting symptoms of net form net blotch (Liu et al. 2011).

Net form net blotch can be managed with multiple methods that include cultural practices, fungicides, and host resistance with combinations of these methods put into an integrated pest management strategy. Crop rotation is an effective method for reducing the primary inoculum, as *P. teres* f. *teres* overwinters on stubble (Liu et al. 2011). Fungicide treatments can be used to mediate the symptoms of net form net blotch but, durable disease resistance still remains the most economically effective method of controlling the disease (Mathre 1997; Shipton et al. 1973; Liu et al. 2011).

Pyrenophora teres f. *teres* diversity

Several *P. teres* populations have been examined to determine the degree of genetic diversity within and between populations. *P. teres* population diversity studies have been conducted using available genetic marker technologies that include RAPDs, AFLPs, RFLPs, DaRTs, and SSRs. *P. teres* f. *teres* populations derived from geographically distinct regions have shown high genetic diversity when compared between the geographical locations (Peever and Milgroom 1994; Serenius et al. 2007; Lemensiek et al. 2010). Peever and Milgroom (1994) examined populations from Germany, New York (USA), North Dakota (USA), and two populations from Alberta (Canada) and found that 46% of the genetic diversity could be attributed to differences between the five populations. Serenius et al. (2007) used two large populations, the MTT population consisting of 167 isolates collected from Finland, Russia, Sweden, the United States, the United Kingdom, Denmark, Canada, the Czech Republic, and Australia; and the SARDI population consisting of 139 isolates collected from Australia. Results from Serenius et al. (2007) indicated a high degree of genetic diversity when comparing geographically distinct populations, aligning with previous studies (Peever and Milgroom 1994). A later study examining populations from South Africa and Australia also confirmed that geographically distinct populations have a high degree of genetic diversity (Lehmensiek et al. 2010). Proximal populations have lower genetic diversity than geographically distinct populations with a range of 5.6-11% of the genetic diversity attributed to variation between populations (Jonsson et al. 2000; Serenius et al. 2005; Leisova et al. 2005; Liu et al. 2012; Peltonen et al. 1996; Campbell 2002; Bogacki et al. 2010; Ficsor et al. 2014; Poudel et al. 2019).The low diversity found in proximal populations juxtaposed with the high levels of genetic diversity found in geographically distinct populations provides evidence of genetic isolation and

a lack of gene flow between geographically distinct populations. Additional studies have examined the diversity of *P. teres* f. *teres* populations over multiple years and have found genetic diversity between annual populations (Leisova et al. 2005; Serenius et al. 2007; Liu et al. 2012; Ficsor et al. 2014; Poudel et all 2019). This genetic diversity is attributed to changes in environments from collection years with causes hypothesized to include changes in barley cultivars being planted, changing cultural practices, or environmental events such as wide scale flooding (Leisova et al. 2005; Serenius et al. 2007; Poudel et al. 2019).

Early studies also aimed to assess whether populations of *P. teres* were reproducing sexually or asexually. Several studies examined linkage disequilibrium that could result in populations that favor asexual reproduction, whereas linkage equilibrium would indicate sexual reproduction was prevalent in the population. Low levels of linkage disequilibrium were found in several studies and indicated that sexual reproduction was an active mode of reproduction in these *P. teres* f. *teres* populations (Peever and Milgroom 1994; Jonsson et al. 2000; Rau et al. 2003; Serenius et al. 2007; Liu et al. 2012; Bogacki et al. 2010). However, there are also studies that identified linkage disequilibrium and indicated that in those populations examined, asexual reproduction was present in the population (Leisova et al. 2005; Serenius et al. 2007). Rau et al. (2005) successfully identified the two mating type loci that constitute the opposing sexes that are required for *P*. *teres* f. *teres* to cross in the field. Now armed with known mating types, studies examining *P*. *teres* f. *teres* populations could examine the ratios of two mating types with even ratios indicating that sexual reproduction was active in the population (Liu et al. 2012; Bogacki et al. 2010; Akhavan et al. 2015).

Linde et al. (2019) examined populations of *P. teres* f. *teres* collected from cultivated barley and wild barley species. The population of isolates collected from cultivated barley was

more diverse than isolates collected from wild barley relatives and no shared multi-locus genotypes were observed between the two subsets (Linde et al. 2019). For comparison, within the isolates collected from cultivated barley, there were nine shared multi-locus genotypes out of the 567 isolates examined (Linde et al. 2019).

Phenotypic diversity of *P. teres* f. *teres* has been assessed using many different populations from around the globe. Khan and Boyd (1969) were the first to describe different pathotypes observed in isolates inoculated on the barley line Beecher. Later, Kahn (1982) described changes in virulence profiles of collected *P. teres* f. *teres* isolates collected in different years on Beecher barley where isolates collected from earlier dates, when Beecher was a commonly planted cultivar, were virulent and isolates collected at later dates, when Beecher was no longer a planted cultivar, were avirulent (Kahn 1982). Many subsequent studies phenotyped collections of *P. teres* f. *teres* on multiple barley genotypes. The number of isolates tested in each study ranged from three to 1,000 isolates and the number of barley lines tested in each study ranged from one to 38 (Tekauz 1990; Steffanson and Webster 1992; Sato and Takeda 1993; Jonsson 1997; Jalli and Robinson 2000; Arabi et al. 2003; Cromey and Parks 2003; Afanasenko et al. 2009; Akhavan et al. 2016). The overall trend observed in these studies was that increasing the number of isolates and lines being tested increased the number of pathotypes observed (Sako and Takeda 1993). Afanasenko et al. (2009) used an international collection of 1,000 *P. teres* f. *teres* isolates inoculated on 14 barley genotypes in order to create a universal set of differential barley lines to assess *P. teres* f. *teres* phenotypic diversity.

Pyrenophora teres f. *teres* virulence

Pathogen virulence is governed by a set of molecules collectively known as effectors. Effectors can be any molecule, including proteins, secondary metabolites, and small RNAs that the pathogen uses to manipulate their host (Toruño et al. 2016).

P. teres f. *teres* produces proteinaceous and non-proteinaceous effectors to manipulate its barley host to gain nutrients. Smedegard-Peterson (1977) isolated and purified two toxins from *P. teres* f. *teres* culture filtrates designated toxin A and toxin B and determined that toxin A was more potent than toxin B via a phytotoxicity assay. Both toxin A and toxin B induced chlorosis within three days followed by the progression to necrosis (Smedegard-Peterson 1977). A third toxin, toxin C, was identified and it was determined that toxin B was a derivative of toxin C (Bach et al. 1979; Friis et al. 1991). Toxin C produced different symptoms than toxin A and toxin B with necrosis being the primary reaction observed (Weiergang et al. 2002). The three identified toxins were screened on diverse barley backgrounds to assess resistance and susceptibility and it was found that the toxins were not isolate-specific and host specificity was absent with lines being either sensitive or insensitive to the toxins (Sharma 1984; Weiergang et al. 2002). Proteinaceous metabolites have also been implicated in *P. teres* f. *teres* virulence on barley. Sarapeleh et al. (2007) determined that isolated proteinaceous metabolites induced host-specific necrosis and that protease treatment inhibited the phytotoxic properties of isolated proteinaceous metabolites. In comparison, low molecular weight non-proteinaceous molecules were found to be non-selective and primarily induced chlorosis (Sarapeleh et al. 2007).

Differing virulence profiles have been identified in a number of *P. teres* f. *teres* populations and this variation is thought to originate from segregating effector repertoires (Liu et al. 2011; Liu et al. 2015; Richards et al. 2016). Using six *P. teres* f. *teres* isolates, Ismail et al.

(2014a) identified morphological differences in conidial germination and appressoria formation between virulent and avirulent isolates. However, extracted effectors from all isolates induced necrosis (Ismail et al. 2014a). Ismail et al. (2014b) subsequently proposed that three effectors; PttXyn11A, PttCHFP1, and PttSP1 showing homology to proteins involved in plant disease interaction, may be involved in the *P. teres* f. *teres* infection process. A later study identified 63 proteins common to virulent isolates using one and two-dimensional polyacrylamide gel electrophoresis representing possible candidate effector genes (Ismail and Able 2016). Ismail and Able (2017) identified a transition between colonization and necrotrophy occurring approximately 48 h after inoculation using *in planta* gene expression analysis of 222 proteins. No functional validation was conducted on these candidate genes/proteins and it is not yet known whether any of these are biologically relevant to the *P. teres* f. *teres*-barley interaction (Ismail and Able 2017).

The first proteinaceous effector to be identified and partially characterized was isolated from intercellular wash fluid of the susceptible barley line Hector after inoculation with *P. teres* f. *teres* isolate 0-1 (Liu et al. 2015). This effector was named PttNE1 and was approximately 6.5-12.3 kDa in size and induced necrosis upon interaction with *SPN1* (Liu et al. 2015). The interaction of PttNE1 and *SPN1* implies that necrotrophic effector triggered susceptibility (NETS) is involved in the interaction and the plant defense response is triggered by PttNE1 (Liu et al. 2015).

PttNE1 was never mapped in *P. teres* f. *teres* but a number of studies have taken advantage of the heterothallic nature of *P. teres* f. *teres* to generate pathogen bi-parental mapping populations (Weiland et al. 1999; Lai et al. 2007; Beattie et al. 2007; Shjerve et al. 2014; Koladia et al. 2017). Each mapping study has revealed unique genomic loci specific to *P. teres* f. *teres*

isolates phenotyped on different host genotypes. The first bi-parental mapping population of *P. teres* f. *teres* was a cross between the Canadian isolate 0-1 and the California isolate 15A (Weiland et al. 1999). This population was used to map a locus associated with avirulence on barley line Harbin coming from isolate 15A. Segregation ratios observed in the progeny of the 0-1 × 15A population appeared as a 1:1 ratio indicating a single gene named *AvrHar*. The 0-1 × 15A population was used again along with supplemental progeny to investigate virulence/avirulence on barley lines Canadian Lake Shore (CLS), Tifang, and Prato. Segregation ratios of inoculated progeny showed a 1:1 (avirulence:virulence) ratio on both Tifang and CLS and a 3:1 ratio of avirulence:virulence on Prato. Interval regression analysis revealed a major effect locus in isolate 15A associated with avirulence on Tifang and CLS while also identifying two loci in isolate 0-1 associated with avirulence on Prato. Interestingly, the genomic region associated with avirulence on barley lines Tifang and CLS corresponded to the previously identified *AvrHar* locus. The two loci associated with avirulence on barley line Prato were named *AvrPra1* and *AvrPra2*. *AvrPra2* also mapped to the same region as *AvrHar* but avirulence was conferred by the opposite parental isolate, 0-1. The co-segregation of *AvrHar* and *AvrPra2* indicated that they were either alleles of the same gene, two closely linked genes, or the same gene conferring avirulence and virulence dependent upon the host genotype (Lai et al. 2007).

Beattie et al. (2007) mapped an avirulence locus on barley line Heartland (*AvrHeartland*) with a mapping population consisting of two differential Canadian isolates. Shjerve et al. (2014) created a bi-parental pathogen population using the two California *P. teres* f. *teres* isolates 15A and 6A and used a restriction-site associated digest – genotype-by-sequencing (RAD-GBS) approach (Leboldus et al. 2015) to construct a genetic linkage map. 15A and 6A show differential reactions on barley lines Kombar and Rika with isolate 15A showing virulence on

Kombar but avirulence on Rika and isolate 6A showing virulence on Rika but avirulence on Kombar. Using the 15A × 6A population, four virulence QTL were identified. Two of the QTL corresponded to virulence coming from isolate 15A on Kombar barley named *VK1* and *VK2*. The other two QTL corresponded to 6A virulence on Rika barley named *VR1* and *VR2* (Shjerve et al. 2014). Most recently, a cross between *P. teres* f. *teres* isolates BB25 of Denmark and FGOH04Ptt-21 of Fargo, ND, USA was used to examine avirulence/virulence on eight barley lines commonly used in differential sets. A total of nine unique QTL were reported across the eight barley lines. Both shared and unique QTL were identified with three of the QTL contributing to over 45% of the variation (R2>0.45) (Koladia et al., 2017). To date, none of the genes underlying QTL identified for *P. teres* f. *teres* virulence/avirulence have been identified.

Barley resistance and susceptibility to *Pyrenophora teres* f. *teres*

Genetic resistance to *P. teres* f. *teres* has been observed to be incredibly complex with dominant resistance and dominant susceptibility identified (Friesen et al. 2007; Mode and Schaller 1958; Abu Qamar et al. 2008; Liu et al. 2015). The presence of both dominant resistance and dominant susceptibility and the discovery of NEs illustrates the complicated interaction between barley and *P. teres* f. *teres* (Liu et al. 2011; Liu et al. 2012; Liu et al. 2015; Koladia et al., 2017).

Barley resistance to *P. teres* f. *teres* was initially shown to be quantitative by Geschelle (1928) and an immense number of mapping studies have identified QTL and marker trait associations. In total, 340 loci have been reported and many of these loci represent large overlapping regions of the barley genome. At least 18 studies investigating resistance and susceptibility to *P. teres* f. *teres* have been done using bi-parental populations and QTL analysis with QTL identified on all barley chromosomes. Resistance and susceptibility to *P. teres* f. *teres*

has been studied at the seedling and adult plant stages of the barley life cycle with instances of effective resistance mapping to the same locus for both life stages (Grewal et al. 2008).

The first resistance locus identified for resistance to *P. teres* was found in a Tifang × Atlas cross with Tifang barley being resistant to isolates of *P. teres* from California, USA (Schaller 1955). The locus identified by Schaller (1955) was designated *Pt1* along with two other loci designated *Pt2* and *Pt3* (Mode and Schaller 1958). The *Pta* locus was later identified and was found to be a distinct locus from previously identified loci based on resistance to the Australian isolate WA-2 (Khan and Boyd 1969). Trisomic analysis was used to characterize the inheritance of *Pt1* and *Pt2* leading to the collapse of *Pt1* and *Pt2* into the *Rpt1* locus given that *Pt1* and *Pt2* could not be separated (Bockelman et al. 1977). *Rpt1* was localized to the long arm of chromosome 3H and subsequent studies identified loci that were collapsed into the *Rpt1* loci (Graner et al. 1996; Raman et al. 2003; Lehmesiek et al. 2007). Dinglasan et al. (2019) mapped an additional resistance locus to the short arm of chromosome 3H in Canadian Lake Shore that is not *Rpt1*. Manninen et al. (2006) reported major and minor effect loci effective against different *P. teres* f. *teres* isolates on chromosome 6H and 1H. A major locus identified as *Rpt3* was found to be located on chromosome 2H (Graner et al. 1996; Franckowiak and Platz 2013; Grewal et al. 2012; Steffenson et al. 1996; Cakir et al. 2003; Raman et al. 2003; Emebiri et al. 2005). The *Rpt7* locus conferred resistance to *P. teres* f. *teres*, and is located on the long arm of chromosome 4H in Halcyon (Raman et al. 2003; Read et al. 2003) and it is not known whether additional QTL reported on chromosome 4H are at the same *Rpt7* locus (Steffenson et al. 1996; Lehmensiek et al. 2007; Grewal et al. 2008; Yun et al. 2005; Richter et al. 1998; König et al. 2014; König et al. 2013; Spaner et al. 1998; Islamovic et al. 2017).

The *Pta* locus was renamed *Rpt5* on chromosome 6H and at least three alleles have been identified that include *Rpt5.f*, *Spt1.r*, and *Spt1.k* (Franckowiak and Platz 2013; Khan and Boyd 1969a; Manninen et al. 2006, Abu Qamar et al. 2008; Koladia et al. 2017b, Richards et al. 2016). Additional loci have been reported as overlapping the *Rpt5* locus and possibly contain additional alleles include *QRpts6L*, *QRpt6*, *AL_QRptt6-1*, *Rpt-Nomini*, and *Rpt-CIho2291* in Halcyon (Raman et al. 2003), TR251 (Grewal et al. 2012), Lavrans (Wonneberger et al. 2017a), Nomini and CIho2291 (O'Boyle et al. 2014). Additionally there are a number of undesignated QTL that may also represent alleles of *Rpt5* as they are in the same genomic region identified in Steptoe (Steffenson et al. 1996), HOR 9088 (Richter et al. 1998), Kaputar, ND11213 (Cakir et al. 2003; Emebiri et al. 2005; Liu et al. 2015), Chevron (Ma et al. 2004), SM89010 (Friesen et al. 2006), M129 (St. Pierre et al. 2010), Baudin (Cakir et al. 2011), WPG8412, Pompadour, Stirling (Gupta et al. 2011), Falcon (Islamovic et al. 2017), H602 (Hisano et al. 2017), and UVC8 (Martin et al. 2018).

The *Rpt5* locus is a complex locus in the centromeric region of chromosome 6H and has been reported by numerous studies using diverse host populations with diverse isolates (Abu Qamar et al. 2008; Adawy et al. 2013; Adhikari et al. 2019; Cakir et al. 2011, 2003; Dontsova et al. 2018; Emebiri et al. 2005; Friesen et al. 2006; Grewal et al. 2012; Grewal et al. 2008; Gupta et al. 2011; Gupta et al. 2010; Koladia et al. 2017a; Liu et al. 2015; Ma et al. 2004; Manninen et al. 2000; Martin et al. 2018; Molnar et al. 2000; Novakazi et al. 2019; O'Boyle et al. 2014; Raman et al. 2003; Rau et al. 2015; Richards et al. 2016; Richter et al. 1998; Rozanova et al. 2019; Spaner et al. 1998; St. Pierre et al. 2010; Steffenson et al. 1996; Wonneberger et al. 2017b; Wonneberger et al. 2017a). The *Rpt5* locus has also been designated susceptibility to *P. teres* 1 (*Spt1*), indicating the dominant nature of susceptibility to *P. teres* f. *teres* in some barley lines

(Richards et al. 2016). Due to the large number of reports identifying both resistance and susceptibility at the 6H locus, it has been proposed that a cluster of genes may reside in this complex genomic region. This genomic landscape is primed for the trench warfare of host-pathogen co-evolution with a diversity of resistances and susceptibilities that pathogens must evolve to evade or take advantage of.

Abu Qamar et al. (2008) found that Rika barley was resistant to *P. teres* f. *teres* isolate 15A but susceptible to isolate 6A. Kombar barley showed the reciprocal reaction to Rika in that Kombar was resistant to 15A and susceptible to 6A. Using a double haploid (DH) population derived from a cross of barley cultivars Rika and Kombar recessive resistance genes *rpt.r* and *rpt.k* were mapped to the two *P. teres* f. *teres* isolates, 6A and 15A. F_2 segregation analysis resulted in a 3:1 (susceptible:resistant), confirming the nature of the susceptibility as dominant (Abu Qamar et al. 2008). The dominant susceptibility to both isolates mapped to the centromeric region of chromosome 6H spanning approximately 5.9 cM (Abu Qamar et al. 2008). The region was further delimited by Liu et al. (2010) using barley expressed sequence tags (ESTs) to identify sequences containing SSRs and subsequently, was delimited to approximately 3.3 cM (Liu et al. 2010). Using immortal critical recombinants (ICRs) of the previously defined *Rpt5*/*Spt1* locus, the region was fine mapped to a ~0.24 cM interval, spanning a ~466 kb genomic region of Brachypodium with 62 annotated genes (Richards et al. 2016). The syntenic genomic region of Brachypodium translated to ~9.5 Mb of physical sequence in the barley genome and contained 39 high-confidence genes. It was hypothesized that the locus contained a single gene, two tightly linked genes, or a cluster of susceptibility genes that interact with effectors to varying specificities (Richards et al. 2016). Six of the genes within the *Rpt5*/*Spt1* locus encoded for immunity receptor-like proteins that could be responsible for a susceptible

response (Richards et al. 2016). An additional gene was mapped to the 6H locus through an interaction with PttNE1 and SPN1, though it was determined that the *SPN1* locus was in fact a distinct locus indicating additional genes are located at this important genomic region near the centromere of chromosome 6H (Liu et al. 2015; Richards et al. 2016).

In addition to bi-parental mapping and QTL analysis examining barley reactions to *P. teres* f. *teres,* seven association mapping studies have been performed (Adhikari et al. 2019; Amezrou et al. 2018; Daba et al. 2019; Novakazi et al. 2019; Richards et al. 2017; Rozanova et al. 2019; Wonneberger et al. 2017b). Studies examining barley reactions to diverse global *P. teres* f. *teres* isolates were performed on the Barley Core Collection (Muñoz-Amatriaín et al. 2014; Richards et al. 2017), Nordic Barley Panel (Wonneberger et al. 2017b), ICARDA AM-2014 Panel (Amezrou et al. 2018), Ethiopian, ICARDA and NDSU Barley Panel (Daba et al. 2019), Siberian barley panel (Rozanova et al. 2019), Ethiopian and Eritrean Barley Collection (Adhikari et al. 2019), and Vavilov Research Institute Collection (Novakazi et al. 2019). Unique marker trait associations identified seven to 31 independent genomic loci associated with disease reactions of barley to *P. teres* f. *teres* with a majority of significant marker trait associations shown to be located in the previously identified 6H locus (Richards et al. 2017). Genome wide association study (GWAS) results thus illustrate the importance of the 6H locus in the interaction of barley and *P. teres* f. *teres*.

Novel approaches to identifying loci associated with barley reactions to *P. teres* f. *teres* have included the use of a nested association mapping (NAM) population (Vatter et al. 2017) and exome QTL-seq (Hisano et al. 2017). Vatter et al. (2017) used a NAM population to study barley resistance to *P. teres* f. *teres* using reaction type (Tekauz 1985) and average ordinate to phenotype the ‘Halle Exotic Barley 25’ population (Maurer et al. 2015). Twenty-four QTL were

identified with both reaction type and average ordinance phenotyping with a QTL on chromosome 2H responsible for the most phenotypic variation that does not correspond to *Rpt3* (Vatter et al. 2017). The *Rpt5*/*Spt1* locus was only minorly identified in using the nested association mapping approach possibly due to few of the founding barley lines used in the HEB-25 population containing resistance alleles to *P. teres* f. *teres* (Vatter et al. 2017). Hisano et al. (2017) used exome QTL-seq in which ten resistant and ten susceptible lines were pooled separately prior to exome capture library construction and sequencing. Three QTL were identified with two QTL on chromosome 3H and one QTL on chromosome 6H from barley line H602 (Hisano et al. 2017). Twice as many SNPs were identified on chromosomes 3H and 6H in the resistant bulk sample compared to the susceptible bulk sample (Hisano et al. 2017). The increase in SNP number within the resistant bulk sample suggests that these loci were undergoing diversifying selection potentially to evade recognition of *P. teres* effectors that were triggering NETS (Hisano et al. 2017; Liu et al. 2015). The sheer number of reports of QTL located at the centromere of chromosome 6H, now referred to as *Rpt5*/*Spt1*, warrants validation and characterization. Both gene-for-gene (Flor 1955; Flor 1971) and inverse gene-for-gene (Friesen et al. 2007) interactions have now been observed in the *P. teres* f. *teres* patho-system, illustrating the complexity of the interaction between *P. teres* f. *teres* and its barley host.

Genomics of *Pyrenophora teres* f. *teres*

The first sequenced *P. teres* f. *teres* genome was of the Canadian *P. teres* f. *teres* isolate 0-1 and was sequenced using Illumina paired-end reads (Ellwood et al. 2010). The size of the 0-1 genome was initially estimated at 41.95 Mb with an average sequence coverage of 20x and confirmed the presence of at least nine chromosomes based on a combination of pulse field gel electrophoresis and germ tube burst visualization (Ellwood et al. 2010). Current genome

assemblies include an additional eight *P. teres* f. *teres* isolates; W1-1, NB29, NB85, NB73 that are all Australian isolates (Syme et al. 2018), 15A and 6A that are California (USA) isolates, FGOH04Ptt-21 that is an isolate from North Dakota (USA), and isolate BB25 from Denmark (Wyatt et al. 2019, Chapter 3). Genome sizes currently range from 46.31 – 51.76 Mb in *P. teres* f. *teres* and it was observed that *P. teres* f. *teres* genomes had undergone rapid expansions due primarily to repetitive elements (Syme et al. 2018; Wyatt et al. 2019, Chapter 3). Contig scaffolding of *P. teres* f. *teres* isolates using optical mapping (Syme et al. 2018) or genetic linkage maps (Wyatt et al. 2019, Chapter 3) confirmed the presence of 12 chromosomes in the *P. teres* f. *teres* genome.

Along with currently published genomes, two marker sets created using the 15A × 6A mapping population and the FGOH04Ptt-21 × BB25 mapping population were used to generate genetic linkage maps that are publicly available (Shjerve et al. 2014; Koladia et al. 2017). Shjerve et al. (2014) generated a genetic map consisting of 18 linkage groups and containing 468 markers. The size of the genetic map was reported as 1,799.77cM with a physical to genetic distance ratio of 23Kb/cM (Shjerve et al. 2014). Koladia et al. (2017a) identified 370 non-redundant SNP markers using a RAD-GBS protocol (Leboldus et al. 2014). A genetic map was constructed from the 370 markers resulting in 16 linkage groups spanning 1,905.81cM (Koladia et al. 2017a). The marker sets and genetic maps are valuable resources in both the investigation of marker trait associations and in evaluating genomic assemblies.

Fungal genomics

The birth of fungal genome sequencing owes its origin to the first fungi to be sequenced, *Saccharomyces cerevisiae,* also known as baker's yeast (Oliver et al. 1992; Goffeau et al. 1996). The sequencing of *S. cerevisiae* kicked off a large number of projects that aimed to sequence

diverse yeast genomes and higher organisms and currently there are more than 3,000 sequenced fungal genomes (Piskur and Langkjaer 2004; Aylward et al. 2017). The model species *Neurospora crassa* was the first filamentous fungus to have a high quality sequenced genome (Galagan et al. 2003). The *N. crassa* genome represented a leap forward in sequencing fungal genomes as the *N. crassa* genome was substantially larger than previously sequenced fungal genomes and its sequencing uncovered important features of fungal genomes, including; chromatin structure, methylation status, and genome defense mechanisms that included repeat-induced point mutation (RIP) (Galagan et al. 2003; Borkovich et al. 2004).

The first plant pathogenic fungi to be sequenced was *Magnaporthe oryzae* that causes the devastating disease rice blast (Dean et al. 2004). The *M. oryzae* genome was sequenced using a whole-genome shotgun sequencing approach by end-sequencing a series of plasmid, fosmid, and bacterial artificial chromosome (BAC) libraries (Dean et al. 2004). Since the sequencing of *M. oryzae*, 191 plant pathogenic fungi have been sequenced from a diverse array of crop species and are publicly available (Aylward et al. 2017), with more being sequenced each year. Long read sequencing technologies such as Pacific Biosciences single molecule real-time (SMRT) sequencing and Oxford Nanopore sequencing are capable of producing reads longer than 1kb and allow the assembly of large contiguous sequences (Ashton et al. 2015; Huddleston et al. 2014; Laszlo et al. 2014; Powers et al. 2013). For fungal plant pathogens, long read sequencing provided the opportunity to finish fungal genomes and produce assemblies containing full chromosomes. Finishing the genomes of plant pathogens has provided key insights into important genomic aspects that include the organization of genes within the genome, repetitive regions and transposable elements, as well as epigenomics, and instances of horizontal gene transfer (Thomma et al. 2016; Friesen et al. 2006; Shi-Kunne et al. 2019; Fouche et al. 2019).

Fungal genomes that have been effectively finished as a result of long read sequencing technologies include *Zymoseptoria tritici* (Goodwin et al. 2011), *Puccinia striiformis* f. sp. *tritici* (Schwessinger et al. 2018), *Verticillium dahliae* (Klosterman et al. 2011), *Pyrenophora tritici-repentis* (Manning et al. 2013; Moolhuijzen et al. 2018) *Parastagonosporum nodorum* (Richard et al. 2018), and both forms of *P. teres* (Wyatt et al. 2018, Chapter 2; Wyatt et al. 2019, Chapter 3; Syme et al. 2018) to name a few.

This wealth of sequencing data for plant pathogenic species has revealed interesting genomic trends within plant pathogen genomes. For example, the average genome size and number of genes of plant pathogens is larger than their non-pathogenic relatives consistent with previous observations that genome size is correlated with pathogenicity as it allows greater niche adaptation of pathogens in their respective hosts (Spanu et al. 2010; Ohm et al. 2012). Within plant pathogens, the basidiomycetes had a larger average haploid genome size relative to the ascomycete pathogen, primarily due to the contributions of rust pathogens with significantly larger genomes (Duplessis et al. 2011; Aylward et al. 2017). Sequencing of fungal genomes has generated important insights into the context of observational and experimental data gathered prior to sequencing and allows hypothesis driven research to be done using genome editing tools (McDonald et al. 2019; Aylward et al. 2017). An example of how fungal genome sequencing has helped understand previous observations comes from the horizontal gene transfer events of the necrotrophic effector ToxA (Friesen et al. 2006, McDonald et al. 2018, McDonald et al. 2019). Friesen et al. (2006) originally identified the presence of the necrotrophic effector ToxA in the pathogen *P. nodorum* and confirmed its horizontal transfer to *P. tritici-repentis*. ToxA was later observed to have been horizontally transferred to the pathogen *Bipolaris sorokiniana* (McDonald et al. 2018; Friesen et al. 2018). With sufficient genome sequencing from a number of isolates

from the species *P. tritici repentis*, *P. nodorum*, and *B. sorokiniana*, the mechanism of these horizontal gene transfer events was determined to be driven by a type II DNA transposon with terminal inverted repeats that incorporated the ToxA gene (McDonald et al. 2019). McDonald et al. (2019) subsequently named this genomic compartment ToxAhAT in order to describe the totality of the mobile element. Importantly, the generation of high quality finished genomes has aided in effector identification in important patho-systems. The wheat pathogen *P. graminis* f. sp. *tritici* finished genome was critical in the cloning of the first effector genes from this system, *AvrSr35* and *AvrSr50* (Duplessis et al. 2011; Salcedo et al. 2017; Chen et al. 2017).

Decreased costs for long read sequencing provides the opportunity to sequence more than a single representative genome for a species to high contiguity. Several studies have taken advantage of this cost decrease to produce finished genomes for the plant pathogenic species *V. dahliae* (Klosterman et al. 2011), *Z. tritici* (Plissoneau et al. 2018), *P. tritici-repentis* (Moolhuijzen et al. 2018), *P. nodorum* (Richards et al. 2018), and *P. teres* f. *teres* (Wyatt et al. 2019, Chapter 3) to name a few. The sequencing of multiple full genomes of a species has produced pan-genomic comparisons wherein the common features of each genome can be classified as belonging to the core genome and features unique to any particular individual or features absent from any individual represent the accessory genome. In the case of *Z. tritici* the accessory genome is primarily comprised of full length chromosomes that can be lost with seemingly no reduction in the fitness of the isolates (Fouche et al. 2018). In other fungal genomes, accessory genes tend to exhibit a presence absence polymorphism with accessory genes correlating in proximity to transposable elements (Stukenbrock et al. 2018). Wide scale genomic comparisons within a species and between species will help to uncover the mechanisms of evolution and host specificity in these complex organisms.

Association mapping in fungi

Association mapping has gained popularity in studies examining marker trait associations in fungi. Gene identification underlying a trait of interest in fungi has historically been accomplished using proteomic approaches, genomic mining, and bi-parental mapping populations generated from crosses of two individuals with differential phenotypes. Bi-parental mapping populations work well for detecting less common alleles, but, the level of diversity and genetic recombination is relatively low and the development of suitable mapping populations takes a considerable amount of time (Zhu et al. 2008; Rafalski 2010). With the falling cost of next-generation sequencing technologies, genome wide association studies (GWAS) are gaining popularity within the fungal community. Association mapping has several advantages over QTL analysis using bi-parental mapping populations that include less time invested in population generation and a higher resolution of marker trait associations largely due to linkage disequilibrium and ancient recombination events in an organism's evolutionary history. Linkage disequilibrium (LD) is the nonrandom correlation between specific alleles at separate loci (Flint-Garcia et al. 2003). Association mapping exploits this phenomenon and offers an alternative to traditional mapping methods by using natural populations of diverse individuals to exploit evolutionary recombination (Zhu et al. 2008). Model selection in association studies is a crucial step in conducting a meaningful analysis. False positive associations are a common problem when conducting a GWAS and using models that correct confounders such as population structure and kinship via principal components analysis or using Bayesian clustering methods help reduce the rate of false positives (Pritchard et al. 2000; Yu et al. 2006; Zhao et al. 2007). Models have also been proposed that use a variation of relationship matrices through altering

clustering methods or compression parameters to increase statistical power (Zhang et al. 2010; Li et al. 2014).

In fungi, association mapping has been applied to an ever growing list of species. Association mapping was first applied to *S. cerevisiae* to identify marker trait associations (MTAs) with mtDNA copy number as well as clinically relevant traits (Connelly and Akey 2012; Muller et al. 2011). The necrotrophic pathogen *Heterobasidion annosum* was subject to a GWAS using 23 sequenced isolates and 33,018 SNPs to identify 12 SNPs associated with eight candidate genes for virulence (Dalman et al. 2013). In *Fusarium graminearum* association mapping was used to identify 50 (MTAs) for aggressiveness, 29 marker trait associations for deoxynivalenol production, and 74 (MTAs) for propiconazole sensitivity (Talas et al. 2016). In *P. nodorum*, association mapping was applied to identify virulence genes (Gao et al. 2016, Richards et al. 2019). Using 191 *P. nodorum* isolates sequenced using Ion-Torrent sequencing following a RAD-GBS protocol, 2,983 SNPs were used to identify the previously identified necrotrophic effector genes *SnToxA* and *SnTox3* as well as a novel locus corresponding to virulence on the wheat line Alsen (Gao et al. 2016). Richards et al. (2019) used whole genome sequencing of 197 isolates to generate 1,026,859 markers across the *P. nodorum* genome and identified novel loci as well as *SnToxA* and *SnTox3*. GWAS analyses that aimed to identify MTAs with virulence have now included plant pathogens such as *Z. tritici* (Hartman et al. 2017), *Botrytis cinerea* (Atwell et al. 2018), *M. orzyae* (Ganeshan et al. 2018), and *Colletotrichum kahawae* (Vieira et al. 2019) with more to surely come as sequencing costs are driven down and bioinformatics tools become more available.

Literature cited

Abu Qamar, M., Liu, Z. H., Faris, J. D., Chao, S., Edwards, M. C., Lai, Z., Franckowiak, J. D., Friesen, T. L. 2008. A region of barley chromosome 6H harbors multiple major genes associated with net type net blotch resistance. Theor. Appl. Genet. 117:1261.

Adawy, S. S., Diab, A. A., Sayed, A. H. I., Shafik, D. I., El-Morsy, S. I., Saker, M. M. 2013. Constuction of genetic linkage map and QTL analysis of net blotch resistance in barley. Int. J. Adv. Biotechnol. Res. 4:348–363.

Adhikari, A., Steffenson, B. J., Smith, M. J., Dill-Macky, R. 2019. Genome-Wide Association Mapping of Seedling Net Form Net Blotch Resistance in an Ethiopian and Eritrean Barley Collection. Crop Science.

Afanasenko, O. S., Jalli, M., Pinnschmidt, H. O., Filatova, O. Platz, G. J. 2009 Development of an international standard set of barley differential genotypes for *Pyrenophora teres* f. *teres*. Plant Pathol. 58:665–676.

Akhavan, A., Turkington, T. K., Kebede, B., Tekauz, A., Kutcher, H. R., Kirkham, C., Xi, K., Kumar, K., Tucker, J. R., Strelkov, S. E. 2015. Prevalence of mating type idiomorphs in *Pyrenophora teres* f. *teres* and *P. teres* f. *maculata* populations from the Canadian prairies. Can. J. Plant Pathol. 37:52-60.

Akhavan, A., Turkington, T.K., Askarian, H., Tekauz, A., Xi, K., Tucker, J.R., Kutcher, H.R., Strelkov, S.E. 2016. Virulence of *Pyrenophora teres* populations in western Canada. Can. J. Plant Pathol. 38:183-196.

Alcorn, J. L. 1988. The taxonomy of '*Helminthosporium*' species. Ann. Rev. Phytopathol. 26:37-56

Amezrou, R., Verma, R. P. S., Chao, S., Brueggeman, R. S., Belqadi, L., Arbaoui, M., Rehman, S., Gyawali, S. 2018. Genome-wide association studies of net form of net blotch resistance at seedling and adult plant stages in spring barley collection. Mol. Breed. 38:58.

Arabi, M. I. E., Al-Safadi, B., Charbaji, T. 2003. Pathogenic variation among isolates of *Pyrenophora teres*, the causal agent of barley net blotch. J. Phytopathol. 151:376–382.

Ashton, P. M., Nair, S., Dallman, T., Rubino, S., Rabsch, W., Mwaigwisya, S., Wain, J., O'grady, J. 2015. MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. Nat. Biotechnol. 33:296.

Atanasoff, D., and Johnson, A. G. 1920. Treatment of cereal seeds by dry heat. J. Agric. Res. 18:379-390

Atwell, S., Corwin, J., Soltis, N., Zhang, W., Copeland, D., Feusier, J., Eshbaugh, R., Kliebenstein, D. J. 2018. Resequencing and association mapping of the generalist pathogen *Botrytis cinerea*. bioRxiv. 489799.

Aylward, J., Steenkamp, E. T., Dreyer, L. L., Roets, F., Wingfield, B. D., and Wingfield, M. J. 2017. A plant pathology perspective of fungal genome sequencing. IMA fungus, 8:1.

Bach, E., Christensen, S., Dalgaard, L., Larsen, P. O., and Olsen, C. E. 1979. Structures, properties and relationship to the aspergillomarasmines of toxins produced by *Pyrenophora teres*. Physiol. Plant Pathol. 14:41-46.

Beattie, A. D., Scoles, G. J., Rossnagel, B. G. 2007. Identification of molecular markers linked to a *Pyrenophora teres* avirulence gene. Phytopathology. 97:842–849

Bockelman, H. E., Sharp, E. L., Eslick, R. F. 1977. Trisomic analysis of genes for resistance to scald and net blotch in several barley cultivars. Can. J. Bot. 55:2142–2148.

Bogacki, P., Keiper, F. J., and Oldach, K. H. 2010. Genetic structure of South Australian Pyrenophora teres populations as revealed by microsatellite analyses. Fungal Biol. 114:834-841.

Borkovich, K. A., Alex, L. A., Yarden, O., Freitag, M., Turner, G. E., Read, N. D., Seiler, S., Bell-Pedersen, D., Paietta, J., Plesofsky, N. and Plamann, M. 2004. Lessons from the genome sequence of *Neurospora crassa*: tracing the path from genomic blueprint to multicellular organism. Microbiol. Mol. Biol. Rev., 68:1-108.

Cakir, M., Gupta, S., Li, C., Hayden, M., Mather, D. E., Ablett, G. A., Platz, G. J., Broughton, S., Chalmers, K. J., Loughman, R., Jones, M. G. K., Lance, R. C. M. 2011. Genetic mapping and QTL analysis of disease resistance traits in the barley population Baudin × AC Metcalfe. Crop Pasture Sci. 62:152–161.

Cakir, M., Gupta, S., Platz, G. J., Ablett, G. A., Loughman, R., Emebiri, L. C., Poulsen, D., Li, C. D., Lance, R. C. M., Galwey, N. W., Jones, M. G. K., Appels, R. 2003. Mapping and validation of the genes for resistance to *Pyrenophora teres* f. *teres* in barley (*Hordeum vulgare L.*). Aust. J. Agric. Res. 54:1369–1377.

Campbell, G. F., & Crous, P. W. 2003. Genetic stability of net × spot hybrid progeny of the barley pathogen *Pyrenophora teres*. Australas. Plant Pathol. 32:283-287.

Campbell, G. F., Crous, P. W., and Lucas, J. A. 1999. *Pyrenophora teres* f. *maculata*, the cause of *Pyrenophora* leaf spot of barley in South Africa. Mycol. Res. 103:257-267

Campbell, G. F., Lucas, J. A., and Crous, P. W. 2002. Evidence of recombination between netand spot-type populations of *Pyrenophora teres* as determined by RAPD analysis. Mycol. Res. 106:602-608

Chen, J., Upadhyaya, N. M., Ortiz, D., Sperschneider, J., Li, F., Bouton, C., Breen, S., Dong, C., Xu, B., Zhang, X. and Mago, R. 2017. Loss of *AvrSr50* by somatic exchange in stem rust leads to virulence for Sr50 resistance in wheat. Science. 358:1607-1610.

Connelly, C. F., and Akey, J. M. 2012. On the prospects of whole-genome association mapping in *Saccharomyces cerevisiae*. Genetics. 191:1345-1353.

Cromey, M. G. and Parks, R. A. 2003. Pathogenic variation in *Dreschlera teres* in New Zealand. N. Z. Plant Protect. 56:251-256.

Daba, S. D., Horsley, R., Brueggeman, R., Chao, S., and Mohammadi, M. 2019. Genome-wide association studies and candidate gene identification for leaf scald and net blotch in barley (*Hordeum vulgare L.*). Plant dis. 103:880-889.

Dalman, K., Himmelstrand, K., Olson, Å., Lind, M., Brandström-Durling, M., and Stenlid, J. 2013. A genome-wide association study identifies genomic regions for virulence in the non-model organism *Heterobasidion annosum* ss. PLoS One. 8:e53525.

Dean, R. A., Talbot, N. J., Ebbole, D. J., Farman, M. L., Mitchell, T. K., Orbach, M. J., Thon, M., Kulkarni, R., Xu, J.R., Pan, H. and Read, N. D. 2005. The genome sequence of the rice blast fungus *Magnaporthe grisea*. Nature. 434:980.

Dinglasan, E., Hickey, L., Ziems, L., Fowler, R., Anisimova, A., Baranova, O., Lashina, N. and Afanasenko, O. S. 2019. Genetic characterization of resistance to *Pyrenophora teres* f. *teres* in the international barley differential Canadian Lake Shore. Front. Plant Sci. 10:326.

Dontsova, A. A., Alabushev, A. V., Lebedeva, M. V., Potokina, E. K. 2018. Analysis of polymorphism of microsatellite markers linked to a long-term net form of net blotch resistance gene in winter barley varieties in the south of Russia. Indian J. Genet. 78:317–323.

Drechsler, C. 1923. Some graminicolous species of *Helminthosporium*. J. Agric. Res. 24:641-740

Duplessis S, Cuomo C. A., Lin Y-C., Aerts A., Tisserant E. 2011. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. Proc. Natl. Acad. Sci. USA. 108:9166-9171.

Duplessis, S., Cuomo, C. A., Lin, Y. C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., Joly, D. L., Hacquard, S., Amselem, J., Cantarel, B. L. and Chiu, R. 2011. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. P. Natl. Acad. Sci. USA, 108:9166-9171.

Ellwood, S. R., Liu, Z., Syme, R. A., Lai, Z., Hane, J. K., Keiper, F., Moffat, C. S., Oliver, R. P., Friesen, T. L. 2010. A first genome assembly of the barley fungal pathogen *Pyrenophora teres* f. *teres*. Genome Biol. 11:R109.

Emebiri, L. C., Platz, C. G., Moody, D. B. 2005. Disease resistance genes in a doubled haploid population of two-rowed barley segregating for malting quality attributes. Aust. J. Agric. Res. 56:49-56.

Ficsor, A., Tóth, B., Varga, J., Csösz, M., Tomcsányi, A., Mészáros, K., Kótai, É. and Bakonyi, J. 2014. Variability of *Pyrenophora teres* f. *teres* in Hungary as revealed by mating type and RAPD analyses. J. Plant Pathol. 96:515-523.

Flint-Garcia, S. A., Thornsberry, J. M., and Buckler, E. S. 2003. Structure of Linkage Disequilibrium in Plants. Annu. Rev. Plant Biol. 54:357-374

Flor, H.H. 1955. Host-parasite interaction in flax rust - its genetics and other implications. Phytopathology 45.

Flor, H.H. 1971. Current Status of the Gene-For-Gene Concept. Annu. Rev. Phytopathol. 9:275–296.

Fouché, S., Badet, T., Oggenfuss, U., Plissonneau, C., Francisco, C. S., and Croll, D. 2019. Stress-driven transposable element de-repression dynamics and virulence evolution in a fungal pathogen. Mol. Biol. Evol.

Franckowiak, J. D., Platz, G. J. 2013. International database for barley genes and barley genetic stocks. Barley Genet. Newsl. 43:48–223.

Friesen, T. L., Faris, J. D., Lai, Z., Steffenson, B. J. 2006. Identification and chromosomal location of major genes for resistance to *Pyrenophora teres* in a doubled-haploid barley population. Genome 49:855-859.

Friesen, T. L., Holmes, D. J., Bowden, R. L., and Faris, J. D. 2018. ToxA is present in the US *Bipolaris sorokiniana* population and is a significant virulence factor on wheat harboring Tsn1. Plant dis. 102:2446-2452.

Friesen, T. L., Meinhardt, S. W., 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. Plant J. 51:681-692.

Friis, P., Olsen, C. E. and Møller, B. L. 1991. Toxin production in *Pyrenophora teres*, the Ascomycete causing the net-spot blotch disease of barley (Hordeum vulgare L.). J. Biol. Chem. 266:329-335.

Galagan, J. E., Calvo, S. E., Borkovich, K. A., Selker, E. U., Read, N. D., Jaffe, D., FitzHugh, W., Ma, L. J., Smirnov, S., Purcell, S. and Rehman, B. 2003. The genome sequence of the filamentous fungus *Neurospora crassa*. Nature. 422:859.

Ganeshan, V. D., Opiyo, S. O., Mutiga, S. K., Rotich, F., Thuranira, D. M., Were, V. M., Ouedraogo, I., Zhou, B., Soanes, D. M., Correll, J. C. and Wang, G. L. 2018. A Genome-Wide Association Study Identifies SNP Markers for Virulence in *Magnaporthe oryzae* Isolates from Sub-Saharan Africa. bioRxiv. p.418509.

Gao, Y., Liu, Z., Faris, J. D., Richards, J., Brueggeman, R. S., Li, X., Oliver, R. P., McDonald, B. A. and Friesen, T. L. 2016. Validation of genome-wide association studies as a tool to identify virulence factors in *Parastagonospora nodorum*. Phytopathol. 106:1177-1185.

Geschele, E.E. 1928. The response of barley to parasitic fungi *Helminthosporium teres* Sacc. Bull. Appl. Bot. Genet. Plant Breed. 19:371-384.

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M. and Louis, E. J. 1996. Life with 6000 genes. Science. 274:546-567.

Goodwin, S. B., M'Barek, S. B., Dhillon, B., Wittenberg, A. H., Crane, C. F., Hane, J. K., Foster, A. J., Van der Lee, T. A., Grimwood, J., Aerts, A. and Antoniw, J. 2011. Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. PLoS Genet. 7:e1002070.

Graner, A., Foroughi-Wehr, B., Tekauz, A. 1996. RFLP mapping of a gene in barley conferring resistance to net blotch (*Pyrenophora teres*). Euphytica. 91:229-234.

Grewal, T. S., Rossnagel, B. G., and Scoles, G. J. 2012. Mapping quantitative trait loci associated with spot blotch and net blotch resistance in a doubled-haploid barley population. Mol. Breed. 30:267-279

Grewal, T. S., Rossnagel, B. G., Pozniak, C. J. and Scoles, G. J. 2008. Mapping quantitative trait loci associated with barley net blotch resistance. Theor. Appl. Genet. 116:529-539.

Gupta, S., Li, C. D., Loughman, R., Cakir, M., Platz, G., Westcott, S., Bradley, J., Broughton, S., Lance, R. 2010. Quantitative trait loci and epistatic interactions in barley conferring resistance to net type net blotch (*Pyrenophora teres* f. *teres*) isolates. Plant Breed. 129:362-368.

Gupta, S., Li, C., Loughman, R., Cakir, M., Westcott, S., Lance, R. 2011. Identifying genetic complexity of 6H locus in barley conferring resistance to *Pyrenophora teres* f. *teres*. Plant Breed. 130:423-429.

Hargreaves, J. A. and Keon, J. P. R. 1983. The binding of isolated mesophyll cells from barley leaves to hyphae of *Pyrenophora teres*. Plant Cell Rep. 2:240-243

Hartmann, F. E., Sánchez-Vallet, A., McDonald, B. A., and Croll, D. 2017. A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. The ISME J. 11:1189.

Hisano, H., Sakamoto, K., Takagi, H., Terauchi, R., Sato, K. 2017. Exome QTL-seq maps monogenic locus and QTLs in barley. BMC Genomics 18:125.

Huddleston, J., Ranade, S., Malig, M., Antonacci, F., Chaisson, M., Hon, L., Sudmant, P. H., Graves, T. A., Alkan, C., Dennis, M. Y. and Wilson, R. K. 2014. Reconstructing complex regions of genomes using long-read sequencing technology. Genome Res. 24:688-696.

Islamovic, E., Bregitzer, P., Friesen, T. L. 2017. Barley 4H QTL confers NFNB resistance to a global set of *P. teres* f. *teres* isolates. Mol. Breed. 37:29.

Ismail, I. A., Able, A. J. 2016. Secretome analysis of virulent *Pyrenophora teres* f. *teres* isolates. Protemics. 16:2625-2636.

Ismail, I. A., Able, A. J. 2017. Gene expression profiling of virulence-associated proteins in planta during net blotch disease of barley. Physiol. Mol. Plant Pathol. 98:69-79.

Ismail, I. A., Godfrey, D., Able, A. J. 2014a. Fungal growth, proteinaceous toxins and virulence of *Pyrenophora teres* f. *teres* on barley. Australas. Plant Pathol. 43:535-546.

Ismail, I. A., Godfrey, D., Able, A. J. 2014b. Proteomic analysis reveals the potential involvement of xylanase from *Pyrenophora teres* f. *teres* in net form net blotch disease of barley. Australas. Plant Pathol. 43:715-726.

Jalli, M. and Robinson, J. 2000. Stable resistance in barley to *Pyrenophora teres* f. *teres* isolates from the Nordic-Baltic region after increase on standard host genotypes. Euphytica. 113:71-77.

Jonsson, R., Bryngelsson, T. and Gustafsson, M. 1997. Virulence studies of Swedish net blotch isolates (*Drechslera teres*) and identification of resistant barley lines. Euphytica. 94:209-218.

Jonsson, R., Sall, T., and Bryngelsson, T. 2000. Genetic diversity for random amplified polymorphic DNA (RAPD) markers in two Swedish populations of *Pyrenophora teres*. Can. J. Plant Pathol. 22:258-264

Jordan, V. W. L. 1981. Aetiology of barley net blotch caused by *Pyrenophora teres* and some effects on yield. Plant Pathol. 30:77-87

Jørgensen, H. J. L., Lübeck, P. S., Thordal-Christensen, H., de Neer-gaard, E. and Smedegård-Petersen, V. 1998. Mechanisms of induced resistance in barley against *Drechslera teres*. Phytopathol. 88:698–70

Khan, T.N. 1982. Changes in barley genotypes grown in Western Australia. Plant Dis. 66:655-656.

Khan, T.N. and Boyd, W.J.R. 1969a. Physiologic specialization in *Drechslera teres*. Aust. J. Biol. Sci. 22:1229-1235.

Klosterman, S. J., Subbarao, K. V., Kang, S., Veronese, P., Gold, S.E., Thomma, B. P., Chen, Z., Henrissat, B., Lee, Y. H., Park, J. and Garcia-Pedrajas, M. D. 2011. Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. PLoS pathogens. 7:e1002137.

Koladia, V. M., Faris, J. D., Richards, J. K., Brueggeman, R. S., Chao, S., Friesen, T. L. 2017a. Genetic analysis of net form net blotch resistance in barley lines CIho 5791 and Tifang against a global collection of *P. teres* f. *teres* isolates. Theor. Appl. Genet. 130:163-173.

Koladia, V. M., Richards, J. K., Wyatt, N. A., Faris, J. D., Brueggeman, R.S., Friesen, T.L. 2017b. Genetic analysis of virulence in the *Pyrenophora teres* f. *teres* population BB25×FGOH04Ptt-21. Fungal Genet. Biol. 107:12-19.

König, J., Perovic, D., Kopahnke, D., Ordon, F. 2013. Development of an efficient method for assessing resistance to the net type of net blotch (*Pyrenophora teres* f. *teres*) in winter barley and mapping of quantitative trait loci for resistance. Mol. Breed. 32:641-650.

König, J., Perovic, D., Kopahnke, D., Ordon, F., 2014. Mapping seedling resistance to net form of net blotch (*Pyrenophora teres* f. *teres*) in barley using detached leaf assay. Plant Breed. 133:356-365.

Lai, Z., Faris, J. D., Weiland, J. J., Steffenson, B. J., Friesen, T. L., 2007. Genetic mapping of *Pyrenophora teres* f. *teres* genes conferring avirulence on barley. Fungal Genet. Biol. 44:323-329.

Laszlo, A. H., Derrington, I. M., Ross, B. C., Brinkerhoff, H., Adey, A., Nova, I. C., Craig, J. M., Langford, K. W., Samson, J. M., Daza, R. and Doering, K. 2014. Decoding long nanopore sequencing reads of natural DNA. Nat.Biotechnol. 32:829.

Leboldus, J. M., Kinzer, K., Richards, J., Ya, Z., Yan, C., Friesen, T. L., Brueggeman, R. 2015. Genotype-by-sequencing of the plant-pathogenic fungi *Pyrenophora teres* and *Sphaerulina musiva* utilizing Ion Torrent sequence technology. Mol. Plant Pathol. 16:623-632.

Lehmensiek, A., Bester, A. E., Sutherland, M. W., Platz, G., Kriel, W. M., Potgieter, G. F. and Prins, R. 2010. Population structure of South African and Australian *Pyrenophora teres* isolates. Plant Pathol. 59:504-515.

Lehmensiek, A., Platz, G. J., Mace, E., Poulsen, D., Sutherland, M. W. 2007. Mapping of adult plant resistance to net form of net blotch in three Australian barley populations. Aust. J. Agric. Res. 58:1191-1197.

Leisova, L., Minarikova, V., Kucera, L., and Ovesna, J. 2005. Genetic Diversity of *Pyrenophora teres* Isolates as Detected by AFLP Analysis. Phytopathol. 153:569-578

Li, M., Liu, X., Bradbury, P., Yu, J., Zhang, Y., Todhunter, R. J., Buckler, E. S., and Zhang, Z. 2014. Enrichment of statistical power for genome-wide association studies. BMC Biol. 12:73

Linde, C. C., & Smith, L. M. 2019. Host specialisation and disparate evolution of *Pyrenophora teres* f. *teres* on barley and barley grass. BMC Evol. Biol. 19:139.

Liu, Z. H., Zhong, S., Stasko, A. K., Edwards, M. C., and Friesen, T. L. 2012. Virulence Profile and Genetic Structure of a North Dakota Population of *Pyrenophora teres* f. *teres*, the Causal Agent of Net Form Net Blotch of Barley. Phytopathol. 102:539-546

Liu, Z., Ellwood, S. R., Oliver, R. P., and Friesen, T. L. 2011. *Pyrenophora teres*: profile of an increasingly damaging barley pathogen. Mol. Plant Pathol. 12:1-19

Liu, Z., Faris, J.D., Edwards, M.C., Friesen, T.L. 2010. Development of expressed sequence tag (EST)-based markers for genomic analysis of a barley 6H region harboring multiple net form net blotch resistance genes. Plant Genome 3:41-52.

Liu, Z., Holmes, D. J., Faris, J. D., Chao, S., Brueggeman, R. S., Edwards, M. C., and Friesen, T. L. 2015. Necrotrophic effector-triggered susceptibility (NETS) underlies the barley-*Pyrenophora teres* f. *teres* interaction specific to chromosome 6H. Mol. Plant Pathol. 16:188-200

Ma, Z. Q., Lapitan, N. L. V, Steffenson, B. 2004. QTL mapping of net blotch resistance genes in a doubled-haploid population of six-rowed barley. Euphytica. 137:291-296.

Manninen, O. M., Jalli, M., Kalendar, R., Schulman, A., Afanasenko, O., Robinson, J. 2006. Mapping of major spot-type and net-type net-blotch resistance genes in the Ethiopian barley line CI 9819. Genome. 49:1564-1571.

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. 3:41-63.

Martin, A., Platz, G. J., de Klerk, D., Fowler, R. A., Smit, F., Potgieter, F. G., Prins, R. 2018. Identification and mapping of net form of net blotch resistance in South African barley. Mol. Breed. 38:53.

Mathre, D. E. 1997. Compendium of Barley Diseases. The American Phytopathological Society. St. Paul.

Maurer, A., Draba, V., Jiang, Y., Schnaithmann, F., Sharma, R., Schumann, E., Kilian, B., Reif, J. C., Pillen, K. 2015. Modelling the genetic architecture of flowering time control in barley through nested association mapping. BMC Genomics. 16:290.

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*. Mol Plant Pathol. 19:432-439.

McDonald, M. C., Taranto, A. P., Hill, E., Schwessinger, B., Liu, Z., Simpfendorfer, S., Milgate, A. and Solomon, P. 2019. Transposon mediated horizontal transfer of the host-specific virulence protein ToxA between three fungal wheat pathogens. bioRxiv. p671446.

McLean, M. S., Howlett, B. J., and Hollaway, G. J. 2009. Epidemiology and control of spot form of net blotch (*Pyrenophora teres* f. *maculata*) of barley: a review. Crop Pasture Sci. 60:303-315

Mode, C. J., Schaller, C. W., 1958. Two additional factors for host resistance to net blotch in barley. Agron. J. 50:15-18.

Molnar, S. J., James, L. E., Kasha, K. J. 2000. Inheritance and RAPD tagging of multiple genes for resistance to net blotch in barley. Genome. 43:224-231.

Moolhuijzen, P., See, P. T., Hane, J. K., Shi, G., Liu, Z., Oliver, R. P., & Moffat, C. S. 2018. Comparative genomics of the wheat fungal pathogen *Pyrenophora tritici-repentis* reveals chromosomal variations and genome plasticity. BMC genomics. 19:279.

Muller, L. A., Lucas, J. E., Georgianna, D. R., and McCusker, J. H. 2011. Genome-wide association analysis of clinical vs. nonclinical origin provides insights into *Saccharomyces cerevisiae* pathogenesis. Mol. Ecol. 20:4085-4097.

Muñoz-Amatriaín, M., Cuesta-Marcos, A., Endelman, J. B., Comadran, J., Bonman, J. M., Bockelman, H. E., Chao, S., Russell, J., Waugh, R., Hayes, P. M., Muehlbauer, G. J. 2014. The USDA Barley Core Collection: Genetic Diversity, Population Structure, and Potential for Genome-Wide Association Studies. PLoS One. 9:e94688.

Murray, G. M. and Brennan, J. P. 2010. Estimating disease losses to the Australian barley industry. Austral. Plant Pathol. 39:85-96

Novakazi, F., Afanasenko, O., Anisimova, A., Platz, G. J., Snowdon, R., Kovaleva, O., Ordon, F. 2019. Genetic analysis of a worldwide barley collection for resistance to net form of net blotch disease (*Pyrenophora teres* f. *teres*). Theor. Appl. Genet. 1:18.

O'Boyle, P. D., Brooks, W. S., Barnett, M. D., Berger, G. L., Steffenson, B. J., Stromberg, E. L., Maroof, M. A. S., Liu, S. Y., Griffey, C. A. 2014. Mapping Net Blotch Resistance in 'Nomini' and CIho 2291 Barley. Crop Sci. 54:2596:2602.

Ohm R. A., Feau N, Henrissat B., Schoch C. L., Horwitz B. A. 2012. Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. PLOS Pathog. 8:e1003037.

Oliver, S. G., Van der Aart, Q. J. M., Agostoni-Carbone, M. L., Aigle, M., Alberghina, L., Alexandraki, D., Antoine, G., Anwar, R., Ballesta, J. P. G., Benit, P. and Berben, G. 1992. The complete DNA sequence of yeast chromosome III. Nature. 357:38.

Peever, T. L. and Milgroom, M. G. 1994. Genetic structure of *Pyrenophora teres* populations determined with random amplified polymorphic DNA markers. Can. J. Bot. 72:915-923

Peltonen, S., Jalli, M., Kammiovirta, K. and Karjalainen, R. 1996. Genetic variation in *Drechslera teres* populations as indicated by RAPD markers. Ann. Appl. Biol. 128:465-477.

Piškur, J., and Langkjær, R. B. 2004. Yeast genome sequencing: the power of comparative genomics. Mol. Microbiol. 53:381-389.

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

Poudel, B., Ellwood, S. R., Testa, A. C., McLean, M., Sutherland, M. W., & Martin, A. 2017. Rare *Pyrenophora teres* hybridization events revealed by development of sequence-specific PCR markers. Phytopathol. 107:878-884.

Poudel, B., Vaghefi, N., McLean, M. S., Platz, G. J., Sutherland, M. W., and Martin, A. 2019. Genetic structure of a *Pyrenophora teres* f. *teres* population over time in an Australian barley field as revealed by Diversity Arrays Technology markers. Plant Pathol.

Powers, J. G., Weigman, V. J., Shu, J., Pufky, J. M., Cox, D., and Hurban, P. 2013. Efficient and accurate whole genome assembly and methylome profiling of E. coli. BMC Genomics. 14: 675.

Pritchard, J. K., Stephens, M., and Donnelly, P. 2000. Inference of Population Structure Using Multilocus Genotype Data. Genet. 155:945-959

Rafalski, J. A. 2010. Association genetics in crop improvement. Curr. Opin. Plant Biol. 13:174-180

Raman, H., Platz, G.J., Chalmers, K. J., Raman, R., Read, B. J., Barr, A. R., Moody, D. B. 2003. Mapping of genomic regions associated with net form of net blotch resistance in barley. Aust. J. Agric. Res. 54:1359-1367.

Rau, D., Brown, A. H. D., Brubaker, C. L., Attene, G., Balmas, V., Saba, E., and Papa, R. 2003. Population genetic structure of *Pyrenophora teres* Drechs. the causal agent of net blotch in Sardinian landraces of barley (*Hordeum vulgare L.*). Theor. Appl. Genet. 106:947-959

Rau, D., Maier, F. J., Papa, R., Brown, A. H. D., Balmas, V., Saba, E., Schaefer, W., and Attene, G. 2005. Isolation and characterization of the mating-type locus of the barley pathogen *Pyrenophora teres* and frequencies of mating-type idiomorphs within and among fungal populations collected from barley landraces. Genome. 48:855-869

Rau, D., Rodriguez, M., Leonarda Murgia, M., Balmas, V., Bitocchi, E., Bellucci, E., Nanni, L., Attene, G., Papa, R. 2015. Co-evolution in a landrace meta-population: two closely related pathogens interacting with the same host can lead to different adaptive outcomes. Sci. Rep. 5:12834.

Read, B.J., Raman, H., McMichael, G., Chalmers, K. J., Ablett, G. A., Platz, G. J., Raman, R., Genger, R. K., Boyd, W. J. R., Li, C. D., Grime, C. R., Park, R. F., Wallwork, H., Prangnell, R., Lance, R. C. M. 2003. Mapping and QTL analysis of the barley population Sloop × Halcyon. Aust. J. Agric. Res. 54:1145-1153.

Richards, J. K., Wyatt, N. A., Liu, Z., Faris, J. D., and Friesen, T. L. 2018. Reference quality genome assemblies of three *Parastagonospora nodorum* isolates differing in virulence on wheat. G3. 8:393-399.

Richards, J., Chao, S., Friesen, T., and Brueggeman, R. 2016. Fine mapping of the barley chromosome 6H net form net blotch susceptibility locus. G3. 6:1809-1818.

Richards, J., Stukenbrock, E., Carpenter, J., Liu, Z., Cowger, C., Faris, J., and Friesen, T. L. 2019. Local adaptation drives the diversification of effectors in the fungal wheat pathogen *Parastagonospora nodorum* in the United States. Plos Genet. 15:e1008223.

Richards, J.K., Friesen, T.L., Brueggeman, R.S. 2017. Association mapping utilizing diverse barley lines reveals net form net blotch seedling resistance/susceptibility loci. Theor. Appl. Genet. 130:915-927.

Richter, K., Schondelmaier, J., Jung, C. 1998. Mapping of quantitative trait loci affecting *Drechslera teres* resistance in barley with molecular markers. Theor. Appl. Genet. 97:1225–1234.

Rozanova, I. V, Lashina, N. M., Mustafin, Z. S., Gorobets, S. A., Efimov, V. M., Afanasenko, O. S., Khlestkina, E. K. 2019. SNPs associated with barley resistance to isolates of *Pyrenophora teres* f. *teres*. BMC Genomics. 20:292.

Salcedo, A., Rutter, W., Wang, S., Akhunova, A., Bolus, S., Chao, S., Anderson, N., De Soto, M. F., Rouse, M., Szabo, L. and Bowden, R. L. 2017. Variation in the *AvrSr35* gene determines Sr35 resistance against wheat stem rust race Ug99. Science. 358:1604-1606.

Sarpeleh, A., Wallwork, H., Catcheside, D. E .A., Tate, M. E. and Able, A. J. 2007. Evidence of involvement of proteinaceous toxins from *Pyrenophora teres* in symptom development of net blotch of barley. Phytopathol. 97:907-915.

Sato, K. and Takeda, K. 1993. Pathogenic variation of *Pyrenophora teres* isolates collected from Japanese and Canadian spring barley. Bull. Res. Inst. Bioresour. Okayama Univ. 1:147-158

Schaller, C. W. 1955. Inheritance of resistance to net blotch of barley. Phytopathol. 45:174-176.

Schwessinger, B., Sperschneider, J., Cuddy, W. S., Garnica, D. P., Miller, M. E., Taylor, J. M., Dodds, P. N., Figueroa, M., Park, R. F. and Rathjen, J. P. 2018. A near-complete haplotype-phased genome of the dikaryotic wheat stripe rust fungus *Puccinia striiformis* f. sp. *tritici* reveals high interhaplotype diversity. MBio. 9:e02275-17.

Serenius, M., Manninen, O., Wallwork, H., and Williams, K. 2007. Genetic differentiation in *Pyrenophora teres* populations measured with AFLP markers. Mycol. Res. 111:213-223

Serenius, M., Mironenko, N., and Manninen, O. 2005. Genetic variation, occurrence of mating types and different forms of *Pyrenophora teres* causing net blotch of barley in Finland. Mycol. Res. 109:809-817

Sharma, H. S. S. 1984. Assessment of the reaction of some spring barley cultivars to *Pyrenophora teres* using whole plants, detached leaves and toxin bioassay. Plant Pathol. 33:371-376.

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 Biol. Evol. 11:855-868.

Shipton, W. A., Khan, T. N., and Boyd, W. J. R. 1973. Net blotch of barley. Rev. Plant Pathol. 52:269-290

Shjerve, R. A., Faris, J. D., Brueggeman, R. S., Yan, C., Zhu, Y., Koladia, V., Friesen, T. L. 2014. Evaluation of a *Pyrenophora teres* f. *teres* mapping population reveals multiple independent interactions with a region of barley chromosome 6H. Fungal Genet. Biol. 70:104-112.

Shoemaker, R. A. 1959. Nomenclature of *Drechslera* and *Bipolaris*, grass parasites segregated from '*Helminthosporium*'. Can. J. Bot. 37:879-887

Smedegård-Petersen, V. 1978. Genetics of heterothallism in *Pyrenophora graminea* and *P. teres*. Trans. Br. Mycol. Soc. 70: 99-102.

Smedegård-Peterson, V. 1977. Isolation of two toxins produced by *Pyrenophora teres* and their significance in disease development of net-spot blotch of barley. Physiol. Plant Pathol. 10:203-211

Spaner, D., Shugar, L. P., Choo, T. M., Falak, I., Briggs, K. G., Legge, W. G., Falk, D. E., Ullrich, S. E., Tinker, N. A., Steffenson, B. J. and Mather, D. E. 1998. Mapping of disease resistance loci in barley on the basis of visual assessment of naturally occurring symptoms. Crop Sci. 38:843-850.

Spanu P. D., Abbott J. C., Amselem J, Burgis T. A., Soanes D. M., S. 2010. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. Science 330: 1543-1546.

St. Pierre, S., Gustus, C., Steffenson, B., Dill-Macky, R., Smith, K. P. 2010. Mapping net form net blotch and Septoria speckled leaf blotch resistance loci in barley. Phytopathol. 100;80-84.

Steffenson, B. J. and Webster, R. K. 1992. Pathotype Diversity of *Pyrenophora teres* f. *teres* on Barley. Phytopathol. 82:170-177

Steffenson, B. J., Hayes, P. M., Kleinhofs, A. 1996. Genetics of seedling and adult plant resistance to net blotch (*Pyrenophora teres* f. *teres*) and spot blotch (*Cochliobolus sativus*) in barley. Theor. Appl. Genet. 92:552-558

Stukenbrock, E. H., and Dutheil, J. Y. 2018. Fine-scale recombination maps of fungal plant pathogens reveal dynamic recombination landscapes and intragenic hotspots. Genetics. 208:1209-1229.

Syme, R. A., Martin, A., Wyatt, N. A., Lawrence, J. A., Muria-Gonzalez, M. J., Friesen, T. L., Ellwood, S. R. 2018. Transposable element genomic fissuring in *Pyrenophora teres* is

associated with genome expansion and dynamics of host–pathogen genetic interactions. Front. Genet. 9:130.

Talas, F., Kalih, R., Miedaner, T., McDonald, B. A. 2016. Genome-wide association study identifies novel candidate genes for aggressiveness, deoxynivalenol production, and azole sensitivity in natural field populations of *Fusarium graminearum*. Mol. Plant Microbe Interact. 29:417-430.

Tekauz, A. 1990. Characterization and distribution of pathogenic variation in *Pyrenophora teres* f. *teres* and *P. teres* f. *maculata* from western Canada. Can. J. Plant Pathol. 12:141-148

Thomma, B. P., Seidl, M. F., Shi-Kunne, X., Cook, D. E., Bolton, M. D., van Kan, J. A., Faino, L. 2016. Mind the gap; seven reasons to close fragmented genome assemblies. Fungal Genet. Biol. 90:24-30.

Toruño, T. Y., Stergiopoulos, I., and Coaker, G. 2016. Plant-pathogen effectors: cellular probes interfering with plant defenses in spatial and temporal manners. Ann. Rev. Phytopathol. 54:419-441.

Van Caeseele, L. and Grumbles, J. 1979. Ultrastructure of the interac-tion between *Pyrenophora teres* and a susceptible barley host. Can. J. Bot. 57:40–47.

Vatter, T., Maurer, A., Kopahnke, D., Perovic, D., Ordon, F., Pillen, K. 2017. A nested association mapping population identifies multiple small effect QTL conferring resistance against net blotch (*Pyrenophora teres* f. *teres*) in wild barley. PloS one. 12:e0186803.

Vieira, Y., Silva, D. N., Varzea, V. M. P., Paulo, O. S., Batista, D. 2019. Genome-wide signatures of selection in *Colletotrichum kahawae* reveal candidate genes potentially involved in pathogenicity and aggressiveness. Front. Microbiol. 10:1374.

Weiergang, I., Jørgensen, H. J. L., Møller, I. M., Friis, P. and Smedegård-Petersen, V. 2002. Correlation between sensitivity of barley to *Pyrenophora teres* toxins and susceptibility to the fungus. Physiol. Mol. Plant Pathol. 60:121-129.

Weiland, J. J., Steffenson, B. J., Cartwright, R. D., Webster, R. K. 1999. Identification of molecular genetic markers in *Pyrenophora teres* f. *teres* associated with low virulence on 'Harbin' barley. Phytopathol. 89:176-181.

Wonneberger, R., Ficke, A., Lillemo, M. 2017a. Mapping of quantitative trait loci associated with resistance to net form net blotch (*Pyrenophora teres* f. *teres*) in a doubled haploid Norwegian barley population. PLoS One. 12:e0175773.

Wonneberger, R., Ficke, A., Lillemo, M. 2017b. Identification of quantitative trait loci associated with resistance to net form net blotch in a collection of Nordic barley germplasm. Theor. Appl. Genet. 130:2025-2043.

Wyatt, N. A., Richards, J. K., Brueggeman, R. S., Friesen, T. L. 2018. Reference assembly and annotation of the *Pyrenophora teres* f. *teres* isolate 0-1. G3. 8:1-8.

Wyatt, N. A., Richards, J., Brueggeman, R. S., Friesen, T. L. 2019. A comparative genomic analysis of the barley pathogen *Pyrenophora teres* f. *teres* identifies sub-telomeric regions as drivers of virulence. Mol. Plant Microbe Interact.

Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Dolebley, J. F., McMullen, M. D., Gaut, B. S., Nielsen, D. M., Holland, J. B., Kresovich, S., Buckler, E. S. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genet. 38:203-208

Yun, S.J., Gyenis, L., Hayes, P. M., Matus, I., Smith, K. P., Steffenson, B. J., G. J. Muehlbauer, G. J. M. 2005. Quantitative trait loci for multiple disease resistance in wild barley. Crop Sci. 45:2563-2572.

Zhang, Z., Ersoz, E., Lai, C., Todhunter, R. J., Tiwari, H. K., Gore, M. A., Bradbury, P. J., Yu, J., Arnett, D. K., Ordovoas, J. M., Buckler, E. S. 2010. Mixed linear model approach adapted for genome-wide association studies. Nature Genet. 42:355-360

Zhao, K., Aranzana, M. J., Kim, S., Lister, C., Shindo, C., Tang, C., Toomajian, C., Zheng, J., Dean, C., Marjoram, P., Nordborg, M. 2007. An Arabidopsis Example of Association Mapping in Structured Samples. PLoS Genet. 3:e4

Zhu, C., Gore, M., Buckler, E. S., and Yu, J. 2008. Status and Prospects of Association Mapping in Plants. Plant Genome. 1:5-20

CHAPTER 2. REFERENCE ASSEMBLY AND ANNOTATION OF THE *PYRENOPHORA TERES* F. *TERES* ISOLATE 0-1[1]

Abstract

Pyrenophora teres f. *teres*, the causal agent of net form net blotch (NFNB) of barley, is a destructive pathogen in barley growing regions throughout the world. Typical yield losses due to NFNB range from 10-40%, however, complete loss has been observed on highly susceptible barley lines where environmental conditions favor the pathogen. Currently, genomic resources for this economically important pathogen are limited to a fragmented draft genome assembly and annotation with limited RNA support of the *P. teres* f. *teres* isolate 0-1. This research presents an updated 0-1 reference assembly facilitated by long read sequencing and scaffolding with the assistance of genetic linkage maps. Additionally, genome annotation was mediated by RNAseq analysis using three infection time points and a pure culture sample resulting in 11,541 high-confidence gene models. The 0-1 genome assembly and annotation presented here now contains the majority of the repetitive content of the genome. Analysis of the 0-1 genome revealed classic characteristics of a 'two-speed' genome, being compartmentalized into GC-equilibrated and AT-rich compartments. The assembly of repetitive AT-rich regions will be important for future investigation of genes known as effectors which often reside in close proximity to repetitive regions. These effectors are responsible for manipulation of the host defense during infection.

[1] The material in this chapter was co-authored by Nathan A. Wyatt, Jonathan K. Richards, Robert S. Brueggeman, and Timothy L. Friesen. Nathan A. Wyatt had primary responsibility for laboratory experiements, computational analysis, and preparation of the manuscript. Jonathan K. Richards provided valuable consultation of computational analysis and proofread and revised the manuscript. Robert S. Brueggeman proofread and revised the manuscript. Timothy L. Friesen advised Nathan A. Wyatt during laboratory experiments and computational analysis and drafted, proofread, and revised the manuscript. This chapter has been published as Wyatt, N. A., Richards, J. K., Brueggeman, R. S., & Friesen, T. L. (2018). Reference assembly and annotation of the *Pyrenophora teres* f. *teres* isolate 0-1. G3: Genes, Genomes, Genetics, 8:1-8.

This updated *P. teres* f. *teres* isolate 0-1 reference genome assembly and annotation provides a robust resource for the examination of the barley-*P. teres* f. *teres* host-pathogen co-evolution.

Introduction

Net form net blotch (NFNB) of barley (*Hordeum vulgare*) is caused by the fungal pathogen *Pyrenophora teres* f. *teres*. Globally, NFNB results in regular yield losses between 10 and 40% with the potential for complete losses in environmental settings favorable to the pathogen, namely, susceptible cultivars with high sustained humidity and the absence of fungicides (Mathre 1997, Liu et al. 2011). Several studies have investigated the genetics of this host-pathogen interaction, utilizing bi-parental mapping populations of both the host and pathogen as well as genome wide association studies (GWAS) in the host (Liu et al. 2011; Shjerve et al. 2014; Carlsen et al. 2017; Koladia et al. 2017a; Richards et al. 2017). These studies have been critical in developing hypothetical models for this pathosystem. These models have proposed that the *P. teres* f. *teres*-barley interaction involves the production of effectors that are involved in manipulating the host to gain an advantage (Koladia et al. 2017a) and that some of these effectors may be recognized by dominant resistance genes (Koladia et al. 2017b), showing hallmarks of an effector triggered susceptibility/effector triggered immunity type model as described by Jones and Dangle (2006) and Chisholm et al. (2006). *P. teres* f. *teres* has also been shown to produce necrotrophic effectors (NE) that are involved in NE triggered susceptibility (Liu et al. 2015; Shjerve et al. 2014) when recognized by dominant host susceptibility genes (Abu Qamar et al. 2008; Liu et al. 2010). Together these studies indicate a complex interaction where selection pressure has been placed on the pathogen to produce different types of effectors to manipulate its host.

Currently, 1,084 fungal genomes have been sequenced and deposited at the National Center for Biotechnology Information (NCBI) (as of August, 2017, source: http://www.ncbi.nlm.nih.gov/genome/browse/), and many of these genomes remain fragmented. The genomic regions responsible for this fragmentation are typically repetitive regions which are common among fungal "two speed" genomes, a term used for genomes that have a distinct bipartite architecture made up of typical gene rich regions and repeat-rich, gene-poor regions thought to be hotbeds for genome evolution (Dong et al. 2015). Genome assemblies that rely on short read technologies struggle to span these repetitive regions as the repeats are often collapsed (Ekblom and Wolf 2014). Effector genes involved in the manipulation of the plant defenses cluster proximal to and within these repeat regions (Thomma et al. 2016) and therefore it is critical that these regions be accounted for in any genome assembly of a plant pathogen being used for effector discovery.

The current reference isolate of *P. teres* f. *teres* is the Canadian isolate 0-1. The original draft genome sequence of 0-1 was reported in 2010 and used paired-end Illumina sequencing at roughly 20× coverage, resulting in an assembly with a total size of 41.95 Mb (Ellwood et al. 2010). This assembly contained 11,799 gene models providing useful tools for the interrogation of the genome, however, the genome still remained fragmented at 6,412 contigs with an N50 of only 30,037 bps. The fragmented nature of this assembly presented obstacles to the study of the genome of *P. teres* f. *teres* especially in regards to map based cloning using bi-parental mapping populations.

New long read sequencing technologies (e.g. Pacific Biosciences (PacBio) single molecule real-time (SMRT) sequencing platform) are capable of producing reads up to 60 kb (Goodwin et al. 2016) and currently two fungal plant pathogens have been fully sequenced using

these long read technologies (Faino et al. 2015; van Kan et al. 2017). These results demonstrate the utility of long read technologies in the production of better assemblies of fungal plant pathogens.

Here we present an updated, reference quality genome assembly and annotation for the *P. teres* f. *teres* isolate 0-1. In order to update the reference assembly and annotation we sequenced a total of 14 SMRT cells at the National Center for Genome Resources (NCGR) and conducted RNAseq using both *in-vitro* and *in-planta* samples. The updated 0-1 assembly is currently in 86 contigs with a total genomic content of 46.5 Mb with a high confidence annotation set of 11,541 gene models. This updated assembly and annotation presents a useful tool for the genomic interrogation of *P. teres* f. *teres* and specifically the investigation of the secretome and effectorome.

Materials and methods

Biological materials and high molecular weight DNA extraction

Pyrenophora teres f. *teres* isolate 0-1 is a Canadian isolate collected from Ontario (Weiland et al. 1999). Fungal tissue for DNA extraction was obtained in a similar manner to Shjerve et al. (2014). Briefly, a single dried fungal plug was placed on a V8-PDA plate and allowed to grow for 5 days after which the culture underwent a 24 hr light and dark cycle to induce sporulation. Spores were then inoculated into Fries media [5 g $(NH_4)_2C_4H_4O_6$, 1 g NH_4NO_3, 0.5 g $MgSO_4$ * $7H_2O$, 1.3 g KH_2PO_4, 5.48 g K_2HPO_4 * $3H_2O$, 30 g Sucrose, 1 g Yeast extract, 2 mL trace element stock solution (167 mg LiCl, 107 mg CuCl * H_2O, 34 mg H_2MoO_4, 72 mg $MnCl_2$ * $4H_2O$, 80 mg $CoCl_2$ * $4H_2O$, ddH_2O to 1L)] and allowed to grow for five days. Five day cultures were blended and inoculated into new Fries media and allowed to grow for another 24 hrs before harvesting.

Harvested tissue was ground under liquid Nitrogen with a mortar and pestle to a fine powder and then placed in a 50 mL conical tube. Next, 25 mL of Qiagen RLT buffer and 150 µL Rnase A at 20 mg/mL was added to the 50 mL conical tube containing fungal tissue. The mixture was homogenized by pipetting and vortexing until well mixed and incubated at 65° for 45 min, mixing at 15 min intervals. The 50 mL conical tube was centrifuged at 3,166 ×*g* for 15 min and the resulting supernatant was split between two Oakridge tubes. Equal volumes of 25:24:1 phenol:chloroform:isoamyl alcohol was added to each tube and mixed gently and thoroughly by rocking. After mixing, tubes were centrifuged in a fixed angle rotor for 20 min at 13,000 ×*g* at room temperature. The aqueous layer was drawn off so as not to disturb the middle phase and placed in a new 50 mL conical tube. 0.4 volumes of sodium acetate and an equal volume of isopropyl alcohol were added to the 50 mL tube and the solution was mixed and incubated at room temperature for 30 min to precipitate the DNA. DNA was removed from the 50 mL conical tube using a glass hook and placed in a clean weigh boat and subsequently rinsed twice using 2-5 mL of freshly prepared 70% ethanol. Ethanol was then pulled off via pipetting and DNA was moved to a 5 mL tube and placed in a lyophilizer for 30 min to dry. DNA was rehydrated with 1 mL of molecular biology grade water and incubated at 4° overnight. DNA was quantified using a Qubit (ThermoFischer Scientific) and sample concentration was adjusted to 1 µg/mL.

Genomic sequencing and de novo assembly

Genomic DNA was shipped on dry ice to NCGR (Santa Fe, NM) for library preparation and sequencing. Whole genome shotgun sequencing was performed at NCGR using the PacBio RSII instrument with a 20 kb size selected library and current P6-C4 chemistry. A total of 14 SMRT cells were sequenced for *P. teres* f. *teres* isolate 0-1.

Raw reads in the form of FASTQ files were input into the Canu assembler (Koren et al. 2017) under default parameters for correction, trimming, and assembly with a genome size estimate of 42 Mb. A second iteration of genome assembly was then done with the same parameters but a larger genome estimate of 46.5 Mb based on the first draft assembly. Pilon v1.21 (Walker et al. 2014) was used to polish the 0-1 assembly to improve local base calling accuracy. Pilon takes the reference assembly FASTA file and a BAM file of the aligned reads as an input to identify miscalled bases, small indels, or large structural mis-assemblies for correction and outputs a new FASTA file containing the polished genome assembly.

Genetic mapping and genome scaffolding

A genetic linkage map was created from the bi-parental population 0-1 × 15A consisting of 120 progeny isolates, 78 isolates were obtained from Lai et al. (2007) and an additional 42 isolates were obtained from another 0-1 × 15A cross following the methods described in Koladia et al. (2017a). Single nucleotide polymorphic (SNP) markers were identified following a RAD-GBS pipeline also described in Koladia et al. (2017a). A total of ten progeny isolates were dropped from the analysis due to large amounts of missing data (>75%) and five isolates were dropped after being identified as parental clones bringing the total to 105 progeny isolates. Linkage mapping was conducted in MapDisto v1.7.9 (Lorieux, 2012) as described in Koladia et al. (2017a) using a LOD threshold of 5.0.

P. teres f. *teres* isolate 0-1 assembled contigs were scaffolded using ALLMAPS (Tang et al. 2015). ALLMAPS takes assembled contigs and generates scaffolds based on coordinates from genetic linkage maps, optical maps, or syntenic maps. To scaffold the 0-1 assembly, linkage maps from the bi-parental populations 0-1 × 15A and the recently published FGOH04-Ptt21 × BB25 population (Koladia et al. 2017a) were input into ALLMAPS and merged to a single

coordinate BED file. The merged coordinate BED file was input with the 0-1 genome FASTA file for scaffolding under default parameters within ALLMAPS.

RNA sequencing and assembly

Cultures and inoculum of *P. teres* f. *teres* isolate 0-1 were prepared as previously described (Koladia et al. 2017a). Seeds of barley line 'Tifang' were sown into a 96-conetainer rack with a border of Tradition barley and grown under greenhouse conditions for approximately two weeks. For each RNAseq sample time point, five individual conetainers containing two seedlings each were selected for inoculation. The second fully extended leaf of each seedling was taped flat to a 24 × 30 cm plastic surface so as to provide a flat surface to evenly coat sample leaves with inoculum. Following inoculation, plants were placed into a lighted mist chamber for 24 hrs with 100% relative humidity. After 24 hrs, plants were moved to a growth chamber with a temperature of 21° and a 12 hr photoperiod. Samples were collected by punching circular leaf discs from pre-designated regions of the leaf using a sterile hole punch. Each leaf was punched a total of five times equating to 50 tissue samples per collected time point. Tissue was immediately flash frozen in liquid nitrogen and stored at -80° until RNA extraction. Liquid cultures of isolate 0-1 were prepared by incubating collected fungal spores in 75 mL of Fries media (Koladia et al. 2017a) for five days. Tissue was harvested, rinsed, flash frozen in liquid nitrogen, and stored at -80° until RNA extraction. RNA sequencing was done using *in-planta* time points of 48, 72, and 96 hrs post inoculation and a sample from pure culture. Both the pure culture and *in-planta* samples were collected in three replicates. mRNA from each sample was extracted using the mRNA Direct Kit (ThermoFisher Scientific) following the manufacturer's protocol. RNAseq library preparation was done with the Illumina Truseq v.3 kit following the manufacturer's protocol. Quality and fragment size distribution of the prepared libraries were examined using an

Agilent DNA chip on a bioanalyzer (Agilent, Santa Clara, CA). Libraries were sequenced on an Illumina Nextseq at the USDA-ARS Small Grains Genotyping Center (Fargo, ND) to produce 150 bp single-end reads.

The output of the sequencing run was parsed with the open source program bcl2fastq2 (Illumina) and reads were input to FastQC for quality inspection (Andrews et al. 2016). Trimmomatic was used to trim raw reads using the parameters HEADCROP:15, ILLUMINACLIP:2:30:10 with Illumina adapter and index sequences provided, and SLIDINGWINDOW:4:15 (Bolger et al. 2014). Trimmed reads were aligned to the 0-1 genome sequence using HISAT and aligned reads were assembled and analyzed using StringTie following the protocol laid out in Pertea et al. (2016). Briefly, reads were aligned to the genome using HISATv2 with the option "max-intronlen=3000" which is suggested for fungal genomes (Ter-Hovhannisyan et al. 2008). Aligned reads were output in SAM format and converted to sorted BAM files. These BAM files were input into StringTie for assembly of transcripts (Pertea et al. 2016).

Genome annotation

Gene models were determined using the Maker2 pipeline (Holt et al. 2011). The Maker2 pipeline incorporates multiple sources of evidence and leverages these to create the most accurate gene models possible. *Ab initio* annotations were provided through Augustus with the model organisms *Neurospora crassa* training set (Stanke et al. 2006) and Genemark-ES v.2 (Ter-Hovhannisyan et al. 2008) that contains a self-training algorithm. Assembled RNAseq transcripts were also provided to the pipeline as a GFF3 file along with external protein evidence from the closely related species *Pyrenophora tritici-repentis* (Manning et al. 2013) and the current NCBI *P. teres* f. *teres* 0-1 annotation (Ellwood et al. 2010) in fasta format. Options within the Maker2

pipeline that were used in the first iteration of annotation were "est2genome=1" and "protein2genome=1" which instruct the pipeline to use evidence from RNAseq data and BLAST results of supplied proteins to build gene models. These gene models were then used to train the *ab initio* annotation program SNAP (Korf 2004) and the pipeline was re-run with the addition of a SNAP training file specific to the 0-1 genome. The output from this second Maker2 pipeline run was then used to re-train SNAP a second time and subsequently re-run to further refine gene models (Holt et al. 2011). To evaluate the quality of the updated 0-1 assembly gene models, RNAseq transcript coverage was calculated using BEDtools 'coverage' (Quinlan et al. 2010) with Maker2 gene models and aligned transcripts input in BED format. RNAseq evidence for a gene model was defined as having >50% transcript coverage of a gene model.

To evaluate the completeness of the assembly's annotated gene models, the program BUSCO was applied to the updated 0-1 genome assembly and annotation. BUSCO utilizes sets of core genes in taxon-specific databases to evaluate the relative completeness of the assembly and annotation. For the purposes of assessing the 0-1 genome, the BUSCO database for Ascomycota was used as it was the most specific data set in the BUSCO databases relating to *P. teres* f. *teres*. The Ascomycota data set contains 1,315 genes curated from 75 different Ascomycota species (Simão et al. 2015). To compare the completeness of the 0-1 updated assembly, three other closely related species were downloaded from NCBI and evaluated with BUSCO; *Pyrenophora tritici repentis Pt-1C-BFP* (ASM14998v1), *Parastagonospora nodorum SN15* (ASM14691v2), and *Leptosphaeria maculans JN3* (ASM23037v1).

RepeatModeler v1.0.11 (Smit et al. 2015) was used to *de novo* annotate repetitive elements in the genome in order to create a custom *P. teres* f. *teres* repeat library. The RepeatModeler *P. teres* f. *teres* repeat library was input into RepeatMasker (Smit and Hubley

2008-2015) alongside the current release of Repbase (v22.10) (Bao et al. 2015) to soft mask identified repetitive elements and output a final annotation of repetitive elements identified in the newly assembled 0-1 *P. teres* f. *teres* genome. The "buildSummary.pl" RepeatMasker script was applied to gather summary statistics for downstream analysis of repetitive elements.

Secretome, effectorome, GC content structure, and repetitive analysis

Secreted proteins were identified using SignalP v4 (Petersen et al. 2011) and output as mature proteins, lacking the signal sequence. Mature secreted proteins were input into EffectorP v1 (Sperschneider et al. 2016) to identify putative effectors in the updated 0-1 annotation set.

OculterCut v1 (Testa et al. 2016) is a tool used to identify GC content patterns in genomes. OculterCut outputs a number of useful statistics which include the average sizes of GC rich and poor regions as well as average gene densities within each region. OculterCut was run with the updated 0-1 genome input as a FASTA file, the updated 0-1 annotation in GFF3 format, and default run parameters. RIPCAL (Hane and Oliver 2008) was implemented to scan for evidence of repeat induced point mutations (RIP) within the repetitive content of the 0-1 genome. Repeat family sequences of greater than 400 bps of the five most common repeat families were subjected to RIPCAL to determine RIP dominance. The TpA/ApT RIP index and the (CpA+TpG)/(ApC+GpT) RIP index were additionally computed and compared to the RIP indices of a randomly parsed set of sequences from the 0-1 genome. Random sequences of similar size were parsed using the custom Perl script supplied in Debyshire et al. (2017) for a total of 50 sequences.

Whole genome alignment

Contigs from the first assembly of *P. teres* f. *teres* isolate 0-1 (Ellwood et al. 2010) were aligned to the newly assembled reference with the alignment program 'nucmer' within the

MUMmer v3.0 package (Delcher et al. 2003) using the 'mum' option to compute maximal unique matches in the references and query sequences. Alignments were converted to a BED file for downstream analysis. Genome coverage for the 12 reference 0-1 scaffolds were calculated using Bedtools v2.26.0 'coverage' (Quinlan et al. 2010) to output a percent coverage of each scaffold.

Genomic windows consisting of 1000 bps of the reference 0-1 assembly were calculated using Bedtools 'makewindows' (Quinlan et al. 2010) in order to compare coverage statistics relative to repeat regions of the genome. Genomic regions containing low to no coverage were identified as having less than 200 bps of overlap within a 1000 bp genomic region. These low to no coverage regions were then compared to regions of the genome harboring repetitive element using Bedtools v2.26.0 'coverage' (Quinlan et al. 2010). Genomic regions containing low to no coverage and also having 50% of the region covered by a repetitive element were output.

Data availability

Sequence and annotation data are available at NCBI GenBank under BioProject PRJNA392275.

Results and discussion

Sequencing and de novo genome assembly

PacBio SMRT sequencing of *P. teres* f. *teres* isolate 0-1 generated a total of 1,148,507 reads from the 14 SMRT cells sequenced with an average read length of 8,051 bps. A total of 9,246,774,161 bps were obtained equating to approximately 200× coverage of the 0-1 genome.

Assembling this data using the Canu assembler yielded a fairly contiguous assembly with 85 total contigs and one contig representing the mitochondrial genome (86 total contigs). The total size of the assembly was approximately 46.5 Mb with an N50 of 1,730,401 bp and an L50

of 11 contigs. This assembly provides a drastic improvement from the previous assembly based on the quality metrics summarized in **Table 2.1**. The use of long read technology allowed for the assembly of low complexity, repeat dense regions which are difficult to assemble using short read technologies. The increased resolution of the genome will aid in the investigation of evolutionarily active regions that are often repeat rich and harbor important genes related to pathogen adaptation.

Table 2.1. Updated 0-1 assembly summary statistics compared to previous 0-1 assembly (Ellwood et al. 2010).

Feature	Updated 0-1 assembly	Previous 0-1 assembly
Genome size	46,508,966	41,957,260
Total contigs	86	6,412
Largest contig	3,573,185	300,442
Smallest contig	27,932	200
Pre-scaffolding L50[a]	11	408
Pre-scaffolding N50[b]	1,730,401	26,790
Contigs <100Kb	33	6,389
GC%[c]	46	48
Telomeres[d]	9	0

[a] Smallest number of contigs whose length equals 50% of the pre scaffolding genome assembly
[b] Length of the smallest contig in an ordered set of contigs corresponding to 50% of the pre-scaffolding assembly length
[c] Overall GC% content of the 0-1 genome assembly
[d] Number of telomeres identified in the 0-1 assembly on the end of contigs

Genetic mapping and genome scaffolding

Ion Torrent RAD-GBS sequencing generated reads for parental isolates 0-1 and 15A, and the 120 progeny. Reads were aligned to the updated *P. teres* f. *teres* 0-1 genome and a total of 284 SNP markers were identified. SNPs were input into MapDisto v1.7.9 following a filtering

process for genetic mapping. A total of 17 linkage groups (LGs) were obtained with sizes that ranged from 9.69 cM to 159.44 cM and totaling a genetic distance of 987.36 cM.

The 0-1 × 15A genetic map appears to be at low resolution by comparison to the FGOH04 × BB25 genetic map consisting of 16 LGs with 685 SNP markers (370 non-redundant markers) spanning a genetic distance of 1905.81 cM (Koladia *et al*. 2017a). To increase the resolution of the genetic map used for scaffolding the updated 0-1 assembly, the 0-1 × 15A SNPs and the FGOH04 × BB25 SNPs were combined into a single coordinate BED file using ALLMAPS (Tang et al. 2015).

Genome scaffolding using ALLMAPS (Tang et al. 2015) was accomplished with the combined 0-1 × 15A and FGOH04 × BB25 genetic maps (Koladia et al. 2017a) input as a single merged coordinate BED file and the updated 0-1 genome assembly in FASTA format. Scaffolding resulted in 12 scaffolds containing 43 of the 85 contigs in the updated 0-1 assembly and represents 91.8% of the total base pairs in the 0-1 assembly (**Table 2.2**). Marker density across the 12 scaffolds equaled 15.3 markers per Mbp with 652 of the markers uniquely anchored in the scaffolds (**Table 2.2**). Un-scaffolded contigs ranged in size from ~28 Kb to ~367 Kb and represented 8.2% of the 0-1 genome. ALLMAPS scaffolding statistics are summarized in **Table 2.2**.

Table 2.2. ALLMAPS genome scaffolding statistics.

Feature	Scaffolded 0-1 assembly
Markers	652
Markers per Mb	15.3
Scaffolded contigs	43
Scaffolded bases	42,704,288
[a]Un-scaffolded contigs	42
[b]Un-scaffolded bases	3,804,678
[c]Scaffold N50	4,379,536
[d]Scaffold L50	5
Genome % scaffolded	91.8%
Total scaffolds	12

[a] Contigs lacking a marker from either linkage map used when scaffolding
[b] Total bases in the unscaffolded contigs
[c] Smallest number of scaffolds whose length equals 50% of genome assembly
[d] Length of the smallest scaffold in an ordered set of scaffolds corresponding to 50% of the assembly length

Genome annotation and assessment

Genome annotation via the Maker2 pipeline yielded 11,541 genes or pseudogenes (**Table 2.3)**. Evidence for the updated 0-1 gene annotations were derived from either imported protein sequences from the current genome annotation of 0-1 (Ellwood et al. 2010) or from the closely related species *Pyrenophora tritici-repentis* (Manning et al. 2013), as well as *ab initio* annotations. RNAseq evidence was present for 72.9% (8,414) of the gene annotations illustrating the high level of confidence for these gene models. Many of the gene models without RNAseq evidence are likely to be involved in the saprotrophic stage of the *P. teres* f. *teres* life cycle due to the samples being collected only from culture and early *in-planta* time points.

Table 2.3. Gene annotation summary statistics.

Parameter	Value
Genes	11,541
Mean gene length	1,470
Max gene length	43,584
Min gene length	60
Mean exons/gene	3
Predicted secreted proteins[a]	1,002
Predicted effectors[b]	282

[a] Proteins harboring predicted signal sequence via SignalP (Petersen et al. 2011)
[b] Secreted proteins predicted to be effectors via EffectorP (Sperschneider et al. 2015)

BUSCO is a tool for evaluating genome and annotation completeness based on the presence of core genes in curated taxon-specific databases (Simão et al. 2015). For the purposes of assessing the completeness of the updated 0-1 reference genome, BUSCO was run utilizing the Ascomycota core gene set which is comprised of a total of 1,315 core genes. Results from running BUSCO on the updated 0-1 genome resulted in 97.8% of genes being present from the core Ascomycota gene set. Of the genes present relating to the core Ascomycota gene set, 97.7% were complete, 0.8% were fragmented, and 1.6% were missing. These results compare well with four other previously sequenced Dothideomycetes including *Pyrenophora tritici-repentis Pt-1C-BFP* (97.6%), *Parastagonospora nodorum SN15* (97.4%), and *Leptospheria maculans JN3* (97.6%) (**Table 2.4)**.

Table 2.4. BUSCO analysis on Assembly and Annotations.

Species	Isolate	BUSCO Library[a]	Complete[b]	Fragmented[b]	Missing[b]
P. teres	*0-1*	Ascomycota	1285	10	20
P. tritici repentis	*Pt-1C-BFP*	Ascomycota	1283	15	17
P. nodorum	*SN15*	Ascomycota	1280	17	18
L. maculans	*JN3*	Ascomycota	1284	10	21

[a] Busco contains custom curated libraries for different taxa. The Ascomycota library consists of 1,315 genes and was used in this analysis
[b] Complete Busco Ascomycota genes identified
[c] Partially identified Busco Ascomycota genes
[d] Missing Busco Ascomycota genes

Functional analysis and evidence of a two-speed genome.

The host-pathogen interaction is directly modulated by a suite of pathogen secreted proteins known as effectors (Franceschetti et al. 2017). Effectors work to modulate plant cell physiology to facilitate pathogen infection. SignalP v4 (Petersen et al. 2011) predicted 1,002 secretion signals from the 11,541 *P. teres* f. *teres* isolate 0-1 annotated genes (**Table 2.3)**, representing the secretome of 0-1. Mature amino acid sequences (lacking secretion signals) were input into EffectorP v1 (Sperschneider et al. 2016) to further differentiate the effectorome from within the predicted secretome resulting in a total of 282 proteins predicted to be effectors (**Table 2.3**). These predicted effectors are likely to be important in the barley-*P. teres* f. *teres* interaction.

Fungal plant pathogens have been shown to have bipartite compartmentalized genomes comprised of gene rich, repeat sparse regions and gene sparse, repeat rich regions. This genomic architecture represents the 'two-speed genome' model which has been observed in a number of plant pathogens (Dong et al. 2015; Faino et al. 2016; Laurent et al. 2017; Rouxel et al. 2017). Genes in close proximity to repeat-rich and gene sparse regions have been shown to undergo higher rates of positive selection, indicating these compartments are evolutionarily active (Raffaele et al. 2010). A fungal genome defense mechanism against duplication events known as repeat-induced point mutation (RIP) may be a contributing factor in the development of AT-rich genome regions. Evolution rates of genes in close proximity to AT-rich regions could be increased through the aid of RIP (Lo Presti et al. 2015; Testa et al. 2016). The program OcculterCut v1 (Testa et al. 2016) was used to examine the overall GC content of the genome. The output of this analysis resulted in two distinct genome categories, one with high GC content (41-100%) and the other with low GC content (0-41%) (**Table 2.5**; **Figure 2.1**). The high GC

content portion constituted ~75% of the genomic content with an average gene density of 306 genes/Mb in contrast to the low GC content region which constitutes 24.9% of the genome with an average gene density of 72.6 genes/Mb (**Table 2.5**).This clear genomic segmentation is a classic representation of the two-speed fungal genome often seen in plant pathogens (Dong et al. 2015; Thomma et al. 2016; Testa et al. 2016).

Table 2.5. OculterCut analysis of *P. teres.* f. *teres*

Feature	High GC content (>41-100%)	Low GC content (0-41%)
Peak GC content[a]	51.6%	32.2%
Percentage of genome[b]	75.1%	24.9%
Average region length (Kb)[c]	58.3	18.3
Number of genes in regions[d]	10,501	797
Gene density (genes/Mb)[e]	306	72.6

[a] Peak GC content in the elevated GC regions of the genome

[b] Proportion of the genome belonging containing a relatively high or low GC content

[c] Average length of identified GC rich and GC poor regions in the genome

[d] Number of genes residing in GC rich and GC poor regions of the genome

[e] Density of annotated genes within the GC rich and GC poor regions of the genome

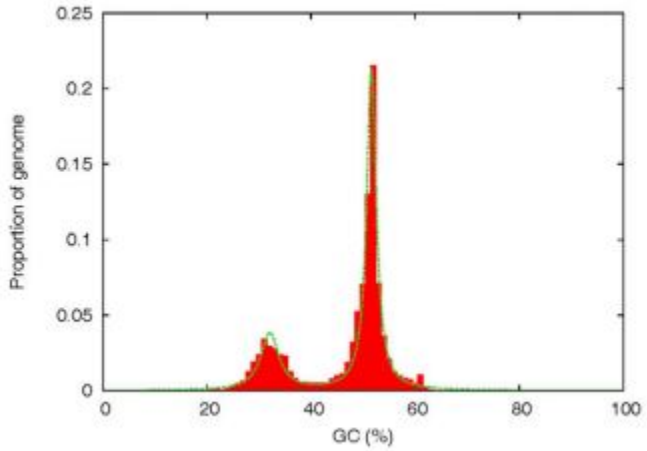

Figure 2.1. OculterCut v1 GC% plot of the *Pyrenophora teres* f. *teres* 0-1 genome. A bimodal distribution is observed in which genome segments fall into the high GC% category (41-100%) or the low GC% category (0-40%).

The OculterCut analysis supports the repetitive analysis output through the RepeatModeler/RepeatMasker pipeline. This repetitive analysis identified 26.7% of the 0-1

genome as being interspersed repeat elements and an additional 5.0% of simple repeats. This equates to roughly 32% of the genome being comprised of repetitive elements, a greater number compared to the closely related species *Pyrenophora tritici-repentis* (16.7%) and *Parastagonospora nodorum* (4.52%) (Manning et al. 2013; Syme et al. 2016). The most numerous repetitive element classes annotated were the LTR-Gypsy elements comprising 9.28% of the genome and DNA-TcMar-Fot1 elements at 5.38% of the genome with an additional 7.81% of the genome belonging to unclassified transposable elements.

Repeat induced point mutation (RIP) is a genomic defense mechanism against transposons that has been identified in a number of fungal species (Singer et al. 1995; Dean et al. 2005; Idnurm et al. 2003; Manning et al. 2013; Syme et al. 2016). RIP involves C:G nucleotide transitions to T:A nucleotides and affects sequences with ~80% identity over at least 400 bps in length creating a bias towards TpA dinucleotides over CpA dinucleotides in RIP affected areas (Hane and Oliver, 2008). RIPCAL RIP indices were calculated for the top five annotated repeat families and compared to a set of randomly extracted DNA sequences of the same size range of the 0-1 genome. RIP indices that show evidence of RIP were defined as values of TpA/ApT > 2.0 and/or (CpA+TpG)/(ApC+GpT) < 0.7 (Galagar et al. 2003). Using this criteria, all five repeat families show evidence of RIP with (CpA+TpG)/(ApC+GpT) < 0.7 but none of the five repeat families show evidence of RIP with TpA/ApT > 2.0. RIPCAL alignment "degenerative consensus" analysis of the five repeat families indicated a RIP dominance of CpT to TpA transitions (TpG to TpA in the reverse complement) and CpT to TpT transitions (TpG to TpT in the reverse complement). Given the indication of RIP affected sequences from only one of the common indices and the high degree of homology observed between members of the five repeat families examined it would seem that RIP is not an efficient process in the *P. teres* f. *teres* isolate

0-1 genome. This is further supported by the increased repetitive content of the 0-1 genome relative to closely related species and reflects similar results observed in *P. tritici-repentis* (Manning et al. 2013), which concluded that if RIP is functional the efficiency is low.

Whole genome alignment

Using a combination of MUMmer v3.0 (Delcher et al. 2003) and Bedtools v2.26.0 (Quinlan et al. 2010) whole genome alignments were calculated and compared between the first draft assembly of *P. teres* f. *teres* isolate 0-1 and the newly assembled reference genome of 0-1. Alignments between the first draft assembly and the 12 reference scaffolds resulted in coverages ranging from 60.5%-80.1% for the 12 reference scaffolds equating to an average of 27.2% missing sequence between the first draft assembly and the current 12 reference scaffolds. This amount of missing data correlates well with the amount of repetitive elements detected in the genome (32%) and in fact 67.1% of the 27.2% missing sequence of the first draft assembly contains an annotated repeat element in the new reference genome of isolate 0-1 (**Figure 2.2**).

The correlation between missing sequence and repetitive elements adds support to previous observations that short read technologies struggle to span low complexity regions (Ekblom and Wolf 2014) and highlights the usefulness of long read technologies such as PacBio (Goodwin et al. 2016). With regards to host-pathogen interactions, long read technologies will aid in understanding effector genes that can be difficult to identify as they are known to associate with low complexity repeat regions (Thomma et al. 2016). Long read technologies present the best method for sequencing and assembling the genomes of fungal plant pathogen species with the goal of understanding the host-pathogen interaction.

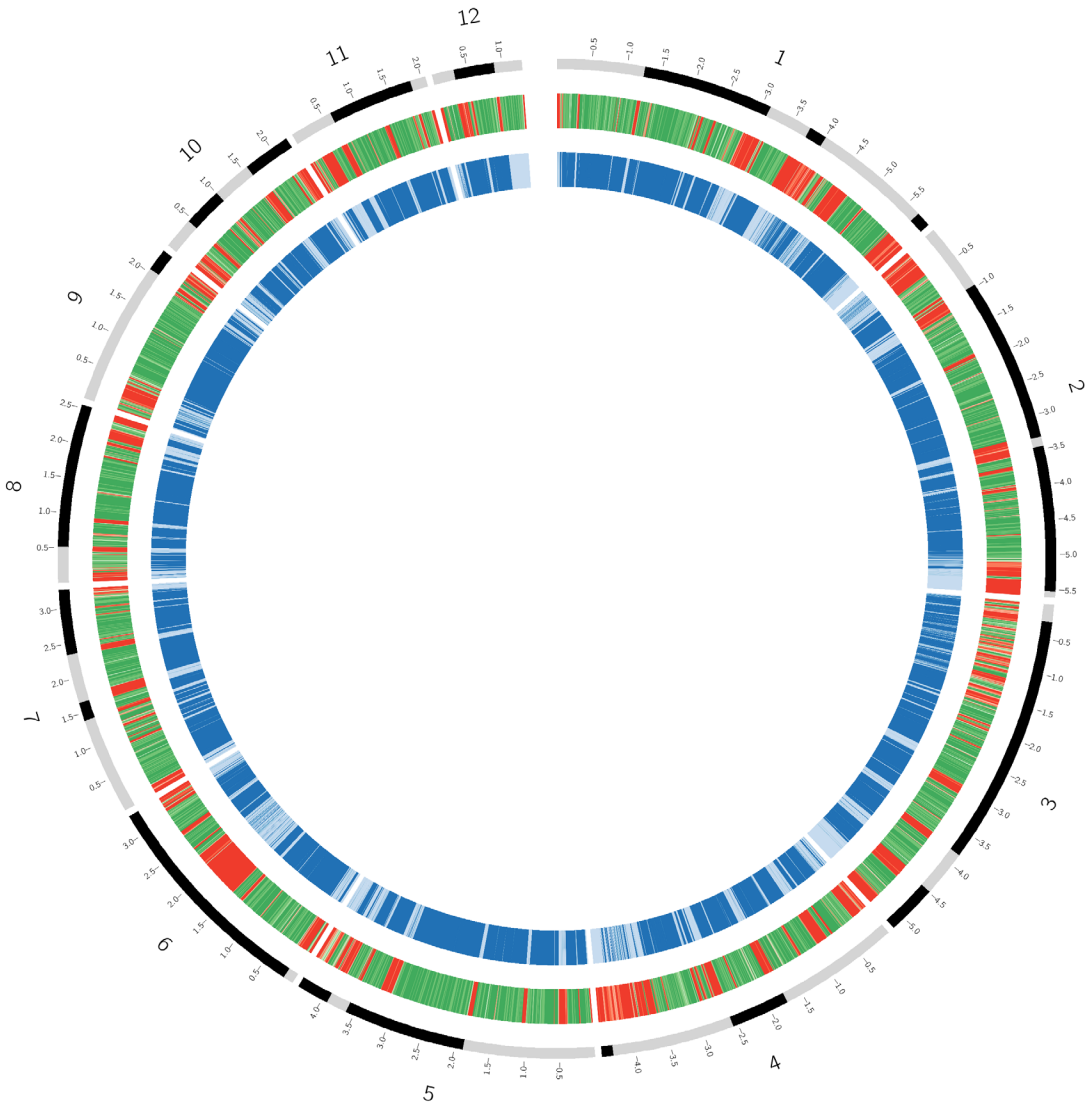

Figure 2.2. Circos plot depicting the relationship between repetitive elements in the 12 current 0-1 scaffolds and the first draft assembly of 0-1 (Ellwood et al. 2010). (A) The outer-most ring represents the twelve 0-1 genomic scaffolds. Large numbers represent scaffold (chromosome) numbers ordered from largest (1) to smallest (12). Small numbers represent size in megabases. Alternating light and dark color patterns within each scaffold represent 0-1 scaffolded contigs (B) A heatmap depicting the density of repetitive elements along the twelve 0-1 scaffolds. Red represents higher repetitive content and green represents less repetitive content. (C) Coverage of the first draft genome assembly of *P. teres* f. *teres* isolate 0-1 (Ellwood et al. 2010) aligned to the 12 current 0-1 scaffolds. Dark blue represents regions of alignment and light blue represents regions of sequence absent in the first draft assembly.

Conclusion

Here, we present an updated assembly and annotation of the barley pathogenic fungus *P. teres* f. *teres* reference isolate 0-1. This improved assembly and annotation provides a higher resolution assembly of the *P. teres* f. *teres* 0-1 reference genome and annotation that now includes a large proportion of the repetitive content within the genome that have been shown to be of evolutionary importance to plant pathogens (Dong et al. 2015). This data set will be particularly useful in investigating effector genes that have been reported to reside in and proximal to evolutionarily active repetitive regions of the genome.

Acknowledgements

Research supported by funding from The North Dakota Barley Council. Authors thank Danielle Holmes for technical assistance and expertise and Dr. Zhaohui Liu for comments on the manuscript. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

Literature cited

Abu Qamar M., Z. H. Liu, J. D. Faris, S. Chao, M. C. Edwards *et al.*, 2008 A region of barley chromosome 6H harbors multiple major genes associated with net type net blotch resistance Theoret. Appl. Genet. 117:1261-1270.

Andrews, S., 2011 FastQC: a quality control tool for high throughput sequence data. *Cambridge, UK: Babraham Institute*.

Bao, W., K. K. Kojima, O. Kohany, 2015 Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob. DNA, 2015;6:11

Bolger, A. M., M. Lohse and B. Usadel, 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), pp.2114-2120.

Carlsen, S. A., A. Neupane, N. A. Wyatt, J. K. Richards, J. D. Faris *et al*., 2017 Characterizing the Pyrenophora teres f. maculata–barley interaction using pathogen genetics. *G3: Genes, Genomes, Genetics*, *7*(8), pp.2615-2626.

Chisholm, S. T., G. Coaker, B. Day and B. J. Staskawicz, 2006 Host-microbe interactions: shaping the evolution of the plant immune response. *Cell*, *124*(4), pp.803-814.

Dong, S., S. Raffaele, and S. Kamoun, 2015 The two-speed genomes of filamentous pathogens: waltz with plants. *Current opinion in genetics & development*, *35*, pp.57-6

Dean, R. A., N. J. Talbot, D. J. Ebbole, M. L. Farman, T. K. Mitchell *et al.* 2005 The genome sequence of the rice blast fungus Magnaporthe grisea. *Nature*, *434*(7036), pp.980-986.

Derbyshire, M., M. Denton-Giles, D. Hegedus, S. Seifbarghy, J. Rollins *et al.* 2017 The complete genome sequence of the phytopathogenic fungus Sclerotinia sclerotiorum reveals insights into the genome architecture of broad host range pathogens. *Genome biology and evolution*, *9*(3), pp.593-618.

Ekblom, R. and J. B. Wolf, 2014 A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary applications*, *7*(9), pp.1026-1042.

Delcher, A. L., S. L. Salzberg and A. M. Phillippy, 2003 Using MUMmer to identify similar regions in large sequence sets. *Current Protocols in Bioinformatics*, pp.10-3.

Ellwood, S. R., Z. Liu, R. A. Syme, Z. Lai, J. K. Hane *et al.,* 2010 A first genome assembly of the barley fungal pathogen Pyrenophora teres f. teres. *Genome biology*, *11*(11), p.R109.

Faino, L., M. F. Seidl, E. Datema, G. C. van den Berg, A. Janssen *et al.,* 2015 Single-molecule real-time sequencing combined with optical mapping yields completely finished fungal genome. *MBio*, *6*(4), pp.e00936-15.

Faino, L., M. F. Seidl, X. Shi-Kunne, M. Pauper, G. C. van den Berg *et al.*, 2016 Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. *Genome research*, *26*(8), pp.1091-1100.

Franceschetti, M., A. Maqbool, M. J. Jiménez-Dalmaroni, H. G. Pennington, S. Kamoun *et al.*, 2017 Effectors of filamentous plant pathogens: commonalities amid diversity. *Microbiology and Molecular Biology Reviews*, *81*(2), pp.e00066-16.

Galagan, J. E., S. E. Calvo, K. A. Borkovich, E. U. Selker, N. D. Read *et al.*, 2003 The genome sequence of the filamentous fungus Neurospora crassa. *Nature*, *422*(6934), pp.859-868.

Goodwin, S., J. D. McPherson, and W. R. McCombie, 2016 Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, *17*(6), pp.333-351.

Hane, J. K. and R. P. Oliver, 2008 RIPCAL: a tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences. *BMC bioinformatics*, *9*(1), p.478.

Holt, C. and M. Yandell, 2011 MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC bioinformatics*, *12*(1), p.491.

Idnurm, A. and B. J. Howlett, 2003 Analysis of loss of pathogenicity mutants reveals that repeat-induced point mutations can occur in the Dothideomycete Leptosphaeria maculans. *Fungal Genetics and Biology*, *39*(1), pp.31-37.

Jones, J. D. and J. L. Dangl, 2006 The plant immune system. *Nature*, *444*(7117), pp.323-329.

Koladia, V. M., J. K. Richards, N. A. Wyatt, J. D. Faris, R. S. Brueggeman and T. L. Friesen, 2017a Genetic analysis of virulence in the *Pyrenophora teres* f. *teres* population BB25× FGOH04Ptt-21. *Fungal Genetics and Biology*.

Koladia, V. M., Faris, J. D., Richards, J.K., Brueggeman, R. S., Chao, S., and Friesen, T. L. 2017b Genetic analysis of net form net blotch resistance in barley lines CIho5791 and Tifang against a global collection of *P. teres* f. *teres* isolates. Theor. Appl. Genet. 130:163-173.

Koren, S., B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman *et al.,* 2017 Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. bioRxiv, p.071282.

Korf, I., 2004 Gene finding in novel genomes. *BMC bioinformatics*, *5*(1), p.59.

Lai, Z., J. D. Faris, J. J. Weiland, B. J. Steffenson and T. L. Friesen, 2007 Genetic mapping of Pyrenophora teres f. teres genes conferring avirulence on barley. *Fungal Genetics and Biology*, *44*(5), pp.323-329.

Laurent, B., C. Palaiokostas, C. Spataro, M. Moinard, E. Zehraoui *et al.*, 2017 High-resolution mapping of the recombination landscape of the phytopathogen Fusarium graminearum suggests two-speed genome evolution. *Molecular plant pathology*.

Liu, Z., S. R. Ellwood, R. P. Oliver, and T. L. Friesen, 2011 Pyrenophora teres: profile of an increasingly damaging barley pathogen. Molecular Plant Pathology, *12*(1), pp.1-19.

Liu, Z., D. J. Holmes, J. D. Faris, S. Chao, R. S. Brueggeman *et al.*, 2015 Necrotrophic effector triggered susceptibility (NETS) underlies the barley-Pyrenophora teres f. teres interaction specific to chromosome 6H. Mol. Plant Pathol. 16, 188-200

Liu, Z. H., J. D. Faris, M. C. Edwards, T. L. Friesen, 2010. Development of expressed sequence tags (EST)-based markers for genomic analysis of a barley 6H region harboring multiple net form net blotch resistance genes. Plant Genome 3:41-52.

Lo Presti, L., D. Lanver, G. Schweizer, S. Tanaka, L. Liang *et al.*, 2015 Fungal effectors and plant susceptibility. *Annual review of plant biology*, *66*, pp.513-545.

Lorieux, M., 2012 MapDisto: fast and efficient computation of genetic linkage maps. *Molecular Breeding*, *30*(2), pp.1231-1235.

Manning, V. A., I. Pandelova, B. Dhillon, L. J. Wilhelm, S. B. Goodwin *et al.*, 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*(1), pp.41-63.

Mathre, D. E., G. D. Kushnak, J. M. Martin, W. E. Grey and R. H. Johnston, 1997 Effect of residue management on barley production in the presence of Net Blotch Disease. *Journal of production agriculture*, *10*(2), pp.323-326.

Pertea, M., D. Kim, G. M. Pertea, J. T. Leek and S. L. Salzberg, 2016 Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols*, *11*(9), pp.1650-1667.

Petersen, T. N., S. Brunak, G. von Heijne and H. Nielsen, 2011 SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods*, *8*(10), pp.785-786.

Quinlan, A. R. and I. M. Hall, 2010 BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26(6): 841-842

Raffaele, S., R. A. Farrer, L. M. Cano, D. J. Studholme, D. MacLean*et al.,* 2010 Genome Evolution Following Host Jumps in the Irish Potato Famine Pathogen Lineage. Science. 330: 1540-1543

Rouxel, T. and M.H. Balesdent, 2017 Life, death and rebirth of avirulence effectors in a fungal pathogen of Brassica crops, Leptosphaeria maculans. New Phytologist, 214(2), pp.526-532.

Richards, J. K., T. L. Friesen, and R. S. Brueggeman, 2017 Association mapping utilizing diverse barley lines reveals net form net blotch seedling resistance/susceptibility loci. Theoret. Appl. Genet. 130:915-927.

Shjerve, R. A., J. D. Faris, R. S. Brueggeman, C. Yan, Y. Zhu *et al.,* 2014 Evaluation of a Pyrenophora teres f. teres mapping population reveals multiple independent interactions with a region of barley chromosome 6H. *Fungal genetics and biology*, *70*, pp.104-112.

Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva and E. M. Zdobnov, 2015 BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, p.btv351.

Singer, M. J. and E. U. Selker, 1995 Genetic and epigenetic inactivation of repetitive sequences in Neurospora crassa: RIP, DNA methylation, and quelling. In *Gene silencing in higher plants and related phenomena in other eukaryotes* (pp. 165-177). Springer Berlin Heidelberg.

Smit, A. F. A., R. Hubley, *RepeatModeler Open-1.0*. 2008-2015 http://www.repeatmasker.org

Smit, A. F. A., R. Hubley, P. Green, *RepeatMasker Open-4.0*. 2013-2015 http://www.repeatmasker.org.

Sperschneider, J., D. M. Gardiner, P. N. Dodds, F. Tini, L. Covarelli *et al.*, 2016 EffectorP: predicting fungal effector proteins from secretomes using machine learning. *New Phytologist*, *210*(2), pp.743-761.

Stanke, M., O. Keller, I. Gunduz, A. Hayes, S. Waack *et al*., 2006 AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic acids research*, *33*(suppl_2), pp.W465-W467.

Syme, R. A., K. C. Tan, J. K. Hane, K. Dodhia, T. Stoll *et al*., 2016 Comprehensive annotation of the Parastagonospora nodorum reference genome using next-generation genomics, transcriptomics and proteogenomics. *PloS one*, *11*(2), p.e0147221.

Tang, H., X. Zhang, C. Miao, J. Zhang, R. Ming *et al*., 2015. ALLMAPS: robust scaffold ordering based on multiple maps. *Genome biology*, *16*(1), p.3.

Ter-Hovhannisyan, V., A. Lomsadze, Y. O. Chernoff and M. Borodovsky, 2008 Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome research, 18(12), pp.1979-1990.

Testa, A. C., R. P. Oliver and J. K. Hane, 2016 OculterCut: a comprehensive survey of AT-rich regions in fungal genomes. *Genome biology and evolution*, *8*(6), pp.2044-2064.

Thomma, B. P., M. F. Seidl, X. Shi-Kunne, D. E. Cook, M. D. Bolton *et al.,* 2016 Mind the gap; seven reasons to close fragmented genome assemblies. *Fungal Genetics and Biology*, *90*, pp.24-30.

Van Kan, J. A., J. H. Stassen, A. Mosbach, T. A. Van Der Lee, L. Faino *et al*., 2017 A gapless genome sequence of the fungus Botrytis cinerea. Molecular plant pathology, 18(1), pp.75-89.

Walker, B. J., T. Abeel, T. Shea, M. Priest, A. Abouelliel *et al.,* 2014 Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS one*, *9*(11), p.e112963.

Weiland, J. J., B. J. Steffenson, R. D. Cartwright and R. K. Webster, 1999 Identification of molecular genetic markers in Pyrenophora teres f. teres associated with low virulence on ‘Harbin’barley. *Phytopathology*, *89*(2), pp.176-181.

CHAPTER 3. A COMPARATIVE ANALYSIS OF THE BARLEY PATHOGEN *PYRENOPHORA TERES* F. *TERES* IDENTIFIES SUB-TELOMERIC REGIONS AS DRIVERS OF VIRULENCE[2]

Abstract

Pyrenophora teres f. *teres* causes net form net blotch of barley and is an economically important pathogen throughout the world. However, *P. teres* f. *teres* is lacking in the genomic resources necessary to characterize the mechanisms of virulence. Recently a high quality reference genome was generated for *P. teres* f. *teres* isolate 0-1. Here, we present the reference quality sequence and annotation of four new isolates and we use the five available *P. teres* f. *teres* genomes for an in-depth comparison resulting in the generation of hypotheses pertaining to the potential mechanisms and evolution of virulence. Comparative analyses were performed between all five *P. teres* f. *teres* genomes examining genomic organization, structural variations, and core and accessory genomic content, specifically focusing on the genomic characterization of known virulence loci and the localization of genes predicted to encode secreted and effector proteins. We showed that 14 of 15 currently published virulence quantitative trait loci (QTL) span accessory genomic regions consistent with these accessory regions being important drivers of host adaptation. Additionally, these accessory genomic regions were frequently found in sub-telomeric regions of chromosomes with 10 of the 14 accessory region QTL localizing to sub-telomeric regions. Comparative analysis of the sub-telomeric regions of *P. teres* f. *teres*

[2] The material in this chapter was co-authored by Nathan A. Wyatt, Jonathan K. Richards, Robert S. Brueggeman, and Timothy L. Friesen. Nathan A. Wyatt had primary responsibility for laboratory experiements, computational analysis, and preparation of the manuscript. Jonathan K. Richards provided valuable consultation of computational analysis and proofread and revised the manuscript. Robert S. Brueggeman proofread and revised the manuscript. Timothy L. Friesen advised Nathan A. Wyatt during laboratory experiments and computational analysis and drafted, proofread, and revised the manuscript. This chapter has been published as Wyatt, N. A., Richards, J., Brueggeman, R. S., & Friesen, T. L. (2019). A comparative genomic analysis of the barley pathogen *Pyrenophora teres* f. *teres* identifies sub-telomeric regions as drivers of virulence. Molecular plant-microbe interactions: MPMI.

chromosomes revealed translocation events where homology was detected between non-homologous chromosomes at a significantly higher rate than the rest of the genome. These results indicate that the sub-telomeric accessory genomic compartments not only harbor most of the known virulence loci, but also that these regions have the capacity to rapidly evolve.

Introduction

The intimate process of co-evolution between plants and their pathogens is often depicted as a form of trench warfare where pathogens evolve to manipulate the host, leading to the acquisition of nutrients and in response, selection pressure is placed on the plant to develop mechanisms to resist the pathogen (Stahl and Bishop 2000). Plants have evolved an innate immune system that recognizes and responds to microbes that successful pathogens must overcome (Cook et al. 2015; Jones and Dangle 2006). This physical interaction of a pathogen with its host often occurs through a set of small secreted proteins termed “effectors”.

Pathogen effectors have evolved to manipulate the host defense response to the benefit of the pathogen, but these effectors also provide targets for plant recognition of the pathogen (Bialas et al. 2017). A pathogen’s effector function is specific to the pathogen’s lifestyle and mode of nutrient acquisition. Obligate biotrophic pathogens typically invade host tissue in a manner designed to evade host recognition to gain nutrient from living tissue. Obligate biotrophs often develop haustoria that are appendages of fungal hyphae used to invaginate plant cells to serve as feeding structures and effector secretion hot spots (Perfect et al. 2001). Conversely, other pathogens, often classified as necrotrophs, can gain nutrient from dead or dying tissue resulting from pathogen induced programmed cell death of plant cells and have evolved mechanisms to deal with the various aspects of the plant defense response (Liu et al. 2012b; Liu et al. 2016). Between these two models is a spectrum of fungi often classified as hembiotrophic

that begin plant colonization in a manner similar to a biotrophic or endophytic interaction, referred to as the biotrophic or symptomless phase, where evasion of plant recognition is important. The symptomless or biotrophic phase is followed by a transition to the necrotrophic phase, characterized by inducing host cell death, colonization, and eventual sporulation (Vleeshouwers et al. 2014).

Effector discovery is hindered by a lack of homology to known proteins and the association of genes that encode effectors with low complexity repeat-rich genomic compartments that are drivers of rapid evolutionary adaptation (Raffaele and Kamoun 2012; Dong et al. 2015; Faino et al. 2016). Pathogens must be able to quickly modify or lose effectors that are co-opted by the host, resulting in effector genes that can be highly polymorphic between members of the same species (Cook et al. 2015). Low complexity regions, that are rich in repeats and harbor transposable elements (TEs), are thought to be dynamic compartments involved in rapid evolution as compared to the more conserved stable gene-dense compartments. Pathogen effector genes often reside in these low complexity regions and consequently are affected by increased structural polymorphism, point mutagenesis, and diversifying selection that are common to these genomic compartments (Raffaele 2010; Rouxel 2011; de Jonge et al. 2013; Dong et al. 2015; Faino et al. 2016, Moller and Stukenbrock 2017).

Generally, in fungi, the lack of available sequence data has been attributed to the difficulties and costs of sequencing and assembling genomes (Thomma et al. 2016). The Pacific Biosciences (PacBio) single molecule real-time (SMRT) sequencing platform is capable of read lengths of up to 60 Kb allowing for the sequencing and assembly of full microbial genomes containing large repetitive elements (Goodwin et al. 2016; Thomma et al. 2016). In recent years, SMRT sequencing has been used to obtain complete genome sequences of a number of fungal

plant pathogen species, demonstrating its utility for fungal plant pathogen genomics (Faino et al. 2015; Thomma et al. 2016; van Kan et al. 2016, Derbyshire et al. 2017; Wyatt et al. 2018; Richards et al. 2018). Finishing fungal genomes opens the door to identifying all of the genes in the genome including previously difficult to identify effector genes residing in low complexity regions. When multiple isolates of the same species can be fully sequenced, the full gene repertoires of the species can be compared and characterized, this is termed the pan-genome. (Hurgobin & Edwards 2017; Tettelin et al. 2005, Vernikos et al. 2015). This pan-genome defines the entire genomic repertoire of a given taxonomic clade and accounts for all possible lifestyle processes of the organisms being observed. Pan-genomes are further defined by their core genomes, representing genes common to all individuals, and their accessory genomes, representing genes that are unique to individuals within the clade. Core genomes often encode proteins associated with basic biological aspects of the entire clade, whereas accessory genomes are thought to encode genes involved in niche adaptation. The intra-specific gene content of the accessory genome is of particular interest to plant-microbe interactions given that the genes responsible for virulence are often not shared among all individuals of a species (Croll et al. 2012). With the decreased cost of sequencing technology, it has become possible to sequence many individuals of a species to compare intraspecific gene content and structural variation. These comparative studies have been done for a number of bacterial species as well as some well-studied crop species. Most recently, pan-genome analyses have been conducted in the fungal plant pathogen species *Zymoseptoria tritici* (Plissonneau et al. 2018)*, Pyrenophora tritici-repentis* (Moolhuijzen et al. 2018)*,* and *Parastagonospora nodorum* (Syme et al. 2018a).

Net form net blotch (NFNB) is a stubble born foliar disease of barley (*Hordeum vulgare*) induced by the fungal pathogen *Pyrenophora teres* f. *teres*. Typical disease losses due to NFNB

have ranged between 10 and 40% with the potential for complete yield loss given environmental conditions favorable to the pathogen, namely, wide planting of a susceptible cultivar, and high humidity or rainfall (Mathre 1997). *P. teres* f. *teres* infection on a susceptible host results in dot-like lesions on the leaf, progressing to longitudinal striations and finally the net-like pattern from which the disease gets its name. Necrotrophic effectors (NEs) were shown to be an important component of the *P. teres* f. *teres* infection cycle (Liu et al. 2015), similar to the closely related species *P. nodorum* and *P. tritici-repentis* that also employ NEs to induce NE triggered susceptibility (NETS) (Faris et al. 2013; Friesen and Faris 2010). In addition to NETS, dominant resistance has also been identified in a number of barley backgrounds that follow a gene-for-gene model. The presence of both dominant resistance and dominant susceptibility and the discovery of NEs illustrates the complicated interaction at play (Liu et al. 2011; Liu et al. 2012a; Liu et al. 2015; Koladia et al. 2017). As of yet, no effector genes have been cloned and characterized in *P. teres* f. *teres* and therefore the mechanism of virulence/pathogenicity is still poorly understood.

Adding to the complexity of *P. teres* f. *teres* virulence is the diversity observed in both local and global pathogen populations. Host genotype specificity was first observed in a set of Australian *P. teres* f. *teres* isolates (Khan and Boyd 1969) and bi-parental pathogen mapping populations have been an important tool used to associate genomic loci of *P. teres* f. *teres* isolates with host specific virulence and avirulence (Weiland et al. *1999;* Lai et al. 2007; Beattie et al. 2007; Liu et al. 2011; Shjerve et al. 2014; Koladia et al. 2017). Currently published mapping populations have been developed from crosses of *P. teres* f. *teres* isolates 0-1 × 15A, isolates 15A × 6A, and isolates FGOH04Ptt-21 × BB25 and report on a total of 15 unique genetic loci contributing to virulence leading to NFNB disease on different barley lines.

The diversity observed in *P*. *teres* f. *teres* bi-parental mapping studies presents an obstacle for effector discovery in *P*. *teres* f. *teres* isolates that have virulence associations not present in the currently published reference isolate 0-1 (Ellwood et al. 2010; Wyatt et al. 2018). With each new bi-parental mapping study published, unique QTL have been identified on common cultivars located in different parts of the *P. teres* f. *teres* genome with very few QTL being identified in all bi-parental mapping studies (Weiland *et al*. 1999; Lai *et al*. 2007; Shjerve *et al*. 2014; Koladia *et al*. 2017). Given the diversity observed in previous mapping studies of avirulence/virulence, it is unlikely that a single isolate would capture a representative sample of the effectorome of *P*. *teres* f. *teres*. To examine a broader sampling of the *P*. *teres* f. *teres* effectorome, we used a pan-genome approach by sequencing and assembling four additional isolates using PacBio SMRT sequencing and annotated their genomes using RNAseq support following the same protocol outlined in Wyatt et al. (2018). Isolates sequenced have all been used previously in biparental mapping studies including the two California, USA, isolates 15A (Weiland et al. 1999; Lai et al. 2007; Shjerve et al. 2014) and 6A (Shjerve et al. 2014); the North Dakota, USA, isolate FGOH04Ptt-21 (Koladia et al. 2017), and the Danish isolate BB25 (Koladia et al. 2017). Using the reference isolate 0-1 (Lai et al. 2007; Ellwood et al. 2010; Wyatt et al. 2018) and the four newly sequenced isolates, a pan genome analyses was performed on the five isolates to evaluate the diversity in genomic architecture and gene content. The genome annotations produced were used to identify chromosome structural variation and parse the effectorome of each isolate for comparison of effector diversity within *P*. *teres* f. *teres* isolates to examine common and unique effectors of the species.

Materials and methods

Biological materials

Five *P. teres* f. *teres* isolates were used in this study, each of which has been included as a parent in a previous virulence mapping study. Isolate 0-1 is the currently published reference genome (Ellwood et al. 2010; Wyatt et al. 2018) and is a Canadian isolate collected in Ontario, Canada (Weiland et al. 1999). Isolates 15A (Steffenson and Webster 1992) and 6A (Wu et al. 2003) were collected in California, USA. Isolate FGOH04Ptt-21 was collected in Fargo, North Dakota, USA (Koladia et al. 2017) and isolate BB25 is an isolate collected in Denmark (kindly provided by Lise Nistrup Jorgensen). These isolates were chosen as a result of their inclusion in previous genetic studies as they differ in virulence and avirulence across barley lines commonly used in differential screening sets. Isolates 15A and 0-1 were first used in a bi-parental mapping study due to their differential reactions on Harbin barley (Weiland et al. 1999) and were later used in a bi-parental mapping study examining the differential reactions of these two isolates on the barley lines Tifang, and Prato (Lai et al. 2007). Isolate 15A was also crossed to isolate 6A to generate a bi-parental mapping population due to their different disease reactions on Kombar and Rika barley, with isolate 15A being virulent on Kombar but avirulent on Rika and isolate 6A being avirulent on Kombar but virulent on Rika (Sherve et al. 2014). Isolates FGOH04Ptt-21 and BB25 were crossed to create a bi-parental mapping population based on the isolates differential disease responses (Koladia et al. 2017) on barley lines Manchurian, Tifang, CI4922, Beecher, Celebration, Pinnacle, Hector, and Stellar (Koladia et al. 2017).

Fungal tissue for DNA extraction was grown for *P. teres* f. *teres* isolates 15A, 6A, FGOH04Ptt-21, and BB25 as described in Koladia et al. 2017 (Appendix). High molecular

weight DNA extractions were completed following the methods outlined in Richards et al. 2018 (Appendix).

Genome sequencing, assembly, and scaffolding

PacBio 20 kb size-selected sequencing libraries were prepared for all isolates and then sequenced on a PacBio RSII instrument at the Mayo Clinic Molecular Biology Core (Rochester, MN). A total of nine SMRT cells were sequenced for each of the four *P. teres* f. *teres* isolates.

Genome assemblies of the four *P. teres* f. *teres* isolates 15A, 6A, FGOH04Ptt-21, and BB25 were assembled in a similar manner to 0-1 (Wyatt et al. 2018). The Canu v1.5 assembler (Koren et al. 2017) was used with raw reads input in FASTQ format for correction, trimming, and assembly. The option 'genomeSize=46.5m' was used based on the *P. teres* f. *teres* reference isolate 0-1 genome size (Wyatt et al. 2018). Genome polishing of each assembly was done with Pilon v1.21 (Walker et al. 2014) using each isolate's genome and each isolate's sequencing reads aligned to the genome in BAM file format. Contigs flagged as potential repeats or flagged as potential non-unique contigs by Canu were manually inspected and excluded.

The genetic linkage map of the bi-parental mapping population of FGOH04Ptt-21 × BB25 (Koladia et al. 2017) was used to scaffold the *P. teres* f. *teres* isolates 15A, 6A, FGOH04Ptt-21, and BB25 with the program ALLMAPS (Tang et al. 2015). ALLMAPS takes assembled contigs and generates scaffolds based on coordinates from genetic linkage maps, optical maps, or syntenic maps. Contigs were ordered according to marker order on linkage groups with 100 'N' nucleotides inserted to represent gaps. Assembled contigs were quality checked by mapping reads back to the assembly in order to assess collapsed or expanded repeat regions and potentially misassembled regions by manual inspection of read coverage. Contig collinearity with the linkage map was uniform with the exception of isolate BB25 that contained

a chromosome fusion that spanned two linkage groups. Reads were mapped back to the BB25 chromosome fusion in order to assess read depth coverage across the fusion point of chromosomes 1 and 2 for confirmation of the chromosome fusion.

Genome assemblies for *P. teres* f. *teres* isolates 15A, 6A, FGOH04Ptt-21, and BB25 were submitted to NCBI Genbank and can be found under accession numbers VBVL00000000 (15A), VFEN00000000 (6A), VBVN00000000 (FGOH04Ptt-21), and VBVM00000000 (BB25). The reference genome for *P. teres* f. *teres* isolate 0-1 was previously submitted to NCBI Genbank and can be found under accession number NPOS00000000.

RNA sequencing and genome annotation

RNA sequencing was done using *in-planta* time points of 48 h, 72 h, and 96 h post inoculation and a sample from liquid culture. Liquid culture and *in-planta* samples were collected in three replicates. Libraries were sequenced on an Illumina Nextseq at the USDA-ARS Small Grains Genotyping Center (Fargo, ND) to produce 150 bp single-end reads. Illumina FASTQ files were quality checked and trimmed to remove adapters and low quality sequence (APPENDIX). Trimmed reads were aligned and assembled into transcripts following the protocol described in Pertea et al. (2016). For each of the four *P. teres* f. *teres* isolates including 15A, 6A, FGOH04Ptt-21, and BB25, RNAseq reads were aligned to each previously assembled genome (APPENDIX).

Genome annotations were compiled using the Maker2 pipeline (Holt et al. 2011) using the same annotation process used in the generation of the 0-1 reference annotation set (Wyatt et al. 2018) (Supplemental Materials and Methods). To evaluate the quality of each *P. teres* f. *teres* isolate's assembled gene models, RNAseq transcript coverage was analyzed with BEDtools using the 'coverage' command (Quinlan et al. 2014). Gene models were determined to have

RNAseq evidence if a gene model had 100% transcript coverage. Genome annotations were subjected to BUSCO v3 analysis using the Ascomycota data set to assess annotation completeness (Simão et al. 2015) (Appendix).

RepeatModeler v1.0.11 (Smit et al. 2015) was used to *de novo* annotate repetitive elements within the genomes. Repetitive elements were then combined from each of the four genomes (15A, 6A, FGOH04Ptt-21, and BB25) and added to the *P. teres* f. *teres* specific repeat elements derived in Wyatt et al. (2018). The RepeatModeler *P. teres* f. *teres* repeat library was input into RepeatMasker (Tarailo-Graovac et al., 2009) alongside the current release of Repbase (v22.10) (Bao et al. 2015) to soft mask identified repetitive elements and output a final annotation of repetitive elements identified in the genomes. The "buildSummary.pl" RepeatMasker script was applied to gather summary statistics for downstream analysis of repetitive elements.

Protein domain prediction, detection of biosynthetic gene clusters, and gene ontology

Interproscan v. 5.25 (Jones et al. 2014) was used to predict protein domains and assign gene ontology (GO) terms for each of the *P. teres* f. *teres* isolates 0-1, 15A, 6A, FGOH04Ptt-21, and BB25. GO terms (Ashburner et al. 2000) were examined for enrichment using a hyper geometric test with a false discovery rate cut-off set to 0.05 using the R package topGO (Alexa and Rhnenfugrer 2010). SignalP v4.1 was used to identify potential protein secretion signals under default parameters (Petersen et al. 2011) and TMHMM v2.0 was used to predict protein transmembrane domains (Krogh et al. 2001).

Biosynthetic gene clusters were detected in the five *P. teres* f. *teres* genomes using the online analysis tool antiSMASH3.0 (Weber et al. 2015). Two analyses were run with different parameters as in Syme et al. (2018b). A simple analysis was first run on each of the five *P. teres*

f. *teres* isolates using unannotated fasta files as input to antiSMASH3.0. Parameters used in the first analysis included the options to run 'KnownClusterBlast', 'SubClusterBlast', 'smCoG analysis', 'ActiveSiteFinder', and 'whole-genome PFAM analysis'. A second run was done and the analysis options 'ClusterFinder' and 'use ClusterFinder algorithm for BGC border prediction' added. Parameters for the 'use ClusterFinder algorithm for BGC border prediction' included setting the minimum cluster size set to five, the minimum number of biosynthesis-related PFAM domains set to five, and a minimum 'ClusterFinder' probability of 80%. Extra features added during the second run included 'Cluster-border prediction based on transcription factor binding sites (CASSIS)' and 'ClusterBlast' (Syme et al. 2018b). The antiSMASH program reports detected number of polyketides (PKs), non-ribosomal peptides (NRPs), PK-NRP hybrids, terpenes, and other (secondary metabolite-like clusters) resulting from the initial simple analysis. An additional report is also generated for ClusterFinder predicted biosynthetic gene clusters.

Comparative pan-genome analysis

Whole genome alignments of the scaffolded and repeat masked genome assemblies were facilitated by the MUMmer suite using the internal programs 'nucmer' for genome alignments (Delcher et al. 2003; Marcais et al. 2018). The four newly sequenced *P. teres* f. *teres* isolates 15A, 6A, FGOH04Ptt-21, and BB25 were aligned to the 0-1 reference genome using the 'nucmer' program (options: –mum –mincluster 100 –minmatch 50) followed by filtering of the resulting alignment files with 'delta-filter' (options: -q -r) to remove any repetitive alignments. Each *P. teres* f. *teres* genome was subjected to analysis using Occultercut v1 to identify clusters of high GC gene dense regions relative to low GC gene sparse regions after identifying and removing mitochondrial sequences (Testa et al. 2016). SNPs were identified by using the Harvest suite (Treangen et al. 2014) program 'parsnp' to align *P. teres* f. *teres* isolates 15A, 6A,

FGOH04Ptt-21, and BB25 to the reference 0-1 assembly (Wyatt et al. 2018). SNPs were input into a principal component analysis (PCA) for the five genomes using the R software package SNPRelate (Zheng et al. 2012). An additional PCA was done using only SNPs from gene coding regions and the resulting eigen values were similar and are supplied in the Appendix.

The program SNPGenie was used to analyze the 0-1 genome and annotation to determine the number of synonymous and non-synonymous sites (Nelson et al. 2015) and the program snpEff was used to calculate the number of mutations classified as synonymous or non-synonymous. The pN/pS ratio was calculated for both the accessory and core gene sets and imported into R for testing significance using a Kruskal-Wallis test. Accessory and core genes were subset and mutations per gene per Kb was calculated by dividing the total mutations in the accessory genes by the length of the corresponding transcripts. Accessory and core genes were checked for intersecting transposable elements using bedtools 'intersect' (Quinilan et al. 2014).

Proteins were clustered into families using the program OrthoFinder (Emms et al. 2015). The pan-genome was constructed based on OrthoFinder blastp result comparisons between the five *P. teres* f. *teres* proteomes (Emms et al. 2015).

Effector prediction and comparative analysis

Effectors were predicted under the criteria of being secreted, as predicted by SignalP v4.1 (Petersen et al. 2011), lacking a transmembrane domain as predicted by TMHMM v2.0 (Krogh et al. 2001), and having a molecular mass less than 50 kDa. The resulting list of small secreted proteins (SSP) was subjected to further analysis.

Further analyses were run on the set of small secreted proteins, including effector prediction using EffectorP v1.0 (Sperschneider et al. 2015), a program that incorporates machine learning to predict fungal effectors from the secretomes using the default cutoff value. ApoplastP

v1.0 (Sperschneider et al. 2017) was also run on the set of SSPs using the default cutoff value. ApoplastP v1.0 is a program that uses machine learning to predict apoplastic localized proteins from non-apoplastic localized proteins. SSPs were further examined for homology both between isolates and within each isolate's genome. This was accomplished by using ncbi-blast+ 'blastp' of the amino acid sequences of each *P. teres* f. *teres* SSP set and protein clustering by the program Orthofinder (Emms et al. 2015).

Data visualization

Circos plots for Figure 3.1, Figure 3.3, and Figure A3 were generated on a local Linux server using perl version 5.24. Figure 1 heatmaps were generated with a window size of 5Kb. Figure 3.1 core and accessory genome frequency plots were graphed as apercent coverage of 5Kb genomic windows. Figure 3.3 ribbons represent single best matches of genes between *P. teres* f. *teres* isolates.

Results

Genome sequencing, assembly, and scaffolding

To assess genome composition and structure of *P. teres* f. *teres*, we generated high quality genome assemblies using PacBio SMRT sequencing for isolates 15A, 6A, FGOH04Ptt-21, and BB25. These isolates were chosen due to their differing disease reactions on commonly used differential barley lines and their inclusion in previously reported genetic studies using pathogen bi-parental mapping populations (Weiland et al. 1999; Lai et al. 2007; Shjerve et al. 2014; Koladia et al. 2017). Sequencing results of the four new *P. teres* f. *teres* isolates compare favorably with previous sequencing efforts for *P. teres* f. *teres* isolate 0-1 which produced 1,148,507 reads with average read lengths of 8,051 bps (Wyatt et al. 2018) (**Table 3.1**).

Assemblies of the sequenced *P. teres* f. *teres* isolates 15A, 6A, FGOH04Ptt-21, and BB25 also compared favorably with the currently published *P. teres* f. *teres* isolate 0-1 assembly having assembled into 163, 52, 42, and 112 contigs, respectively (**Table 3.1**). Scaffolding of the resulting assemblies using the 16 linkage groups of the FGOH04Ptt-21 × BB25 linkage maps (Koladia et al. 2017) produced 12 total scaffolds for *P. teres* f. *teres* isolates 15A, 6A, and FGOH04Ptt-21 and produced 11 scaffolds for *P. teres* f. *teres* isolate BB25 (**Table 3.1**). This is a reduced number of scaffolds compared to the number of linkage groups due to assembled genomic sequence coverage across multiple linkage groups. *P. teres* f. *teres* has previously been reported to have 12 chromosomes and the 12 and 11 scaffolds in the four newly sequenced isolates represent near full chromosomes based on alignments to previously assembled *P. teres* f. *teres* genomes (Wyatt et al. 2018; Syme et al. 2018b). Interestingly, the total genome size of the *P. teres* f. *teres* isolates ranged from 46.5 Mb to 50.6 Mb (**Table 3.1**). Additional assembly metrics are summarized in Table 1 with the currently published *P. teres* f. *teres* isolate 0-1 assembly (Wyatt et al. 2018) shown as a reference metric for quality comparison between the four newly sequenced isolates 15A, 6A, FGOH04Ptt-21, and BB25 (**Table 3.1**).

Table 3.1. Summary statistics for *P. teres* f. *teres* genome assemblies.

Genome sequencing	0-1	15A	6A	FGOH04Ptt-21	BB25
SMRTcells[a]	14	9	9	9	9
Reads[b]	1,148,507	1,191,458	952,789	1,010,420	1,145,857
Mean read length	8,051	8,674	7,227	9,927	7,666
Coverage	199	222	148	216	189
Assembly statistics					
Total genome size (Mb)	46.5	45.3	48.6	49.7	50.6
Scaffolds[c]	12	12	12	12	11
Markers[d]	648	369	369	369	369
Scaffolded bases (Mb)	42.4	36.4	46.3	47.6	46.6
Percent scaffolded bases	91.1%	80.4%	95.2%	95.8%	92.1%
N50 scaffolds (Mb)[e]	4.163	2.996	4.076	4.128	5.743
L50 scaffolds[f]	6	6	6	5	4
Pre-scaffolding contigs[g]	85	163	52	42	115
N50 contigs (Mb)	1.730	0.769	3.608	4.128	3.759
Un-scaffolded contigs	42	74	36	16	55

a. The number of Pacific Biosciences RSII instrument SMRT cells used to sequence each *P. teres* f. *teres* isolate.
b. Total corrected reads generated and input into the Canu assembly program.
c. Number of scaffolds generated utilizing the established genetic map (Koladia et al. 2017).
d. Number of markers used to establish scaffolds.
e. Length of the smallest scaffold in an ordered set of scaffolds corresponding to 50% of the genomes total size.
f. Smallest number of scaffolds whose length equals 50% of the genome total size.
g. The total number of assembled un-gapped contigs .

Genome annotation

To compare gene content between the five *P. teres* f. *teres* isolates we generated high confidence gene models using RNA sequencing data from each isolate. The total number of annotated genes ranged from 11,551 (6A) to 12,183 (15A) gene models between the five *P. teres* f. *teres* isolates with RNAseq evidence for the gene models ranging from 74.2% (0-1) to 98.8% (FGOH04Ptt-21). *P. teres* f. *teres* isolates 0-1, 15A, and 6A were indexed, pooled, and sequenced on a single NextSeq run and isolates FGOH04Ptt-21 and BB25 were indexed, pooled, and sequenced on an additional NextSeq run. Differences in RNAseq coverage between isolates

is likely due to the increased depth of sequencing for isolates FGOH04Ptt-21 and BB25. The high number of genes with RNAseq evidence in FGOH04Ptt-21 and BB25 helped to refine gene models in the genome annotations of 0-1, 15A, and 6A as the full set of gene models for all *P. teres* f. *teres* isolates were used to refine gene models for each individual *P. teres* f. *teres* genome. BUSCO analysis runs on each of the *P. teres* f. *teres* genomes using the Ascomycota gene set identified complete genes for 97.4% of the genes for isolate 0-1, 98.1% for isolate 15A, 96.1% for isolate 6A, 98.0% for isolate FGOH04Ptt-21, and 97.8% for isolate BB25. Consistent with genome annotations of other fungal pathogens, many predicted genes do not encode known pfam domains with the number of genes encoding a pfam protein domain ranging between 60.6% (15A) and 62.2% (6A and BB25) (**Table3.2**).

Table 3.2. Summary statistics for *P. teres* f. *teres* genome assemblies.

Isolate	0-1	15A	6A	FGOH04Ptt-21	BB25
Genes[a]	11,573	12,183	11,551	11,557	11,986
Mean gene length (bp)	1,621	1,643	1,616	1,633	1,634
Mean protein length (aa)	476	477	470	477	477
Mean gene density ($Mb^{-1)}$)	248	251	222	217	220
RNA evidence (100% cov) [b]	74.23%	77.80%	78.11%	98.75%	94.96%
Protein functions[c]					
% Proteins with PFAM domain	62.1	60.6	62.2	62.2	61.2
Predicted secreted proteins[d]	1039	1066	1036	1053	1062
Small secreted proteins (<50 kDa)	562	589	575	570	583
Apoplastic proteins[e]	540	570	551	563	572
Effectors[f]	197	205	201	207	201
Genome repeats					
Percent repetitive content[g]	31.90%	27.50%	34.30%	31.80%	35.60%
Low GC% peak[h]	32.10%	30.50%	31.00%	32.10%	30.40%
High GC% peak[i]	51.60%	51.60%	51.60%	51.60%	51.60%
Low complexity gene density (genes/Mb)[j]	2.04	1.14	1.23	1.2	1.28
High complexity gene density (genes/Mb)[k]	328	323	328	328	323
Biosynthetic gene clusters[l]					
PKs	12	11	13	12	10
NRPs	40	45	50	45	50
PK-NRP	4	4	3	4	4
Terpenes	4	3	2	3	4
Other	2	2	3	3	2
ClusterFinder	29	27	30	30	26
Total	91	92	101	97	96

a. Total gene models output by the Maker2 pipeline (Holt et al., 2011).
b. RNA evidence represents the number of gene models where full transcript length was covered by RNAseq evidence. Isolates 0-1, 15A, and 6A were pooled and sequenced and isolates FGOH)4Ptt-21 and BB25 were pooled and sequenced.
c. Protein functions were predicted using Interproscan v5.25 (Jones et al., 2014).
d. Proteins were determined to be secreted by the presence of a signal peptide as predicted by SignalP v4.1 (Petersen et al., 2001).
e. Apoplastic proteins were predicted by the machine learning software ApoplastP v1.0 (Sperschneider et al., 2017).
f. Effector protein were predicted by the machine learning software EffectorP v1.0 (Sperschneider et al., 2015).
g. The percentage of the genome containing repetitive elements was output from RepeatMasker (Tarailo-Graovac et al., 2009).
h. Low GC% peak represents the lowest GC% calculated for each genome as output by OculterCut v1 (Testa et al., 2016).
i. High GC% peak represents the highest GC% calculated for each genome as output by OculterCut v1 (Testa et al., 2016).
j. Low complexity regions were identified by OculterCut v1 (Testa et al., 2016) and represent genomic regions with a <41 GC%
k. High complexity regions were identified by OculterCut v1 (Testa et al., 2016) and represent genomic regions with a >41 GC%
l. Biosynthetic gene clusters were predicted using antiSMASH (Weber et al., 2015) and are broken down in to polyketide synthases (PKs), non-ribosomal peptide synthetase (NRPs), hybrid polyketide non-ribosomal peptides (PK-NRPs), terpenes, other, and predicted biosynthetic clusters represented by ClusterFinder.

Secreted proteins and secondary metabolites are important in fungal biology, specifically with how fungi interact with their environments. Plant pathogens employ these molecules as

virulence factors and we therefore identified both the genes encoding secreted proteins and biosynthetic gene clusters. The total number of predicted secreted proteins for the five *P. teres* f. *teres* isolates ranged from 1,036 (6A) to 1,066 (15A) (**Table 3.2**). Total biosynthetic gene clusters identified in the five *P. teres* f. *teres* isolates ranged from 91 (0-1) to 101 (6A) (**Table 3.2**). Relatively consistent numbers of each class of biosynthetic gene clusters were detected in each of the five *P. teres* f. *teres* isolates (**Table 3.2**).

Annotated repetitive content of the five *P. teres* f. *teres* genomes ranged from ~27.5% (15A) to ~35.6% (BB25) of the genome (**Table 3.2**). Analysis using Occultercut v1 (Testa et al. 2016) revealed a bipartite genome structure divided into high complexity, gene dense regions and low complexity, gene sparse regions (**Figure 3.1**). Gene density in low complexity regions ranged from 1.14 genes/Mb to 2.04 genes/Mb and gene density in high complexity regions ranged from 323 genes/Mb to 328 genes/Mb (**Table 3.2**). A previous examination of the repeat content of *P. teres* f. *teres* found the most numerous identifiable transposable element families to be the DNA transposable element Tcl-Mariner and the long terminal repeat (LTR) Gypsy element, though a large proportion of annotated repeats belonged to yet unclassified repeat families (Syme et al. 2018b, Wyatt et al. 2018). Repeat elements were also annotated and compared for the five *P. teres* f. *teres* isolates in this study. Overall, the results of the repeat analysis in this study was similar to that of Syme et al. (2018b) in finding the largest proportion of annotated transposable elements belonging to either the DNA Tcl-Mariner transposable element or the LTR-Gypsy element with the LTR-Gypsy element comprising the largest proportion in each of the five *P. teres* f. *teres* genomes (**Figure 3.2**). A large proportion of the annotated repeats in the five *P. teres* f. *teres* isolates examined remain unclassified as was also previously noted by Syme et al. (2018b) (**Figure 3.2**). Numbers of annotated repetitive elements

in the five examined genomes were consistent although isolate 15A had consistently lower numbers of annotated repeats likely due to the lower quality of assembly relative to isolates 0-1, 6A, FGOH04Ptt-21, and BB25 (**Figure 3.2; Table 3.1**).

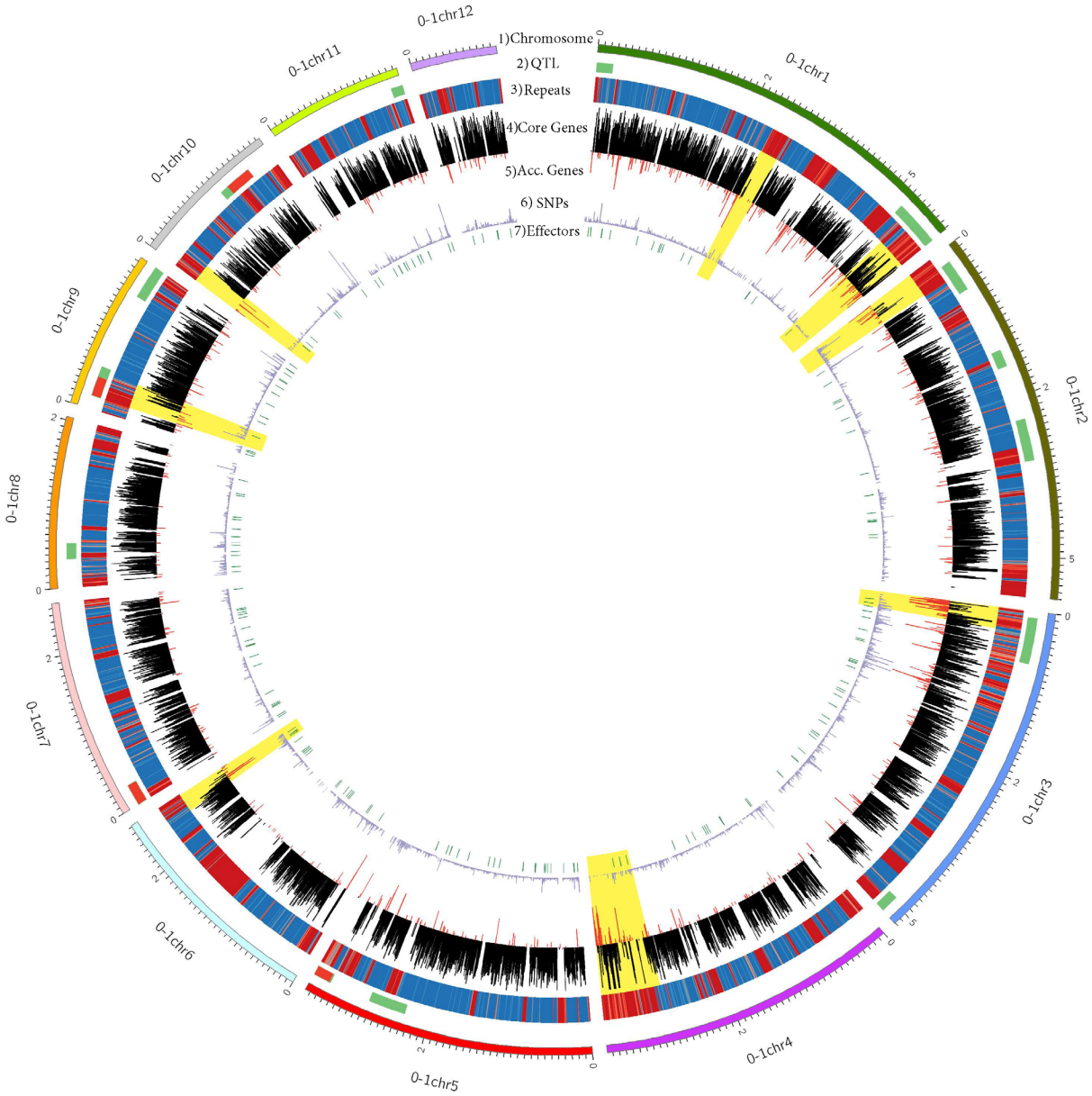

Figure 3.1. Circos plot showing the genomic landscape of *P. teres* f. *teres* isolate 0-1. **1**) The outermost track of the Circos plot represents the scaffolded chromosomes of isolate 0-1. **2**) The second track shown in green represents published locations of QTL (Lai et al., 2007; Sherve et al., 2014; Koladia et al., 2017). When multiple QTL are detected at a single locus, red is used to mark the overlapping region. **3**) The third track represents a heatmap of repetitive elements in the 0-1 genome where red indicates more repetitiveness and blue denotes repeat sparse and gene dense regions. **4**) The fourth track of the plot represents locations and density of core genes shown in black. **5**) The fifth track represents the locations and density of accessory genes in the 0-1 genome shown in red. **6**) The sixth track shows the density of SNPs identified between the five *P. teres* f. *teres* isolates **7**) The seventh track shows the location of effectors predicted by EffectorP (Sperschneider et al., 2017) shown in green. Yellow blocks highlight exemplary accessory compartments in the 0-1 genome that show more repetitiveness and an increase in accessory gene content.

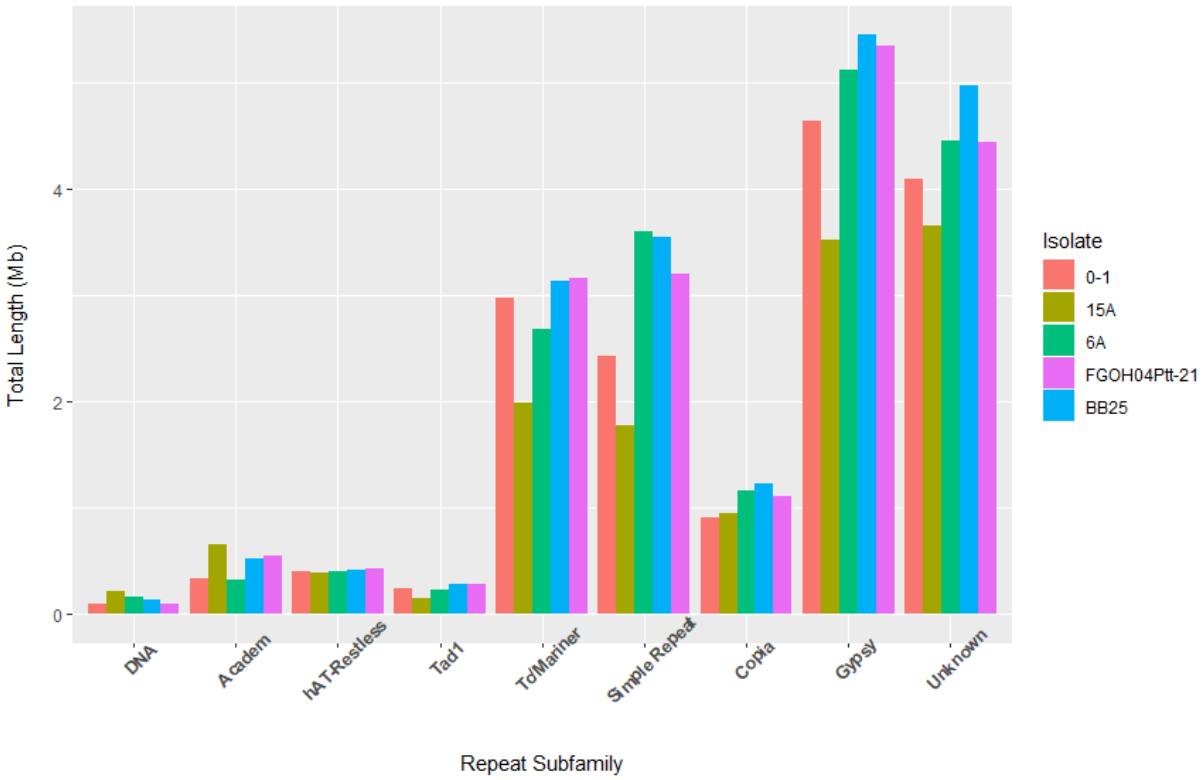

Figure 3.2. The total length of each type of repeat subfamily expressed in Mbs from each of the five *P. teres* f. *teres* genomes. Amongst all five *P. teres* f. *teres* isolates, the most numerous repeat subfamilies are the Tc/Mariner DNA elements and the Gypsy LTR elements, though a large portion of each genome is comprised of simple repeats and largely unknown/unclassified repetitive elements. *P. teres* f. *teres* isolate 15A shows a reduced number of repeats relative to the other four isolates and this is attributed to uncaptured sequence.

Comparative pan-genome analysis

To examine macro synteny between *P. teres* f. *teres* isolates, the four newly sequenced isolates 15A, 6A, FGOH04Ptt-21, and BB25 were aligned to the reference 0-1 genome. A high degree of collinearity was observed between the reference isolate 0-1 and isolates 15A, 6A, and FGOH04Ptt-21 along the twelve established chromosomes (**Figure 3.3**). The alignment between reference isolate 0-1 and isolate BB25 showed an apparent chromosome fusion between chromosomes 1 and 2 in the BB25 isolate (**Figure 3.4**).The fusion of *P. teres* f. *teres* isolate BB25 chromosome 1 and chromosome 2 also generated a small mini-chromosome comprised of the end of the two chromosomes. It has been previously shown that *P. teres* f. *teres*

chromosomes 1 and 2 share synteny along the entire length of *P. tritici-repentis* chromosome 1 (Syme et al. 2018b). Therefore, the genomes of reference isolate 0-1 and BB25 were aligned to the *P. tritici-repentis* isolate M4 genome to assess the relatedness of the chromosome 1 and 2 fusion in BB25. The pattern of synteny remained relatively consistent between *P. teres* f. *teres* isolates 0-1 and BB25 compared to *P. tritici-repentis* isolate M4 (**Figure 3.4**). We interpret this to mean that the isolate BB25 chromosome fusion is a recent event and not the result of ancestral inheritance or an interspecies hybridization event. A total of 87,189 SNPs were identified between *P. teres* f. *teres* reference isolate 0-1 and the four newly sequenced *P. teres* f. *teres* isolates 15A, 6A, FGOH04Ptt-21, and BB25 for use in principal component analysis (Figure 5H). The first principal component accounted for 38% of the variability and the second principal component accounted for 26% of the variability. These results indicate that substantial genetic differentiation exists between isolates, with the exception of 0-1 and FGOH04Ptt-21 clustering closer together.

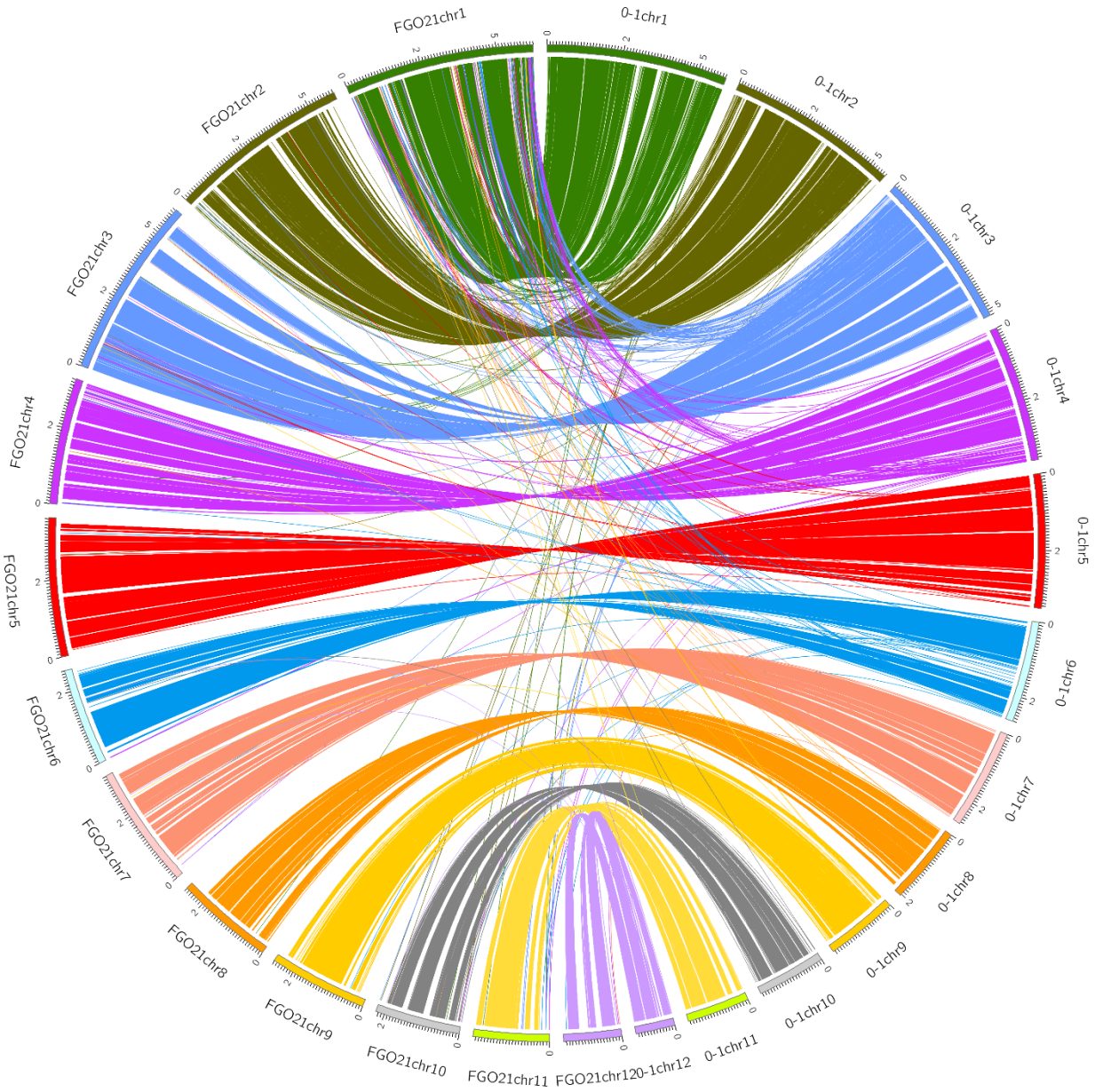

Figure 3.3. Whole genome synteny plot between *P. teres* f. *teres* isolates 0-1 and FGOH04Pt-21. Bars comprising the outer ring represent individual chromosomes labeled with size (Mb) with tick marks measuring 100 kb. Each ribbon extending from one of the 0-1 chromosomes (0-1chr#) to one of the FGOH04Ptt-21 chromosomes (FGO21chr#) represents a single 0-1 gene's best hit in the FGOH04Ptt-21 genome. Whole genome synteny plots between *P. teres* f. *teres* isolate 0-1 and isolates 15A, 6A, and BB25 can be found in the appendix.

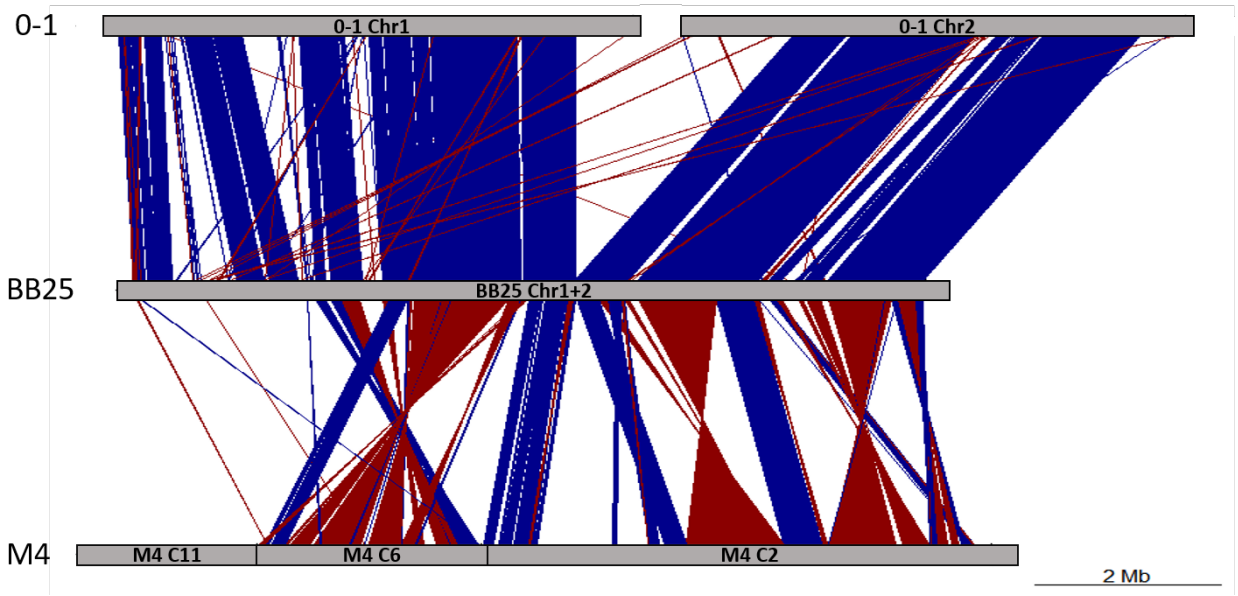

Figure 3.4. Alignments between *P. teres* f. *teres* isolate 0-1 chromosomes 1 and 2 (top), *P. teres* f. *teres* isolate BB25 fusion chromosome 1-2 (middle), and *P. tritici-repentis* isolate M4 chromosome 1 (bottom). Alignments shown are greater than 2 kb in length and at least 90% identity. Alignments shown in blue represent alignments in the forward orientation and alignments shown in red represent inversions. The pattern of synteny is conserved between the two *P. teres* f. *teres* isolates and both *P. teres* f. *teres* isolates show similar rearrangements relative to *P. tritici-repentis*.

To assess shared and unique protein functions of the five *P. teres* f. *teres* isolates 0-1, 15A, 6A, FGOH04Ptt-21, and BB25, proteins were clustered into homologous families, termed orthogroups, using the program Orthofinder (Emms et al. 2015). The total number of non-redundant protein family orthogroups identified was 12,073. Of the 12,073 orthogroups 10,444 (86.5%) were found to be shared amongst all five *P. teres* f. *teres* genomes and constitute the core genome. A total of 10,161 of the 10,444 core orthogroups represented single copy orthogroups with each *P. teres* f. *teres* isolate represented, leaving 283 multi-copy orthogroups with shared and unique gene family expansions (**Supplementary file 1**). Of the 1,629 proteins resulting from the accessory genome, 750 proteins were found to be shared between at least two isolates and 879 proteins were only present in one of the five *P. teres* f. *teres* isolates (**Figure 3.5A**). The highest number of genes unique to one isolate was observed in 15A (350) and the

remaining four *P. teres* f. *teres* isolates 0-1, 6A, FGOH04Ptt-21, and BB25 had 103, 118, 76, and 244 unique genes, respectively (**Figure 3.5A**).

For further analysis, we subset the pangenome into two categories, the core genome comprised of genes represented in all five genomes and the accessory genome comprised of all genes represented in four or fewer genomes. The average length of amino acid sequence was significantly different between core and accessory genes with core genes encoding longer proteins, 484 amino acids on average, and accessory genes encoding proteins with an average amino acid length of 372 (Kruskal-Wallis test, $P<0.0001$, **Figure 3.5D**). The number of proteins containing a conserved protein domain was also found to be significantly different, with 64.2% of the genes belonging to the core genome encoding proteins containing conserved domains, whereas the proportion of accessory proteins harboring conserved domains was only 28.4% (**Figure 3.5C**). Secreted proteins were examined and found to constitute a significantly higher proportion of the core genome when compared to the accessory genome with the core genome containing 9.3% secreted proteins and the accessory genome containing 4.6% secreted proteins (**Figure 3.5F**). Predicted effector proteins, as predicted by EffectorP (Sperschider et al. 2016), were also examined and no significant difference in the proportion of encoded effector proteins was found between the core and accessory gene sets with the proportions of each being 1.7% (**Figure 3.5G**). Transposable elements have been shown to be important in pathogen genome evolution and were examined for their proximity to core and accessory genes. A significant difference in transposable element proximity was identified with accessory genes clustering nearest transposable elements. On average, transposable elements were found to be 12.6 Kb away from accessory genes and 26.3 Kb away from core genes (**Figure 3.5E**).

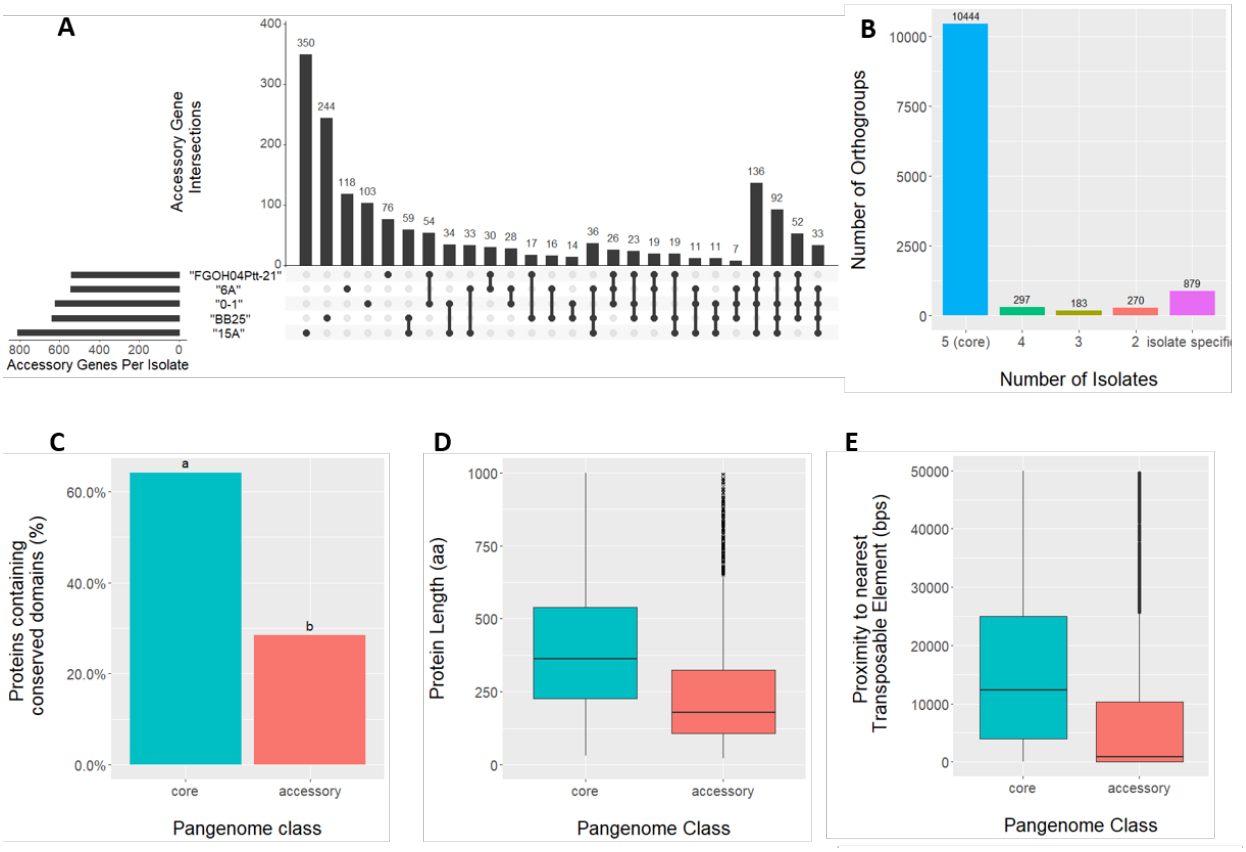

Figure 3.5. Comparative genomics summary statistics assessed between different gene and protein classes in the *P. teres* f. teres genomes. Panel **A,** is an UpSet diagram representing the number of shared accessory orthogroups between each isolate. The bar graph on top represents the number of proteins in the group and the lower panel indicates which isolates are being compared, with dots indicating the isolate is present and lines drawn between dots to indicate the shared presence. Panel **B,** represents the number of OrthoFinder (Emms *et* al. 2015) determined orthogroups belonging to each isolate. Orthogroups with a protein stemming from each of the five *P. teres* f. *teres* isolates were classified as core proteins and represent the largest group. Orthogroups with proteins stemming from less than five genomes were classified as accessory, and singleton if the orthogroup was derived from just one genome. Panel **C,** shows the different proportions of core and accessory proteins that contain conserved protein domains with letters 'a' and 'b' above bar plots denoting significantly different groups (Kruskal-Wallis test, P<0.0001). Panel **D,** shows the difference in protein amino acid sequence length between core and accessory proteins (Kruskal-Wallis test, P<0.0001). Panel **E,** shows the difference in the distance to the nearest transposable element (TE) compared between core and accessory proteins (Kruskal-Wallis test, P<0.0001). Panel **F,** shows the different proportions of core and accessory proteins that contain a signal peptide with letters 'a' and 'b' above bar plots denoting significantly different groups (Kruskal-Wallis test, P<0.0001). Panel **G,** shows the proportion of the core and accessory proteins that are predicted effectors and there was no significant difference found. Panel **H,** shows a principal component analysis that was conducted using >89,156 SNPs.

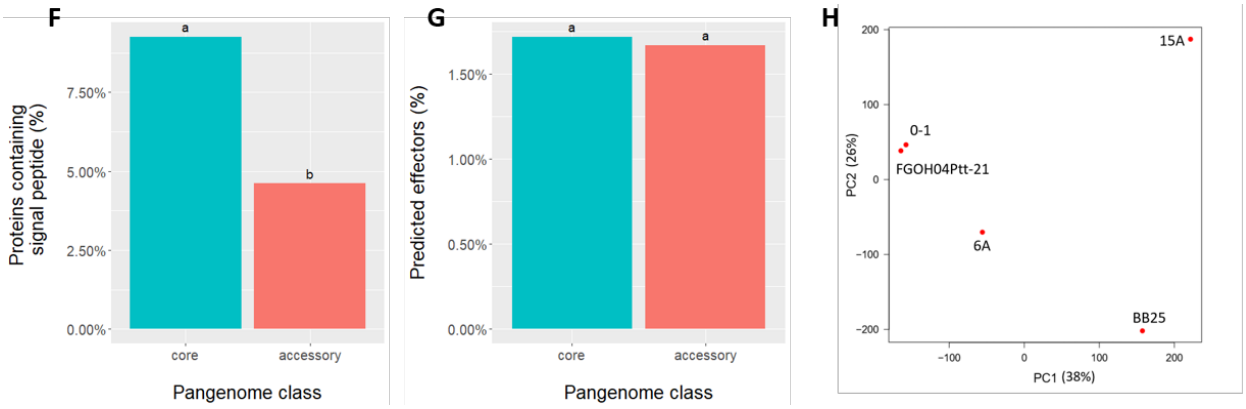

Figure 3.5. Comparative genomics summary statistics assessed between different gene and protein classes in the P. teres f. teres genomes (continued). Panel A, is an UpSet diagram representing the number of shared accessory orthogroups between each isolate. The bar graph on top represents the number of proteins in the group and the lower panel indicates which isolates are being compared, with dots indicating the isolate is present and lines drawn between dots to indicate the shared presence. Panel B, represents the number of OrthoFinder (Emms et al. 2015) determined orthogroups belonging to each isolate. Orthogroups with a protein stemming from each of the five P. teres f. teres isolates were classified as core proteins and represent the largest group. Orthogroups with proteins stemming from less than five genomes were classified as accessory, and singleton if the orthogroup was derived from just one genome. Panel C, shows the different proportions of core and accessory proteins that contain conserved protein domains with letters 'a' and 'b' above bar plots denoting significantly different groups (Kruskal-Wallis test, $P<0.0001$). Panel D, shows the difference in protein amino acid sequence length between core and accessory proteins (Kruskal-Wallis test, $P<0.0001$). Panel E, shows the difference in the distance to the nearest transposable element (TE) compared between core and accessory proteins (Kruskal-Wallis test, $P<0.0001$). Panel F, shows the different proportions of core and accessory proteins that contain a signal peptide with letters 'a' and 'b' above bar plots denoting significantly different groups (Kruskal-Wallis test, $P<0.0001$). Panel G, shows the proportion of the core and accessory proteins that are predicted effectors and there was no significant difference found. Panel H, shows a principal component analysis that was conducted using >89,156 SNPs.

To assess if core and accessory genomic compartments were evolving at different rates we examined a number of diversity statistics relative to the reference *P. teres* f. *teres* isolate 0-1. We subset the 87,189 SNPs into the relative core and accessory marker sets and found that the mutation rate in accessory genes is slightly elevated at 2.34 SNPs/Kb relative to core genes at 2.04 SNPs/Kb. Additionally, we found that the mutations that do occur in accessory genes are on average twice as likely to be nonsynonymous (missense/silent ratio: 1.3357) relative to the core genome (missense/silent ratio: 0.9168). The calculated pN/pS ratio for core genes was 0.29 and

the pN/pS ratio for accessory genes was 0.45 and were significantly different (Kruskal-Wallis test, P<0.0001). Transposable elements were implicated in mutation by insertion into a gene or its promotor, and given the expansion of transposable elements in the *P. teres* f. *teres* genome, we examined the rate of transposable element insertion in genes and their promotor regions between core and accessory gene sets. We found that the proportion of genes in the accessory genome that had a transposable element inserted in the promotor or coding region ranged from 24.7% to 36.8% and the core genome ranged from 4.3% to 11.4%. Differences in the type of transposable elements involved in the insertions were observed with LTR-Gypsy, TcMariner-Fot1, hAT-Restless, and DNA-academy transposable elements being inserted in the accessory genome at a higher proportion relative to the core genome.

Gene ontology enrichment analysis was performed on the core and accessory gene sets to further characterize putative functions of the pan-genome. A total of 86 molecular function and 108 biological process gene ontology terms were identified as significantly enriched ($p < 0.05$) within the *P. teres* f. *teres* core genome (**Supplementary file 2**). Core gene functions showing significant enrichment included basic cellular functions and metabolic processes along with typical housekeeping functions (**Supplementary file 2**). Gene ontology enrichment analysis of *P. teres* f. *teres* accessory genes identified significant enrichments for functions involved in small molecule binding including nucleotide binding, ion binding, and carbohydrate binding. Accessory genes were also enriched in gene ontologies associated with several metabolic processes as well as DNA integration, transposition, and recombination likely driven by the proliferation of transposable elements in the *P. teres* f. *teres* genome detected in the genome annotation (**Supplementary file 2**).

We identified 283 multi-copy orthogroups representing paralogs of expanded gene families ranging in size from 6 to 203 genes per orthogroup (**Table 3.3**; **Supplementary file 1**). Many of these expanded gene families comprise significantly different gene counts between isolates. For example, the largest family, OG0000000, contains 203 genes with per isolate counts ranging from a single protein in isolate FGOH04Ptt-21 to 104 proteins in isolate 6A (**Table 3.3; Supplementary file 1**). Several of the largest orthogroups contain genes with functions in secondary metabolite production and gene functions associated with mechanisms of transposition (**Table 3.3**). As an example, the annotated protein domains of OG0000000 included AMP-dependent synthetase/ligase domains, phosphopantetheine binding ACP domains, and condensation domains, often found in proteins involved in the synthesis of polyketides. Annotated protein domains of the second largest and fifth largest groups, OG0000001 and OG0000004, included retrotransposon gag domains, hAT family C-terminal dimerization regions, and ribonuclease H-like domains associated with retro elements. Four orthogroups belonging to a single isolate were identified and represent intragenomic duplication events. The four orthogroups were OG0000214 belonging to isolate 15A and comprising seven genes, OG0000119 belonging to isolate FGOH04Ptt-21 with eleven genes, and orthogroups OG0000098 and OG0000124 belonging to isolate 0-1 and having twelve and ten genes, respectively. Annotated functions of these isolate-specific orthogroups showed functions related to transposition similar to the larger expanded orthogroups.

Table 3.3. Orthogroup summary statistics for the five *P. teres* f. *teres* isolates

OrthoGroup	0-1 proteins	15A proteins	6A proteins	BB25 proteins	FGOH04Ptt-21 proteins	Total proteins	Associated functions
OG0000000	3	42	104	53	1	203	Secondary metabolite synthesis
OG0000001	23	23	17	11	20	94	Transposable elements
OG0000002	62	11	2	8	4	87	Secondary metabolite synthesis
OG0000003	31	1	14	10	28	84	No known function
OG0000004	2	36	24	2	16	80	Transposable elements
OG0000005	30	5	0	0	41	76	No known function
OG0000006	0	12	0	2	61	75	No known function
OG0000007	6	38	2	24	3	73	No known function
OG0000008	16	15	12	12	11	66	No known function
OG0000009	5	11	4	33	10	63	Transposable elements
OG0000010	15	11	10	7	12	55	No known function
OG0000011	4	20	8	14	9	55	Transposable elements
OG0000012	11	16	9	10	8	54	No known function
OG0000013	10	11	9	12	11	53	No known function
OG0000014	10	16	0	17	10	53	No known function
OG0000015	10	13	8	11	8	50	Epigenetic regulation

a. OrthoGroups were output from OrthoFinder (Emms et al., 2015) and groups larger than 50 proteins are shown here.
b. Associated functions were interpreted based on results obtained from Interproscan for the genes in each OrthoGroup. OrthoGroups with proteins lacking a conserved domain as predicted by Interproscan were labeled 'No known function'.

Effector prediction and comparative analysis

Effector proteins are the suite of proteins employed by plant pathogens to manipulate their respective plant hosts and we anticipated the five *P. teres* f. *teres* isolates in this study to have different effector repertoires based on previously published genetic analyses using these isolates (Lai et al. 2007; Shjerve et al. 2014; Koladia et al. 2017). The total number of secreted proteins for each of the five *P. teres* f. *teres* isolates ranged from 1,036 secreted proteins (6A) to 1,066 secreted proteins (15A) and, after filtering for a size of less than 50 kDa, the number of small secreted proteins (SSPs) identified ranged from 562 (0-1) to 589 (15A) (**Table 3.2**). This set of small secreted proteins was further analyzed using the machine learning programs ApoplastP, designed to predict proteins localized to the apoplast, and EffectorP, designed to

predict effector-like proteins. ApoplastP predicted between 95.8-98.7% of SSPs to be localized to the apoplast and EffectorP predicted between 34.4-36.3% of SSPs to have effector-like protein qualities (**Table 3.2**). The difference between the proportion of SSPs predicted to be localized to the apoplast versus predicted to be effectors can be attributed to the greater specificity of EffectorP. Not all proteins secreted into the apoplast would be classified as an effector and examples would include the multitude of carbohydrate active enzymes that assist in plant cell wall degradation. Comparative analysis of the set of SSPs for the five *P. teres* f. *teres* isolates showed that 2,728 of the 2,879 SSPs belong to the core genome (94.7%) and only 21 SSPs were identified in a single isolate. Of the 21 SSPs unique to a single isolate, eight are predicted to be effectors by EffectorP and eight are predicted to be apoplastic by ApoplastP, however only four of the 21 isolate-specific SSPs are both predicted to be apoplastic and effectors. Protein locus identifiers for the 21 isolate-specific SSPs can be found in the appendix (**Table A1**). All 21 isolate-specific SSPs were found to be expressed *in planta* and possibly constitute proteins important for each isolate's differential response on different barley genotypes.

Comparisons between predicted effector proteins and secreted non-effectors showed significant differences between amino acid sequence length and proximity to transposable elements with predicted effectors having shorter amino acid length and clustering in closer proximity to transposable elements (Kruskal-Wallis $P<0.0001$). Gene ontology enrichment analysis was performed on the set of SSPs and a total of 56 molecular function and 96 biological process gene ontologies were found to be enriched ($p<0.05$) (**Supplementary file 2**). Among the enriched gene ontologies for the set of SSPs were terms associated with hydrolase activity, carbohydrate binding, cell wall degradation and carbohydrate active enzymes (CAZymes) (**Supplementary file 2**). Gene ontology enrichment analyses was further performed on the set of

predicted effector genes to identify any effector specific enrichment. A total of 19 molecular function and 23 biological process gene ontology terms were found to be enriched (**Supplementary file 2**). Notable among the enriched effector gene ontology terms were hydrolase activity, carbohydrate binding and degradation, regulation of ion transport, and chitin binding. Proteins with hydrolase activity and carbohydrate binding and degradation activity may have important functions in the necrotrophic infection stage of the *P. teres* f. *teres* life cycle and chitin binding proteins have been shown to be involved in protecting and hiding pathogens from plant defense responses (van den Burg et al. 2006; De Jonge et al. 2010; Liu et al. 2016) (**Supplementary file 2**).

Discussion

Genomic plasticity in P. teres *f.* teres

In this study we present reference quality genome assemblies of four additional *P. teres* f. *teres* isolates including 15A, 6A, FGOH04Ptt-21, and BB25 (**Table 3.1**). Each of these isolates have been included in at least one biparental mapping study that investigated the genetics of avirulence/virulence in *P. teres* f. *teres* because of their diverse disease reactions on differential barley lines (Weiland *et* al. 1999; Lai et al. 2007; Shjerve et al. 2014; Koladia et al. 2017). Whole genome comparisons of synteny between the four newly sequenced isolates and the current *P. teres* f. *teres* reference sequence 0-1 (Ellwood et al. 2010; Wyatt et al. 2018) revealed evidence of a highly plastic genome structure with conserved synteny among matching chromosomes with large breaks at repeat rich genomic compartments. These repeat rich genomic compartments were common in sub-telomeric regions of the *P. teres* f. *teres* chromosomes (**Figure 3.1**, **Figure 3.6**). Previous genomic analysis of *P. teres* f. *teres* isolates has found more repetitive content relative to the closely related species *P. tritici-repentis* and *Parastagonospora*

nodorum as well as the other form of the species *P. teres* f. *maculata* (Moolhuijzen et al. 2018; Richards et al. 2018; Syme et al. 2018b). The increase in repetitive content of *P. teres* f. *teres* has facilitated a type of genomic fissuring that led to a genomic architecture where regions of highly repetitive DNA are undergoing increased evolutionary speeds relative to the more GC-equilibrated gene-rich regions of the genome (Raffaele and Kamoun 2012; Dong et al. 2015; Faino et al. 2016). Our data, using the five *P. teres* f. *teres* isolates examined in this study, show the importance of repeat dense accessory genomic compartments in the context of previously published disease related quantitative trait loci (QTL). Almost all published *P. teres* f. *teres* disease QTL span accessory compartments and are in sub-telomeric regions of the genome (Figure 1, Figure 6) (Lai et al. 2007; Shjerve et al. 2014; Koladia et al 2017). This makes sense given that these accessory compartments have been shown to be important in the host-pathogen interaction, as they are evolutionary hotspots and often harbor effector proteins (Croll et al. 2012; Thomma et al. 2016; Franceschetti et al. 2017; Seidl et al. 2017; Bertazzoni et al. 2018).

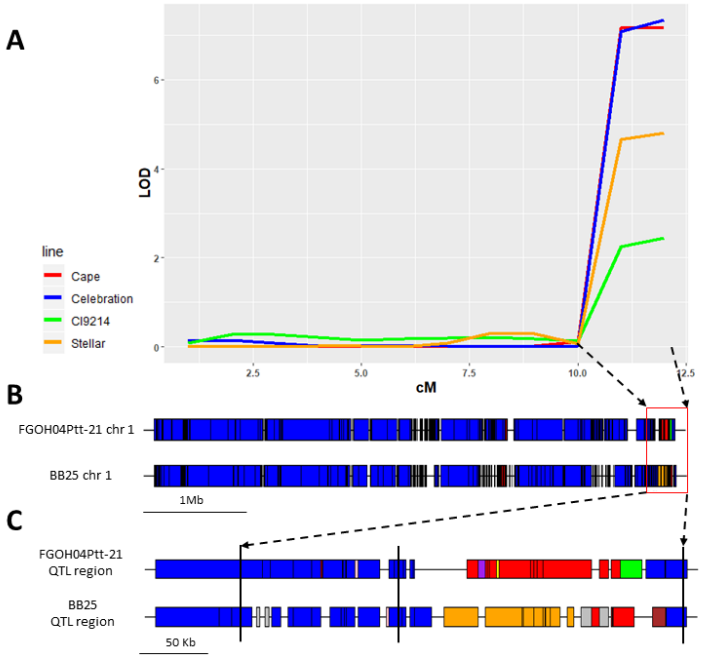

Figure 3.6. A representative example (Chromosome 1) of nonhomologous chromosomal exchange prevalent at the telomeres of *Pyrenophora teres* f. *teres*. **A**, a quantitative trait locus (QTL) analysis showing a significant association with disease reaction on the barley lines Cape, Celebration, CI9214, and Stellar. LOD score is indicated on the y-axis and centimorgan (cM) distance is indicated on the x-axis. Arrows directed from the QTL region in **A** to the red box in **B** designate the underlying genomic region of *Pyrenophora teres* f. *teres* chromosome 1. **B**, represents the pattern of synteny to the reference *P. teres* f. *teres* isolate 0-1 chromosome 1 along the full length of the chromosome between isolates FGOH04Ptt-21 and BB25. Syntenic regions between matching chromosomes (i.e. chromosome 1 to chromosome 1) are shown in blue. Syntenic regions between non-matching chromosomes are indicated with various colors. Gaps in the alignments are represented by gaps along chromosome 1's syntenic representation. The red box at the end of the chromosome 1 represents the identified QTL region with black dashed lines with arrows extending to **C** which indicates the genomic region underlying the QTL region identified. **C**, represents the synteny between isolate 0-1 and FGOH04Ptt-21 and BB25, respectively. Vertical black bars indicate the location of the three associated SNP markers that defining the identified QTL region. Synteny between matching chromosomes (i.e. chromosome 1 to chromosome 1) is represented in blue with the different colors representing synteny to sequences not present on isolate 0-1 chromosome 1. Different colors were used to identify different chromosome syntenies. Importantly, the pattern of synteny does not match between isolates FGOH04Ptt-21 and BB25 as is shown by the large red and green block of synteny in isolate FGOH04Ptt-21 and the large orange and red blocks in isolate BB25. Red blocks indicate homology to chromosome 11, green blocks indicate homology to chromosome 2, and orange blocks indicate homology to chromosome 3.

Whole genome alignments were assessed between the five *P. teres* f. *teres* isolates to examine the degree of intraspecies synteny and genome plasticity (**Figure 3.3**; **Figure A1**). A

chromosome fusion between chromosomes 1 and 2 was discovered in *P. teres* f. *teres* isolate BB25 relative to the reference isolate 0-1. Previous work examining whole genome alignments between *P. teres* f. *teres* and the closely related species *P. tritici-repentis* showed that a similar genome structure was present in *P. tritici-repentis* in which *P. teres* f. *teres* chromosomes 1 and 2 have synteny along the entire length of *P. tritici-repentis* chromosome 1 (Syme et al. 2018b). Isolate BB25's fused chromosome 1-2 was aligned to the newly published genomic sequence of *P. tritici-repentis* isolate M4 (Moolhijzen et al. 2018) to determine if this chromosome fusion was a result of ancestral inheritance. It was found that although *P. teres* f. *teres* chromosome 1 and 2 were fused in isolate BB25, the genomic synteny was more conserved between other *P. teres* f. *teres* isolates and exhibited the same pattern of rearrangements relative to *P. tritici-repentis* (**Figure 3.4**). We interpret this to mean that the fusion between chromosome 1 and 2 in *P. teres* f. *teres* isolate BB25 to be a recent event and not inherited through ancestry between *P. teres* f. *teres* and *P. tritici-repentis*. The chromosome fusion in isolate BB25 also created a short chromosome from the remainder of the two fused chromosomes. Unlike mini-chromosomes and accessory chromosomes from other plant pathogen species it does not appear that this short chromosome represents a disposable part of the genome as the sequence content is present in all five *P. teres* f. *teres* isolates. Chromosome length polymorphisms were identified between *P. teres* f. *teres* isolates 0-1 and 15A using pulse field gel electrophoresis (Elwood et al. 2010) and it was recently reported that a chromosome fusion was detected in *P. tritici-repentis* isolate M4 relative to the reference *P. tritici-repentis* isolate Pt-1C-BFP (Moolhuijzen et al. 2018). It is possible that chromosome breakage/fusion events are a common feature of the *Pyrenophora* genus, resulting in chromosome length polymorphisms and chromosome fusion events and represents an interesting area for further investigation.

Chromosome synteny appears relatively conserved along matching chromosomes between *P. teres* f. *teres* isolates (**Figure 3.3**; **Figure A1**). Breaks in chromosome collinearity occur primarily around highly repetitive sequences in the *P. teres* f. *teres* genomes. Repetitive sequences and genome rearrangements between *P. teres* f. *teres* isolates appeared at a higher frequency toward the ends of chromosomes within sub-telomeric regions where genomic synteny was observed to switch between matching chromosomes and all other chromosomes in the genome (**Figure 3.3**, **Figure 3.6**). A possible mechanism driving these observed structural rearrangements are telomere-centromere breakage-fusion-bridge cycles (Croll et al. 2013) and recombination between non-homologous chromosomes mediated by the high density of repetitive elements that are in accessory genomic compartments, particularly in sub-telomeric regions (Argueso et al. 2008; Bzymek et al. 2001; Raskina et al. 2008).

Accessory genomic compartments and pathogenicity

Previous genomic work in the closely related species *Zymoseptoria tritici* and *P. tritici-repentis* have examined multiple isolates within the same species in a pan-genomic approach to assess the core and accessory gene content of the species (Plissonneau et al. 2018; Moolhuijzen et al. 2018). Core gene content is representative of genes that are present in all members of the species and are likely required for the survival of the species. Accessory gene content includes genes that are absent in members of the same species and are thought to be involved in niche processes. Accessory genomic compartments have been previously shown to undergo rapid evolution relating to high mutation rates, copy number polymorphisms, and frequent ectopic recombination mediated by increased frequency of repetitive elements (Croll et al. 2012). These accessory genomic compartments can be as large as small chromosomes and in the case of *Z. tritici*, include several accessory chromosomes. Though we did not observe accessory

chromosomes in *P*. *teres* f. *teres*, there is strong indication of accessory genomic regions within the genome (**Figure 3.1**). These accessory genomic regions were frequently identified on the ends of *P. teres* f. *teres* chromosomes where breaks in synteny occur and synteny between non-homologous chromosomes occur. These accessory genomic regions are likely acting in a similar manner to accessory and mini chromosomes found in other fungal species. The *P*. *teres* f. *teres* accessory regions exhibit common characteristics that include high density repetitive content, low density gene content, an increased presence of accessory genes, and the decreased presence of core genes (**Figure 3.1**). Accessory genes also exhibited hallmarks of diversifying selection as they had an increased mutational rate and increased pN/pS ratio relative to core genes. Further adding to the increased evolutionary rate of accessory genes was the frequent intersection of accessory genes with transposable elements. The intersection of accessory genes and transposable elements makes sense in *P. teres* f. *teres* given the genomic architecture of accessory genomic compartments described above. In addition to accessory genes clustering in accessory regions, it was observed that a high proportion of secondary metabolite gene clusters were also present in accessory regions of the genome.

Genomic accessory regions and accessory chromosomes have been associated with pathogenicity in multiple plant pathogenic species and genes implicated in virulence have been identified in these regions (Bertazzoni et al. 2018). In the plant pathogen *Leptosphaeria maculans* the *AvrLm11* effector was identified on an accessory chromosome and conferred virulence on *Brassica rapa* cultivars lacking the *Rlm11* gene (Balesdent et al. 2013). The *Magnaporthe oryzae* effector *AvrPik* was also identified on a small 1.6 Mb accessory chromosome (Luo et al. 2007; Chen et al. 2013). To date there have been no published effector genes in *P*. *teres* f. *teres* but there are currently 15 published QTL regions that have been

associated with virulence/avirulence. Of the 15 published QTL, 14 QTL reside in genomic regions that would be categorized as accessory regions and 10 of the 14 QTL are specifically located in sub-telomeric regions of the genome (**Figure 3.1**, **Figure 3.6**) (Weiland et al. 1999; Lai et al. 2007, Shjerve et al. 2014; Koladia et al. 2017). This demonstrates the importance of these rapidly evolving accessory regions in the pathogenicity of *P. teres* f. *teres* (**Figure 3.1**, **Figure 3.6**).

Gene family expansions in P. teres *f.* teres

Gene orthology analysis of the five *P. teres* f. *teres* isolates revealed gene family expansions both within single and multiple *P. teres* f. *teres* isolates relative to other *P. teres* f. *teres* isolates examined in this study (**Supplementary file 1**). Annotated gene functions of expanded gene families within *P. teres* f. *teres* revolved around either, genes related to TEs or, genes relating to the production of secondary metabolites. The association of secondary metabolite synthesis related proteins in the expanded gene families was a common pattern and was observed in other orthologous groups often with a single isolate contributing the majority of the associated homologs (**Supplementary file 1**). Interestingly, the gene family expansions involving proteins related to secondary metabolite synthesis were present in differing numbers in the five *P. teres* f. *teres* isolates with some clusters being absent from isolates. These secondary metabolite orthogroups may represent niche specific secondary metabolites. Further, QTL regions overlap predicted secondary metabolite clusters pointing to their possible involvement in the disease cycle of *P. teres* f. *teres*. Earlier research on *P. teres* noted the production of phytotoxic molecules that contributed to the disease as they were associated with chlorosis (Smedegard-Petersen 1977; Coval et al. 1990; Bach 1979; Keon and Hargreaves 1983). A recent analysis comparing the number of biosynthetic gene clusters identified between the two forms of

P. teres (f. *teres* vs. f. *maculata*) using Australian isolates *P. teres* f. *teres* W1-1 and *P. teres* f. *maculata* isolate SG1 found significantly more biosynthetic gene clusters in the *P. teres* f. *teres* isolate (97 total biosynthetic gene clusters) (Syme et al. 2018b). Results of the current study correlate well with that of Syme et al. (2018b) and establish that *P. teres* f. *teres* possesses a wide array of biosynthetic gene clusters (**Table 3.2**). *P. teres* f. *teres* biosynthetic gene clusters and their interaction during infection remain an important and unexplored area of research. While further examination of *P. teres* f. *teres* biosynthetic gene clusters is needed, our results confirm that *P. teres* f. *teres* possesses an enormous capacity to produce biologically active secondary metabolites (Syme et al. 2018b; Coval et al. 1990; Smedegard-Petersen 1977).

The second notable observation of the gene orthology analysis was the presence of TE-related proteins in the expanded gene families. Protein domains found in the top expanded gene families included the retrotransposon gag domain, hAT family C-terminal dimerization region, and the ribonuclease H-like domain. The presence of these protein domains in the identified expanded gene families adds support to previous results of annotated repetitive elements in the *P. teres* f. *teres* genome. Previous examinations of the *P. teres* f. *teres* repetitive content showed large repeat expansions relative to *P. teres* f. *maculata* (Syme et al. 2018b).

Main conclusions and prospects

This study presents four new near complete genome sequences of *P. teres* f. *teres* with gene model annotations and annotated repetitive elements. The four newly sequenced genomes of *P. teres* f. *teres* isolates 15A, 6A, FGOH04Ptt-21, and BB25 were compared to the published reference genome sequence of *P. teres* f. *teres* isolate 0-1 and an intraspecies pan-genomic comparison was performed. Importantly, accessory genomic regions were identified in the *P. teres* f. *teres* genome that represent evolutionarily active genome compartments. Accessory

genomic regions represented breaks in genome synteny and held higher numbers of accessory genes and transposable elements. This analysis has set the foundation for understanding the genomic landscape of *P. teres* f. *teres* and the mechanisms involved in this pathogen's evolution. These resources will also facilitate future research in the barley-*P. teres* f. *teres* pathosystem providing a foundation for future research examining effector biology, secondary metabolites, transposable elements, and evolutionary population genomics.

Acknowledgments

The authors would like to acknowledge Danielle Holmes for technical assistance. The authors would also like to acknowledge Eva Stukenbrock for valuable suggestions and critical review of the manuscript. This research was supported by the North Dakota Barley Council, National Science Foundation Grant Number: # 1759030, and NIFA-AFRI grant number: #2018-67014-28491.

Availability of data and materials

P. teres f. *teres* genomes and gene models are available at NCBI under BioProject PRJNA434142. Genome assemblies for *P. teres* f. *teres* isolates 15A, 6A, FGOH04Ptt-21, and BB25 were submitted to NCBI Genbank and can be found under accession numbers VBVL00000000 (15A), VFEN00000000 (6A), VBVN00000000 (FGOH04Ptt-21), and VBVM00000000 (BB25). The reference genome for *P. teres* f. *teres* isolate 0-1 was previously submitted to NCBI Genbank and can be found under accession number NPOS00000000. RNA sequencing data was deposited at NCBI SRA and can be found under accession numbers SRR9856883 (0-1 3-day liquid culture), SRR9856882 (0-1 *in-planta* 48h post inoculation), SRR9856885 (0-1 *in-planta* 72h post inoculation), SRR9856884 (0-1 *in-planta* 96h post inoculation); SRR9875069 (15A 3 day liquid culture), SRR9875070 (15A *in-planta* 48h post

inoculation), SRR9875071 (15A *in-planta* 72h post inoculation), SRR9875072 (15A *in-planta* 96h post inoculation); SRR9875078 (6A 3 day liquid culture), SRR9875073 (6A *in-planta* 48h post inoculation), SRR9875076 (6A *in-planta* 72h post inoculation), SRR9875077 (6A *in-planta* 96h post inoculation); SRR9875067 (FGOH04Ptt-21 3 day liquid culture), SRR9875068 (FGOH04Ptt-21 *in-planta* 48h post inoculation), SRR9875079 (FGOH04Ptt-21 *in-planta* 72h post inoculation), SRR9875080 (FGOH04Ptt-21 *in-planta* 96h post inoculation); SRR9875074 (BB25 3 day liquid culture), SRR9875075 (BB25 *in-planta* 48h post inoculation), SRR9875065 (BB25 *in-planta* 72h post inoculation), SRR9875066 (BB25 *in-planta* 96h post inoculation).

Literature cited

Alexa A, Rahnenfuhrer J. 2010. topGO: enrichment analysis for gene ontology. R package version. 2(0).

Andrews S. 2016. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Argueso, J. L., Westmoreland, J., Mieczkowski, P. A., Gawel, M., Petes, T. D., and Resnick, M. A. 2008. Double-strand breaks associated with repetitive DNA can reshape the genome. Proc. Natl. Acad. Sci. U.S.A. 105:11845-11850.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., and Harris, M. A. 2000. Gene Ontology: tool for the unification of biology. Nat. Genet. 25:25.

Bach, E., Christensen, S., Dalgaard, L., Larsen, P. O., Olsen, C. E., Smedegård-Petersen, V. 1979. Structures, properties and relationship to the aspergillomarasmines of toxins produced by Pyrenophora teres. Physiol. Plant Pathol. 14:41–46.

Balesdent, M. H., Fudal, I., Ollivier, B., Bally, P., Grandaubert, J., Eber, F., Chèvre, A. M., Leflon, M., and Rouxel, T. 2013. The dispensable chromosome of Leptosphaeria maculans shelters an effector gene conferring avirulence towards Brassica rapa. New Phytol. 198:887-898.

Bao, W., Kojima, K. K., and Kohany, O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob. DNA 6:11.

Beattie, A. D., Scoles, G. J., and Rossnagel, B. G. 2007. Identification of molecular markers linked to a Pyrenophora teres avirulence gene. Phytopathology 97:842-849.

Bertazzoni, S., Williams, A., 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.Mol. Plant Microbe Interact. 31:779-788.

Bialas, A., Zess, E. K., De la Concepcion, J. C., Franceschetti, M., Pennington, H. G., Yoshida, K., Upson, J. L., Chanclud, E., Wu, C.H., Langner, T., and Maqbool, A. 2017. Lessons in effector and NLR biology of plant-microbe systems. bioRxiv171223.

Bolger, A. M., Lohse, M. and Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30:2114-2120.

Bzymek, M. and Lovett, S. T. 2001. Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. Proc. Natl. Acad. Sci. U.S.A. 98:8319-8325.

Chen, C., Lian, B., Hu, J., Zhai, H., Wang, X., Venu, R. C., Liu, E., Wang, Z., Chen, M., Wang, B., and Wang, G. L. 2013. Genome comparison of two Magnaporthe oryzae field isolates reveals genome variations and potential virulence effectors. BMC genomics, 14:887.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., and Ruden, D. M. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly 6:80-92.

Cook, D. E., Mesarich, C. H., and Thomma, B. P. 2015. Understanding plant immunity as a surveillance system to detect invasion. Annu. Rev. Phytopathol. 53:541-563.

Coval, S. J., Hradil, C. M., Lu, H. S., Clardy, J., Satouri, S., and Strobel, G. A. 1990. Pyrenoline-A and-B, two new phytotoxins from Pyrenophora teres. Tetrahedron Lett. 31:2117-2120.

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

Croll, D., Zala, M., and McDonald, B. A. 2013. Breakage-fusion-bridge cycles and large insertions contribute to the rapid evolution of accessory chromosomes in a fungal pathogen. PLoS Genet. 9:e1003567.

Faino, L., Seidl, M. F., Datema, E., van den Berg, G. C., Janssen, A., Wittenberg, A. H., and Thomma, B. P. 2015. Single-molecule real-time sequencing combined with optical mapping yields completely finished fungal genome. MBio 6:e00936-15.

Delcher, A. L., Salzberg, S. L., and Phillippy, A. M. 2003. Using MUMmer to identify similar regions in large sequence sets. Curr. Protoc. Bioinformatics 1:10-3.

Derbyshire, M., Denton-Giles, M., Hegedus, D., Seifbarghy, S., Rollins, J., van Kan, J., Seidl, M. F., Faino, L., Mbengue, M., Navaud, O., and Raffaele, S. 2017. The complete genome sequence of the phytopathogenic fungus Sclerotinia sclerotiorum reveals insights into the genome architecture of broad host range pathogens. Genome Biol. Evol. 9:593-618.

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

Ellwood, S. R., Liu, Z., Syme, R. A., Lai, Z., Hane, J. K., Keiper, F., Moffat, C. S., Oliver, R. P., and Friesen, T. L. 2010. A first genome assembly of the barley fungal pathogen Pyrenophora teres f. teres. Genome Biol. 11:R109.

Emms, D. M. and Kelly, S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16:157.

Faino, L., Seidl, M. F., Datema, E., van den Berg, G. C., Janssen, A., Wittenberg, A. H., and Thomma, B. P. 2015. Single-molecule real-time sequencing combined with optical mapping yields completely finished fungal genome. MBio 6:e00936-15.

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 Res. 26:1091-1100.

Faris, J. D., Liu, Z., and Xu, S. S. 2013. Genetics of tan spot resistance in wheat. Theor. Appl. Genet. 126:2197-2217.

Franceschetti, M., Maqbool, A., Jiménez-Dalmaroni, M. J., Pennington, H. G., Kamoun, S., and Banfield, M. J. 2017. Effectors of filamentous plant pathogens: commonalities amid diversity. Microbiol. Mol. Biol. Rev. 81:e00066-16.

Friesen, T. L. and Faris, J. D. 2010. Characterization of the wheat-Stagonospora nodorum disease system: what is the molecular basis of this quantitative necrotrophic disease interaction? Can. J. Plant Pathol. 32:20-28.

Flor, H.H. 1971. Current status of the gene-for-gene concept. Ann. Rev. Phytopathol. 9:275-296.

Goodwin, S., McPherson, J. D., and McCombie W. R. 2016 Coming of age: ten years of next-generation sequencing technologies. Nat. Rev. Genet. 17:333-351.

Grewal, T. S., Rossnagel, B. G., Pozniak, C. J., and Scoles, G. J. 2008. Mapping quantitative trait loci associated with barley net blotch resistance. Theor. Appl. Genet.116:529-539.

de Guillen, K., Ortiz-Vallejo, D., Gracy, J., Fournier, E., Kroj, T., and Padilla, A. 2015. Structure analysis uncovers a highly diverse but structurally conserved effector family in phytopathogenic fungi. PLoS Patho. 11:e1005228.

Holt, C. and Yandell, M., 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC bioinformatics 12:491.

Hurgobin, B. and Edwards, D. 2017. SNP discovery using a pangenome: has the single reference approach become obsolete? Biology. 6:21.

Jones, J. D. and Dangl, J. L. 2006. The plant immune system. Nature 444:323.

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., and Pesseat, S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236-1240.

de Jonge, R., van Esse, H. P., Kombrink, A., Shinya, T., Desaki, Y., Bours, R., van der Krol, S., Shibuya, N., Joosten, M. H., and Thomma, B. P. 2010. Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants. Science 329:953-955.

de Jonge, R., Bolton, M. D., Kombrink, A., van den Berg, G. C., Yadeta, K. A., and Thomma, B. P. 2013. Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. Genome Res. 23:1271-1282.

Van Kan, J. A., Stassen, J. H., Mosbach, A., Van Der Lee, T. A., Faino, L., Farmer, A. D., Papasotiriou, D. G., Zhou, S., Seidl, M. F., Cottam, E., and Edel, D. 2017. A gapless genome sequence of the fungus Botrytis cinerea. Mol. Plant Patholo. 18:75-89.

Keon, J. P. R., Hargreaves, J. A. 1983. A cytological study of the net blotch disease of barley caused by Pyrenophora teres. Physiol. Plant Pathol. 22:321-IN14.

Khan, T. N. and Boyd, W. J. R. 1969. Environmentally induced variability in the host reaction of barley to net blotch. Aust. J. Biol. Sci. 22:1237-1244.

Kim, D., Langmead, B., and Salzberg, S. L. 2015. HISAT: a fast spliced aligner with low memory requirements. Nat. Methods, 12:357.

Koladia, V. M., Richards, J. K., Wyatt, N. A., Faris, J. D., Brueggeman, R. S., and Friesen, T. L. 2017. Genetic analysis of virulence in the Pyrenophora teres f. teres population BB25× FGOH04Ptt-21. Fungal Genet. Biol. 107:12-19.

Kombrink, A. and Thomma, B. P. 2013. LysM effectors: secreted proteins supporting fungal life. PLoS Pathog. 9:e1003769.

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 Res. 27:722-736.

Korf, I. 2004. Gene finding in novel genomes. BMC Bioinformatics 5:59.

Krogh, A., Larsson, B., Von Heijne, G., and Sonnhammer, E. L. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305:567-580.

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 Res. 19:1639-1645.

Lai, Z., Faris, J. D., Weiland, J. J., Steffenson, B. J., and Friesen, T. L. 2007. Genetic mapping of Pyrenophora teres f. teres genes conferring avirulence on barley. Fungal Genet. Biol. 44:323-329.

Liu, Z., Ellwood, S.R., Oliver, R.P., and Friesen, T.L. 2011. Pyrenophora teres: profile of an increasingly damaging barley pathogen. Mol. Plant Pathol. 12:1-19.

Liu, Z., Zhang, Z., Faris, J. D., Oliver, R. P., Syme, R., McDonald, M. C., McDonald, B. A., Solomon, P. S., Lu, S., Shelver, W. L., and Xu, S. 2012a. The cysteine rich necrotrophic effector SnTox1 produced by Stagonospora nodorum triggers susceptibility of wheat lines harboring Snn1. PLoS Pathog. 8:e1002467.

Liu, Z. H., Zhong, S., Stasko, A. K., Edwards, M. C., and Friesen, T. L. 2012b. Virulence profile and genetic structure of a North Dakota population of Pyrenophora teres f. teres, the causal agent of net form net blotch of barley. Phytopathology. 102:539-546.

Liu, Z., Holmes, D. J., Faris, J. D., Chao, S., Brueggeman, R. S., Edwards, M. C., and Friesen, T. L. 2015. Necrotrophic effector-triggered susceptibility (NETS) underlies the barley–Pyrenophora teres f. teres interaction specific to chromosome 6H. Mol. Plant Pathol. 16:188-200.

Liu, Z., Gao, Y., Kim, Y. M., Faris, J. D., Shelver, W. L., Wit, P. J., Xu, S. S., and Friesen, T. L., 2016. SnTox1, a Parastagonospora nodorum necrotrophic effector, is a dual-function protein that facilitates infection while protecting from wheat-produced chitinases. New Phytol. 211:1052-1064.

Luo, C. X., Yin, L. F., Ohtaka, K., and Kusaba, M. 2007. The 1.6 Mb chromosome carrying the avirulence gene AvrPik in Magnaporthe oryzae isolate 84R-62B is a chimera containing chromosome 1 sequences. Mycol. Res. 111:232-239.

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. 3:41-63.

Marçais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L., and Zimin, A. 2018. MUMmer4: A fast and versatile genome alignment system. PLoS Comput. Biol. 14:e1005944.

Mathre, D. E., Kushnak, G. D., Martin, J. M., Grey W. E., and Johnston, R. H. 1997. Effect of residue management on barley production in the presence of Net Blotch Disease. J. Prod. Agric. 10:323-326.

Möller, M. and Stukenbrock, E. H. 2017. Evolution and genome architecture in fungal plant pathogens. Nat. Rev. Microbiol. 15:756.

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. 19:279.

Nelson, C. W., Moncla, L. H. and Hughes, A. L. 2015. SNPGenie: estimating evolutionary parameters to detect natural selection using pooled next-generation sequencing data. Bioinformatics, 31(22), pp.3709-3711.

Perfect, S. E. and Green, J. R. 2001. Infection structures of biotrophic and hemibiotrophic fungal plant pathogens. Mol. Plant Pathol. 2:101-108.

Pertea, M., Pertea, G.M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nature Biotechnol. 33:290.

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., and Salzberg, S. L. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nature Protoc. 11:1650.

Petersen, T. N., Brunak, S., von Heijne, G. and Nielsen, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods. 8:785.

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

Quinlan, A. R. 2014. BEDTools: the Swiss-army tool for genome feature analysis. Curr. Protoc. Bioinformatics. 11-12.

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.

Raffaele, S. and Kamoun, S., 2012. Genome evolution in filamentous plant pathogens: why bigger can be better. Nat. Rev. Microbiol., 10:417.

Raskina, O., Barber, J. C., Nevo, E., and Belyayev, A. 2008. Repetitive DNA and chromosomal rearrangements: speciation-related events in plant genomes. Cytogenet Genome Res. 120:351-357.

Richards, J. K., Wyatt, N. A., Liu, Z., Faris, J. D., and Friesen, T. L. 2018. Reference quality genome assemblies of three Parastagonospora nodorum isolates differing in virulence on wheat. G3. 8:393-399.

Rouxel, T., Grandaubert, J., Hane, J. K., Hoede, C., Van de Wouw, A. P., Couloux, A., Dominguez, V., Anthouard, V., Bally, P., Bourras, S., and Cozijnsen, A. J. 2011. Effector

diversification within compartments of the Leptosphaeria maculans genome affected by Repeat-Induced Point mutations. Nat. Commun. 2:202.

Sánchez-Vallet, A., Fouché, S., Fudal, I., Hartmann, F.E., Soyer, J. L., Tellier, A., and Croll, D. 2018. The Genome Biology of Effector Gene Evolution in Filamentous Plant Pathogens. Ann. Rev. Phytopathol. 56:21-40.

Seidl, M. F. and Thomma, B. P. 2017. Transposable elements direct the coevolution between plants and microbes. Trends Genet. 33:842-851.

Shjerve, R. A., Faris, J. D., Brueggeman, R. S., Yan, C., Zhu, Y., Koladia, V., and Friesen, T. L. 2014. Evaluation of a Pyrenophora teres f. teres mapping population reveals multiple independent interactions with a region of barley chromosome 6H. Fungal Genet Biol. 70:104-112.

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.

Smedegård-Petersen, V. 1977. Isolation of two toxins produced by Pyrenophora teres and their significance in disease development of net-spot blotch of barley. Physiol. Plant Pathol. 10:203-211.

Smit, A. F. A. and Hubley, R., 2008. RepeatModeler Open-1.0. Available from http://www.repeatmasker.org.

Sperschneider, J, Gardiner, D. M., Dodds, P. N., Tini, F., Covarelli, L., Singh KB, Manners, J. M., Taylor, J. M. 2015 EffectorP: Predicting Fungal Effector Proteins from Secretomes Using Machine Learning. New Phytol. 210:743-761.

Sperschneider, J., Dodds, P. N., Singh, K. B., and Taylor, J. M. 2017. ApoplastP: prediction of effectors and plant proteins in the apoplast using machine learning. bioRxiv. 182428.

Stahl, E. A. and Bishop, J. G. 2000. Plant–pathogen arms races at the molecular level. Curr. Opin.Plant Biol. 3:299-304.

Stanke, M., Schöffmann, O., Morgenstern, B., and Waack, S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics. 7:62.

Steffenson, B. J. and Webster, R. K. 1992. Pathotype diversity of Pyrenophora teres f. teres on barley. Phytopathology. 82:170-177.

Syme, R. A., Tan, K. C., Rybak, K., Friesen, T. L., McDonald, B. A., Oliver, R. P., and Hane, J. K. 2018a. Pan-Parastagonospora Comparative Genome Analysis—Effector Prediction and Genome Evolution. Genome Biol. Evol. 10:2443-2457.

Syme, R., Martin, A., Wyatt, N. A., Lawrence, J., Muria-Gonzalez, M., Friesen, T. L., and Ellwood, S. 2018b. Transposable element genomic fissuring in Pyrenophora teres is associated with genome expansion and dynamics of host-pathogen genetic interactions. Front. Genet. 9:130.

Tang, H., Zhang, X., Miao, C., Zhang, J., Ming, R., Schnable, J. C., Schnable, P. S., Lyons, E., and Lu, J. 2015. ALLMAPS: robust scaffold ordering based on multiple maps. Genome Biol. 16:3.

Tarailo-Graovac, M. and Chen, N. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr. Protoc. Bioinformatics. 5:4-10.

Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y. O., and Borodovsky, M. 2008. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res. 18:1979-1990.

Testa, A. C., Oliver, R. P., and Hane, J. K. 2016. OcculterCut: a comprehensive survey of AT-rich regions in fungal genomes. Genome Biol. Evol. 8:2044-2064.

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". Proc. Nat. Acad. Sci. 102:13950-13955.

Thomma, B. P., Seidl, M. F., Shi-Kunne, X., Cook, D. E., Bolton, M. D., van Kan, J. A., and Faino, L. 2016. Mind the gap; seven reasons to close fragmented genome assemblies. Fungal Genet Biol. 90:24-30.

Treangen, T. J., Ondov, B. D., Koren, S., and Phillippy, A. M. 2014. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol. 15:524.

van den Burg, H. A., Harrison, S. J., Joosten, M. H., Vervoort, J., and de Wit, P. J. 2006. Cladosporium fulvum Avr4 protects fungal cell walls against hydrolysis by plant chitinases accumulating during infection. Mol. Plant Microbe Interact. 19:1420-1430.

Vernikos, G., Medini, D., Riley, D. R. and Tettelin, H. 2015. Ten years of pan-genome analyses. Curr. Opin Microbiol. 23:148-154.

Vleeshouwers, V. G. and Oliver, R. P. 2014. Effectors as tools in disease resistance breeding against biotrophic, hemibiotrophic, and necrotrophic plant pathogens. Mol. Plant Microbe Interact. 27:196-206.

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., and Earl, A. M., 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PloS One. 9:e112963.

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., Lee, S. Y., Fischbach, M. A., Müller, R., Wohlleben, W., and Breitling, R. 2015. antiSMASH 3.0 - a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res. 43:W237-W243.

Weiland, J. J., Steffenson, B. J., Cartwright, R. D., and Webster, R. K. 1999. Identification of molecular genetic markers in Pyrenophora teres f. teres associated with low virulence on 'Harbin'barley. Phytopathology. 89:176-181.

Wu, H. L., Steffenson, B. J., Zhong, S., Li, Y. and Oleson, A. E. 2003. Genetic variation for virulence and RFLP markers in Pyrenophora teres. Can. J. Plant Pathol. 25:82-90.

Wyatt, N. A., Richards, J. K., Brueggeman, R. S., and Friesen, T. L. 2018. Reference assembly and annotation of the Pyrenophora teres f. teres isolate 0-1. G3. 8:1-8.

Yoshida, K., Saitoh, H., Fujisawa, S., Kanzaki, H., Matsumura, H., Yoshida, K., Tosa, Y., Chuma, I., Takano, Y., Win, J., and Kamoun, S. 2009. Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen Magnaporthe oryzae. Plant Cell. 21:1573-1591.

Zheng, X., Levine D., Shen, J., Gogarten, S., Laurie, C., Weir, B. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics. 28:3326-3328.

Zipfel, C. 2009. Early molecular events in PAMP-triggered immunity. Curr. Opin. Plant Biol. 12:414-420.

CHAPTER 4. CRISPR-CAS9-RNP MEDIATED VALIDATION OF THE *PYRENOPHORA TERES* F. *TERES* EFFECTOR AVRHAR IDENTIFIED USING A GENOME WIDE ASSOCIATION STUDY

Abstract

Pyrenophora teres f. *teres* is the causal agent of the stubble born foliar disease net form net blotch of barley. *P. teres* f. *teres* is prevalent globally across all barley growing regions and is the most devastating foliar disease of barley. Though economically important, the molecular mechanisms *P. teres* f. *teres* uses to cause disease are poorly understood and investigations into the mechanisms of disease were previously hindered by a lack of genomic tools. Using a pathogen bi-parental mapping population and a natural population for QTL analysis and GWAS, respectively, we identified a candidate gene for the previously mapped *AvrHar*. QTL analysis identified a locus extending off the end of *P. teres* f. *teres* chromosome 5 and GWAS analysis identified significant associations with a gene encoding a small secreted protein. The candidate *AvrHar* gene was validated using CRISPR-Cas9-RNP gene editing to induce a gene disruption in parental isolates 15A and 0-1. Disruption of *AvrHar* in isolate 15A did not result in a phenotypic change while disruption of the 0-1 allele resulted in a complete loss of pathogenicity. Homologs to *AvrHar* were identified in several plant pathogenic fungi infecting different grass crops and within the natural *P. teres* f. *teres* population seven different alleles were identified. Here we present the first identification of an effector gene from *P. teres* f. *teres* and validate its virulence function using CRISPR-Cas9-RNP gene editing and highlight the usefulness of QTL analyses and GWAS.

Introduction

Net form net blotch (NFNB) is a stubble born foliar disease of barley (*Hordeum vulgare*) induced by the fungal pathogen *Pyrenophora teres* f. *teres* and is prevalent globally across all barley growing regions. Typical disease loss due to NFNB has ranged between 10 and 40% with the potential for complete crop loss given environmental conditions favorable to the pathogen, namely, wide planting of a susceptible cultivar, and high humidity or rainfall (Mathre 1997). *P. teres* f. *teres* infection on a susceptible host initially results in dot-like lesions on the leaf, progressing to longitudinal striations and finally the net-like pattern from which the disease gets its name. *P. teres* f. *teres* directly penetrates host cells without forming haustoria and after establishing itself, begins to kill host tissue as infection proceeds (reviewed in Liu et al. 2011)

Necrotrophic effectors (NEs) are important features of the *P. teres* f. *teres* infection cycle (Liu et al. 2015) in a similar manner to the closely related species *Parastagonospora nodorum* and *Pyrenophora tritici-repentis*. These pathogens employ NEs to induce necrotrophic effector triggered susceptibility (NETS) causing hallmarks of the plant defense response that in turn favor pathogens (Faris et al. 2013; Friesen and Faris 2010). In addition to NETS, dominant resistance has also been identified in a number of barley backgrounds that follow a gene-for-gene model (Koladia et al. 2017). The presence of both dominant resistance and dominant susceptibility with the discovery of NEs illustrates the complicated interaction at play in the barley-*P. teres* f. *teres* pathosystem (Liu et al. 2011; Liu et al. 2012; Liu et al. 2015; Koladia et al. 2017). To date, no effector genes have been validated and characterized in *P. teres* f. *teres* and the mechanisms of virulence and pathogenicity are poorly understood. Complicating the investigation of the mechanisms of *P. teres* f. *teres* virulence is the diversity observed in both local and global pathogen populations. Host genotype specificity was first observed in a set of Australian *P. teres*

f. *teres* isolates (Khan and Boyd 1969) and bi-parental pathogen mapping populations have been an important tool used to associate genomic loci of *P. teres* f. *teres* isolates with host specific virulence and avirulence (Weiland et al. 1999; Lai et al. 2007; Beatie et al. 2007; Liu et al. 2011; Shjerve et al. 2014; Koladia et al. 2017). Each mapping study has revealed unique genomic loci specific to *P. teres* f. *teres* isolates phenotyped on different hosts. The first bi-parental mapping population of *P. teres* f. *teres* was a cross between the Canadian isolate 0-1 and the California isolate 15A (Weiland et al. 1999). This population was used to map a locus associated with avirulence on barley line Harbin coming from isolate 15A. Segregation ratios of virulence:avirulence observed in the progeny of the 0-1 × 15A population appeared as a 1:1 ratio indicating a single gene named *AvrHar* was segregating for virulence. The 0-1 × 15A population was used again along with supplemental progeny to investigate virulence/avirulence on barley lines Canadian Lake Shore (CLS), Tifang, and Prato (Lai et al. 2007). Segregation ratios of inoculated progeny showed a 1:1 ratio for avirulence:virulence on both Tifang and CLS and a 3:1 ratio of avirulence:virulence on Prato. Interval regression analysis revealed a major-effect locus in isolate 15A associated with avirulence on Tifang and CLS while also identifying two loci in isolate 0-1 associated with avirulence on Prato. The genomic region associated with avirulence on barley lines Tifang and CLS corresponds to the previously identified *AvrHar* locus (Weiland et al. 1999). The two loci associated with avirulence on barley line Prato were named *AvrPra1* and *AvrPra2*. *AvrPra2* also mapped to the same region as *AvrHar* but avirulence was conferred by the opposite parental isolate 0-1. The co-segregation of *AvrHar* and *AvrPra2* indicated that they were either alleles of the same gene, two closely linked genes, or the same gene conferring avirulence and virulence dependent upon the host genotype (Lai et al. 2007). Beattie et al. (2007) mapped an avirulence locus on the barley cultivar Heartland ($Avr_{Heartland}$)

with a mapping population consisting of two differential Canadian isolates. Shjerve et al. (2014) created a bi-parental pathogen population using the two California *P. teres* f. *teres* isolates 15A and 6A. 15A and 6A show differential reactions on barley lines Kombar and Rika with isolate 15A showing virulence on Kombar but avirulence on Rika and isolate 6A showing virulence on Rika but avirulence on Kombar. Using the 15A × 6A population, four QTL associated with virulence were identified. Two of the QTL named *VK1* and *VK2* corresponded to virulence coming from isolate 15A on Kombar barley. The other two QTL, named *VR1* and *VR2* corresponded to 6A virulence on Rika barley (Shjerve et al. 2014). Most recently, a cross between *P. teres* f. *teres* isolates BB25 collected in Denmark and FGOH04Ptt-21 collected in Fargo, ND, USA was used to examine avirulence/virulence on eight barley lines commonly used in differential sets. A total of nine unique QTL were reported across the eight barley lines. Both shared and unique QTL were identified with three of the QTL accounting for over 45% of the variation (R^2>0.45) (Koladia et al. 2017).

Genomic investigation of the QTL regions identified in previous bi-parental mapping studies was hindered by a lack of genomic resources necessary to identify candidate genes. The first *P. teres* f. *teres* draft genome was sequenced using 20x Illumina paired-end sequencing to produce a genome that consisted of 6,412 contigs accounting for 41.95 Mb of sequence (Ellwood et al. 2010). With the advent of long-read sequencing technology such as Pacific Biosciences (PacBio) single molecular real-time (SMRT) it became feasible to fully sequence plant pathogens and finish their genomes (Thomma et al. 2016; Goodwin et al. 2016). The reference *P. teres* f. *teres* isolate 0-1 genome was fully sequenced and scaffolded into the 12 chromosomes of the genome using PacBio SMRT sequencing providing the first reference quality genome in this species (Wyatt et al. 2018, Chapter 2). Following the publication of the 0-1 reference genome

several other *P. teres* f. *teres* isolates were sequenced and scaffolded using long read technologies that included Australian isolates W1-1, NB29, NB73, and NB85 (Syme et al. 2018) and isolates 15A, 6A, FGOH04Ptt-21, and BB25 (Wyatt et al. 2019, Chapter 3). A comparative analysis of the five isolates used in previous bi-parental mapping studies 0-1, 15A, 6A, FGOH04Pttt-21, and BB25 identified accessory genomic compartments undergoing significant diversification. These accessory genomic compartments tended to localize in sub-telomeric regions of the genome and overlapped with 14 of 15 previously published QTL (Weiland et al. 1999; Lai et al. 2007; Shjerve et al. 2014; Koladia et al. 2017; Wyatt et al. 2019, Chapter 3). The variance observed in accessory genomic compartments coinciding with previously identified QTL highlights the need for more than one representative genome for *P. teres* f. *teres* species (Wyatt et al. 2019, Chapter 3). Multiple reference genomes will be crucial to characterizing the genes underlying previously published QTL (Weiland et al.1999; Lai et al. 2007; Shjerve et al. 2014; Koladia et al. 2017).

Genome wide association studies (GWAS) are highly effective at identifying significant marker-trait associations in plant pathogenic species (Talas et al. 2016; Talas et al. 2015; Gao et al. 2016; Yoshinda et al 2009; Richards et al. 2019). Typically, GWAS relies on variation in a natural population and ancient recombination events in a species history to identify regions with significant linkage disequilibrium and association with a particular phenotype (Hill and Robertson 1968; Nordberg and Tavare 2002; Zhu et al. 2008). The reliance on a natural population makes GWAS approaches significantly faster than bi-parental approaches if a natural population is readily available. For *Parastagonospora nodorum*, a GWAS approach was used to identify associations with the previously cloned effectors SnToxA and SnTox3 (Gao et al. 2016, Richards et al. 2019). Extreme linkage disequilibrium decay was found by Gao et al. (2016), an

advantage to GWAS analysis given sufficient marker coverage, and was further highlighted by Richards et al. (2019) using whole genome resequencing for variant calling. The first effector gene cloned from the wheat pathogen *Zymoseptoria tritici* was first identified using a GWAS approach (Zhong et al. 2017; Brunner and McDonald 2018) and later characterized from a bi-parental mapping population (Kema et al. 2018) thus highlighting the utility of each approach.

In this study we used a pathogen bi-parental mapping population and a natural population for QTL analysis and association mapping, respectively, to identify a candidate gene for *AvrHar* (Weiland et al. 1999), and we validated this gene using CRISPR-Cas9 gene editing.

Materials and methods

Population development

P. teres f. *teres* isolate 0-1 was collected from a barley field in Ontario, Canada (Tekauz et al. 1990) and isolate 15A was collected from a barley field in Solano County, California (Weiland et al. 1999). *P. teres* f. *teres* progeny isolates from a cross of parental isolates 15A and 0-1 were produced as described in Weiland et al. (1999) and obtained from Brian Steffenson, University of Minnesota. Additional progeny isolates were created by crossing parental isolates 15A and 0-1 following the methods described in Koladia et al. (2017). Single ascospores were collected and allowed to grow for two rounds of single spore isolation to ensure genetic purity. In total 78 isolates from Lai et al. (2007) and 40 new progeny isolates from the 15A × 0-1 population were obtained totaling 118 isolates.

Bi-parental progeny genotyping

DNA was extracted from the 118 15A × 0-1 progeny using a modified CTAB extraction method as described in Koladia et al. (2017). A restriction-associated-digest genotype-by-sequencing (RAD GBS) method was used to create sequencing libraries (Baried et al. 2008;

Leboldus et al. 014; Carlsen et al. 2017; Koladia et al. 2017). Genomic libraries were generated, and sequencing was performed with four runs on an Ion Torrent PGM system (Life Technologies). A SNP calling pipeline was modified from methods described in Carlsen et al. (2017) with polymorphism being identified in relation to the reference 0-1 genome (Wyatt et al. 2018, Chapter 2). Annotated SNPs were filtered based on previously published filtering criteria, namely, having a read depth of 3 or more, having a phred-scaled probability of 999, having less than 30% missing data, and an allele frequency of 25-75% (Shjerve et al. 2014; Carlsen et al. 2017; Koladia et al. 2017). Genotypic data for all progeny isolates were examined for clonal progeny and missing data. All progeny isolates were found to be non-clonal and retained more than 70% of their genotypic data after the cutoff criteria was applied.

Genetic map construction

Markers obtained from the SNP calling pipeline and filtering process were put into MapDisto v2.0 (Heffelfinger et al. 2017) for construction of genetic linkage maps. The "find groups" command was used to generate initial linkage groups using a LOD min = 7.0 and r_{MAX}=0.3. Following initial linkage mapping the functions "check inversions" and "ripple order" were used to obtain the best order of markers for each linkage group. The "drop locus" command was used to remove problematic markers that expanded the map size by greater than 3 cM. Co-segregating markers were removed from the final genetic map by removing markers with the least data. Linkage groups were named in accordance with the chromosome designations of the 0-1 reference genome and when a chromosome was represented by more than one linkage group the linkage groups were named consecutively (i.e. if two linkage groups represented chromosome 1 they were named LG 1.1 and LG 1.2) (Wyatt et al. 2018, Chapter 2). Marker order was examined to determine the accuracy of the genetic map in relation to the physical

sequence of the 0-1 reference genome and with the exception of co-segregating markers that were filtered out, there were no miss associations detected along the length of the reference chromosomes.

Bi-parental progeny phenotyping

Tifang barley seeds were sown into conetainers (Stuwe and Sons, Inc., Corvallis, OR, USA) with three seeds in each cone and grown for 14-16 days until the secondary leaves were fully extended. *P. teres* f. *teres* progeny isolates from the 15A × 0-1 bi-parental population were grown and inoculated as described in Lai et al. (2007) and Koladia et al. (2017). Briefly, progeny isolates were inoculated onto V8-PDA (150 mL V8 juice, 3 g $CaCO_3$, 10 g Difco PDA, 10 g agar, 850 mL H_2O) plates and allowed to grow for 5 days. After this initial growth period, plates were moved into a light bank for a 24 h light cycle followed by a dark cycle for 24 h at 15 °C to induce sporulation. Spores were collected from V8-PDA plates by flooding the plate with distilled water and scraping the surface with an inoculating loop. Inoculum was diluted to a concentration of 2000 spores/mL and Tween 20 was added to each set of inoculum at a quantity of 1 drop per 50 mL. Barley seedlings grown 14-16 days in the greenhouse were inoculated with progeny isolate inoculum using an air sprayer (Huskey, model #HDS790) until the leaves were covered with a heavy mist of inoculum but prior to run-off (Friesen et al 2006; Lai et al. 2007; Koladia et al. 2017). Inoculated plants were placed in a mist chamber with 100% relative humidity with light for 24 h at 21°C. After 24 h, plants were moved into a growth chamber maintained at 21°C with a 12 h photoperiod until disease was evaluated at 7 days post inoculation. A total of three replicates were completed for each progeny isolate on Tifang barley. Phenotypic scores for each isolate were averaged for the three replicates and used as the final data for QTL analysis.

QTL analysis

Phenotypic data for disease reactions on Tifang barley, based on the 1 to 10 disease rating (Tekauz 1985), were put into MapDisto v2.0 (Heffelfinger et al. 2017) and output formatted for the QTL mapping software Qgene4.3 (Joehanes and Nelson 2008). Critical logarithm of odds (LOD) was calculated from a permutation test of 1000 permutations (Shjerve et al. 2014; Carlsen et al. 2017; Koladia et al. 2017). Composite interval mapping (CIM) was used with co-factors to identify significant QTL as described in Carlsen et al. (2017).

Candidate gene identification

Candidate genes were identified by examining the genomic region between flanking markers of the *AvrHar* QTL that extends off the end of the linkage group corresponding to 0-1 chromosome 5. Genes within the QTL region were scored based on the presence of a signal sequence detected by SignaP v4.0 (Peterson et al. 2011), being small in size (<50 kDa), and being polymorphic between parental isolates 15A and 0-1. Candidate genes were put into the program Disulfind (Ceroni et al. 2006) to evaluate the potential of each candidate gene to form disulfide bonds between cysteine residues. Candidate proteins were also submitted to the machine learning programs EffectorP 1.0 (Sperschneider et al. 2016), EffectorP 2.0 (Sperschneider et al. 2018) and ApoplastP (Sperschneider et al. 2018) to examine whether the proteins contained common features of effectors and whether or not the protein was predicted to localize to the apoplastic space.

Whole genome sequencing of a P. teres *f.* teres *natural population*

A whole genome sequencing approach was chosen to genotype a global collection of *P. teres* f. *teres* isolates. The *P. teres* f. *teres* population chosen consisted of 73 isolates collected in North Dakota (U.S.A), 57 isolates collected in Morocco, 21 isolates collected in France, 19

isolates collected in Australia, 16 isolates collected in Azerbaijan, five isolates collected in Denmark, two isolates collected in California (U.S.A) , and a single Canadian isolate 0-1.

DNA extractions were conducted using a modified protocol from Wyatt et al. (2018) (Chapter 2). Briefly, dried fungal plugs were inoculated into 60 mL of Fries medium and allowed to grow for five days. Five day old cultures were blended and inoculated into new Fries medium and incubated overnight. Fungal tissue was then collected, lyophilized, and homogenized using a drill and a pestle drill bit after freezing with liquid nitrogen. Homogenized tissue was then placed in a 15 mL conical tube and 3 mL of lysis buffer was added (Wyatt et al. 2018, Chapter 2) and homogenized. Samples were then placed in a 65°C water bath for 45 minutes with mixing every 15 minutes followed by centrifugation at 3100 ×g for 20 minutes. The supernatant was poured from each 15 mL tube to a new 15 mL tube and 1/10 volume of 3 M sodium acetate (pH 5.2) and an equal volume of isopropyl alcohol was added. Samples were mixed well, and precipitated DNA was removed from the 15 mL tube using a glass hook to a 2 mL tube. Samples were then rinsed twice with 70% ethanol , removing the 70% ethanol between rinses. Samples were pulse centrifuged for five minutes at 17,000 ×g in a benchtop centrifuge to pellet DNA and 70% ethanol was remove. DNA pellets were allowed to dry for five minutes in a fume hood and 700 uL of TE buffer was added. The sample was then incubated at 65°C for 30 minutes and then incubated at room temperature until the pellet dissolved. After the pellet had dissolved, samples were centrifuged for 20 minutes at 17,000 ×g and 500 uL were removed to a new 1.5mL tube. DNA was enzymatically fragmented using dsDNA fragmentase (New England Biolabs) for isolates collected in North Dakota (U.S.A.), Morocco, and Australia. Isolates collected from France, Azerbaijan, and Denmark were fragmented mechanically using a Sonic Dismembrator (fisherscientific Part No. FB-705) set for a run time of 28 minutes with a 10 second “ON” and 20

second “OFF” cycle at an amplitude of 75% of maximum. Whole genome sequencing libraries were prepared using the NEBNext Ultra II Library kit (New England Biolabs) according to the recommended protocol. NEBNext Multiplex Oligos for Illumina were used to uniquely index libraries and were subsequently sequenced at the Beijing Genome Institute (BGI) on an Illumina HiSeq 4000.

Natural population marker development

Illumina sequencing data obtained from BGI was uploaded to the USDA SciNet high performance computing cluster for quality confirmation and genetic marker development. Sequencing data was first checked for low quality sequences and adapter content using the program FastQC (Andrews 2010). Sequencing data was then quality trimmed using the program Trimmomatic v0.36 (Bolger et al. 2014) using a minimum length of 36 base pairs and providing a custom text file containing a collection of Illumina adapter sequences. Sequence data was then aligned to the reference quality genomes of 0-1 (Wyatt et al. 2018, Chapter 2) and 15A (Wyatt et al. 2019, Chapter 3) to identify genetic markers. Sequence alignments were accomplished using the program BWA-MEM (Li and Durbin 2009) generating per sample sequence alignment (SAM) files. SAM files were converted to binary alignment (BAM) files and indexed using the program samtools (Li et al. 2009). BAM files were then used to call SNPs and INDELs from the alignment data using the genome analysis toolkit (GATK) programs. First, GATK “HaplotypeCaller” was used with the option -ERC GVCF to create per sample variant call files (VCFs) that were then combined using GATK “CombineGVCFs” into a single VCF file. Next, GATK “GenotypeGVCFs” was used to jointly genotype all samples and produce a final VCF for quality filtering. The final VCF was filtered based on GATK best practices guide. Heterozygote genotypes were converted to missing data in the data set as *P. teres* f. *teres* is a haploid organism

and these genotypes are likely due to sequencing error. Marker sets generated by variant calling against the 0-1 and 15A genomes were compared using VCFtools (Danecek et al. 2011) and similar numbers of markers were identified at greater than 99% of the loci.

Phenotyping of the natural population

A total of 29 barley differential lines were selected for phenotyping that included the barley lines Beecher, CI5791, Algerian, Atlas, Canadian Lake Shore, Tifang, CI9214, Harbin, Hazera, Ming, CI11458, Kombar, Manchurian, Prato, Rika, Cape, CI4922, Manchuria, Hector, Heartland, NDB112, SM89010, Rabat, Pinnacle, AC Metcaf, Robust, Lacey, PI67381, and Tradition. Isolates were inoculated in three replicates on the differential set. Planting and inoculations were performed as described above for the 15A × 0-1 bi-parental population.

Genome wide association study

Markers generated for the natural population were converted from VCF format to HapMap format to be input into the program GAPIT (Tang et al. 2016). GAPIT is a program written in the R language used for association mapping. Phenotypic data was input with genotypic data and a genome wide association analysis was conducted for each of the phenotyped barley lines using a naive model and a mixed linear model (MLM) accounting for three principal components, kinship, and fixed effects. A minor allele frequency (MAF) was set to 5%. Significant loci were subject to a Bonferroni correction (Bonferroni 1936) as a strict measure of significance and a Benjamini-Hochberg correction (Benjamini and Hochberg 1995) as a less stringent measure of significance.

AvrHar alleles and orthologs

Sequencing data of the *P. teres* f. *teres* natural population was assembled using the genome assembly program SPAdes (Bankevich et al. 2012). A database of all sequenced isolates

was subsequently created using NCBI-blast+ “makeblastdb” command and the database was queried using the genomic sequence of the *AvrHar* candidate gene with 1,000 base pairs up stream of the start codon representing the putative promotor sequence (Camacho et al. 2009). Blast results were converted into a BED file and BEDTools ‘getfasta’ was used to extract the genomic sequence of the *AvrHar* candidate gene from each assembled isolate in the natural population (Quinlan 2014). Genomic sequences were then uploaded to Multalin (Corpet 1988). Aligned sequences were reduced to nonredundant haplotypes manually. The *AvrHar* candidate gene was used in a BLAST analysis against the NCBI nonredundant database to identify proteins with homology in other species. Cut off values used for inclusion into homology were 75% query coverage and 60% homology.

CRISPR-Cas9 mediated gene editing of the AvrHar *locus*

A CRISPR-Cas9 mediated gene editing protocol was adapted from Foster et al. (2018) to edit genes in *P. teres* f. *teres*. After identifying the *AvrHar* candidate gene the genomic sequence was input into the online program E-Crisp that identifies CRISPR-Cas9 targets and protospacer adjacent motifs (PAMs) (Heigwer et al. 2014). Optimal target sequences identified by E-Crisp were ordered and used as input into the EnGen sgRNA Synthesis Kit (New England BioLabs E3322) to produce small guide RNAs for complexing with purified EnGen Spy Cas9-NLS protein (New England BioLabs M0646). For each fungal transformation, purified Cas9-NLS and sgRNA were combined in a 20 µL reaction at a 1:1 molar ratio using 2 µl (3.21 µg) of Cas9-NLS and incubated at 25°C for 15 minutes. In addition to the precomplexed sgRNA-Cas9-NLS product, a double stranded donor DNA template was introduced at a concentration of 2-6 µg in 30 µL. Donor DNA is supplied as a template for the homologous recombination repair pathway that will be induced following a double stranded cut created by Cas9. The donor DNA was

created by amplifying a hygromycin resistance cassette using primers containing 40 bps of homology added that surrounded the Cas9 cut site in the desired *P. teres* f. *teres* cut site. Donor DNA and precomplexed sgRNA-Cas9-NLS were added to 100µl of *P. teres* f. *teres* protoplast cells.

To produce fungal protoplasts *P. teres* f. *teres* cultures were started by adding five dried fungal plugs to 60 mL of Fries medium. Cultures were allowed to incubate for five days at which time they were blended into fine particulates and transferred to fresh Fries media where they were allowed to incubate for 2 h. Following the 2 h incubation, tissue was collected and washed with protoplasting buffer (1 M NaCl, 10 mM MgCl, 10 mM KH_2PO_4, 10 mM K_2HPO_4) and 0.8 g of tissue were added to a deep well Petri dish. A lysing solution was created using 0.6 g Lysing enzyme from *Trichoderma harzianum* (Sigma Life Science #L1412-25G) and 0.1 g β-Glucanase from *Trichoderma longibrachiantum* (Sigma Life Science #G4423-100G) dissolved in 40 mL of protoplasting buffer. A volume of 40 mL of Lysing enzyme was added to each Petri dish containing fungal tissue and allowed to incubate with agitation for 5.5 h at 30°C. After incubation, protoplasts were collected in a sterile beaker by straining through double layered Miracloth (EMD Millipore Corp. #475855-1R) to remove undigested fungal tissue. Protoplasts were then pelleted by centrifugation for 5 minutes at 2000 ×g and brought up to a total volume of 10 mL using protoplasting buffer. Protoplasts were counted using a hemocytometer, pelleted, and volumes were adjusted for a concentration of 1×10^7-10^8 cells in 100 µL in a 4:1 STC:50%PEG solution ([1.2 M Sorbitol, 10 mM Tris-HCl pH 7.5, 10 mM $CaCl_2$]:[0.15 M Polyethylene glycol, 10 mM Tris-HCl, 10 mM $CaCl_2H_2O$]).

Protoplasts were added to a 15 mL tube and mixed with precomplexed sgRNA-Cas9-NLS and donor DNA, gently mixed and allowed to incubate on ice for 20 minutes. After

incubation, 50% PEG was added to the mixture in three steps using volumes of 100 µL, 300 µL, and 600 µL for a total volume of 1 mL with gentle mixing after each addition. The transformation mixture was then allowed to incubate at room temperature for 20 minutes. STC was then added to the transformation mixture in a stepwise fashion using volumes of 1 mL, 3 mL, and 4 mL with gentle mixing after each addition. The transformation mixture was then centrifuged at 2000 ×g for five minutes to pellet the protoplasts. The supernatant was then pipetted off leaving a volume of 0.4 mL remaining in the 15 mL tube. Pelleted protoplasts were then brought up to 2 mL in a two-step addition of 800 µL of regeneration medium (1 M Sucrose, 1% volume Yeast Extract, 1% volume Tryptone) and incubated with gentle agitation for 2 h. Protoplasts were then added to petri dishes and overlaid with regeneration medium agar containing 100 mg/mL Hygromycin B (Invitrogen #10687070) and placed in a 30°C incubator. Colonies observed to be growing after two days were cut from the agar and placed on PDA plates containing 100 mg/mL Hygromycin B (Invitrogen #10687070). After four days of outgrowth on PDA-Hygromycin B, plates were moved through a 24 h light/dark cycle to induce sporulation as described above for spore inoculations. Following sporulation, single conidia were collected from each colony and placed on a V8-PDA plate and allowed to grow for five days. After five days of growth, plates were moved through a 24 h light/dark cycle to induce sporulation. Half of the tissue on a single plate was scraped into a 2 mL tube for DNA extraction using a Biosprint 15 (Qiagen). And the remaining half plate was turned into dried fungal plugs for long term storage. Genomic DNA was extracted using a Biosprint 15 using the manufacturer's protocol (Qiagen). DNA extracted from hygromycin resistant colonies were checked for the proper insertion using PCR by amplifying an internal segment of the hygromycin resistance cassette and also by amplifying a DNA fragment using primers outside of the inserted

hygromycin resistance cassette and a primer internal to the hygromycin resistance cassette. These two PCR reactions verify the presence of the hygromycin resistance cassette and the proper location of the cassette in the genome. Colonies identified as true edits were inoculated on the barley differential sets to assess changes in phenotype following protocols outlined above for isolate sporulation and phenotyping.

To produce genomic co-edits at two loci the procedure above was used as described with a precomplexed Cas9 and sgRNA for each locus to be edited. This type of co-editing involved inserting a resistance cassette at a benign locus in the genome and performing a different gene edit at the locus of interest. Colonies containing the resistance cassette were then sent for Sanger sequencing (Eurofins) of the gene edited locus to confirm the presence of the gene edit.

Results

Bi-parental progeny genotyping and genetic map construction

To generate markers for QTL analysis I used a RAD-GBS approach with a bi-parental pathogen population generated by crossing the *P. teres* f. *teres* isolates 15A and 0-1. A total of 118 isolates were genotyped and the resulting sequencing data produced 75,112 SNP markers using the 0-1 genome as a reference (Wyatt et al. 2018, Chapter 2). After filtering, a total of 274 markers were used with the linkage mapping program MapDisto v2.0. Linkage mapping resulted in 15 linkage groups (LGs) and allowed the identification of co-segregating markers that were subsequently removed to produce 115 non-redundant SNP markers (**Table 4.1**). The total genetic distance of the resulting genetic map was 946.35 cM with LGs ranging in size from 9.69 to 159.44 cM (**Table 4.1**). The number of non-redundant markers for each linkage group ranged from 3 to 18 with average marker densities ranging from one marker every 8,998 bp to 386,134 bp (**Table 4.1**). The linkage groups spanned 57.2% of the reference 0-1 genome with 11 of 12

chromosomes represented to some degree (**Table 4.1**). Chromosome 10 was not represented in the marker set and thus was not covered by a linkage group. The proportion of each chromosome covered by any single linkage group ranged from 8.39% (Chromosome 12) to 92.65% (Chromosome 6) (**Table 4.1**).

Table 4.1. Summary statistics for the 15 linkage groups of the 15A × 0-1 bi-parental mapping population.

LG[a]	Markers (non-redundant)[b]	Genetic distance (cM)	Chromosome	Sequence length (bp)	Chromosome coverage[c]
1	43 (18)	76.09	3	3,172,935	60.96%
2	36 (17)	159.44	2	5,285,447	94.23%
3	35 (14)	69	6	3,074,477	92.65%
4	20 (12)	126.16	1	3,162,219	53.44%
5	13 (8)	95.81	9	2,063,170	87.00%
6	13 (6)	42.28	3	1,026,792	19.73%
7	13 (6)	44.89	4	1,361,526	30.37%
8	11 (6)	53.42	7	1,727,321	52.77%
9	7 (5)	86.81	11	915,993	44.54%
10	27 (5)	51.55	4	1,678,433	37.43%
11	8 (5)	58.08	8	841,640	32.96%
12	5 (4)	23.82	5	502,696	11.48%
13	15 (3)	35.57	1	1,158,401	19.57%
14	20 (3)	9.69	5	523,815	11.96%
15	8 (3)	13.74	12	107,975	8.39%
Total	274(115)	946.35	11	26,602,840	57.20%

a. Linkage groups (LG) of the 15A × 0-1 bi-parental mapping population
b. Total marker numbers with the number of non-redundant markers after removing co-segregating markers in parenthesis
c. Chromosome coverage is calculated as the total number of bases between end markers on a linkage group divided by the total length of the respective chromosome.

Bi-parental progeny phenotyping

Progeny isolates from the 15A × 0-1 cross were inoculated onto the barley line Tifang to assess disease reactions. Parental isolate 15A is consistently avirulent on Tifang with parental isolate 0-1 giving the reciprocal reaction of virulence on Tifang. Phenotyping data of the 40

additional isolates was combined with the phenotyping from Lai et al. (2007). Sixty-nine progeny isolates had a disease score of less than five and 49 had a score of greater than or equal to five on Tifang barley and therefore were classified as avirulent and virulent, respectively. These results fit a 1:1 ratio (χ^2=3.38, p=0.066) and agree with previous results for this population indicating that a single major gene is controlling virulence/avirulence on Tifang barley (Lai et al. 2007).

QTL analysis

To identify regions of the genome associated with pathogen virulence/avirulence, a QTL analysis was performed using the genetic map and phenotypic data from the 15A × 0-1 pathogen bi-parental population. A LOD value threshold (α=0.05) obtained after a permutation test employing 1,000 permutations resulted in values of 3.75 for Tifang phenotyping data. A single major QTL was identified extending off the end of chromosome 5 for Tifang with a LOD score of 38.7 and was responsible for 72% of the disease variance (**Figure 4.1**). This Tifang locus corresponds to the locus identified as *AvrHar* (Weiland et al. 1999; Lai et al. 2007).

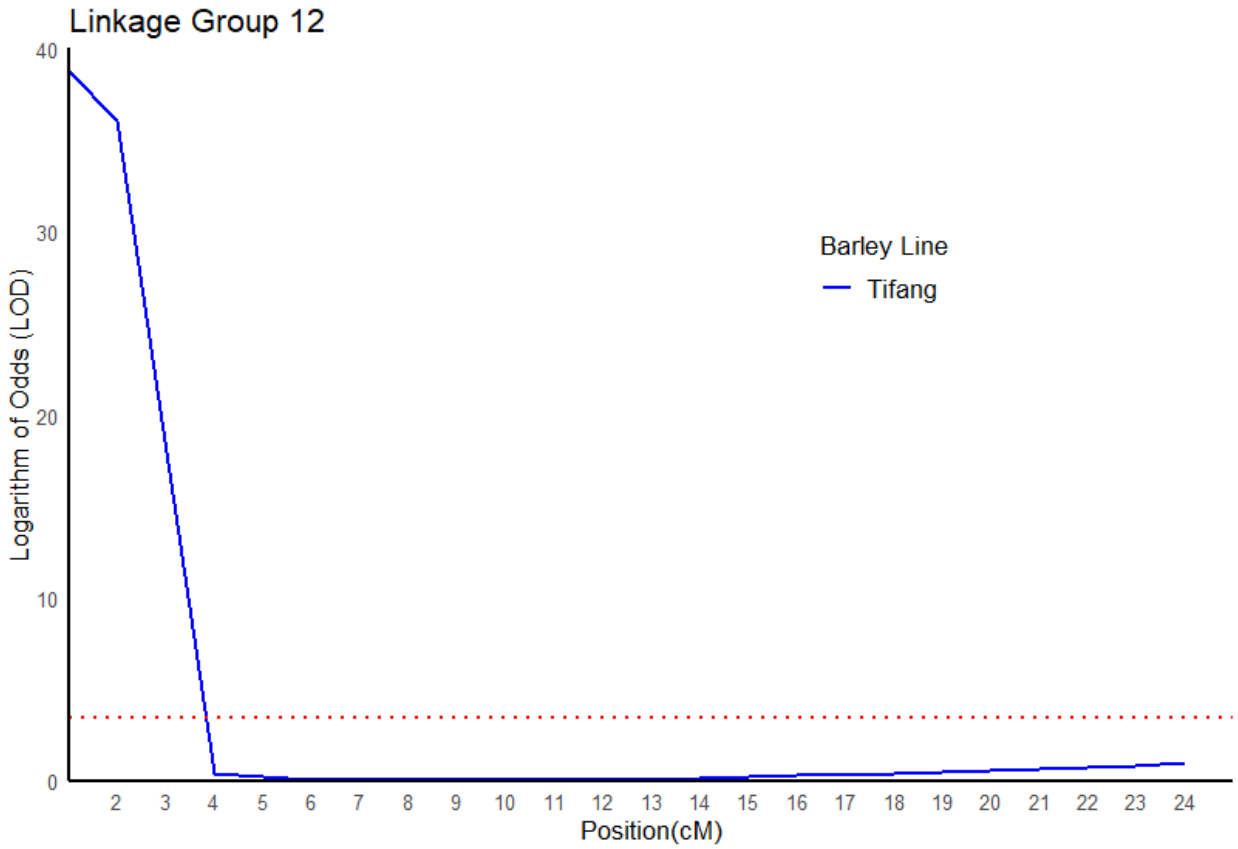

Figure 4.1. Quantitative trait loci (QTL) analysis results of the 15A × 0-1 pathogen bi-parental mapping population for response on Tifang barley represented in blue. A single QTL was identified on linkage group 12 corresponding to the sub-telomeric region of chromosome 5 having a LOD score of 38.7. LOD scores are represented on the y-axis with centimorgan (cM) position represented on the x-axis. A significance threshold of 3.75 (α=0.05) was determined after a permutation test using 1,000 permutations and is depicted as the horizontal red dashed line.

AvrHar *candidate gene identification*

To identify candidate genes within the *AvrHar* QTL region, prioritizing criteria were set. Candidate gene criteria included genes located within the bounds of the QTL region, having polymorphism between *P. teres* f. *teres* isolates 15A and 0-1, having a secretion signal as predicted by SignalP v4.0 (Peterson et al. 2014), and encoding a small protein (<50 kDa). The *AvrHar* QTL region extends off the end of chromosome 5 and encompasses a 300 Kb sub-

telomeric region. Telomeric repeats (TTAGGG) were identified on the end of chromosome 5 in the 0-1 *P. teres* f. *teres* reference genome. Within the 300 Kb region there are 73 total genes with seven of those genes having a signal peptide and encoding proteins <50 kDa in size representing effector candidate genes. Four of the seven effector candidate genes were deprioritized due to a lack of polymorphism between parental isolates 15A and 0-1 resulting in 3 candidate small secreted effector proteins within the *AvrHar* QTL region. Two of the candidate genes did not contain any predicted protein domains and had homology only to hypothetical proteins in the NCBI non-redundant database. The third candidate gene was predicted to encode a cutinase.

GWAS population marker development

In order to generate a sufficient marker set for a GWAS analysis we conducted full genome sequencing of 194 *P. teres* f. *teres* isolates. Sequencing data was aligned to the reference 0-1 genome and, following the GATK best practices for quality filtering, a total of 681,262 SNPs and small INDELs were identified for all isolates. Average distance between markers equaled 69 bp with the largest gap between adjacent markers being 65,296 located on chromosome 1. No significant gaps in marker coverage were detected across all chromosomes with marker coverage extending into the sub telomeric regions of each chromosome.

Markers were subsequently filtered for a minor allele frequency of 5% leaving 282,240 markers to be used in the GWAS analysis. Marker densities ranged from 4.8 markers/Kb (chromosome 1) to 7.8 markers/Kb (chromosome 9). The 282,240 markers were using in a principal component analysis to observe the underlying population structure using three principal components. Three principal components (PCs) were used as they each represented greater than 5% of the variation within the data set. Principal component 1 accounted for 22.2% of the variation, PC2 accounted for 8.0% of the variation, and PC3 accounted for 6.5% of the variation

with a total of 36.7% of the variation encompassed (**Figure 4.2**). Principal component 1 separated isolates based on geographical location with US and Moroccan isolates showing the greatest separation and isolates from French clustering tightly in the middle of PC1 (**Figure 4.2**). Isolates from Australia and Azerbaijan spaced randomly between US and Moroccan populations along PC1 (**Figure 4.2**). Along PC2 the primary separation was between the French isolates and all other isolates with the exception of three isolates from Azerbaijan and a single isolate from Denmark which clustered with the French isolates (**Figure 4.2**). Along PC3, isolates from the US, Morocco, and France cluster together with isolates from Australia, Denmark, and Azerbaijan separating (**Figure 4.2**).

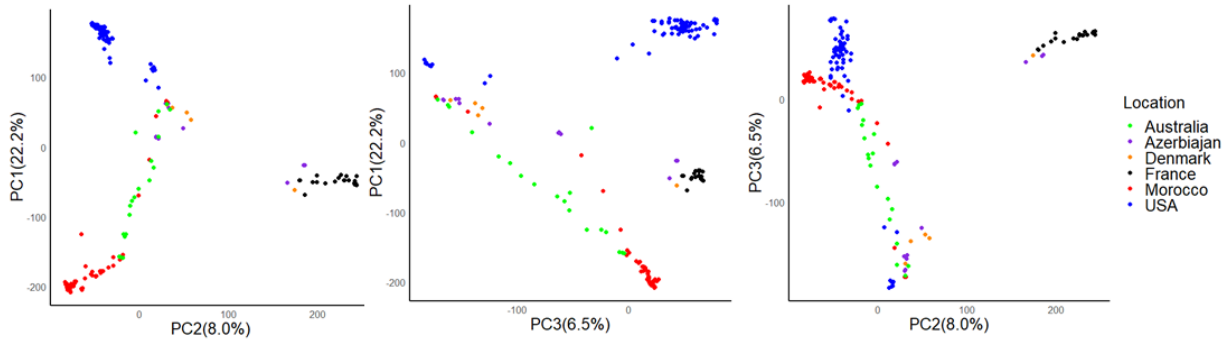

Figure 4.2. Pricipal componenet (PC) plots for the three principal components of the *P. teres* f. *teres* natural population. The left panel depicts the relationship between PC1 (y-axis) accounting for 22.2% of the variation and PC2 (x-axis) accounting for 8.0% of the variation. The middle panel depicts the relationship between PC1 and PC3 accounting for 6.5% of the variation. The right panel depicts the relationship between PC3 (y-axis) and PC2 (x-axis).

Pyrenophora teres *f.* teres *genome wide association study*

To identify significant marker-trait associations for *P. teres* f. *teres* virulence/avirulence, the GWAS natural population was phenotyped on 28 barley differential lines. Included in the differential set were barley lines Harbin, Tifang, and Canadian Lake Shore that were previously used in studies to identify the *AvrHar* locus (Weiland et al. 1999; Lai et al. 2007).

A significant association was identified for Tifang barely on the end of chromosome 5 corresponding to the same location previously identified as the locus for *AvrHar* (**Figure 4.3**) (Weiland et al. 1999; Lai et al. 2007). This association was also detected on barley lines Canadian Lake Shore and Harbin. For Tifang, the most significant associations occurred in the 5' promotor region of a small secreted protein that represented one of the candidate genes previously identified using the 15A × 0-1 bi-parental mapping population QTL analysis.

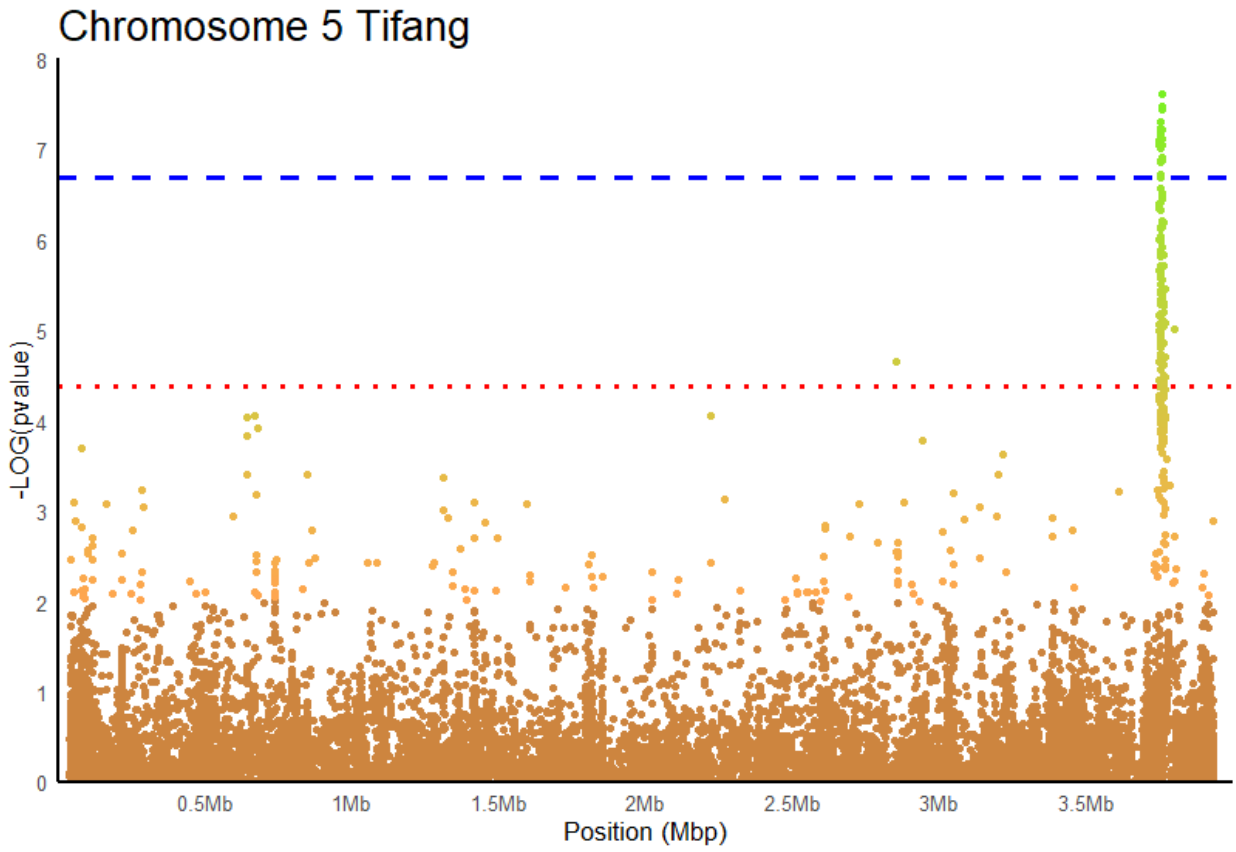

Figure 4.3. Manhattan plot for the marker-trait associations identified on *P. teres* f. *teres* chromosome 5 using phenotyping on Tifang barley. The y-axis represents the -LOG(p-value) determined by GAPIT and the x-axis represents the genomic positions of chromosome 5. Marker significance is depicted by a continuous color scale with brown representing low -LOG(p-value) scores not associated with the Tifang phenotyping and green representing high -LOG(p-value) scores that are associated with the Tifang phenotyping. The horizontal blue dashed line represents a strict significance threshold calculated using a Bonferroni correction. The horizontal dotted red line represents a less stringent significance threshold calculated using a false discovery rate.

AvrHar *candidate gene, haplotypes, and homology analysis*

The candidate gene for *AvrHar* was selected from the reference 0-1 *P. teres* f. *teres* genome and encoded a small secreted protein with no known protein domains and only showed homology to hypothetical proteins in the NCBI non-redundant database. The *AvrHar* candidate gene encoded a 165 amino acid protein with a secretion signal, cleaved at the 16th amino acid, creating a 139 amino acid mature protein. The candidate gene contained four cysteine residues that were predicted to form two disulfide bonds shown to be important for protein stability of effectors (Liu et al. 2012). The candidate gene was not predicted to be an effector by EffectorP v1.0 or v2.0 (Sperschneider et al. 2016; Sperschneider et al. 2018a) but was predicted to be secreted into the apoplastic space by ApoplastP (Sperschneider et al. 2018b).

Parental isolates 15A and 0-1 were examined to evaluate parental polymorphism in the candidate gene for *AvrHar* residing beneath the Tifang GWAS marker trait association. It was found that isolate 15A had a six base pair, in-frame, deletion that resulted in the deletion of an aspartic acid and an alanine and a single synonymous SNP in the coding region and a single SNP in the intron. The 1 Kb promotor region of this gene was also examined for polymorphism and it was found that there were 12 total SNPs and a seven bp indel where the deletion had happened in isolate 0-1.

To assess the diversity of the *AvrHar* candidate gene within the natural population, the genomes of all sequenced isolates were assembled and compiled into a BLAST compatible database. The majority of isolates assembled into a single contig covering the promotor and genic region of the *AvrHar* candidate gene. A total of 187 sequences were extracted from the database and aligned to assess the number of alleles present in the natural population. Seven alleles were identified and average disease reactions on Tifang barley were compared between

the seven different alleles. Allele groups 1 and 2 had average disease reaction scores of 6.8 each on Tifang with the remaining allele groups 3 through 7 having average disease reaction scores of less than 3.1 (**Figure 4.4**). Comparing virulent and avirulent alleles revealed that many of the genic polymorphisms are shared between virulent and avirulent alleles while the two virulent alleles share none of the polymorphisms observed in the avirulent isolates promotor regions (**Figure 4.4**).

To identify any orthologous genes in closely related species, the *AvrHar* candidate gene was subject to BLAST analysis to the NCBI non-redundant database. Homologs of *AvrHar* were identified in several grass pathogens. Homologs were identified in *P. teres* f. *maculata* the causal agent spot form net blotch of barley, *Pyrenophora tritici-repentis* the causal agent of wheat tan spot, *Leptosphaeria maculans* causal agent of black leg disease of *Brassica* crops, *Setosphaeria turcica* causal agent of northern corn leaf blight, and multiple species of *Bipolaris* that infect wheat, rice, and maize (**Figure 4.5**) . The wheat pathogen *Parastagonospora nodorum* also contained a homolog but at a lower percent identity to the other plant pathogens. Each pathogen identified as having a homolog of *AvrHar* is a foliar pathogen of a grass species with the exception of the *Brassica* pathogen *L. maculans*.

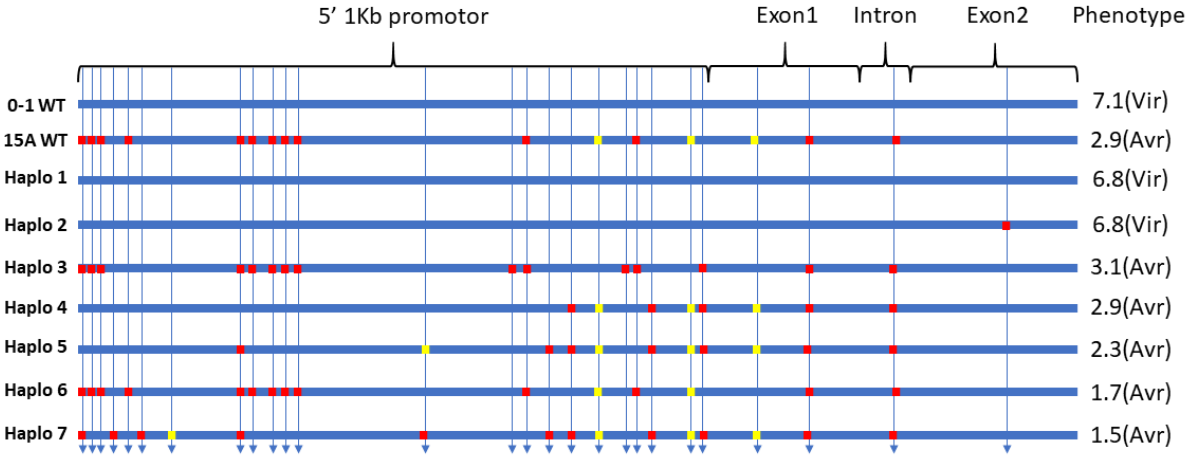

Figure 4.4. Schematic representation of the haplotypes present in the natural population for *AvrHar*. The genomic region 1,000 bp upstream of the *AvrHar* start site is included with the gene region. Annotated regions are depicted at the top of the figure representing the 5' 1Kb promotor region, the exon 1 region, the intron region, and the exon 2 region. Parental isolate 0-1, belonging to haplo-group 1, and parental isolate 15A, belonging to haplo-group 6, are shown above the seven haplo-groups. Each haplotype is then shown below with red squares indicating a SNP and yellow squares indicating an INDEL. Average disease scores of parental isolates 0-1 and 15A on Tifang, as well as the average disease scores for each haplotype, are shown on the right side of the figure.

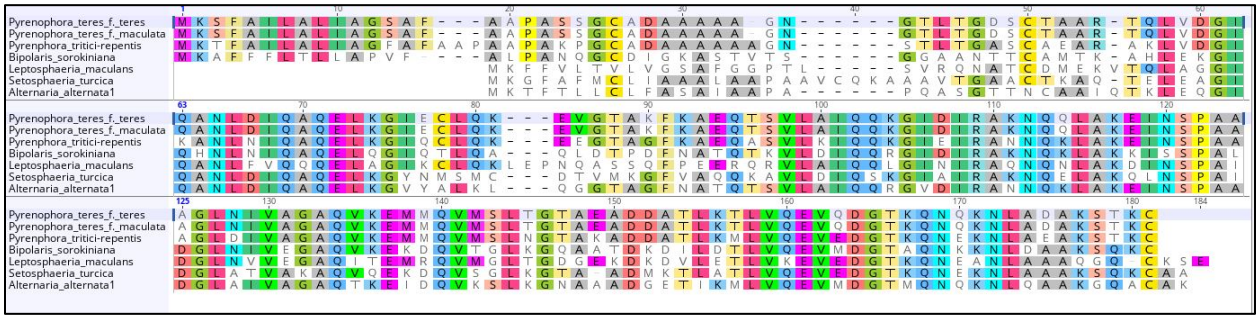

Figure 4.5. Multiple alignment of *AvrHar* orthologs from other plant pathogen species that include *P. teres* f. *teres* shown at the top, *P. teres* f. *maculata* that causes spot form net blotch of barley, *Pyrenophora tritici-repentis* that causes wheat tan spot, multiple species of *Bipolaris* that infect wheat, rice, and maize with *Bipolaris sorokiniana* shown, *Leptosphaeria maculans* causal agent of black leg disease of *Brassica* crops, *Setosphaeria turcica* causal agent of northern corn leaf blight, and *Alternaria alternata* that can infect many host species.

CRISPR-Cas9 mediated gene editing of the AvrHar locus

In order to validate the *AvrHar* candidate gene, we pursued a CRISPR-Cas9 ribonucleoprotein (CRISPR-Cas9-RNP) mediated gene editing protocol to disrupt the *AvrHar*

candidate gene in the parental isolates 0-1 and 15A by inserting a hygromycin resistance cassette into the genic region. Disruption of the candidate gene in isolate 15A resulted in no detectable change in phenotype (Figure 6). Disruption of the candidate gene in isolate 0-1 resulted in a complete loss of virulence on Tifang barley (Figure 6). However, disruption of the *AvrHar* gene in isolate 0-1 resulted in a loss of pathogenicity on all barley lines tested from the differential barley set.

To assess the causal polymorphisms determining the different phenotypes in the *AvrHar* gene, CRISPR-Cas9-RNP co-edits were performed by inserting a hygromycin resistance cassette at a different location than the *AvrHar* gene while also editing the alleles of *AvrHar.* Gene edits of the coding sequence of *AvrHar* to switch the alleles in isolates 15A and 0-1 to each other's respective allele did not result in phenotypic changes. Gene editing in the promotor sequence of the 15A and 0-1 alleles have not yet been successfully completed. To further validate *AvrHar* gene disruptions were performed with isolate LDN07Ptt-5, which belongs to the same allele group as parental isolate 0-1. Gene edits of *AvrHar* in isolate LDN07Ptt-5 resulted in a loss of pathogenicity on all barley lines tested in concordance with results of gene disruptions in isolate 0-1.

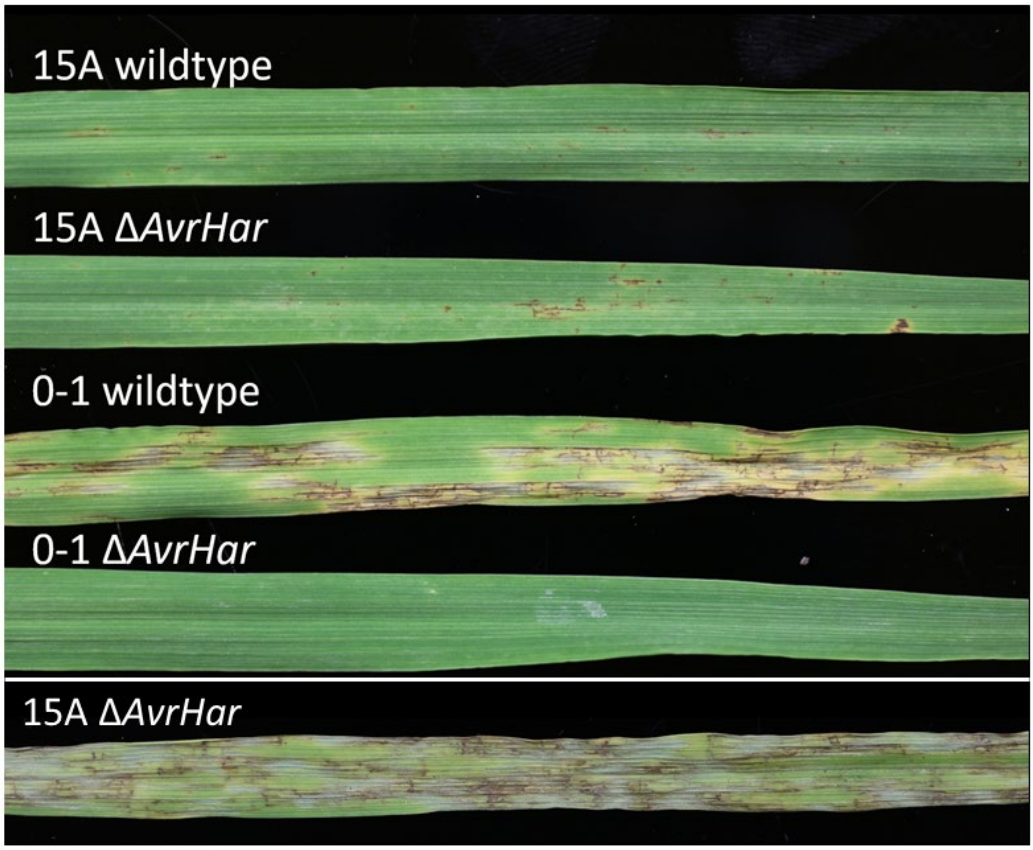

Figure 4.6. Phenotypic results of wildtype and *AvrHar* gene disruptions from isolates 15A and 0-1. Wild type 15A inoculated on Tifang barley showing the typical avirulent response is depicted at the top of the picture. The second leaf from the top is isolate 15A Δ*AvrHar,* containing a gene disruption in the *AvrHar* gene, showing an avirulent phenotype similar to wild type 15A on Tifang barley. The third leaf shows a wild type reaction of isolate 0-1 on Tifang barley with noticeable virulence. The fourth leaf shows 0-1 Δ*AvrHar*, containing a gene disruption in the *AvrHar* gene, showing an avirulent reaction on Tifang barley. The fifth leaf shows the phenotype of 15A Δ*AvrHar* inoculated on the susceptible barley line Kombar, confirming 15A Δ*AvrHar* as a virulent isolate.

Discussion

P. teres *f.* teres *bi-parental mapping and QTL analysis*

P. teres f. *teres* pathogen bi-parental mapping populations are valuable tools for identifying regions associated with virulence/avirulence on differential barley lines (Weiland et al. 1999; Lai et al. 2007; Shjerve et al. 2014; Koladia et al. 2017). *AvrHar* was previously identified using the 15A × 0-1 pathogen population segregating for virulence/avirulence on Harbin and Tifang barley (Weiland et al. 1999; Lai et al 2007). In this study we used a RAD-

GBS approach on the 15A × 0-1 population and the newly generated reference genome for *P. teres* f. *teres* isolate 0-1 to generate a genetic map (Wyatt et al. 2018, Chapter 2). The genetic map covered 57.20% of the 0-1 reference genome with at least one linkage group present on 11 of 12 chromosomes. Chromosome 10 was not represented in the genetic map likely due to short comings of the RAD-GBS method as polymorphisms have been identified between isolates 15A and 0-1 across all chromosomes (Wyatt et al 2019, Chapter 3).

The segregation ratio observed for the 15A × 0-1 progeny indicated that a single gene was conditioning virulence/avirulence on the barley line Tifang. With the new genetic map physically anchored to the *P. teres* f. *teres* 0-1 reference genome (Wyatt et al. 2018, Chapter 2), a significant QTL was identified extending off the end of chromosomes 5 corresponding to differences in virulence/avirulence on Tifang barley that matched the locus previously identified as *AvrHar* (Weiland et al 1999; Lai et al, 2007). Using the 0-1 reference genome and annotation, three candidate genes were identified using effector gene criteria. Two of the three candidate genes had homology only to hypothetical proteins when subject to BLAST analysis against the NCBI non-redundant database and the third gene had homology to cutinase genes. The two candidate genes having homology only to hypothetical proteins were selected for validation as the primary candidates for *AvrHar*. These two candidate genes were both secreted, polymorphic between parental isolates 0-1 and 15A, and were located very near to the most closely associated marker from the QTL analysis indicating that they could be the causal gene conditioning avirulence/virulence on Tifang barley.

Genome wide association study

Genome wide association study analysis is a powerful tool to identify significant marker trait associations in plant pathogenic species. Fungal plant pathogens exhibit extreme linkage

disequilibrium decay with decay occurring in as little as 600 bp in *Zymoseptoria tritici* and 6.5 Kb reported for *Parastagonospora nodorum* (Hartmann et al. 2017, Richards et al. 2019). The extreme linkage disequilibrium in fungal plant pathogens is a double-edged sword in that insufficient marker coverage can result in missed associations however, when associations are detected they are often within or near the causal gene. To ensure enough marker coverage in the *P. teres* f. *teres* natural population we pursued a whole genome sequencing protocol using Illumina paired end sequencing. The total number of SNP/INDEL markers generated from this data was 681,262 and filtering for a five percent minor allele frequency left 282,240 markers to be used in a GWAS analysis. Having 282,240 markers translates to a marker density for each chromosome ranging from a marker every 4.8 to 7.8 Kb with markers observed at adjacent base pairs where the marker density was the highest.

A principal component analysis was conducted using the 282,240 markers from the *P. teres* f. *teres* natural population. Three PCs were used with PC1 accounting for 22.2% of the total variation with isolates forming a continuum along PC1 (**Figure 4.2**). Isolates from the United States and Morocco cluster at opposing ends of PC1 with isolates from Australia, Azerbaijan, Denmark bridging the gap between the United States and Morocco (**Figure 4.2**). Isolates from France also grouped in the central region of PC1 but form a tighter cluster than isolates from Australia and Azerbaijan (**Figure 4.2**). Principal component two primarily forms as a result of isolates from France clustering apart from all other isolates except for a single isolate from Denmark and three Azerbaijan isolates indicating that the isolates from France are genetically distinct from the other isolates in this study (**Figure 4.2**). European populations thus represent a potentially untapped resource for additional genetic diversity that could be used in future GWAS studies. The clustering of the United States and Moroccan isolates along PC2 and PC3 may

indicate gene flow between these two populations and possess an interesting question for future investigation as previous studies examining the genetic diversity of geographically distinct populations have observed no gene flow between populations (**Figure 4.2**) (Peever and Milgroom 1994; Serenius et al. 2007; Lehmensick et al. 2010; Rau et al. 2003; Akhaven et al. 2015).

GWAS analysis was performed using the *P. teres* f. *teres* marker set using Tifang phenotyping data identified a significant association located on the sub-telomeric region of chromosome 5. The GWAS association on Tifang barley was located in the genomic locus previously identified as containing *AvrHar*. The exact same association was also identified on Harbin, and Canadian Lake Shore barley. The identical marker-trait associations identified for Tifang, Harbin, and Canadian Lake Shore confirmed that the results of previous mapping studies using these barley lines have mapped the same association on the end of chromosome 5 (Weiland et al. 1999; Lai et al. 2007). The two most strongly associated markers were a SNP marker and an INDEL marker located seven bases and 18 bases upstream of the start site of a gene previously identified in the bi-parental mapping as a candidate for *AvrHar.* This highlights the advantage of GWAS over bi-parental QTL analysis in identifying marker-trait associations at higher resolution.

CRISPR-Cas9 gene editing in P. teres *f.* teres

The gene editing tool CRISPR-Cas9 has brought on a new age in genome engineering. Recent work in the model fungal plant pathogen species *Magnaporthe oryzae* demonstrated the ability to induce specific mutations using a CRISPR-Cas9 ribonucleoprotein (CRISPR-Cas9-RNP) to induce double stranded breaks in DNA that could be repaired following the homologous recombination repair pathway supplied with a double stranded donor DNA (Foster et al. 2019).

In order to validate the *AvrHar* candidate gene, a gene disruption was created by inducing a double stranded break in the coding region of *AvrHar* and providing a donor DNA for homologous recombination containing a hygromycin resistance cassette. In this way I was able to both disrupt the gene of interest, *AvrHar*, and select colonies containing the proper insert. *AvrHar* gene disruptions in isolate 15A did not result in a change in phenotype as would be expected if *AvrHar* was an avirulence gene as its name suggests (**Figure 4.6**) (Weiland et al 1999). *AvrHar* gene disruptions in isolate 0-1 resulted in severe phenotypic changes that prevented spores from infecting leaf tissue on all barley lines tested (**Figure 4.6**). This result indicated that *AvrHar* encoded a protein involved in virulence and not avirulence.

To determine the causal polymorphisms between isolates 15A and 0-1, CRISPR-Cas9-RNP gene edits were performed to introduce a selectable marker at a targeted benign locus while simultaneously inducing a second double stranded break with a second CRISPR-Cas9-RNP and donor DNA designed for editing mutations in *AvrHar*. Complete gene replacements were accomplished in this way to replace the 0-1 and 15A alleles with the opposing allele, respectively. Neither editing the 0-1 allele to be the coding equivalent of 15A nor editing the 15A allele to be the 0-1 allele resulted in a phenotypic change. These results indicated that the promotor region of this gene may be the location of the causal polymorphism and gene editing of the immediate upstream promotor region would produce important information as to the control of this important virulence factor. This result is not surprising given the strongest marker-trait associations occurred in the genomic region immediately upstream of the start site of *AvrHar*. Additionally, RNAseq data previously obtained from *in planta* infection time points identified sequencing reads mapping to *AvrHar* in isolate 0-1 but, no transcription was detected in isolate 15A (Wyatt et al. 2018, Chapter 2; Wyatt et al. 2019, Chapter 3). The differences in detected

transcripts for *AvrHar* in isolates 0-1 and 15A presents further evidence implicating the promotor region as the location of causal polymorphisms.

Many plant pathogenic species exhibit presence absence polymorphisms in their respective effector repertoires (Bertazzonie et al. 2018). To examine if *AvrHar* presence/absence polymorphisms were observed in the *P. teres* f. *teres* natural population all sequences were assembled. *AvrHar* was found in a complete copy for every isolate having sufficient coverage at the *AvrHar* locus. Within the natural population seven alleles were identified with varying degrees of polymorphism and phenotypic variation. Haplogroups 1 and 2 that include parental isolate 0-1 were observed to be virulent on Tifang barley while the remaining haplogroups 3 through 7 were observed to be avirulent (**Figure 4.4**). Examination of the common polymorphisms between virulent and avirulent alleles showed a correlation with polymorphisms occurring in the promotor region of *AvrHar* alleles (**Figure 4.4**). Further gene editing experiments are needed to fully understand the causal polymorphism in the *AvrHar* alleles.

AvrHar orthologs were identified in several related plant pathogenic species (**Figure 4.5**). Many of the orthologs belong to other grass pathogens that infect rice, wheat, and corn as well as belonging to the *Brassica* pathogen *L. maculans*. Orthologs were only identified in Dothideomycete pathogens and *AvrHar* may represent a conserved pathogenicity effector protein. Functional characterization of *AvrHar* in the closely related species *P. teres* f. *maculata* and *P. tritici-repentis* as well as the other plant pathogens will yield useful insights into the conserved nature of *AvrHar*. *AvrHar* may even represent a target for engineering durable resistance to multiple plant pathogenic species.

Future prospects

Further experiments are needed to fully validate and functionally characterize *AvrHar*. Gene editing using CRISPR-Cas9-RNPs will facilitate fast and efficient interrogation of several important questions raised in the identification of *AvrHar*. Why are some alleles of *AvrHar* dispensable for certain isolates, like isolate 15A , and what critical function is *AvrHar* fulfilling in isolates such as 0-1? What is compensating for the critical function of *AvrHar* in isolates that lack *AvrHar* but are still pathogenic? What is the role of *AvrHar* orthologs in other plant pathogenic species and can durable resistance be engineered to target *AvrHar* in crops species?

Literature cited

Akhavan, A., Turkington, T. K., Kebede, B., Tekauz, A., Kutcher, H. R., Kirkham, C., Xi, K., Kumar, K., Tucker, J. R. and Strelkov, S. E. 2015. Prevalence of mating type idiomorphs in *Pyrenophora teres* f. *teres* and *P. teres* f. *maculata* populations from the Canadian prairies. Can. J. Plant Pathol. 37:52-60.

Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A. and Johnson, E. A. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PloS one, 3:e3376.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D. and Pyshkin, A. V. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19:455-477.

Beattie, A. D., Scoles, G. J., Rossnagel, B. G. 2007. Identification of molecular markers linked to a *Pyrenophora teres* avirulence gene. Phytopathol.y 97:842–849

Benjamini, Y., and Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Royal Stat. Soc. 57:289-300.

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

Bolger, A. M., Lohse, M., Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinform. 30:2114-2120.

Bonferroni, C. 1936. Teoria statistica delle classi e calcolo delle probabilita. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze. 8:3-62.

Brunner, P. C., & McDonald, B. A. 2018. Evolutionary analyses of the avirulence effector AvrStb6 in global populations of *Zymoseptoria tritici* identify candidate amino acids involved in recognition. Mol. Plant Pathol. 19:1836-1846.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. 2009. BLAST+: architecture and applications. BMC Bioinform. 10:421.

Carlsen, S. A., Neupane, A., Wyatt, N. A., Richards, J. K., Faris, J. D., Xu, S. S., Brueggeman, R. S. and Friesen, T. L. 2017. Characterizing the *Pyrenophora teres* f. *maculata*–barley interaction using pathogen genetics. G3. 7:2615-2626.

Ceroni, A., Passerini, A., Vullo, A., & Frasconi, P. 2006. DISULFIND: a disulfide bonding state and cysteine connectivity prediction server. Nucl Acids Res. 34:W177-W181.

Corpet, F. 1988. Multiple sequence alignment with hierarchical clustering. Nucl. Acids Res. 16:10881-10890.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T. and McVean, G. 2011. The variant call format and VCFtools. Bioinform. 27:2156-2158.

Ellwood, S. R., Liu, Z., Syme, R. A., Lai, Z., Hane, J. K., Keiper, F., Moffat, C. S., Oliver, R. P., Friesen, T. L. 2010. A first genome assembly of the barley fungal pathogen *Pyrenophora teres* f. *teres*. Genome Biol. 11:R109.

Faris, J. D., Liu, Z., & Xu, S. S. 2013. Genetics of tan spot resistance in wheat. Theor. Appl. Genet. 126:2197-2217.

Foster, A. J., Martin-Urdiroz, M., Yan, X., Wright, H. S., Soanes, D. M., & Talbot, N. J. 2018. CRISPR-Cas9 ribonucleoprotein-mediated co-editing and counterselection in the rice blast fungus. Sci. Reports. 8:14355.

Friesen, T. L., Faris, J. D., Lai, Z., & Steffenson, B. J. 2006. Identification and chromosomal location of major genes for resistance to *Pyrenophora teres* in a doubled-haploid barley population. Genome. 49:855-859.

Gao, Y., Liu, Z., Faris, J. D., Richards, J., Brueggeman, R. S., Li, X., Oliver, R. P., McDonald, B. A. and Friesen, T. L. 2016. Validation of genome-wide association studies as a tool to identify virulence factors in *Parastagonospora nodorum*. Phytopathol. 106:1177-1185.

Goodwin, S. B., M'Barek, S. B., Dhillon, B., Wittenberg, A. H., Crane, C. F., Hane, J. K., Foster, A. J., Van der Lee, T. A., Grimwood, J., Aerts, A. and Antoniw, J. 2011. Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. PLoS Genet. 7:e1002070.

Hartmann, F. E., McDonald, B. A., & Croll, D. 2018. Genome-wide evidence for divergent selection between populations of a major agricultural pathogen. Mol. Ecol. 27:2725-2741.

Heffelfinger, C., Fragoso, C. A., & Lorieux, M. 2017. Constructing linkage maps in the genomics era with MapDisto 2.0. Bioinform. 33:2224-2225.

Heigwer, F., Kerr, G., & Boutros, M. 2014. E-CRISP: fast CRISPR target site identification. Nat. Meth. 11:122.

Hill, W. G., & Robertson, A. 1968. Linkage disequilibrium in finite populations. Theor. Appl. Genet. 38:226-231.

Joehanes, R., & Nelson, J. C. 2008. QGene 4.0, an extensible Java QTL-analysis platform. Bioinform. 24:2788-2789.

Kema, G. H., Gohari, A. M., Aouini, L., Gibriel, H. A., Ware, S. B., van Den Bosch, F., Manning-Smith, R., Alonso-Chavez, V., Helps, J., M'Barek, S. B. and Mehrabi, R. 2018. Stress and sexual reproduction affect the dynamics of the wheat pathogen effector AvrStb6 and strobilurin resistance. Nat. Genet. 50:375.

Khan, T. N. and Boyd, W. J. R. 1969a Physiologic specialization in *Drechslera teres*. Aust. J. Biol. Sci. 22;1229-1235.

Koladia, V. M., Richards, J. K., Wyatt, N. A., Faris, J. D., Brueggeman, R. S., & Friesen, T. L. 2017. Genetic analysis of virulence in the Pyrenophora teres f. teres population BB25× FGOH04Ptt-21. Fungal Genet. Biol. 107:12-19.

Kolattukudy, P. E. 1985. Enzymatic penetration of the plant cuticle by fungal pathogens. Ann. Rev. Phytopathol. 23: 223-250.

Lai, Z., Faris, J. D., Weiland, J. J., Steffenson, B. J., & Friesen, T. L. 2007. Genetic mapping of *Pyrenophora teres* f. *teres* genes conferring avirulence on barley. Fungal Genet. Biol. 44:323-329.

Leboldus, J. M., Kinzer, K., Richards, J., Ya, Z., Yan, C., Friesen, T. L., & Brueggeman, R. 2015. Genotype-by-sequencing of the plant-pathogenic fungi *P yrenophora teres* and S*phaerulina musiva* utilizing Ion Torrent sequence technology. Mol. Plant Pathol. 16:623-632.

Lehmensiek, A., Bester, A. E., Sutherland, M. W., Platz, G., Kriel, W. M., Potgieter, G. F. and Prins, R. 2010 Population structure of South African and Australian *Pyrenophora teres* isolates. Plant Pathol. 595:504– 515.

Li, H., & Durbin, R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinform. 25:1754-1760.

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

Liu, Z., Ellwood, S. R., Oliver, R. P., and Friesen, T. L. 2011. *Pyrenophora teres*: profile of an increasingly damaging barley pathogen. Mol. Plant Pathol. 12:1-19

Liu, Z., Holmes, D. J., Faris, J. D., Chao, S., Brueggeman, R. S., Edwards, M. C., and Friesen, T. L. 2015. Necrotrophic effector-triggered susceptibility (NETS) underlies the barley-Pyrenophora teres f. teres interaction specific to chromosome 6H. Mol. Plant Pathol. 16:188-200

Liu, Z., Zhang, Z., Faris, J. D., Oliver, R. P., Syme, R., McDonald, M. C., McDonald, B. A., Solomon, P. S., Lu, S., Shelver, W. L. and Xu, S. 2012. The cysteine rich necrotrophic effector SnTox1 produced by *Stagonospora nodorum* triggers susceptibility of wheat lines harboring *Snn1*. PLoS Pathog. 8:p.e1002467.

Mathre, D. E. 1997. Compendium of Barley Diseases. The American Phytopathological Society. St. Paul

Nordborg, M., & Tavaré, S. 2002. Linkage disequilibrium: what history has to tell us. Trends Genet. 18:83-90.

Peever, T. L. and Milgroom, M. G. 1994. Genetic structure of Pyrenophora teres populations determined with random amplified polymorphic DNA markers. Can. J. Bot. 72:915-923

Petersen, T. N., Brunak, S., Von Heijne, G., & Nielsen, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Meth. 8:785.

Quinlan, A. R. 2014. BEDTools: the Swiss-army tool for genome feature analysis. Curr. Prot. Bioinformatics, 47:11-12.

Rau, D., Brown, A. H. D., Brubaker, C. L., Attene, G., Balmas, V., Saba, E., and Papa, R. 2003. Population genetic structure of *Pyrenophora teres* Drechs. the causal agent of net blotch in Sardinian landraces of barley (*Hordeum vulgare* L.). Theor. Appl. Genet. 106:947-959

Richards, J., Stukenbrock, E., Carpenter, J., Liu, Z., Cowger, C., Faris, J., & Friesen, T. L. 2019. Local adaptation drives the diversification of effectors in the fungal wheat pathogen *Parastagonospora nodorum* in the United States. PLOS Genet. e657007.

Serenius, M., Manninen, O., Wallwork, H., and Williams, K. 2007. Genetic differentiation in *Pyrenophora teres* populations measured with AFLP markers. Mycol. Res. 111:213-223

Shjerve, R. A., Faris, J. D., Brueggeman, R. S., Yan, C., Zhu, Y., Koladia, V., Friesen, T. L. 2014. Evaluation of a *Pyrenophora teres* f. *teres* mapping population reveals multiple independent interactions with a region of barley chromosome 6H. Fungal Genet. Biol. 70:104–112.

Sperschneider, J., Dodds, P. N., Gardiner, D. M., Singh, K. B., & Taylor, J. M. 2018a. Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. Mol. Plant Pathol. 192094-2110.

Sperschneider, J., Dodds, P. N., Singh, K. B., & Taylor, J. M. 2018b. ApoplastP: prediction of effectors and plant proteins in the apoplast using machine learning. New Phytol. 217:1764-1778.

Sperschneider, J., Gardiner, D. M., Dodds, P. N., Tini, F., Covarelli, L., Singh, K. B., Manners, J. M. and Taylor, J. M. 2016. EffectorP: predicting fungal effector proteins from secretomes using machine learning. New Phytol. 210:743-761.

Syme, R. A., Martin, A., Wyatt, N. A., Lawrence, J. A., Muria-Gonzalez, M. J., Friesen, T. L., & Ellwood, S. R. 2018. Transposable element genomic fissuring in *Pyrenophora teres* is associated with genome expansion and dynamics of host–pathogen genetic interactions. Front. Genet. 9:130.

Talas, F., & McDonald, B. A. 2015. Genome-wide analysis of *Fusarium graminearum* field populations reveals hotspots of recombination. BMC genomics, 16:996.

Talas, F., & McDonald, B. A. 2015. Significant variation in sensitivity to a DMI fungicide in field populations of *Fusarium graminearum*. Plant Pathol. 64:664-670.

Talas, F., Kalih, R., Miedaner, T., & McDonald, B. A. 2016. Genome-wide association study identifies novel candidate genes for aggressiveness, deoxynivalenol production, and azole sensitivity in natural field populations of *Fusarium graminearum*. Mol. Plant-Microbe Interact. 29:417-430.

Talas, F., Kalih, R., Miedaner, T., & McDonald, B. A. 2016. Genome-wide association study identifies novel candidate genes for aggressiveness, deoxynivalenol production, and azole sensitivity in natural field populations of *Fusarium graminearum*. Mol. Plant Microbe Interact. 29:417-430.

Tang, Y., Liu, X., Wang, J., Li, M., Wang, Q., Tian, F., Su, Z., Pan, Y., Liu, D., Lipka, A. E. and Buckler, E. S. 2016. GAPIT version 2: an enhanced integrated tool for genomic association and prediction. Plant Genome. 9:1.

Tekauz, A. 1985. A numerical scale to classify reactions of barley to *Pyrenophora teres*. Can. J. Plant Pathol. 7:181-183.

Tekauz, A. 1990. Characterization and distribution of pathogenic variation in *Pyrenophora teres* f. *teres* and *P. teres* f. *maculata* from western Canada. Can. J. Plant Pathol. 12:141-148.

Thomma, B. P., Seidl, M. F., Shi-Kunne, X., Cook, D. E., Bolton, M. D., van Kan, J. A., & Faino, L. 2016. Mind the gap; seven reasons to close fragmented genome assemblies. Fungal Genet. Biol. 90, 24-30.

Weiland, J. J., Steffenson, B. J., Cartwright, R. D., Webster, R. K. 1999. Identification of molecular genetic markers in *Pyrenophora teres* f. *teres* associated with low virulence on 'Harbin' barley. Phytopathol. 89:176–181.

Wyatt, N. A., Richards, J. K., Brueggeman, R. S., & Friesen, T. L. 2018. Reference assembly and annotation of the *Pyrenophora teres* f. *teres* isolate 0-1. G3. 8:1-8.

Wyatt, N. A., Richards, J., Brueggeman, R. S., & Friesen, T. L. 2019. A comparative genomic analysis of the barley pathogen *Pyrenophora teres* f. *teres* identifies sub-telomeric regions as drivers of virulence. Mol. Plant Microbe Interact.

Yoshida, K., Saitoh, H., Fujisawa, S., Kanzaki, H., Matsumura, H., Yoshida, K., Tosa, Y., Chuma, I., Takano, Y., Win, J. and Kamoun, S. 2009. Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *Magnaporthe oryzae*. Plant Cell. 21:1573-1591.

Zhong, Z., Marcel, T. C., Hartmann, F. E., Ma, X., Plissonneau, C., Zala, M., Ducasse, A., Confais, J., Compain, J., Lapalu, N. and Amselem, J. 2017. A small secreted protein in *Zymoseptoria tritici* is responsible for avirulence on wheat cultivars carrying the Stb6 resistance gene. New Phytol. 214:619-631.

Zhu, C., Gore, M., Buckler, E. S., & Yu, J. 2008. Status and prospects of association mapping in plants. Plant Genome. 1:5-20.

APPENDIX. CHAPTER 3 SUPPLEMENTARY MATERIALS AND METHODS

Extraction of high molecular weight DNA

Lysis buffer was prepared by combining 6.5 mL of Buffer A (350 mM sorbitol, 5mM EDTA, and 100 mM Tris-Cl), 6.5 mL of Buffer B (50 mM EDTA, 2000 mM NaCl, 200 mM Tris-Cl, and 2% CTAB), 2.6 mL of Buffer C (5% N-lauroylsarcosine), and 1.75 mL of 1% polyvinylpyrrolidone (PVP). A total of 500 mg of lyophilized tissue was added to a 50 mL conical tube (two tubes per isolate), followed by homogenization in 25 mL lysis buffer and 150 µL RNase A (20 mg/mL). The samples were incubated for 30-45 minutes at 65°, mixing every 15 minutes. A volume of 8.25 mL of 5 M potassium acetate (pH 7.5) was added to each tube, mixed by inversion, and incubated on ice for 30 minutes. Samples were centrifuged for 20 minutes at approximately 3,100 ×g at 4° and the aqueous phase was transferred to a new 50 mL conical tube. A 0.1 volume of 3 M sodium acetate (pH 5.2) and equal volume of room temperature isopropyl alcohol were added to each tube, mixed by inversion, and incubated at room temperature for 5 minutes. Precipitated DNA was collected with a glass hook and subsequently rinsed with 5 mL of freshly prepared 70% ethanol. The ethanol was removed; the DNA was rinsed and transferred to a new 1.5 mL tube. Excess ethanol was removed by pipetting and the DNA pellet freeze dried for 5-10 minutes. DNA was re-suspended in 500 µL H_2O and incubated at 65° for 30 minutes, followed by incubation at 4° overnight. Samples were then centrifuged at 2,000 × g for 2 minutes and the supernatants of samples from the same isolate were combined into new 15 mL conical tubes using a large bore pipette tip. High-molecular weight DNA was then purified using the Qiagen Genomic-Tip 100/G kit according to the manufacturer's recommended protocol.

RNAseq tissue collection

Cultures and inoculum of *P. teres* f. *teres* isolates 15A, 6A, FGOH04Ptt-21, and BB25 were prepared as described (Wyatt et al. 2018). Isolates 15A and 6A were inoculated on susceptible barley lines Kombar and Rika, respectively. Isolates FGOH04Ptt-21 and BB25 were each inoculated on the susceptible barley line Pinnacle. Seeds of barley lines Kombar, Rika, and Pinnacle were sown into a 96-conetainer rack surrounded by a border of the barley cultivar Tradition and grown under greenhouse conditions for approximately two weeks. For each RNAseq sample, five individual conetainers containing two seedlings each were selected for inoculation. The second fully extended leaf of each seedling was taped flat to a 24 × 30 cm plastic surface to provide an even surface to coat sample leaves with inoculum. After inoculation, plants were placed into a lighted mist chamber for 24 hrs with 100% relative humidity. After 24 hrs, plants were moved to a growth chamber with a temperature of 21° and a 12 hr photoperiod. Samples were collected by punching circular leaf discs from pre-designated regions of the leaf using a sterile 5mm hole-punch. Each leaf was punched a total of five times equating to 50 tissue samples per collected time point. Tissue was immediately flash frozen in liquid nitrogen and stored at -80° until RNA extraction. Liquid cultures of *P. teres* f. *teres* isolates 15A, 6A, FGOH04Ptt-21, and BB25 were prepared by incubating collected fungal spores in 75mL of Fries media (Koladia et al. 2017) for five days. Tissue was harvested, rinsed with sterile distilled water, flash frozen in liquid nitrogen, and stored at -80° until RNA extraction.

RNAsequencing

mRNA from each sample was extracted using the mRNA Direct Kit (ThermoFisher Scientific) following the manufacturer's protocol. RNAseq library preparation was done with the Illumina Truseq v.3 kit following the manufacturer's protocol and Truseq v.3 barcodes were

used to barcode each timepoint. The three replicates for each time point were pooled and given the same barcode. Quality and fragment size distribution of the prepared libraries were examined using an Agilent DNA chip on a bioanalyzer (Agilent, Santa Clara, CA). The output of the Illumina sequencing run was parsed with the program bcl2fastq2 (Illumina) and reads were input to FastQC for quality inspection (Andrews et al. 2016). Trimmomatic v0.36 was used to trim raw reads using the parameters HEADCROP:15, ILLUMINACLIP:2:30:10 with Illumina adapter and index sequences provided, and SLIDINGWINDOW:4:15 (Bolger et al. 2014). Trimmed reads aligned and assembled into transcripts following the protocol described in Pertea *et al*., (2016). FASTQ files of RNAseq reads were aligned using HISAT2 v2.0.5 (Kim et al. 2015). For each of the four *P*. *teres* f. *teres* isolates; 15A, 6A, FGOH04Ptt-21, and BB25, RNAseq reads were aligned to each previously assembled genome. The resulting alignment files were converted to BAM files and input to StringTie v1.3.2 for transcript assembly (Pertea et al. 2015).

Genome annotation

The Maker2 pipeline compiles multiple forms of evidence to infer the best gene models possible. Each of the four *P*. *teres* f. *teres* isolates 15A, 6A, FGOH04Ptt-21, and BB25 were annotated using the same procedure. *Ab initio* annotations were generated from the programs Augustus, utilizing the *Neurospora crassa* training set (Stanke et al. 2006), and Genemark-ES v.2 that contains a self-training algorithm for *ab initio* gene annotation (Ter-Hovhannisyan et al. 2008). These *ab initio* gene annotations were passed into the Maker2 pipeline in GFF3 file format with previously assembled transcripts derived from RNAseq of each respective isolate also in GFF3 format. Additionally, external protein evidence was supplied from the reference *P*. *teres* f. *teres* isolate 0-1 (Wyatt et al. 2018, Chapter 2) and *Pyrenophora tritici-repentis* isolate Pt-1C-BFP (Manning et al. 2013) in FASTA format. The Maker2 pipeline was run in three

iterations. In the first Maker2 iteration, options within the pipeline were "est2genome=1" and "protein2genome=1" which instruct the pipeline to use evidence from RNAseq data and BLAST results of supplied proteins to build gene models. Resulting gene models were then used to train SNAP, a trainable *ab initio* annotation program (Korf 2004). The second iteration of the Maker2 pipeline was then run with the SNAP training file to refine *ab initio* gene models which were lacking transcript or protein evidence. The output of the second Maker2 iteration was then used to retrain SNAP and rerun the Maker2 pipeline to further refine gene models and produce final annotations with the additional input of all gene models from all *P. teres* f. *teres* isolates.

Principal component analysis

A principal component analysis was performed on the single nucleotide polymorphisms identified between the five *P. teres* f. *teres* isolates using SNPs from the entire genome and just the coding regions of genes. The results of these two PCA were highly similar and the results from the whole genome SNPs PCA was reported. Princible components of the SNPs from the full genome encompassed 40.03%, 23.50%, and 21.41% of the variation for PC1, PC2, and PC3 respectively. Principal components of the SNPs from genic regions accounted for 40.07%, 23.93%, and 21.18% of the variation for PC1, PC2, and PC3 respectively. R2 of Eigenvectors compared between the Full Genome SNPS and the Genic SNPS: 0.989521 (0-1), 0.977594 (15A), 0.899027 (6A), 0.988017 (BB25), 0.996939 (FGOH04Ptt-21).

Additional supplementary files are supplied as supplementary file 1 containing gene orthology statistics and supplementary file 2 containing gene ontology enrichment statistics for the different classes of genes. (See supplementary files)

Table A1. The set of 21 small secreted proteins (SSPs) identified as unique to a single isolate of the five *P. teres* f. *teres* isolates in this study. The first column lists the geneID, the second column identifies the isolate the SSP was identified in, the third and fourth column show if the gene was identified by EffectorP or ApoplastP as belonging to those groups respectively (Y=yes, N=No).

Protein ID	Isolate	EffectorP prediction	ApoplastP prediction
CFE70_00011078	0-1	Effector	Apoplastic
DOD00_00000226	15A	Effector	Apoplastic
DOD00_00000227	15A	Effector	Non-Apoplastic
DOD00_00002227	15A	Effector	Apoplastic
DOD00_00003280	15A	Non-effector	Apoplastic
DOD00_00009748	15A	Non-effector	Apoplastic
DOD00_00011497	15A	Non-effector	Non-Apoplastic
DOD00_00011689	15A	Non-effector	Non-Apoplastic
DOD00_00011863	15A	Non-effector	Non-Apoplastic
DOD00_00011930	15A	Non-effector	Non-Apoplastic
DOD00_00012046	15A	Non-effector	Non-Apoplastic
DOD00_00012104	15A	Non-effector	Non-Apoplastic
DOD01_00000324	6A	Non-effector	Non-Apoplastic
DOD01_00003053	6A	Non-effector	Apoplastic
DOD01_00010782	6A	Effector	Non-Apoplastic
DOD03_00000109	FGOH04Ptt-21	Effector	Non-Apoplastic
DOD02_00001664	BB25	Non-effector	Non-Apoplastic
DOD02_00001806	BB25	Non-effector	Non-Apoplastic
DOD02_00002680	BB25	Effector	Non-Apoplastic
DOD02_00004633	BB25	Effector	Apoplastic
DOD02_00002680	BB25	Non-effector	Apoplastic

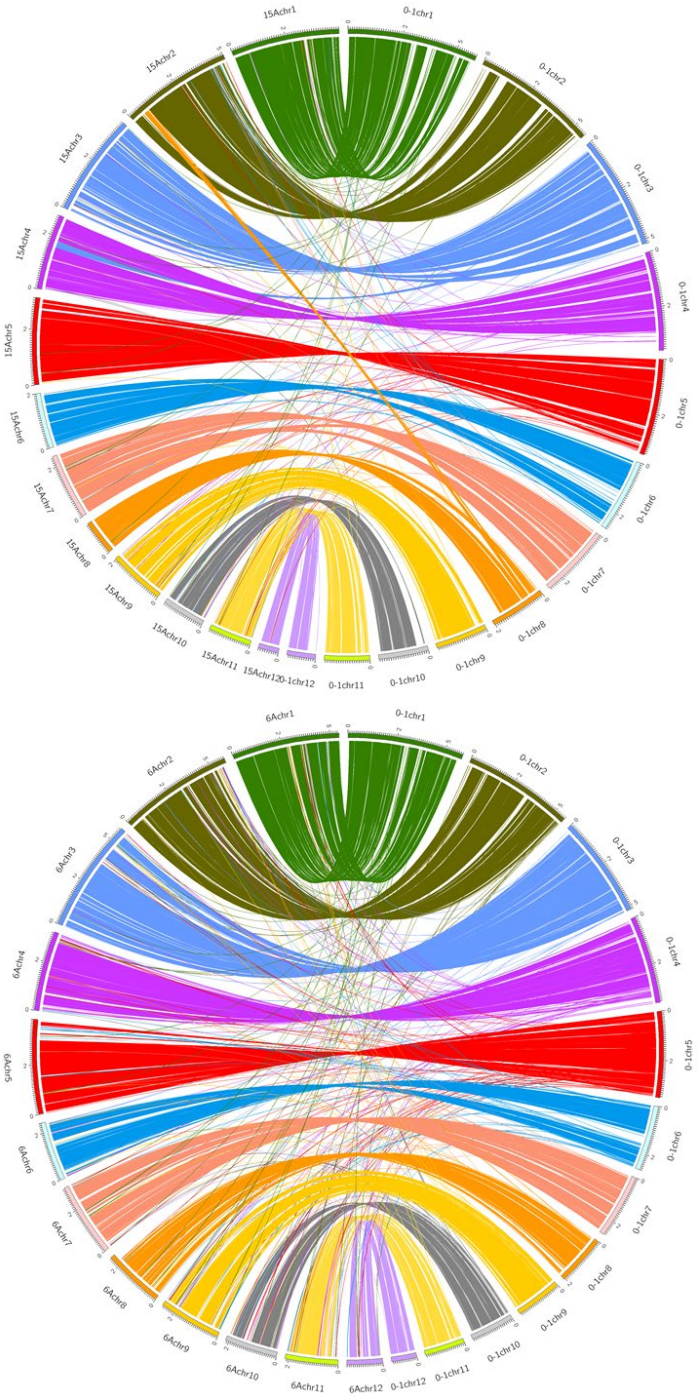

Figure A1. Whole genome synteny plot between *P. teres* f. *teres* isolates 0-1 and 15A. Bars comprising the outer ring represent individual chromosomes labeled with size (Mb) with tick marks measuring 100 kb. Each ribbon extending from one of the 0-1 chromosomes (0-1chr#) to one of the FGOH04Ptt-21 chromosomes (FGO21chr#) represents a 5 Kb syntenic region from 0-1's best hit in the 15A genome.

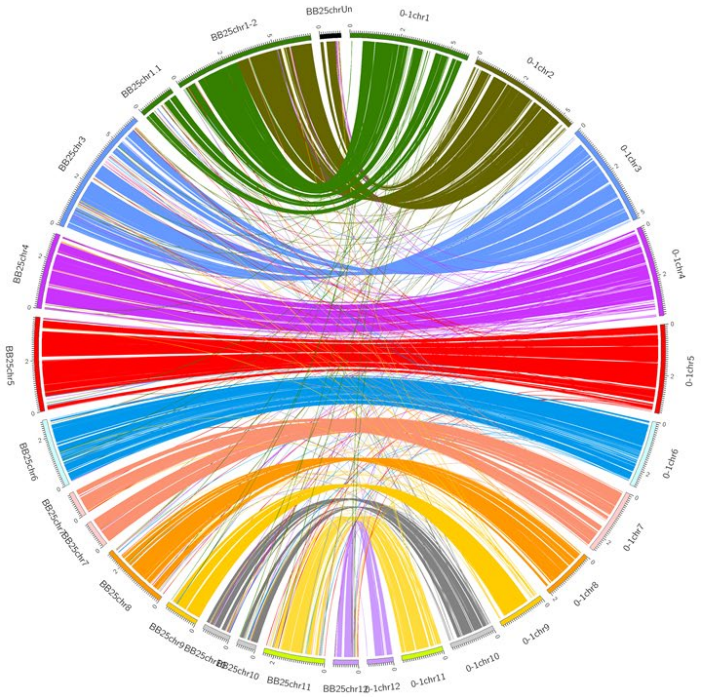

Figure A1. Whole genome synteny plot between P. teres f. teres isolates 0-1 and 15A (continued). Bars comprising the outer ring represent individual chromosomes labeled with size (Mb) with tick marks measuring 100 kb. Each ribbon extending from one of the 0-1 chromosomes (0-1chr#) to one of the FGOH04Ptt-21 chromosomes (FGO21chr#) represents a 5 Kb syntenic region from 0-1's best hit in the 15A genome.

Literature cited

Bolger, A. M., Lohse, M., and Usadel, B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120.

Kim, D., Langmead, B., and Salzberg, S. L. 2015. HISAT: A fast spliced aligner with low memory requirements. Nat. Methods 12:357-360.

Korf, I. 2004. Gene finding in novel genomes. BMC Bioinformatics 5:59.

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., Holman, W. H., Kodira, C. D., Martin, J., Oliver, R. P., Robbertse, B., Schackwitz, W., Schwartz, D. C., Spatafora, J. W., Turgeon, B. G., Yandava, C., Young, S., Zhou, S., Zeng, Q., Grigoriev, I. V., Ma, L. J., and Ciuffetti, L. M. 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 3:41-63.

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. 33:290-295.

Stanke, M., Sch¨offmann, O., Morgenstern, B., and Waack, S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics 7:62.

Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y. O., and Borodovsky, M. 2008. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res. 18:1979-1990.