Population genomics analysis of US *Fusarium graminearum* isolates and identification of the genetic basis of fitness traits

by

Upasana Dhakal

B.S., Tribhuvan University, 2015
M.S., University of Hawai‘i at Mānoa, 2018

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Plant Pathology
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2023

Abstract

Fusarium graminearum is a filamentous ascomycete fungus that causes Fusarium head blight (FHB) disease. FHB is present throughout the world and is one of the major constraints to wheat and barley production. In North America, *F. graminearum* is the major *Fusarium* species causing FHB, although many other *Fusarium* species are known to cause the disease. In the United States, *F. graminearum* exists in populations correlated to trichothecene chemotypes such as DON, 3ADON, 15ADON, NX-2 and NIV. Even though *F. graminearum* is one of the most well-studied plant pathogens, large-scale population level studies in *F. graminearum* are limited. Population level studies help define the population structure and diversity of the pathogen in the field and detect the direction and amount of gene flow between the populations, which is crucial in making disease management decisions. Genomic patterns in isolates across populations also can be harnessed to study how the populations are evolving and the influence of natural selection in pathogen adaptation. Genotypic variation uncovered through population genomics studies in combination with phenotypic variation can be used to identify the genetic determinants of pathogen traits using quantitative genomics approaches.

The broader goal of my research is to identify the genetic basis of pathogen fitness traits in *F. graminearum.* Growth rate, sexual fertility, aggressiveness, mycotoxin production, and fungicide sensitivity are traits of economic interest and the focus of my research projects. I attempted to identify the genetic basis of *F. graminearum* fitness through the indirect approach of scanning genomes of *F. graminearum* populations to detect the signs of recent positive natural selection. The assumption was that the regions under the influence of positive natural selection are likely to provide fitness advantage to the fungus. I identified 594 and 200 genomic regions in three *F. graminearum* populations under the influence of selection using two bioinformatics

software packages. The regions include many genes with unknown function and some genes with roles in plant-microbe interaction, fungicide/drug resistance, mycotoxin production, sexual spore production, transportation and genes that code for components of membranes and cellular organelles. Other population genomic analyses conducted identified frequent recombination, a moderate level of gene flow and high evolutionary potential in *F. graminearum* populations from the U.S.

I also used a more direct approach to detect the genetic basis of fitness traits in *F. graminearum* using genome wide association (GWAS) mapping. I identified a total of 13 significant quantitative trait nucleotides (QTNs) for growth rate, 30 for sexual fertility, 48 for propiconazole sensitivity, 37 for tebuconazole sensitivity, 8 for DON production *in vitro*, 3 for 15ADON production and 2 for aggressiveness in two wheat genotypes. Some of these QTNs fall on or are in close proximity to the genes with known roles for the respective trait. However, many of the QTNs are novel and provide new candidates for functional studies.

Finally, I sequenced the genomes of four isolates of *F. graminearum* using Oxford Nanopore sequencing, assembled the genomes *de novo* and used the assembled genomes to identify structural rearrangements and infer their role in *F. graminearum* adaptation and fitness. I detected a total of 87 inversions, 159 translocations, 245 duplications, 58,489 insertions and 34,102 deletions. Regions with high recombination rates are associated with structural rearrangements, and a significant proportion of inversions, translocations, and duplications overlap with the repeat content of the genome. Structural rearrangements in *F. graminearum* play an important role in shaping pathogen-host interactions by introducing presence-absence polymorphisms in secondary metabolite clusters and predicted effector genes, shuffling genomic segments and changing protein products and gene regulation.

Using GWAS and selection scans, I have identified candidate loci potentially determining fitness traits in *F. graminearum*. These loci provide new candidates for functional studies and, once verified, can be used as markers to monitor change in the genetic composition of field populations with respect to fitness-related traits such as fungicide sensitivity, aggressiveness and mycotoxin production.

Population genomics analysis of US *Fusarium graminearum* isolates and identification of the genetic basis of fitness traits

by

Upasana Dhakal

B.S., Tribhuvan University, 2015
M.S., University of Hawai‘i at Mānoa, 2018

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Plant Pathology
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2023

Approved by:

Major Professor
Dr. Christopher Toomajian

Copyright

Abstract

Fusarium graminearum is a filamentous ascomycete fungus that causes Fusarium head blight (FHB) disease. FHB is present throughout the world and is one of the major constraints to wheat and barley production. In North America, *F. graminearum* is the major *Fusarium* species causing FHB, although many other *Fusarium* species are known to cause the disease. In the United States, *F. graminearum* exists in populations correlated to trichothecene chemotypes such as DON, 3ADON, 15ADON, NX-2 and NIV. Even though *F. graminearum* is one of the most well-studied plant pathogens, large-scale population level studies in *F. graminearum* are limited. Population level studies help define the population structure and diversity of the pathogen in the field and detect the direction and amount of gene flow between the populations, which is crucial in making disease management decisions. Genomic patterns in isolates across populations also can be harnessed to study how the populations are evolving and the influence of natural selection in pathogen adaptation. Genotypic variation uncovered through population genomics studies in combination with phenotypic variation can be used to identify the genetic determinants of pathogen traits using quantitative genomics approaches.

The broader goal of my research is to identify the genetic basis of pathogen fitness traits in *F. graminearum.* Growth rate, sexual fertility, aggressiveness, mycotoxin production, and fungicide sensitivity are traits of economic interest and the focus of my research projects. I attempted to identify the genetic basis of *F. graminearum* fitness through the indirect approach of scanning genomes of *F. graminearum* populations to detect the signs of recent positive natural selection. The assumption was that the regions under the influence of positive natural selection are likely to provide fitness advantage to the fungus. I identified 594 and 200 genomic regions in three *F. graminearum* populations under the influence of selection using two bioinformatics

software packages. The regions include many genes with unknown function and some genes with roles in plant-microbe interaction, fungicide/drug resistance, mycotoxin production, sexual spore production, transportation and genes that code for components of membranes and cellular organelles. Other population genomic analyses conducted identified frequent recombination, a moderate level of gene flow and high evolutionary potential in *F. graminearum* populations from the U.S.

I also used a more direct approach to detect the genetic basis of fitness traits in *F. graminearum* using genome wide association (GWAS) mapping. I identified a total of 13 significant quantitative trait nucleotides (QTNs) for growth rate, 30 for sexual fertility, 48 for propiconazole sensitivity, 37 for tebuconazole sensitivity, 8 for DON production *in vitro*, 3 for 15ADON production and 2 for aggressiveness in two wheat genotypes. Some of these QTNs fall on or are in close proximity to the genes with known roles for the respective trait. However, many of the QTNs are novel and provide new candidates for functional studies.

Finally, I sequenced the genomes of four isolates of *F. graminearum* using Oxford Nanopore sequencing, assembled the genomes *de novo* and used the assembled genomes to identify structural rearrangements and infer their role in *F. graminearum* adaptation and fitness. I detected a total of 87 inversions, 159 translocations, 245 duplications, 58,489 insertions and 34,102 deletions. Regions with high recombination rates are associated with structural rearrangements, and a significant proportion of inversions, translocations, and duplications overlap with the repeat content of the genome. Structural rearrangements in *F. graminearum* play an important role in shaping pathogen-host interactions by introducing presence-absence polymorphisms in secondary metabolite clusters and predicted effector genes, shuffling genomic segments and changing protein products and gene regulation.

Using GWAS and selection scans, I have identified candidate loci potentially determining fitness traits in *F. graminearum*. These loci provide new candidates for functional studies and, once verified, can be used as markers to monitor change in the genetic composition of field populations with respect to fitness-related traits such as fungicide sensitivity, aggressiveness and mycotoxin production.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank all the people who have been with me and supported me in my PhD journey. I would like to express my sincere gratitude to my major professor, Dr. Chris Toomajian for giving me an opportunity to work on his research projects. I want to thank him for being a wonderful mentor and for always supporting, encouraging, and trusting me. I would also like to thank the members of my supervisory committee, Drs. John Leslie, Eric De Wolf, and Ari Jumpponen for their constructive feedback and guidance.

I am very thankful to the members of Leslie Lab, Amy Beyer, Bruce Ramundo, and Jaideep Mallick for their help with day-to-day lab work . I would also like to thank members of Toomajian Lab, Emily, Sandhya, and Sumit for their help with the laboratory and greenhouse experiment. I would like to express my gratitude to Dr. Sanzhen Liu for bioinformatics-related help, Dr. Paul St Amand for helping me set up the fungicide assay plates, Tim Todd for helping me with statistical analysis, Dr. Harold Trick, Dr. Jessica Rupp, and Dr. Shahideh Nouri for providing space for greenhouse experiments. I want to acknowledge the DON testing labs at North Dakota State University and Virginia Tech for testing my samples for mycotoxins.

I am very thankful to the Department of Plant Pathology and the College of Agriculture for the many awards and scholarships I have received over these years. I want to express my special thanks to the US Wheat and Barley Scab Initiative for funding my research. I am very grateful to my friends here in Manhattan, elsewhere in the US, and back home in Nepal for being a wonderful support system. I am immensely grateful to my family for their unconditional love and support, especially to my father and mother for all the sacrifices they have made in life so that I and my sisters could get the best education. Thank you all!

Chapter 1 – Introduction

1.1 *Fusarium graminearum* and Fusarium head blight

Fusarium graminearum is a filamentous ascomycete fungus in the order Hypocreales. The genus *Fusarium* contains over 300 species (O'Donnell et al., 2015), many of which cause devastating plant diseases. Some *Fusarium* species are opportunistic pathogens of humans and other animals. *Fusarium graminearum* causes Fusarium head blight (FHB) of wheat, barley, rice, and other cereal crops; stalk rot, kernel rot, and ear rot of corn, and seedling and root rot of wheat and corn (Goswami and Kistler, 2004).

FHB of wheat is one of the major diseases of wheat in North America and other wheat-growing areas in world. FHB was first described in England in 1884 (Stack, 2003). Since then, the disease has spread to all major continents and is one of the major constraints in wheat production worldwide. In North America, *F. graminearum* is the major causal agent of FHB (Goswami and Kistler, 2004). Wheat heads are susceptible to infection from anthesis to the soft dough stage and most susceptible at anthesis. Infected heads become prematurely bleached. Grains in infected heads are shriveled, discolored, and lightweight. When the weather conditions are favorable (high humidity during anthesis), FHB can destroy the crop in a matter of a few weeks (McMullen et al., 1997).

Besides *F. graminearum,* FHB is caused also by other *Fusarium* species like *F. asiaticum, F. culmorum,* and *F. poae* (Dill-Macky and Jones, 2000; Shah et al., 2018). *F. asiaticum* is the leading cause of FHB in parts of Asia. The United States first experienced FHB epidemics between the early 1910s to late 1930s (McMullen et al., 1997). The disease has re-emerged with major epidemics happening during the last three decades (McMullen et al., 2012, 1997). *Fusarium graminearum* infections not only reduce the grain yield but also contaminate

the grains with mycotoxins. Deoxynivalenol (DON) and its derivatives 3- acetyl deoxynivalenol (3ADON) and 15- acetyl deoxynivalenol (15ADON) are major mycotoxins produced by *F. graminearum.* Mycotoxin contaminated grains are unsuitable for human and animal consumption. Mycotoxin contamination results in discounted grain prices and adds to the loss incurred by farmers and stake holders. An economic analysis conducted by (Nganje et al., 2004) estimated the direct loss caused by FHB to be about $2.5 billion, and about $2 billion dollars were incurred as additional indirect losses in the United States between 1993 and 2001. Similarly, FHB caused $1.47 billion in direct and indirect loss in the United States in 2015-16 alone (Wilson et al., 2017).

1.2 Fusarium head blight disease cycle

The *Fusarium* species causing FHB survive as mycelia in the residues of small grain crops (Leplat et al., 2013). When the weather conditions are favorable (relative humidity >90% for an extended period and temperature between 15-30°C) resting mycelia will grow rapidly and produce perithecia (Xu, 2003). Corn residues are known to support formation of many more perithecia than wheat residue, as corn leaves behind more residue than wheat (Champeil et al., 2004). The primary inoculum for head blight is ascospores forcibly discharged from the perithecia (Wegulo et al., 2015; Xu, 2003). The ascospores can be airborne and travel long distances before they settle due to rain or gravitational force (Francl et al., 1999; Maldonado-Ramirez et al., 2005). Wheat is susceptible to infection when the plants are at anthesis to the soft dough stage, but are most susceptible at anthesis (Dill Macky, 2010). The spores landing on the anthers can germinate and grow down to the ovary and caryopsis (Bai and Shaner, 2004; Parry et al., 1995). The spores landing on the floral cavity germinate and enter the plant tissue through stomata in floral tissue (Pritsch et al., 2000). The disease starts on one or multiple spikelets and

progresses from one spikelet to another spikelet in the head (Goswami and Kistler, 2004). As the fungus moves from spikelet to spikelet through the rachis, invaded tissue dies and infected heads appear bleached (McMullen et al., 2012; Shah et al., 2018). Affected spikelets may have small, deformed grain or may fail to produce grain altogether (McMullen et al., 2012; Shah et al., 2018). Infected grains may have a reddish discoloration due to pigment production by the fungus (Desjardins, 2006). Susceptible cultivars are more prone to infection than are moderately resistant cultivars. Spread of the pathogen in the heads is reduced in resistant cultivars (Schroeder and Christensen, 1963). Secondary spread within the field is less important in the case of FHB because of the relatively narrow infection window of the pathogen (Wegulo et al., 2015). After the crop is harvested, the fungus survives in the infected residue and lodged heads that were not harvested. Infected grains can start a new disease cycle if they are planted the following season. Between wheat growing seasons, the fungus also survives on corn residue in the field. The disease cycle begins again when the weather conditions become favorable. The disease cycle begins again when the weather conditions become favorable.

1.3 Mycotoxins and chemotypes

The major mycotoxins produced by *F. graminearum* are trichothecenes and zearalenone (Chen et al., 2019). *Fusarium graminearum* isolates can be classified into chemotypes based on the dominant variant of the trichothecene mycotoxin produced by the isolate, which is determined by gas chromatography – mass spectrometry (GC-MS). *Fusarium graminearum* trichothecenes fall into type A or type B based on their structure. Type B trichothecenes, including deoxynivalenol (DON), 3-acetyl deoxynivalenol (3ADON), 15-acetyl deoxynivalenol (15ADON), and nivalenol (NIV), are characterized by the presence of a keto group at C-8 (Chen et al., 2019) and have historically been present in the US. DON, also known as vomitoxin,

interferes with protein synthesis and mitochondrial functions in eukaryotes (Szécsi and Bartók, 1995). Consumption of DON-contaminated feed/grains causes feed refusal and vomiting in animals (Ramírez Albuquerque et al., 2021). In 2015, a novel trichothecene 3α-acetoxy-7α,15-dihydroxy-12,13-epoxytrichothec-9-ene (NX-2) was identified as being produced by some isolates collected from Minnesota (Varga et al., 2015). NX-2 is a type A trichothecene and lacks a keto group at the C-8 position (Varga et al., 2015). NX-3, the deacetylated form of NX-2, has similar protein synthesis inhibitory capabilities as DON (Varga et al., 2015). Most of the genes in trichothecene biosynthesis are clustered into a single cluster of 12 genes known as the *TRI* locus (Chen et al., 2019). Besides the *TR*I locus, the *TRI1-TRI6* and *TRI101* loci also play a role in trichothecene biosynthesis (Alexander et al., 2009; Gale et al., 2005). Preliminary determination of chemotype of an isolate is often done by PCR assays that target variation in the *TRI* genes, although this determination, at times, does not match the chemotype identified via GC-MS (Crippin et al., 2019).

1.4 Population structure

Fusarium graminearum isolates from the US exist in populations correlated to trichothecene chemotypes. Previously, *F. graminearum* isolates sampled from the central and eastern US were considered part of a single diverse population with moderate isolation by distance, and isolates in different states just subsamples of this population (Zeller et al., 2004). Later, Gale et al. (2007) identified two distinct populations (15MW and 3UMW) correlated to trichothecene chemotypes 15ADON and 3ADON, respectively, using restriction fragment length polymorphism profiles of 713 isolates collected from nine Upper Midwest states between 1999 and 2000. The population structure is imperfectly correlated to trichothecene chemotype, however. Population 15MW is comprised of a majority of 15ADON isolates and a few 3ADON

isolates. Some 15ADON isolates, possibly admixed, could not be clearly placed into any population. Populations 15MW and 3UMW were later renamed NA1 and NA2 (Liang et al., 2014). In addition to the NA1 and NA2 populations, Kelly and Ward (2018) identified a third population of *F. graminearum* from the US correlated with a newly identified trichothecene chemotype, NX-2. The population was named NA3. Admixed isolates that could not be placed into any population were found in this study too. These isolates have proportions of ancestry shared among all three populations. Population NA1 is the largest and most diverse of the three known populations (Gale et al., 2007; Kelly and Ward, 2018). Population NA3 is speculated to be native to the US based on the high diversity observed in the population (Kelly and Ward, 2018). NA2 population is less diverse than NA1 and NA3 and is thought to have been introduced into the US at some point (Kelly and Ward, 2018). Populations NA1 and NA2 coexist in different parts of the country, while population NA3 is present in North Dakota, South Dakota, Minnesota (Liang et al., 2014), and New York (Lofgren et al., 2018) where the two other populations (NA1 and NA2) are also present. Other than the current restricted geographic distribution of the NA3 population, there is no obvious reproductive barrier between populations NA1, NA2 and NA3. Over the past two decades, a shift in *F. graminearum* chemotype frequency has been observed in some parts of the US, with isolates of the 3ADON chemotype increasing in frequency compared to the 15ADON chemotype (Liang et al., 2014). The reason for this shift in chemotype frequency remains unclear, although some publications indicate that 3ADON isolates are more fit compared to the 15ADON isolates (Puri and Zhong, 2010). Though not well studied compared to the three major populations from the central and northern US, the Gulf Coast population comprising 3ADON and NIV trichothecene chemotype isolates (Starkey et al., 2007)

and the Southern Louisiana population predominantly consisting of isolates of the NIV chemotype (Gale et al., 2011) have been described.

1.5 *Fusarium graminearum* genome and genomics

The *F. graminearum* genome was one of the first *Fusarium* genomes to be sequenced. It was originally sequenced using Sanger sequencing, and assembled into 433 contigs and 31 super contigs (Cuomo et al., 2007). In 2015, King et al. (2015) re-sequenced *F. graminearum* isolate PH-1 using Illumina paired-end sequencing and assembled a complete reference genome (38 Mb) with four chromosomes and one circular mitochondrial genome. The four chromosomes are 11.7 Mb, 8.9 Mb, 7.7 Mb, and 9.4 Mb in size, respectively. The mitochondrial genome is 95 kb long. Unlike *F. graminearum*, most other more distantly related *Fusarium* species like *F. verticilloides* have 12 chromosomes, each smaller in size than *F. graminearum* chromosomes (Navale et al., 2022). *Fusarium graminearum* chromosomes are believed to have originated following chromosome fusion in its ancestors (Cuomo et al., 2007; Ma et al., 2010). The reference PH-1 genome is predicted to have 14,164 genes. The genome of *F. graminearum* and some other filamentous fungi are referred to as two speed genomes, which are characterized by evolution of different compartments at different rates (Dong et al., 2015). The gene-sparse, repeat-rich regions of the genome are enriched for effector genes and rapidly evolve to adapt to changing environments, while the gene dense regions evolve slowly (Dong et al., 2015). In *F. graminearum*, the fast genome is also characterized by a higher rate of recombination, high SNP density and is enriched in genes with a role in plant microbe interaction (Ma et al., 2010). Kelly and Ward (2018) determined that orphan genes, genes missing in reference genome PH-1 but present in one or more *F. graminearum* isolates or highly divergent compared to PH-1, tend to reside in the faster evolving regions near the chromosome ends which are rich in repeat elements.

Some of the orphan genes also are differentially conserved among the populations, suggesting the role of these genes in population-specific adaptation.

The genome of *F. graminearum* reference isolate PH-1 has a very low repeat and transposable element content, as these elements only cover 0.24 percent of the genome (King et al., 2015). The low proportion of repeat sequences in PH-1 is due to repeat-induced point mutation (RIP) operating in *F. graminearum* (Cuomo et al., 2007; King et al., 2015). RIP detects the duplicated sequences during the sexual cycle (after fertilization and before karyogamy) and all duplicated copies are inundated with C to T transition mutations, making the copies nonfunctional (Ikeda et al., 2002). Some other *Fusarium* species have much more repeat content. Twenty-four percent of the *F. oxysporum* genome (isolate 4287) is made up of repetitive sequences (King et al., 2015; Ma et al., 2010). Lineage-specific chromosomes rich in host-specific pathogenicity genes described in other *Fusarium* species (Ma et al., 2010) are absent in *F. graminearum*.

1.6 Population genomics

Large-scale genomics studies dealing with *F. graminearum* populations are limited. Kelly and Ward (2018) re-sequenced 60 *F. graminearum* isolates collected from different parts of the US and Canada and investigated the population structure, genome-wide diversity of *F. graminearum* populations, and genomic regions showing signs of natural selection. In addition to populations NA1 and NA2, they reported a new population, NA3, correlated with the new NX-2 chemotype. Admixed isolates with ancestry shared among two or more populations were also reported. The NA1 population exhibits the highest average nucleotide diversity (0.0013) followed by NA3 (0.001) and NA2 (0.0006). These values are consistent with the hypothesis that population NA1 is endemic to the US while the NA2 population has recently clonally expanded

following a potential transcontinental introduction (Kelly and Ward, 2018; Ward et al., 2008). The NA3 population is thought to have recently evolved from a type-B trichothecene producing ancestor with a significant change of selection pressure at the *TRI1* locus and is likely endemic to Southern Canada and the Northern US (Kelly et al., 2016). Frequent identification of isolates that cannot be assigned into populations suggests recombination between isolates of two or more populations. However, the demographic history of *F. graminearum* populations and extent of gene flow between these populations is yet to be studied. Demographic modeling can also shed light into the size change and time since these populations diverged from their common ancestors. Populations evolve independently following separation from the most recent common ancestor. Beneficial new genomic variants with a role in survival, fitness and adaptation rapidly increase in frequency as they are positively selected. The identification of genomic regions under the influence of positive natural selection sheds light into the traits that are advantageous and indirectly provides the set of loci to use as candidates for functional studies which can identify new targets for disease management. Kelly and Ward (2018) identified 14 genomic regions, including the *TRI* locus, with reduced nucleotide diversity in one or more populations and increased population-wide divergence. The regions under the influence of selection were 10 kb to 30 kb in size and overlapped broader regions high in diversity and recombination. These 14 outlier regions contain core genes like membrane proteins and transcription factors as well as accessory genes with roles in membrane transport, the production of secondary metabolites, and gene regulation.

Patterns of nucleotide diversity and divergence in genomic loci can be used to identify genomic regions under the influence of selection (Nielsen et al., 2005; Sabeti et al., 2007; Stephan, 2019). When a beneficial variant goes to fixation, the variants in linkage disequilibrium

with the beneficial variant also move toward fixation. This phenomenon, called genetic hitchhiking, leads to reduced diversity around the region being positively selected for, known as a selective sweep (Nielsen et al., 2005). Selective sweeps can be hard or soft depending on the extent of variation around the selected site before the selection occurs, and result in differences in the amount of variation reduced in the region post-sweep. Hard sweeps completely eliminate the genetic variation in the region, while soft sweeps increase the frequency of one or more variants found on different genetic backgrounds, and so allow for the preservation of some variation while the beneficial variant is increasing in frequency and is in the process of fixation (Stephan, 2019). Software programs that search for the features associated with positive natural selection within a population, and in comparisons across populations, are available. The composite likelihood method implemented in Sweepfinder2 calculates the likelihood of selection for a particular genomic locus based on the deviation of the site frequency spectrum observed for the locus relative to the genomewide site frequency spectrum (Nielsen et al., 2005). The composite likelihood ratio is calculated by comparing the obtained likelihood to the likelihood of the model that does not allow for selection. While composite likelihood methods are robust to demographic changes and can detect relatively old signals, they have less power to detect soft sweeps. Cross population scans, such as those carried out with the XP-EHH statistic, compare extended haplotype homozygosity across populations (Sabeti et al., 2007). This is based on the principle that the selection events confer local adaptation, and selected regions are likely to be highly differentiated among populations. So, this test compares the haplotype homozygosity in two populations, considering one population as the test population and the other as the reference population. The ratio of haplotype homozygosity tells us if the locus has been selected for. A higher haplotype homozygosity in the test population compared to the reference population

indicates positive selection. Cross population selection scans are robust in detecting soft sweeps. These tools allow us to systematically scan genomes of *F. graminearum* populations to identify both hard and soft sweeps to build a comprehensive list of regions under the influence of recent positive natural selection.

1.7 Structural variants and their roles in fungal genomes

Chromosomal rearrangements - inversion, translocation, duplication, deletion and insertion - are large events caused by two or more double-stranded breaks in the chromosomes. Inversions flip the orientation of the chromosomal segment and can reverse the gene order. Translocations move chromosome segments to a different location in the same or a different chromosome. Both inversion and translocation preserve the DNA content of the genome (Hurles et al., 2008). However, chromosomal breaks associated with rearrangements, including inversion and translocation, are repaired by an error-prone method of DNA repair: non-homologous end joining. Repair of breaks in coding sequences can change protein products and make them non-functional and can change the patterns of gene expression (Hurles et al., 2008; Spielmann et al., 2018). Duplication, deletion and insertion, unlike inversion and translocation, change the DNA content and can introduce copy number variants and presence-absence polymorphisms in the genome (Spielmann et al., 2018).

While inversion and translocation themselves might not have harmful consequences, recombination between individuals carrying different alleles of inversion or translocation polymorphisms will produce progeny with deficient or duplicated segments (Shaikh, 2017). A single crossover within the inverted region leads to deficiency or duplications in the progeny, and double crossovers, which can balance DNA content, are rare unless the inversion involves a very large chromosomal segment. Progenies deficient in essential genes are not viable. For this

reason, inversions act as recombination suppressors (Boideau et al., 2022; Zhang et al., 2021). Inversions suppressing recombination have been reported from many species, including *Saccharomyces* (Dresser et al., 1994) and *Arabidopsis* (Ederveen et al., 2015; Zapata et al., 2016). Structural variants play an important role in shaping the recombination landscape, as reduction of recombination in the inverted regions is associated with increased recombination in the remainder of the chromosome (Ederveen et al., 2015; Termolino et al., 2019). Recombination within duplicated segments can lead to unequal crossing over and copy number variation of the genes within these segments in progenies (Shaikh, 2017). In maize, the *Rp1* locus that determines the resistance to maize rust is present in multiple copies and often leads to unequal recombination and susceptible progenies (Chavan et al., 2015; Hulbert and Bennetzen, 1991; Sun et al., 2001).

Most of the rearrangements originate following double-stranded breaks during meiotic recombination and transposable/repeat element activities in the genome (Mehrabi et al., 2017; Stukenbrock and Croll, 2014). Genomic features like GC content, transposable/repeat elements, recombination frequency and gene content are major predictors for the origin of structural rearrangements at a particular genomic locus in *Zymoseptoria tritici* and *Arabidopsis* (Badet et al., 2021). Chromosomal rearrangements are important in adaptation and evolution in organisms.

Even though small variants, *e.g*., single nucleotide polymorphisms (SNPs), have been well studied in *F. graminearum*, the systematic study of large-scale variants (structural rearrangements) in *F. graminearum* and other plant pathogenic fungi is limited (Kelly and Ward, 2018; Liang et al., 2014; Wang et al., 2017). This is mostly due to the fact that besides for the reference isolate PH-1, chromosomal-level assemblies were not available for other isolates until recently. Sometimes, the sizes of the structural rearrangements are on the order of hundreds of

kilobases or megabases, and it is not feasible to detect those large changes using genomes assembled into hundreds or thousands of small contigs. Long read sequencing technologies by Oxford Nanopore and PacBIO are facilitating the assembly of contiguous genomes for non-reference isolates. This will allow for the study of large-scale genomic variants and their role in fungal fitness and evolution. Structural rearrangements play an important role in determining phenotypic diversity, as they can change gene copy numbers, change gene expression and change gene products when coding sequences are interrupted by the chromosomal break points of the rearrangements. A few studies on structural rearrangements of plant pathogenic fungi have elucidated the roles these variants play in adaptation, fitness and genome evolution. In the powdery mildew pathogen *Erysiphe nectar*, copy number variation in *EnCYP51* genes is associated with fungal growth in the presence of DMI fungicide (Jones et al., 2014). The wheat pathogen *Zymoseptoria tritici* has ~14% of the SNPs associated with propiconazole resistance overlapping the structurally rearranged regions of the genome (Badet et al., 2021). The rice blast pathogen *Magnaporthe oryzae* has dispensable mini chromosomes present in some isolates which are rich in virulence factors, and the emergence of these chromosomes is facilitated by the rearrangement and duplication of core chromosome segments (Langner et al., 2021).

Species causing FHB in the *F. graminearum* species complex have high presence-absence diversity in secondary metabolite synthesis clusters. Secondary metabolites are small non-essential molecules with a wide range of functions like virulence, competition with other microbial populations, toxicity, and pigmentation, and can be essential for adaptation (Keller, 2019). The presence-absence diversity is due to multiple independent losses driven by complex rearrangement events facilitated by transposable elements or gains through horizontal gene transfer (Tralamazza et al., 2019). Frequent chromosomal rearrangement in the wheat leaf rust

fungus, *Puccinina triticina* is the major factor for the rapid divergence of *Pucciniales* fungi (Li et al., 2023). The presence and distribution of structural variants in *F. graminearum* genomes and their role in adaptation, fitness and evolution remains to be explored.

1.8 Important saprophytic and pathogenic traits in *Fusarium graminearum*

1.8.1 Growth rate

Fusarium graminearum survives as a saprophyte in crop residues when there are no hosts available for infection. A faster growth rate helps the fungus to rapidly colonize the residues and survive between the crop cycles. Growth rate of *F. graminearum* isolates on PDA and production of ascospores on corn stalks are significantly correlated (Spolti et al., 2014a). *Fusarium graminearum* growth rate is influenced by temperature, water availability, oxygen availability, and the types of residues on which it is growing. *Fusarium graminearum* mycelial growth is optimum at 25°C and water activity between 0.950 and 0.995 (Ramirez et al., 2006). *Fusarium graminearum* can survive on crop residues 20 to 25 cm into the soil for years but only grows well in residues on the soil surface (Champeil et al., 2004). Acidic and alkaline conditions limit the mycelial growth in *F. graminearum* (Thompson et al., 1993). Crop residues rich in carbon (high C:N ratio), like maize stem and wheat spikelet, favor growth whereas residues with a low C:N ratio, like wheat stem, support perithecia production (Khonga and Sutton, 1988). The quantity of crop residues available also influences the amount of growth.

1.8.2 Sexual fertility

Ascospores produced in corn and wheat residues from the previous growing seasons are the primary source of inoculum for FHB disease in a new growing season. Ascospores are produced in perithecia, the sexual fruiting bodies, and forcibly discharged into the air. Infection is started when a released ascospore landing on a wheat spikelet germinates. Thus, disease in the

field depends on the ability of the population to produce perithecia and ascospores (sexual fertility). Like the growth rate, sexual fertility is influenced by temperature, light, water availability, oxygen, and type of crop residue. Perithecia are formed between 2°C and 20°C (Inch and Gilbert, 2003) and ascospores are produced optimally between 25°C and 28°C (Sutton, 1982). Higher water activity favors ascospore production. Exposure to light is necessary for the formation of perithecial initials and their maturation as well as ascospore production (Gilbert and Tekauz, 2000; Sutton, 1982). When damaged kernels were placed at the soil surface, 5 cm deep and 10 cm deep into the soil, perithecia were observed on kernels at all depths, but ascospores were produced only on the kernels left on the soil surface (Sutton, 1982) likely due to the sunlight exposure on the surface. Residues towards the later stages of decomposition with depleted nutrient status support perithecia formation in contrast to fresh residues favoring mycelial growth (Leplat et al., 2013). Residues with a low C:N ratio like wheat stem promote perithecia formation while high C:N ratio residues promote growth (Khonga and Sutton, 1988). The total number of perithecia formed in the field also depends upon the type and quantity of available residues. Maize leaves behind more residues compared to wheat, thus more perithecia and ascospores are produced in wheat fields preceded by a maize crop (Morel 1996; Vilain 1989). However, wheat and barley residues support the production of higher ascospores per gram of residue compared to maize and other grasses (Pereyra and Dill-Macky, 2008). In the lab, sexual fertility is commonly assessed by growing *F. graminearum* on carrot agar plates (Leslie and Summerell, 2007).

1.8.3 Mycotoxin production

Fusarium graminearum isolates produce trichothecene mycotoxins. Deoxynivalenol (DON), 3-acetyl deoxynivalenol (3ADON), 15-acetyl deoxynivalenol (15ADON), and nivalenol

(NIV) are the major trichothecenes produced by *F. graminearum* (Miller et al., 1991). Trichothecenes, a class of secondary metabolites, are small sesquiterpenoid molecules with heterocyclic structures (Chen et al., 2019). Trichothecenes are not essential for survival but have an antimicrobial role and are important in helping the fungus go through biological stress and competition against other co-inhabiting microbes (Chen et al., 2019). The trichothecenes produced by *F. graminearum* also have a role in the virulence of the isolate. In the absence of *TRI5*, a gene that codes for an enzyme catalyzing the 1st step of trichothecene biosynthesis, *F. graminearum* can infect the wheat spike but the disease does not progress much beyond the point of infection (Proctor et al., 1995).

1.8.4 Aggressiveness

Fusarium graminearum survives saprophytically between crop hosts. Perithecia are formed on the surface of wheat/maize residues and release ascospores into the air when matured. These airborne ascospores are the primary source of inoculum for FHB (Goswami and Kistler, 2004; Parry et al., 1995). The ascospores enter through the stamens and grow down the stamen filaments. The infection progresses as the fungus grows from the infected spikelet to the healthy one through the rachis node and rachis internode (Brown et al., 2010). *Fusarium graminearum* isolates vary a lot in terms of aggressiveness. Isolates of 3ADON chemotype are reported to be more virulent compared to 15ADON isolates but there is no agreement on this in the literature (Spolti et al., 2014; Von der Ohe et al., 2010)

1.8.5 Fungicide sensitivity

In addition to cultural management practices like growing resistant cultivars, crop rotation, and crop residue management, fungicide application during anthesis is an important component of integrated FHB management. Demethylation inhibitor (DMI) fungicides like

propiconazole, metconazole, prothiconazole, and tebuconazole have been used to manage FHB in the US since the mid-1990s (McMullen et al., 2012). DMI fungicides inhibit the synthesis of ergosterol, an important component of fungal cell walls by binding to 14-a demethylases (the product of the CYP51 gene). Ergosterol synthesis is encoded by members of the CYP51 gene family in fungi. *Fusarium graminearum* has three CYP51 paralogs, *CYP51A*, *CYP51B*, and *CYP51C*, encoding 14-a demethylases (Liu et al., 2011). Intensive application of these fungicides has led to the development of fungicide resistance in some *F. graminearum* field isolates/populations collected from Asia, Europe, and the USA (Klix et al., 2007; Spolti et al., 2014b; Yin et al., 2009). Application of a mixture of two or more DMI fungicides is a common practice in disease management. Cross resistance of isolates to DMI fungicides tebuconazole and metconazole (Spolti et al., 2012) and between metconazole, tebuconazole, ipconazole, epoxiconazole, and difenoconazole (Duan et al., 2018) has been observed in some studies. Liu et al. (2022), however, report no cross-resistance between prothiconazole and tebuconazole.

1.9 Genetic determinants of saprophytic and pathogenic traits

Fusarium graminearum is one of the widely studied plant pathogenic fungi. Many functional genetics studies in this fungus have characterized the roles of different genes in survival, growth, reproduction, and fitness. Faster growth enables the fungus to rapidly colonize host cells and produce more ascospores. Among the genes identified to have a role in growth of *F. graminearum* are a lot of kinases and serine/threonine proteins, phosphatases and genes related to vacuolar sorting, cell division, transcription factors among others. The *F. graminearum* genome has over 100 protein kinase genes. Kinases phosphorylate cellular proteins and are important in signal transduction. Wang et al. (2011) studied the role of genes in the kinome of *F. graminearum* by generating knockout mutants. They identified 32 of the kinase genes that have a

critical role in growth, as mutants had abnormal and reduced growth, colony morphology and branching patterns. Growth in knockout mutants for two kinases, *GzSNF1* and *MGV1,* resulted in over 30% reduction in growth. Phosphatases, unlike kinases, have a role in dephosphorylation and also serve an important role in signal transduction. Phosphatases from *F. graminearum* have also been systematically studied by generation of deletion mutants. Deletion mutants for 22 phosphatases grew 20% slower compared to a wild type isolate (Yun et al., 2015). Vacuolar sorting proteins are components of the endosomal sorting complex required for transport (ESCRT) and play an important role in endocytosis. Mutants for vacuolar sorting proteins FGRAMPH1_01G00781 and FGRAMPH1_01G17487 are defective in growth phenotype likely due to the disruption of endocytosis (Xie et al., 2019). Son et al. (2011) identified 10 transcription factors regulating growth rate in *F. graminearum*, and some of these transcription factors are essential for the regulation of multiple traits.

Fusarium graminearum is a homothallic fungus. Sexual reproduction in *F. graminearum* is controlled by two mating loci (MAT). The *MAT-1-1* locus has three genes, *MAT-1-1-1*, *MAT-1-1-2* and *MAT-1-1-3*. Similarly, the *MAT-1-2* locus has two genes, *MAT-1-2-1* and *MAT-1-2-3*. Mutants for *MAT-1-2-1* produce no perithecia whereas mutant *MAT-1-1-1*, *MAT-1-1-2* and *MAT-1-1-3* isolates form perithecia smaller than the wild type isolates with no asci and ascospores (Kim et al., 2012). Reproduction is a complex process determined by the interaction of MAT loci and other transcription factors in *F. graminearum*. A SRF type transcription factor (FGRAMPH1_01G10593) interacts with *MAT-1-1-1* and regulates perithecium formation and phialide formation (Yang et al., 2015). Besides the master mating type loci, many kinase proteins are responsible for ascospore production and discharge in *F. graminearum*, similar to their role in growth rate. They include cAMP-dependent protein kinase type 3, serine/threonine protein

kinases, ste kinases, and camk kinases (Wang et al., 2011). *Gpmk1* is a MAPK associated protein kinase from *F. graminearum,* and disruption of this gene makes the isolates sterile (Jenczmionka et al., 2003). Mutants for *gpmk1* are also unable to make aerial mycelia and conidia, and sterility of these mutants could be associated with their inability to form aerial mycelia. Protein kinases associated with *gpmk1* MAPK pathways are also essential for sexual fertility in *F. graminearum* (Wang et al., 2011). Some other kinases like FGRAMPH1_01G03687, FGRAMPH1_10563, and FGRAMPH1_16417 are essential for ascospore discharge (Wang et al., 2011).

Genes coding for cell wall degrading enzymes, secondary metabolites, lipases, membrane-bound proteins, ATP binding transporters, and small secreted molecules are known to play a role in the virulence of *F. graminearum* (Rauwane et al., 2020). The mycotoxin DON is one of the major determinants of virulence in *F. graminearum*. Isolates defective in DON synthesis are restricted within the primary infection site (spikelet) and cannot move from one spikelet to another (Jansen et al., 2005; Proctor et al., 1995; Proctor et al., 1997). DON induces the production of reactive oxygen species which in turn induces programmed cell death. Increased virulence in DON-producing isolates may be due to programmed cell death facilitating the necrotrophic/intracellular colonization of the cells (Desmond et al., 2008). Intracellular colonization of non-lignified cells in the rachilla and rachis node is an important part of the pathway for the spikelet-to-spikelet colonization of wheat heads by *F. graminearum* (Brown et al., 2010). Similarly, the secondary metabolite fusaoctaxin A facilitates cell-to-cell penetration by *F. graminearum* and has a role in virulence, although the mechanism remains unknown (Jia et al., 2019). Cell wall degrading enzymes are important arsenals deployed by necrotrophic pathogens to successfully colonize their host tissues. Cell wall degrading enzymes like amylases, cellulases, proteases, and lipases facilitate initial infection and proliferation (Jenczmionka and

Schäfer, 2005). The Gpmk1 MAP kinase of *Fusarium* regulates the production of endoglucanase, xylanases, and protease enzymes (Jenczmionka and Schäfer, 2005). *Fusarium graminearum* derives nutrients from the cells macerated by these cell wall degrading enzymes. A secreted lipase, *FGL1*, is known to have a role in virulence in addition to the initial infection of the spikelet (Voigt et al., 2005). Disease severity was greatly reduced in an *FGL1* mutant strain in both wheat and maize (Voigt et al., 2005). ATP binding transporters and major family super transporters (MFS) can confer virulence to plant pathogens by helping them to transport toxic compounds produced by host plants, transport toxic compounds produced by pathogens into the host and protect the pathogen from its own toxins by compartmentalizing them (Alexander et al., 1999; Callahan et al., 1999; Skov et al., 2004). These transporters may serve such a role in *F. graminearum*. Disruption of a polyketide synthase gene, *PKS2*, in *F. graminearum* reduced virulence and mycelial growth (Gaffoor et al., 2005).

In *F. graminearum*, genes related to trichothecene biosynthesis are located in three genomic loci. 12 out of 15 genes identified to have a role in trichothecene biosynthesis are clustered in a single locus called the *TRI* locus, while another two genes cluster in the *TRI1-TRI16* locus and the remaining gene, *TRI101*, is not clustered with other genes in this pathway (Alexander et al., 2009; Gale et al., 2005). The *TRI* locus consists of *TRI* and *TRI10*, the major regulators of trichothecene biosynthesis genes, cytochrome 450 monooxygenases, hydrolases, and acetyltransferases for oxygenation, hydrolyzation, and transferring acetyl groups, respectively, to form intermediate and final products during trichothecene biosynthesis. Allelic variation in *TRI* genes is associated with variation in trichothecene biosynthesis in *F. graminearum* isolates. In DON producers, calonectrin is responsible for the hydroxylation of the C-7 and C-8 positions, and the hydroxylation of C-8 only in NX-2 producers is associated with

differences in the *TRI1* locus (Chen et al., 2019). Similarly, differential expression of *TRI8* is speculated to be associated with differential acetylation of the C-3 and C-15 positions in 3ADON and 15ADON producers (Alexander et al., 2009). Likewise, *TRI7* and *TRI13* are non-functional in DON producing isolates but functional in NIV producers where *TRI13* is responsible for C-4 hydroxylation (Alexander et al., 2009). The *TRI12* gene codes for a major facilitator transporter and works as an efflux pump to export trichothecene (Menke et al., 2012). In *F. graminearum*, the endoplasmic reticulum is remodeled into toxisomes where trichothecene is synthesized (Chen et al., 2019). It is speculated that other transporters along with *TRI12* play a role in exporting and compartmentalization of trichothecene, while the specific transporters and their roles are yet to be identified. It is highly likely that there are other genes besides *TRI* genes yet to be identified with major roles in trichothecene synthesis.

Resistance to azole fungicides has been reported in other fungal species besides *F. graminearum*. Mutation in CYP51 genes, overexpression of CYP51 genes, and transportation of fungicide out of fungal cells are some of the known mechanisms for azole resistance. Mutations in CYP51 are associated with azole fungicide resistance in the fungus, *Mycospharella graminicola* (Leroux et al., 2007). Repeated sequences upstream of CYP51 genes regulating overexpression of CYP51 genes in resistant isolates have been identified in the fungi *Blumeriella jaapii* and *Penicillium digitatum* (Hamamoto et al., 2000; Ma et al., 2006). In addition to overexpression, transportation of fungicides out of the fungal plasma membrane through efflux pumps is another mechanism conferring fungicide resistance in fungi (Hayashi et al., 2003; Reimann and Deising, 2005). Different DMI fungicides have a differential binding preference to different CYP51 paralogs in *F. graminearum* (Liu et al., 2011). *CYP51A* disruption mutants were sensitive to 7 DMIs including propiconazole and triadimefon whereas *CYP51C* disruption

mutants were sensitive to tebuconazole and four other DMIs but not to propiconazole (Liu et al., 2011). However, in field isolates, mutations in *CYP51A* and *CYP51C* genes are not correlated to DMI resistance (Talas and Mcdonald, 2015a). Talas and Mcdonald (2015a) also did not find any mutations likely regulating the overexpression of genes in the promoter regions for all three *F. graminearum* CYP51 paralogs. The contribution of genes outside the CYP51 gene family in DMI resistance and overexpression regulation of CYP51 genes by other transcription factors has not yet been explored in *F. graminearum*. Talas and Mcdonald (2015a), however, found high variability and heritability (0.97) in fungicide sensitivity among *F. graminearum* isolates. This opens the possibility of using genetic mapping in identifying the genetic basis of fungicide resistance/susceptibility in *F. graminearum*.

Most of the information on the roles of different *F. graminearum* genes in determining different traits come from studies that focus on specific subsets of genes, such as the kinome, phosphatome, effector repertoire and so on. Candidate gene-based approaches are likely to miss important genes and pathways as this approach is biased. Studies that scan the entire genome for genetic determinants of traits without any prior assumption or bias are limited in *F. graminearum*. Genome wide association studies (GWAS) scan the genomes free of any hypothesis, and the genome is screened without any inclination to a specific region or type of gene or variant (Kitsios and Zintzaras, 2009). GWAS in humans has identified previously unknown genes and pathways in many human diseases (Manolio et al., 2008).

1.10 Genome wide association

Genome-wide association studies (GWAS) associate the genetic variation (allele state) to the phenotypic variation (quantitative values for traits) to identify the genetic determinants of the trait. Genome-wide association studies have been successfully used in fungi for example to

identify the genetic determinants of aggressiveness (Atwell et al., 2018; Hartmann et al., 2017; Miedaner et al., 2021) and mycotoxin production (Miedaner et al., 2021). Talas et al. (2012) conducted GWAS with over 100 isolates of *F. graminearum* collected from Germany and identified 50, 29, and 74 significant SNPs associated with aggressiveness, DON production, and propiconazole sensitivity, respectively. The success of genome-wide association studies depends on the genetic architecture of the trait. When traits are governed by a few common genetic variants with large effects, the genetic loci determining the traits can be identified with a small sample size (Korte and Ashley, 2013). Usually, the phenotype/trait is the outcome of many common genetic variants with small effects and requires a large sample size to successfully map the trait using GWAS (Bush and Moore, 2012). Rare genetic variants with small effects are most difficult to associate and a very large sample size is necessary to successfully map these variants to the trait. Successful GWAS also depends on the patterns of recombination within the sampling population and linkage disequilibrium (LD) decay. Frequent recombination within the population promotes the rapid decay of LD with physical distance in the genome. The LD decay of the sample determines the resolution of association mapping as well as influences the marker density necessary for robust analysis. For a species where LD decays quickly, the resolution of mapping is better (Pang et al., 2020). Whereas, if LD extends to long distances, it would be possible to map the locus affecting a trait, but the resolution will be poor. The marker density necessary to detect most of the causal mutations/loci is directly proportional to the rate of LD decay (Zhou et al., 2012)**.** We need greater marker density in the region with a faster rate of LD decay because we are likely to miss the causal QTL if two markers are separated by a longer distance as none of them will be in LD with the causal mutation/locus (Pang et al., 2020). Even though *F. graminearum* is a homothallic fungus, it undergoes frequent outcrossing in nature. Previous

studies have reported rapid decay of LD in *F. graminearum,* indicating a fine mapping resolution in GWAS in this fungus (Talas and McDonald, 2015b). When individuals for GWAS are selected from more than one population, population structure due to non-random mating between the populations can result in false positive associations (Sul et al., 2018). Thus, it is a common practice to correct the population structure when GWAS involves more than one population. The kinship matrix or coefficient of relatedness among the individuals under study is incorporated into linear mixed models to account for the underlying population structure (François et al., 2016). Missing data due to poor sequencing or the nature of sequencing can be imputed to increase the power of GWAS (Korte and Ashley, 2013). Significant association of a marker with the phenotype of interest means the marker is either causal or is in LD with the causal locus (Sul et al., 2018).

1.11 References

Alexander, N.J., McCormick, S.P., Hohn, T.M., 1999. TRI12, a trichothecene efflux pump from *Fusarium sporotrichioides*: Gene isolation and expression in yeast. Mol. Gen. Genet. 261, 977–984. https://doi.org/10.1007/s004380051046

Alexander, N.J., Proctor, R.H., McCormick, S.P., 2009. Genes, gene clusters, and biosynthesis of trichothecenes and fumonisins in *Fusarium*. Toxin Rev. 28, 198–215. https://doi.org/10.1080/15569540903092142

Atwell, S., Corwin, J.A., 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.

Badet, T., Fouché, S., Hartmann, F.E., Zala, M., Croll, D., 2021. Machine-learning predicts genomic determinants of meiosis-driven structural variation in a eukaryotic pathogen. Nat. Commun. 12, 3551. https://doi.org/10.1038/s41467-021-23862-x

Bai, G., Shaner, G., 2004. Management and resistance in wheat and barley to fusarium head blight. Annu. Rev. Phytopathol. 42, 135–161. https://doi.org/10.1146/annurev.phyto.42.040803.140340

Boideau, F., Richard, G., Coriton, O., Huteau, V., Belser, C., Deniot, G., Eber, F., Falentin, C., Ferreira de Carvalho, J., Gilet, M., Lodé-Taburel, M., Maillet, L., Morice, J., Trotoux, G., Aury, J.M., Chèvre, A.M., Rousseau-Gueutin, M., 2022. Epigenomic and structural events preclude recombination in *Brassica napus*. New Phytol. 234, 545–559. https://doi.org/10.1111/nph.18004

Brown, N.A., Urban, M., van de Meene, A.M.L., Hammond-Kosack, K.E., 2010. The infection biology of *Fusarium graminearum*: Defining the pathways of spikelet to spikelet colonisation in wheat ears. Fungal Biol. 114, 555–571. https://doi.org/10.1016/j.funbio.2010.04.006

Bush, W.S., Moore, J.H., 2012. Chapter 11: Genome-Wide Association Studies. PLoS Comput. Biol. 8, e1002822. https://doi.org/10.1371/journal.pcbi.1002822

Callahan, T.M., Rose, M.S., Meade, M.J., Ehrenshaft, M., Upchurch, R.G., 1999. CFP, the putative cercosporin transporter of *Cercospora kikuchii*, is required for wild type cercosporin production, resistance, and virulence on soybean. Mol. Plant-Microbe Interact. 12, 901–910. https://doi.org/10.1094/MPMI.1999.12.10.901

Champeil, A., Doré, T., Fourbet, J.F., 2004. Fusarium head blight: Epidemiological origin of the effects of cultural practices on head blight attacks and the production of mycotoxins by

Fusarium in wheat grains. Plant Sci. 166, 1389–1415. https://doi.org/10.1016/j.plantsci.2004.02.004

Chavan, S., Gray, J., Smith, S.M., 2015. Diversity and evolution of Rp1 rust resistance genes in four maize lines. Theor. Appl. Genet. 128, 985–998. https://doi.org/10.1007/s00122-015-2484-2

Chen, Y., Kistler, H.C., Ma, Z., 2019. *Fusarium graminearum* Trichothecene Mycotoxins: Biosynthesis, Regulation, and Management. Annu. Rev. Phytopathol. 57, 15–39. https://doi.org/10.1146/annurev-phyto-082718-100318

Crippin, T., Renaud, J.B., Sumarah, M.W., David Miller, J., 2019. Comparing genotype and chemotype of *Fusarium graminearum* from cereals in Ontario, Canada. PLoS One 14, e0216735. https://doi.org/10.1371/journal.pone.0216735

Cuomo, C.A., Güldener, U., Xu, J.R., Trail, F., Turgeon, B.G., Di Pietro, A., Walton, J.D., Ma, L.J., Baker, S.E., Rep, M., Adam, G., Antoniw, J., Baldwin, T., Calvo, S., Chang, Y.L., DeCaprio, D., Gale, L.R., Gnerre, S., Goswami, R.S., Hammond-Kosack, K., Harris, L.J., Hilburn, K., Kennell, J.C., Kroken, S., Magnuson, J.K., Mannhaupt, G., Mauceli, E., Mewes, H.W., Mitterbauer, R., Muehlbauer, G., Münsterkötter, M., Nelson, D., O'Donnell, K., Ouellet, T., Qi, W., Quesneville, H., Roncero, M.I.G., Seong, K.Y., Tetko, I. V., Urban, M., Waalwijk, C., Ward, T.J., Yao, J., Birren, B.W., Kistler, H.C., 2007. The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. Science (80-.). 317, 1400–1402. https://doi.org/10.1126/science.1143708

Desjardins, A.E., 2006. Fusarium mycotoxins: chemistry, genetics, and biology., Fusarium mycotoxins: chemistry, genetics and biology.

Desmond, O.J., Manners, J.M., Stephens, A.E., Maclean, D.J., Schenk, P.M., Gardiner, D.M., Munn, A.L., Kazan, K., 2008. The Fusarium mycotoxin deoxynivalenol elicits hydrogen peroxide production, programmed cell death and defence responses in wheat. Mol. Plant Pathol. 9, 435–445. https://doi.org/10.1111/j.1364-3703.2008.00475.x

Dill-Macky, R., Jones, R.K., 2000. The effect of previous crop residues and tillage on Fusarium head blight of wheat. Plant Dis. 84, 71–76. https://doi.org/10.1094/PDIS.2000.84.1.71

Dill-Mackey, R., 2010. Fusarium head blight (scab). In Compendium of Wheat Diseases and Pests. American Phytopathological Society, APS press, St. Paul, 34e36

Dong, S., Raffaele, S., Kamoun, S., 2015. The two-speed genomes of filamentous pathogens: Waltz with plants. Curr. Opin. Genet. Dev. 35, 57–65. https://doi.org/10.1016/j.gde.2015.09.001

Dresser, M.E., Ewing, D.J., Harwell, S.N., Coody, D., Conrad, M.N., 1994. Nonhomologous synapsis and reduced crossing over in a heterozygous paracentric inversion in *Saccharomyces cerevisiae*. Genetics 138, 633–647. https://doi.org/10.1093/genetics/138.3.633

Duan, Y., Li, M., Zhao, H., Lu, F., Wang, J., Zhou, M., 2018. Molecular and biological characteristics of laboratory metconazole-resistant mutants in *Fusarium graminearum*. Pestic. Biochem. Physiol. 152, 55–61. https://doi.org/10.1016/j.pestbp.2018.08.011

Ederveen, A., Lai, Y., van Driel, M.A., Gerats, T., Peters, J.L., 2015. Modulating crossover positioning by introducing large structural changes in chromosomes. BMC Genomics 16, 1–12. https://doi.org/10.1186/s12864-015-1276-z

Francl, L., Shaner, G., Bergstrom, G., Gilbert, J., Pedersen, W., Dill-Macky, R., Sweets, L., Corwin, B., Jin, Y., Gallenberg, D., Wiersma, J., 1999. Daily inoculum levels of

Gibberella zeae on wheat spikes. Plant Dis. 83, 662–666. https://doi.org/10.1094/PDIS.1999.83.7.662

François, O., Martins, H., Caye, K., Schoville, S.D., 2016. Controlling false discoveries in genome scans for selection. Mol. Ecol. 25, 454–469. https://doi.org/10.1111/mec.13513

Gaffoor, I., Brown, D.W., Plattner, R., Proctor, R.H., Qi, W., Trail, F., 2005. Functional analysis of the polyketide synthase genes in the filamentous fungus *Gibberella zeae* (Anamorph *Fusarium graminearum*). Eukaryot. Cell 4, 1926–1933. https://doi.org/10.1128/EC.4.11.1926-1933.2005

Gale, L.R., Bryant, J.D., Calvo, S., Giese, H., Katan, T., O'Donnell, K., Suga, H., Taga, M., Usgaard, T.R., Ward, T.J., Kistler, H.C., 2005. Chromosome complement of the fungal plant pathogen *Fusarium graminearum* based on genetic and physical mapping and cytological observations. Genetics 171, 985–1001. https://doi.org/10.1534/genetics.105.044842

Gale, L.R., Harrison, S.A., Ward, T.J., O'Donnell, K., Milus, E.A., Gale, S.W., Kistler, H.C., 2011. Nivalenol-type populations of *Fusarium graminearum* and *F. asiaticum* are prevalent on wheat in southern Louisiana. Phytopathology 101, 124–134. https://doi.org/10.1094/PHYTO-03-10-0067

Gale, L.R., Ward, T.J., Balmas, V., Kistler, H.C., 2007. Population subdivision of *Fusarium graminearum* sensu stricto in the upper midwestern United States. Phytopathology 97, 1434–1439. https://doi.org/10.1094/PHYTO-97-11-1434

Gilbert, J., Tekauz, A., 2000. Review: Recent developments in research on Fusarium head blight of wheat in Canada. Can. J. Plant Pathol. 22, 1–8. https://doi.org/10.1080/07060660009501155

Goswami, R.S., Kistler, H.C., 2004. Heading for disaster: *Fusarium graminearum* on cereal crops. Mol. Plant Pathol. 5, 515–525. https://doi.org/10.1111/J.1364-3703.2004.00252.X

Hamamoto, H., Hasegawa, K., Nakaune, R., Lee, Y.J., Makizumi, Y., Akutsu, K., Hibi, T., 2000. Tandem repeat of a transcriptional enhancer upstream of the sterol 14α-demethylase gene (CYP51) in *Penicillium digitatum*. Appl. Environ. Microbiol. 66, 3421–3426. https://doi.org/10.1128/AEM.66.8.3421-3426.2000

Hartmann, F.E., Sánchez-Vallet, A., McDonald, B.A., Croll, D., 2017. A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. ISME J. 11, 1189–1204. https://doi.org/10.1038/ismej.2016.196

Hayashi, K., Schoonbeek, H. jan, De Waard, M.A., 2003. Modulators of membrane drug transporters potentiate the activity of the DMI fungicide oxpoconazole against *Botrytis cinerea*. Pest Manag. Sci. 59, 294–302. https://doi.org/10.1002/ps.637

He, X.L., Li, Q., Peng, W.H., Zhou, J., Cao, X.L., Wang, D., Huang, Z.Q., Tan, W., Li, Y., Gan, B.C., 2017. Intra- and inter-isolate variation of ribosomal and protein-coding genes in *Pleurotus*: Implications for molecular identification and phylogeny on fungal groups. BMC Microbiol. 17, 1–9. https://doi.org/10.1186/s12866-017-1046-y

Hulbert, S.H., Bennetzen, J.L., 1991. Recombination at the Rp1 locus of maize. MGG Mol. Gen. Genet. 226, 377–382. https://doi.org/10.1007/BF00260649

Hurles, M.E., Dermitzakis, E.T., Tyler-Smith, C., 2008. The functional impact of structural variation in humans. Trends Genet. 24, 238–245. https://doi.org/10.1016/j.tig.2008.03.001

Ikeda, K.I., Nakayashiki, H., Kataoka, T., Tamba, H., Hashimoto, Y., Tosa, Y., Mayama, S., 2002. Repeat-induced point mutation (RIP) in *Magnaporthe grisea*: Implications for its

sexual cycle in the natural field context. Mol. Microbiol. 45, 1355–1364. https://doi.org/10.1046/j.1365-2958.2002.03101.x

Inch, S.A., Gilbert, J., 2003. Survival of *Gibberella zeae* in Fusarium-damaged wheat kernels. Plant Dis. 87, 282–287. https://doi.org/10.1094/PDIS.2003.87.3.282

Jansen, C., Von Wettstein, D., Schäfer, W., Kogel, K.H., Felk, A., Maier, F.J., 2005. Infection pattern in barley and wheat spikes inoculated with wild-type and trichodiene synthase gene disrupted *Fusarium graminearum*. Proc. Natl. Acad. Sci. U. S. A. 102, 16892–16897. https://doi.org/10.1073/pnas.0508467102

Jenczmionka, N.J., Maier, F.J., Lösch, A.P., Schäfer, W., 2003. Mating, conidiation and pathogenicity of *Fusarium graminearum*, the main causal agent of the head-blight disease of wheat, are regulated by the MAP kinase gpmk1. Curr. Genet. 43, 87–95. https://doi.org/10.1007/s00294-003-0379-2

Jenczmionka, N.J., Schäfer, W., 2005. The Gpmk1 MAP kinase of *Fusarium graminearum* regulates the induction of specific secreted enzymes. Curr. Genet. 47, 29–36. https://doi.org/10.1007/s00294-004-0547-z

Jia, L.J., Tang, H.Y., Wang, W.Q., Yuan, T.L., Wei, W.Q., Pang, B., Gong, X.M., Wang, S.F., Li, Y.J., Zhang, D., Liu, W., Tang, W.H., 2019. A linear nonribosomal octapeptide from *Fusarium graminearum* facilitates cell-to-cell invasion of wheat. Nat. Commun. 10, 922. https://doi.org/10.1038/s41467-019-08726-9

Jones, L., Riaz, S., Morales-Cruz, A., Amrine, K.C.H., McGuire, B., Gubler, W.D., Walker, M.A., Cantu, D., 2014. Adaptive genomic structural variation in the grape powdery mildew pathogen, *Erysiphe necator*. BMC Genomics 15, 1–8. https://doi.org/10.1186/1471-2164-15-1081

Keller, N.P., 2019. Fungal secondary metabolism: regulation, function and drug discovery. Nat. Rev. Microbiol. 17, 167–180. https://doi.org/10.1038/s41579-018-0121-1

Kelly, A., Proctor, R.H., Belzile, F., Chulze, S.N., Clear, R.M., Cowger, C., Elmer, W., Lee, T., Obanor, F., Waalwijk, C., Ward, T.J., 2016. The geographic distribution and complex evolutionary history of the NX-2 trichothecene chemotype from *Fusarium graminearum*. Fungal Genet. Biol. 95. https://doi.org/10.1016/j.fgb.2016.08.003

Kelly, A.C., Ward, T.J., 2018. Population genomics of *Fusarium graminearum* reveals signatures of divergent evolution within a major cereal pathogen. PLoS One 13, e0194616. https://doi.org/10.1371/journal.pone.0194616

Khonga, E.B., Sutton, J.C., 1988. Inoculum production and survival of *Gibberella zeae* in maize and wheat residues. Can. J. Plant Pathol. 10, 232–239. https://doi.org/10.1080/07060668809501730

Kim, H.K., Cho, E.J., Lee, S., Lee, Y.S., Yun, S.H., 2012. Functional analyses of individual mating-type transcripts at MAT loci in *Fusarium graminearum* and *Fusarium asiaticum*. FEMS Microbiol. Lett. 337, 89–96. https://doi.org/10.1111/1574-6968.12012

King, R., Urban, M., Hammond-Kosack, M.C.U., Hassani-Pak, K., Hammond-Kosack, K.E., 2015. The completed genome sequence of the pathogenic ascomycete fungus *Fusarium graminearum*. BMC Genomics 16, 1–21. https://doi.org/10.1186/s12864-015-1756-1

Kitsios, G.D., Zintzaras, E., 2009. Genome-wide association studies: hypothesis-"free" or "engaged"? Transl. Res. 154, 161–164. https://doi.org/10.1016/j.trsl.2009.07.001

Klix, M.B., Verreet, J.A., Beyer, M., 2007. Comparison of the declining triazole sensitivity of *Gibberella zeae* and increased sensitivity achieved by advances in triazole fungicide development. Crop Prot. 26, 683–690. https://doi.org/10.1016/j.cropro.2006.06.006

Korte, A., Ashley, F., 2013. The advantages and limitations of trait analysis with GWAS : a review Self-fertilisation makes Arabidopsis particularly well suited to GWAS. Plant Methods 9, 1–9.

Langner, T., Harant, A., Gomez-Luciano, L.B., Shrestha, R.K., Malmgren, A., Latorre, S.M., Burbano, H.A., Win, J., Kamoun, S., 2021. Genomic rearrangements generate hypervariable mini-chromosomes in host-specific isolates of the blast fungus. PLoS Genet. 17, e1009386. https://doi.org/10.1371/JOURNAL.PGEN.1009386

Leplat, J., Friberg, H., Abid, M., Steinberg, C., 2013. Survival of *Fusarium graminearum*, the causal agent of Fusarium head blight. A review. Agron. Sustain. Dev. 33, 97–111. https://doi.org/10.1007/s13593-012-0098-5

Leroux, P., Albertini, C., Gautier, A., Gredt, M., Walker, A.S., 2007. Mutations in the CYP51 gene correlated with changes in sensitivity to sterol 14α-demethylation inhibitors in field isolates of *Mycosphaerella graminicola*. Pest Manag. Sci. 63, 688–698. https://doi.org/10.1002/ps.1390

Leslie, J.F., Summerell, B.A., 2007. The Fusarium Laboratory Manual, The Fusarium Laboratory Manual. https://doi.org/10.1002/9780470278376

Li, C., Qiao, L., Lu, Y., Xing, G., Wang, X., Zhang, G., Qian, H., Shen, Y., Zhang, Y., Yao, W., Cheng, K., Ma, Z., Liu, N., Wang, D., Zheng, W., 2023. Gapless Genome Assembly of *Puccinia triticina* Provides Insights into Chromosome Evolution in *Pucciniales*. Microbiol. Spectr. 11, e02828-22. https://doi.org/10.1128/spectrum.02828-22

Liang, J.M., Xayamongkhon, H., Broz, K., Dong, Y., McCormick, S.P., Abramova, S., Ward, T.J., Ma, Z.H., Kistler, H.C., 2014. Temporal dynamics and population genetic structure

of *Fusarium graminearum* in the upper Midwestern United States. Fungal Genet. Biol. 73, 83–92. https://doi.org/10.1016/j.fgb.2014.10.002

Liu, J., Jiang, J., Guo, X., Qian, L., Xu, J., Che, Z., Chen, G., Liu, S., 2022. Sensitivity and Resistance Risk Assessment of *Fusarium graminearum* from Wheat to Prothioconazole. Plant Dis. 106, 2097–2104. https://doi.org/10.1094/PDIS-12-21-2684-RE

Liu, X., Yu, F., Schnabel, G., Wu, J., Wang, Z., Ma, Z., 2011. Paralogous cyp51 genes in *Fusarium graminearum* mediate differential sensitivity to sterol demethylation inhibitors. Fungal Genet. Biol. 48, 113–123. https://doi.org/10.1016/j.fgb.2010.10.004

Lofgren, L., Riddle, J., Dong, Y., Kuhnem, P.R., Cummings, J.A., Del Ponte, E.M., Bergstrom, G.C., Kistler, H.C., 2018. A high proportion of NX-2 genotype strains are found among *Fusarium graminearum* isolates from northeastern New York State. Eur. J. Plant Pathol. 150, 791–796. https://doi.org/10.1007/s10658-017-1314-6

Ma, L.J., Van Der Does, H.C., Borkovich, K.A., Coleman, J.J., Daboussi, M.J., Di Pietro, A., Dufresne, M., Freitag, M., Grabherr, M., Henrissat, B., Houterman, P.M., Kang, S., Shim, W.B., Woloshuk, C., Xie, X., Xu, J.R., Antoniw, J., Baker, S.E., Bluhm, B.H., Breakspear, A., Brown, D.W., Butchko, R.A.E., Chapman, S., Coulson, R., Coutinho, P.M., Danchin, E.G.J., Diener, A., Gale, L.R., Gardiner, D.M., Goff, S., Hammond-Kosack, K.E., Hilburn, K., Hua-Van, A., Jonkers, W., Kazan, K., Kodira, C.D., Koehrsen, M., Kumar, L., Lee, Y.H., Li, L., Manners, J.M., Miranda-Saavedra, D., Mukherjee, M., Park, G., Park, J., Park, S.Y., Proctor, R.H., Regev, A., Ruiz-Roldan, M.C., Sain, D., Sakthikumar, S., Sykes, S., Schwartz, D.C., Turgeon, B.G., Wapinski, I., Yoder, O., Young, S., Zeng, Q., Zhou, S., Galagan, J., Cuomo, C.A., Kistler, H.C., Rep,

M., 2010. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. Nature 464, 367–373. https://doi.org/10.1038/nature08850

Ma, Z., Proffer, T.J., Jacobs, J.L., Sundin, G.W., 2006. Overexpression of the 14α-demethylase target gene (CYP51) mediates fungicide resistance in *Blumeriella jaapii*. Appl. Environ. Microbiol. 72, 2581–2585. https://doi.org/10.1128/AEM.72.4.2581-2585.2006

Maldonado-Ramirez, S.L., Schmale, D.G., Shields, E.J., Bergstrom, G.C., 2005. The relative abundance of viable spores of *Gibberella zeae* in the planetary boundary layer suggests the role of long-distance transport in regional epidemics of Fusarium head blight. Agric. For. Meteorol. 132, 20–27. https://doi.org/10.1016/j.agrformet.2005.06.007

Manolio, T.A., Brooks, L.D., Collins, F.S., 2008. A HapMap harvest of insights into the genetics of common disease. J. Clin. Invest. 118, 1590–1605. https://doi.org/10.1172/JCI34772

McMullen, M., Bergstrom, G., De Wolf, E., Dill-Macky, R., Hershman, D., Shaner, G., Van Sanford, D., 2012. A unified effort to fight an enemy of wheat and barley: Fusarium head blight. Plant Dis. 96, 1712–1728. https://doi.org/10.1094/PDIS-03-12-0291-FE

McMullen, M., Jones, R., Gallenberg, D., 1997. Scab of wheat and barley: A re-emerging disease of devastating impact. Plant Dis. 81, 1340–1348. https://doi.org/10.1094/PDIS.1997.81.12.1340

Mehrabi, R., Mirzadi Gohari, A., Kema, G.H.J., 2017. Karyotype Variability in Plant-Pathogenic Fungi. Annu. Rev. Phytopathol. 55, 483–503. https://doi.org/10.1146/annurev-phyto-080615-095928

Menke, J., Dong, Y.N., Kistler, H.C., 2012. *Fusarium graminearum* Tri12p influences virulence to wheat and trichothecene accumulation. Mol. Plant-Microbe Interact. 25, 1408–1418. https://doi.org/10.1094/MPMI-04-12-0081-R

Miedaner, T., Vasquez, A., Castiblanco, V., Castillo, H.E., Foroud, N., Würschum, T., Leiser, W., 2021. Genome-wide association study for deoxynivalenol production and aggressiveness in wheat and rye head blight by resequencing 92 isolates of *Fusarium culmorum*. BMC Genomics 22, 1–16. https://doi.org/10.1186/s12864-021-07931-5

Miller, J.D., Greenhalgh, R., Wang, Y., Lu, M., 1991. Trichothecene Chemotypes of Three *Fusarium* Species. Mycologia 83, 121–130. https://doi.org/10.2307/3759927

Navale, V.D., Sawant, A.M., Gowda, V.U., Vamkudoth, K.R., 2022. Assembly, Annotation, and Comparative Whole Genome Sequence of *Fusarium verticillioides* Isolated from Stored Maize Grains. Pathogens 11, 810. https://doi.org/10.3390/pathogens11070810

Nganje, W.E., Kaitibie, S., Wilson, W.W., Leistritz, L.F., Bangsund, D.A., 2004. Economic impacts of Fusarium head blight in wheat and barley: 1993-2001, Agribusiness & Applied Economics.

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M.J., Clark, A.G., Bustamante, C., 2005. Genomic scans for selective sweeps using SNP data. Genome Res. 15, 1566–1575. https://doi.org/10.1101/gr.4252305

O'Donnell, K., Ward, T.J., Robert, V.A.R.G., Crous, P.W., Geiser, D.M., Kang, S., 2015. DNA sequence-based identification of *Fusarium*: Current status and future directions. Phytoparasitica 43, 583–595. https://doi.org/10.1007/s12600-015-0484-z

Pang, Y., Liu, C., Wang, D., St. Amand, P., Bernardo, A., Li, W., He, F., Li, L., Wang, L., Yuan, X., Dong, L., Su, Y., Zhang, H., Zhao, M., Liang, Y., Jia, H., Shen, X., Lu, Y., Jiang, H., Wu, Y., Li, A., Wang, H., Kong, L., Bai, G., Liu, S., 2020. High-Resolution Genome-wide Association Study Identifies Genomic Regions and Candidate Genes for Important

Agronomic Traits in Wheat. Mol. Plant 13, 1311–1327. https://doi.org/10.1016/j.molp.2020.07.008

Parry, D.W., Jenkinson, P., McLeod, L., 1995. Fusarium ear blight (scab) in small grain cereals—a review. Plant Pathol. 44, 207–238. https://doi.org/10.1111/j.1365-3059.1995.tb02773.x

Pereyra, S.A., Dill-Macky, R., 2008. Colonization of the residues of diverse plant species by *Gibberella zeae* and their contribution to fusarium head blight inoculum. Plant Dis. 92, 800–807. https://doi.org/10.1094/PDIS-92-5-0800

Pritsch, C., Muehlbauer, G.J., Bushnell, W.R., Somers, D.A., Vance, C.P., 2000. Fungal development and induction of defense response genes during early infection of wheat spikes by *Fusarium graminearum*. Mol. Plant-Microbe Interact. 13, 159–169. https://doi.org/10.1094/MPMI.2000.13.2.159

Proctor, R.H., Hohn, T.M., McCormick, S.P., 1997. Restoration of wild-type virulence to Tri5 disruption mutants of *Gibberella zeae* via gene reversion and mutant complementation. Microbiology 143, 2583–2591. https://doi.org/10.1099/00221287-143-8-2583

Proctor, R.H., Hohn, T.M., McCormick, S.P., 1995. Reduced virulence of *Gibberella zeae* caused by disruption of a trichothecene toxin biosynthetic gene. Mol. Plant-Microbe Interact. 8, 593–601. https://doi.org/10.1094/MPMI-8-0593

Puri, K.D., Zhong, S., 2010. The 3ADON population of *Fusarium graminearum* found in North Dakota is more aggressive and produces a higher level of DON than the prevalent 15ADON population in spring wheat. Phytopathology 100, 1007–1014. https://doi.org/10.1094/PHYTO-12-09-0332

Ramírez Albuquerque, D., Patriarca, A., Fernández Pinto, V., 2021. Can discrepancies between *Fusarium graminearum* trichothecene genotype and chemotype be explained by the influence of temperature in the relative production of 3-ADON and 15-ADON? Fungal Biol. 125, 153–159. https://doi.org/10.1016/j.funbio.2019.09.004

Ramirez, M.L., Chulze, S., Magan, N., 2006. Temperature and water activity effects on growth and temporal deoxynivalenol production by two Argentinean strains of *Fusarium graminearum* on irradiated wheat grain. Int. J. Food Microbiol. 106, 291–296. https://doi.org/10.1016/j.ijfoodmicro.2005.09.004

Rauwane, M.E., Ogugua, U. V., Kalu, C.M., Ledwaba, L.K., Woldesemayat, A.A., Ntushelo, K., 2020. Pathogenicity and virulence factors of *Fusarium graminearum* including factors discovered using next generation sequencing technologies and proteomics. Microorganisms 8, 305. https://doi.org/10.3390/microorganisms8020305

Reimann, S., Deising, H.B., 2005. Inhibition of efflux transporter-mediated fungicide resistance in *Pyrenophora tritici-repentis* by a derivative of 4′-hydroxyflavone and enhancement of fungicide activity. Appl. Environ. Microbiol. 71, 3269–3275. https://doi.org/10.1128/AEM.71.6.3269-3275.2005

Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., Schaffner, S.F., Lander, E.S., Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., Pasternak, S., Wheeler, D.A., Willis, T.D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, Hongbin, Zhao, Hui, Zhou, J., Gabriel, S.B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M.,

Gupta, S., Moore, J., Nguyen, H., Onofrio, R.C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Yan, Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Shen, Yayun, Sun, W., Wang, Haifeng, Wang, Yi, Wang, Ying, Xiong, X., Xu, L., Waye, M.M.Y., Tsui, S.K.W., Xue, H., Wong, J.T.F., Galver, L.M., Fan, J.B., Gunderson, K., Murray, S.S., Oliphant, A.R., Chee, M.S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J.F., Phillips, M.S., Roumy, S., Sallée, C., Verner, A., Hudson, T.J., Kwok, P.Y., Cai, D., Koboldt, D.C., Miller, R.D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L.C., Mak, W., You, Q.S., Tam, P.K.H., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., Tanaka, T., Tsunoda, T., Deloukas, P., Bird, C.P., Delgado, M., Dermitzakis, E.T., Gwilliam, R., Hunt, S., Morrison, J., Powell, D., Stranger, B.E., Whittaker, P., Bentley, D.R., Daly, M.J., De Bakker, P.I.W., Barrett, J., Chretien, Y.R., Maller, J., McCarroll, S., Patterson, N., Pe'Er, I., Price, A., Purcell, S., Richter, D.J., Saxena, R., Sham, P.C., Stein, L.D., Krishnan, L., Smith, A.V., Tello-Ruiz, M.K., Thorisson, G.A., Chakravarti, A., Chen, P.E., Cutler, D.J., Kashuk, C.S., Lin, S., Abecasis, G.R., Guan, W., Li, Y., Munro, H.M., Qin, Z.S., Thomas, D.J., McVean, G., Auton, A., Bottolo, L., Cardin, N., Eyheramendy, S., Freeman, C., Marchini, J., Myers, S., Spencer, C., Stephens, M., Donnelly, P., Cardon, L.R., Clarke, G., Evans, D.M., Morris, A.P., Weir, B.S., Johnson, T.A., Mullikin, J.C., Sherry, S.T., Feolo, M., Skol, A., Zhang, H., Matsuda, I., Fukushima, Y., MacEr, D.R., Suda, E., Rotimi, C.N., Adebamowo, C.A., Ajayi, I., Aniagwu, T., Marshall, P.A., Nkwodimmah, C., Royal, C.D.M., Leppert, M.F., Dixon, M., Peiffer, A., Qiu, R., Kent, A., Kato, K., Niikawa, N., Adewole, I.F., Knoppers, B.M., Foster, M.W., Clayton, E.W., Watkin, J., Muzny, D.,

Nazareth, L., Sodergren, E., Weinstock, G.M., Yakub, I., Birren, B.W., Wilson, R.K., Fulton, L.L., Rogers, J., Burton, J., Carter, N.P., Clee, C.M., Griffiths, M., Jones, M.C., McLay, K., Plumb, R.W., Ross, M.T., Sims, S.K., Willey, D.L., Chen, Z., Han, H., Kang, L., Godbout, M., Wallenburg, J.C., L'Archevêque, P., Bellemare, G., Saeki, K., Wang, Hongguang, An, D., Fu, H., Li, Q., Wang, Z., Wang, R., Holden, A.L., Brooks, L.D., McEwen, J.E., Guyer, M.S., Wang, V.O., Peterson, J.L., Shi, M., Spiegel, J., Sung, L.M., Zacharia, L.F., Collins, F.S., Kennedy, K., Jamieson, R., Stewart, J., 2007. Genome-wide detection and characterization of positive selection in human populations. Nature 449, 913–918. https://doi.org/10.1038/nature06250

Schroeder, H.W., Christensen, J.J., 1963. Factors affecting resistance of wheat to scab caused by *Gibberella zeae*. Phytopathology 53, 831–838.

Shah, L., Ali, A., Yahya, M., Zhu, Y., Wang, S., Si, H., Rahman, H., Ma, C., 2018. Integrated control of Fusarium head blight and deoxynivalenol mycotoxin in wheat. Plant Pathol. 67, 532–548. https://doi.org/10.1111/ppa.12785

Shaikh, T.H., 2017. Copy Number Variation Disorders. Curr. Genet. Med. Rep. 5, 183–190. https://doi.org/10.1007/s40142-017-0129-2

Skov, J., Lemmens, M., Giese, H., 2004. Role of a *Fusarium culmorum* ABC transporter (FcABC1) during infection of wheat and barley. Physiol. Mol. Plant Pathol. 64, 245–254. https://doi.org/10.1016/j.pmpp.2004.09.005

Son, H., Seo, Y.S., Min, K., Park, A.R., Lee, J., Jin, J.M., Lin, Y., Cao, P., Hong, S.Y., Kim, E.K., Lee, S.H., Cho, A., Lee, S., Kim, M.G., Kim, Y., Kim, J.E., Kim, J.C., Choi, G.J., Yun, S.H., Lim, J.Y., Kim, M., Lee, Y.H., Choi, Y. Do, Lee, Y.W., 2011. A phenome-based functional analysis of transcription factors in the cereal head blight fungus,

Fusarium graminearum. PLoS Pathog. 7, e1002310. https://doi.org/10.1371/journal.ppat.1002310

Spielmann, M., Lupiáñez, D.G., Mundlos, S., 2018. Structural variation in the 3D genome. Nat. Rev. Genet. 19, 453–467. https://doi.org/10.1038/s41576-018-0007-0

Spolti, P., de Jorge, B.C., Del Ponte, E.M., 2012. Sensitivity of *Fusarium graminearum* causing head blight of wheat in Brazil to tebuconazole and metconazole fungicides. Trop. Plant Pathol. 37, 419–423. https://doi.org/10.1590/S1982-56762012000600007

Spolti, P., Del Ponte, E.M., Cummings, J.A., Dong, Y., Bergstrom, G.C., 2014a. Fitness attributes of *Fusarium graminearum* isolates from wheat in New York possessing a 3-ADON or 15-ADON trichothecene genotype. Phytopathology 104, 513–519. https://doi.org/10.1094/PHYTO-07-13-0206-R

Spolti, P., Del Ponte, E.M., Dong, Y., Cummings, J.A., Bergstrom, G.C., 2014b. Triazole sensitivity in a contemporary population of *Fusarium graminearum* from New York wheat and competitiveness of a tebuconazole-resistant isolate. Plant Dis. 98, 607–613. https://doi.org/10.1094/PDIS-10-13-1051-RE

Stack, R.W., 2003. History of Fusarium head blight with emphasis on North America. In: Leonard, K.J., Bushnell, W.R. (Eds.), Fusarium Head Blight of Wheat and Barley. APS Press, St. Paul, pp. 1e34

Starkey, D.E., Ward, T.J., Aoki, T., Gale, L.R., Kistler, H.C., Geiser, D.M., Suga, H., Tóth, B., Varga, J., O'Donnell, K., 2007. Global molecular surveillance reveals novel Fusarium head blight species and trichothecene toxin diversity. Fungal Genet. Biol. 44, 1191–1204. https://doi.org/10.1016/j.fgb.2007.03.001

Stephan, W., 2019. Selective Sweeps. Genetics 211, 5–13.

Stukenbrock, E.H., Croll, D., 2014. The evolving fungal genome. Fungal Biol. Rev. 28, 1–12. https://doi.org/10.1016/j.fbr.2014.02.001

Sul, J.H., Martin, L.S., Eskin, E., 2018. Population structure in genetic studies: Confounding factors and mixed models. PLoS Genet. 14, 1–22. https://doi.org/10.1371/journal.pgen.1007309

Sun, Q., Collins, N.C., Ayliffe, M., Smith, S.M., Drake, J., Pryor, T., Hulbert, S.H., 2001. Recombination between paralogues at the Rp1 rust resistance locus in maize. Genetics 158, 423–438. https://doi.org/10.1093/genetics/158.1.423

Sutton, J.C., 1982. Epidemiology of wheat head blight and maize ear rot caused by *Fusarium graminearum*. Can. J. Plant Pathol. 4, 195–209. https://doi.org/10.1080/07060668209501326

Szécsi, Á., Bartók, T., 1995. Trichothecene chemotypes of *Fusarium graminearum* isolated from corn in Hungary. Mycotoxin Res. 11, 85–92. https://doi.org/10.1007/BF03192069

Talas, F., Mcdonald, B.A., 2015a. Significant variation in sensitivity to a DMI fungicide in field populations of *Fusarium graminearum*. Plant Pathol. 64, 664–670. https://doi.org/10.1111/ppa.12280

Talas, F., McDonald, B.A., 2015b. Genome-wide analysis of *Fusarium graminearum* field populations reveals hotspots of recombination. BMC Genomics 16, 1–12. https://doi.org/10.1186/s12864-015-2166-0

Talas, F., Würschum, T., Reif, J.C., Parzies, H.K., Miedaner, T., 2012. Association of single nucleotide polymorphic sites in candidate genes with aggressiveness and deoxynivalenol production in *Fusarium graminearum* causing wheat head blight. BMC Genet. 13, 1–10. https://doi.org/10.1186/1471-2156-13-14

Termolino, P., Falque, M., Aiese Cigliano, R., Cremona, G., Paparo, R., Ederveen, A., Martin, O.C., Consiglio, F.M., Conicella, C., 2019. Recombination suppression in heterozygotes for a pericentric inversion induces the interchromosomal effect on crossovers in Arabidopsis. Plant J. 100, 1163–1175. https://doi.org/10.1111/tpj.14505

Thompson, D.P., Metevia, L., Vessel, T., 1993. Influence of pH Alone and in Combination with Phenolic Antioxidants on Growth and Germination of Mycotoxigenic Species of *Fusarium* and *Penicillium*. J. Food Prot. 56, 134–138. https://doi.org/10.4315/0362-028x-56.2.134

Tralamazza, S.M., Rocha, L.O., Oggenfuss, U., Corrêa, B., Croll, D., Rose, L., 2019. Complex Evolutionary Origins of Specialized Metabolite Gene Cluster Diversity among the Plant Pathogenic Fungi of the *Fusarium graminearum* Species Complex. Genome Biol. Evol. 11, 3106–3122. https://doi.org/10.1093/gbe/evz225

Varga, E., Wiesenberger, G., Hametner, C., Ward, T.J., Dong, Y., Schöfbeck, D., Mccormick, S., Broz, K., Stückler, R., Schuhmacher, R., Krska, R., Kistler, H.C., Berthiller, F., Adam, G., 2015. New tricks of an old enemy: Isolates of *Fusarium graminearum* produce a type A trichothecene mycotoxin. Environ. Microbiol. 17, 2588–2600. https://doi.org/10.1111/1462-2920.12718

Voigt, C.A., Schäfer, W., Salomon, S., 2005. A secreted lipase of *Fusarium graminearum* is a virulence factor required for infection of cereals. Plant J. 42, 364–375. https://doi.org/10.1111/j.1365-313X.2005.02377.x

Von der Ohe, C., Gauthier, V., Tamburic-Ilincic, L., Brule-Babel, A., Fernando, W.G.D., Clear, R., Ward, T.J., Miedaner, T., 2010. A comparison of aggressiveness and deoxynivalenol production between Canadian *Fusarium graminearum* isolates with 3-acetyl and 15-

acetyldeoxynivalenol chemotypes in field-grown spring wheat. Eur. J. Plant Pathol. 127, 407–417. https://doi.org/10.1007/s10658-010-9607-z

Wang, C., Zhang, S., Hou, R., Zhao, Z., Zheng, Q., Xu, Q., Zheng, D., Wang, G., Liu, H., Gao, X., Ma, J.W., Kistler, H.C., Kang, Z., Xu, J.R., 2011. Functional analysis of the kinome of the wheat scab fungus *Fusarium graminearum*. PLoS Pathog. 7, e1002460. https://doi.org/10.1371/journal.ppat.1002460

Wang, Q., Jiang, C., Wang, C., Chen, C., Xu, J.R., Liu, H., 2017. Characterization of the two-speed subgenomes of *Fusarium graminearum* reveals the fast-speed subgenome specialized for adaption and infection. Front. Plant Sci. 8. https://doi.org/10.3389/fpls.2017.00140

Ward, T.J., Clear, R.M., Rooney, A.P., O'Donnell, K., Gaba, D., Patrick, S., Starkey, D.E., Gilbert, J., Geiser, D.M., Nowicki, T.W., 2008. An adaptive evolutionary shift in Fusarium head blight pathogen populations is driving the rapid spread of more toxigenic *Fusarium graminearum* in North America. Fungal Genet. Biol. 45, 473–484. https://doi.org/10.1016/j.fgb.2007.10.003

Wegulo, S.N., Baenziger, P.S., Hernandez Nopsa, J., Bockus, W.W., Hallen-Adams, H., 2015. Management of Fusarium head blight of wheat and barley. Crop Prot. 73, 100–107. https://doi.org/10.1016/j.cropro.2015.02.025

Wilson, W.W., McKee, G., Nganje, W., Dahl, B., Bangsund, D., 2017. Economic impact of USWBSI's Scab Initiative to reudce FHB, Agribusiness and applied economics.

Xie, Q., Chen, A., Zhang, Y., Yuan, M., Xie, W., Zhang, C., Zheng, W., Wang, Z., Li, G., Zhou, J., 2019. Component interaction of ESCRT complexes is essential for endocytosis-

dependent growth, reproduction, DON production and full virulence in *Fusarium graminearum*. Front. Microbiol. 10, 180. https://doi.org/10.3389/fmicb.2019.00180

Xu, X., 2003. Effects of environmental conditions on the development of Fusarium ear blight. Eur. J. Plant Pathol. https://doi.org/10.1023/A:1026022223359

Yang, C., Liu, H., Li, G., Liu, M., Yun, Y., Wang, C., Ma, Z., Xu, J.R., 2015. The MADS-box transcription factor FgMcm1 regulates cell identity and fungal development in *Fusarium graminearum*. Environ. Microbiol. 17, 2762–2776. https://doi.org/10.1111/1462-2920.12747

Yin, Y., Liu, X., Li, B., Ma, Z., 2009. Characterization of sterol demethylation inhibitor-resistant isolates of *Fusarium asiaticum* and *F. graminearum* collected from wheat in China. Phytopathology 99, 487–497. https://doi.org/10.1094/PHYTO-99-5-0487

Yun, Y., Liu, Z., Yin, Y., Jiang, J., Chen, Y., Xu, J.R., Ma, Z., 2015. Functional analysis of the *Fusarium graminearum* phosphatome. New Phytol. 207, 119–134. https://doi.org/10.1111/nph.13374

Zapata, L., Ding, J., Willing, E.M., Hartwig, B., Bezdan, D., Jiao, W.B., Patel, V., James, G.V., Koornneef, M., Ossowski, S., Schneeberger, K., 2016. Chromosome-level assembly of *Arabidopsis thaliana* Ler reveals the extent of translocation and inversion polymorphisms. Proc. Natl. Acad. Sci. U. S. A. 113, E4042–E4060. https://doi.org/10.1073/pnas.1607532113

Zeller, K.A., Bowden, R.L., Leslie, J.F., 2004. Population differentiation and recombination in wheat scab populations of *Gibberella zeae* from the United States. Mol. Ecol. 13, 563–571. https://doi.org/10.1046/j.1365-294X.2004.02098.x

Zhang, L., Reifová, R., Halenková, Z., Gompert, Z., 2021. How important are structural variants for speciation? Genes (Basel). 12, 1084. https://doi.org/10.3390/genes12071084

Zhou, H., Muehlbauer, G., Steffenson, B., 2012. Population structure and linkage disequilibrium in elite barley breeding germplasm from the United States. J. Zhejiang Univ. Sci. B 13, 438–451. https://doi.org/10.1631/jzus.B1200003

Chapter 2 – Population genomics of *Fusarium graminearum* isolates from the Americas

2.1 Abstract

Fusarium head blight (FHB) is a major disease of wheat and barley worldwide and is caused by different species in the genus *Fusarium*, *Fusarium graminearum* being the most important. We conducted population genomics analyses using SNPs obtained through genotyping by sequencing of over 500 isolates of *F. graminearum* from the upper Midwest, New York, Louisiana, and Uruguay. PCA and Structure analyses group our isolates into four previously described populations: NA1, NA2, Southern Louisiana (SLA) and Gulf Coast (GC). Some isolates were not assigned to populations because of mixed ancestry. Population structure was defined by toxin genotype and geographic origin. The NA1, NA2, and SLA populations are differentiated (F_{ST} 0.385 – 0.551) but the presence of admixed isolates indicates that the populations are not reproductively isolated. Patterns of linkage disequilibrium (LD) decay suggest frequent recombination within populations. *Fusarium graminearum* populations from the US have great evolutionary potential given the high recombination rate and a large proportion of admixed isolates. The NA1, NA2, and Southern Louisiana (SLA) populations separated from their common ancestral population roughly at the same time in the past and are evolving with moderate levels of subsequent gene flow between them. Genome-wide selection scans in all three populations revealed outlier regions with the strongest signatures of recent positive natural selection. These outlier regions include many genes with unknown function and some genes with known roles in plant-microbe interaction, fungicide/drug resistance, cellular transport and genes that are related to cellular organelles. Only a very small proportion of outlier regions are shared

as outliers among the three populations, suggesting unique host-pathogen interactions and environmental adaptation.

2.2 Introduction

Fusarium graminearum is the major pathogen causing FHB in North America. In addition to the direct yield loss because of the infection, farmers also suffer price discounts caused by trichothecene mycotoxin contamination of the infected grains. Deoxynivalenol (DON) and its acetylated forms (3ADON and 15ADON), nivalenol (NIV) and NX-2 are some of the trichothecene mycotoxins produced by *F. graminearum*, and isolates are classified into chemotypes based on the major mycotoxin produced by the isolate (Goswami and Kistler, 2004; McMullen et al., 2012; Nganje et al., 2004). Initially in the US, *F. graminearum* populations were considered samples of a single homogeneous, random mating population currently known as North American 1 (NA1) (Zeller et al., 2004). In 2007 Gale et al. (2007) resolved two populations in a sample of *F. graminearum* isolates collected from the Midwest, rejecting the single homogenous population hypothesis. Currently, multiple populations correlated to trichothecene chemotype have been described, likely facilitated by better sampling, advancement in sequencing technologies that provide improved resolution of genetic relatedness, and actual changes in population dynamics that have occurred. Populations NA1 and NA2 are dominated by isolates with 15ADON and 3ADON chemotypes, respectively (Gale et al., 2007; Kelly and Ward, 2018). Population NA3 is the newest population to be identified, and is mainly composed of NX-2 isolates (Kelly and Ward, 2018). The Gulf Coast and Southern Louisiana populations are relatively understudied and are almost exclusively found in the southern US (Gale et al., 2011; Starkey et al., 2007). Isolates with the NIV chemotype make up the Southern Louisiana population (Gale et al., 2011). However, admixed isolates that cannot be assigned to populations

and isolates that do not conform to chemotype-based population assignments are known (Gale et al., 2007; Kelly and Ward, 2018). The NA1 population is the largest, most diverse, and probably native to North America, while NA2 is believed to have been introduced relatively recently (Kelly and Ward, 2018). The NA2 population is less diverse than NA1 and NA3, which is consistent with the hypothesis of its clonal expansion following its introduction from elsewhere (Kelly and Ward, 2018). The NA1, NA2 and NA3 populations currently coexist in several northern US states bordering Canada. We expect the distinction between these coexisting populations to continue to blur due to increased gene flow, as *F. graminearum* isolates outcross with one another in nature. The demographic history of these populations has not been systematically studied. Such studies can provide valuable information on how *F. graminearum* populations are evolving in the US and baseline demographic parameters for future studies.

Over the past decade and a half, *F. graminearum* chemotype shifts in the Western and Maritime provinces of Canada and US states bordering Canada have been documented (Liang et al., 2014; Puri and Zhong, 2010; Ward et al., 2008). These reports document a 14 to 15-fold increase in frequency of 3ADON isolates between the late 90s and early 2000s. However, the cause of this chemotype shift remains unknown. Superior fitness (*e.g.*, higher virulence, higher DON production, fungicide resistance, increased ascospore production) could cause natural selection to favor one group over another. The results of a few studies evaluating the fitness of 3ADON and 15ADON isolates are not consistent. Ward et al. (2008) and Puri and Zhong (2010) report 3ADON isolates to be more aggressive, produce more DON, and have higher fecundity and growth rate. In contrast, Spolti et al. (2014a) found no significant fitness difference between strains with 3ADON and 15ADON chemotypes. Differing geographic origins of the populations could be contributing to the variation in these conclusions. Information on genomic regions

under the influence of selection could help identify different selective forces operating in *F. graminearum* populations. Population genetic analyses were used to identify 14 divergent genomic regions, including the core *TRI* cluster, as containing signatures of natural selection (Kelly and Ward, 2018). In order to detect the unusually divergent regions, these analyses scanned the genome in small segments, window by window, and screened for windows with extreme values in the following population genetic measures: F_{ST}, a measure of population differentiation, Tajima's D, which compares two estimators of a population genetic diversity parameter and can spotlight regions with unusual allele frequency distributions, and nucleotide divergence (Dxy), a simple measure of genetic distance.

We used single nucleotide polymorphisms (SNPs) derived from genotyping by sequencing (GBS) to study population structure and patterns of LD decay, model the demographic history of populations, and identify genomic regions under the influence of positive natural selection. Our results suggest a high evolutionary potential of *F. graminearum* populations given the frequent sexual recombination within field populations and moderate gene flow between populations, and suggest the need for frequent sampling to monitor changing population dynamics. We also identified genomic regions under selection that could be providing fitness advantages to the populations. These regions provide new candidates for functional characterization to investigate their potential roles in pathogen fitness differences.

2.3 Materials and Methods

2.3.1 Fungal isolates, sequencing and variant identification

Fusarium graminearum isolates used for this study were obtained from other researchers and were primarily collected from New York, the Upper Midwest, Louisiana and Uruguay (see Supplemental file 2.1 for further details). GBS libraries were prepared using the *F. graminearum*

isolates as described previously (Fumero et al., 2021). All GBS libraries were sequenced using 100 bp single-end reads on an Illumina HiSeq2000 (Fumero et al., 2021). Trimmed reads were aligned to the PH-1 reference genome (NCBI BioProject: PRJEB5475) using default settings on BWA mem (Li and Durbin, 2009). The sequencing library was generated using restriction digestion so that the start and end of sequencing reads often were not random but rather fixed based on restriction cut sites. This process should cause many reads to appear duplicated, and thus duplicates were neither marked nor removed. Genome analysis toolkit, GATK 4.1.8.1 (McKenna et al., 2010), was used for variant calling. GATK HaplotypeCaller was used to call variants in GVCF mode. Briefly, variants were called separately for each sample and genomic variant call format (GVCFs) files were consolidated. This was followed by joint genotyping using "GenotypeGVCFs". No community-curated standard SNP set is available for *F. graminearum*, so the SNPs obtained from joint genotyping were filtered to create a set of known SNPs for base quality score recalibration (BQSR). Sites with QD <20, MQ <55, MQRankSum <= -10.0 && MQRankSum >= 10.0 and ReadPosRankSum <= -10.0 && ReadPosRankSum >= 10.0 were filtered out. Three rounds of BQSR were performed. At the end of each round, variants were called on recalibrated BAM files using GATK HaplotypeCaller, followed by joint genotyping. Variants were filtered to select SNPs, and SNPs were hard filtered based on the above criteria. The SNP set obtained after hard filtering was used as the known set for the subsequent round of BQSR. For variant quality score recalibration (VQSR), a set of SNPs common to the output files created after applying the hard filter on VCF files from the first, second, and third rounds of BQSR was used as the truth set. The training set was obtained by performing hard filtering (QD <10, MQ <55, MQRankSum <= -10.0 && MQRankSum >= 10.0 and ReadPosRankSum <= -10.0 && ReadPosRankSum >= 10.0) on the SNP set obtained after

the third round of BQSR. GATK VariantRecalibrator was used for variant quality score recalibration. The SNP set (no filters applied) obtained after running GenotypeGVCFs on BAM files after the third round of BQSR was used as the input for variant quality score recalibration. The model obtained using VariantRecalibrator was used to select SNPs after setting truth-sensitivity-filter-level to 99.0.

Pairwise differences between isolates were calculated using the VQSR output. Isolates that appear clonal, *i.e.*, different at fewer than 0.15% of genotyped SNP, were classified into 72 clonal groups. Only one isolate per clonal group was retained. A total of 101 clonal isolates were dropped. Also, the isolates that differed at 13% or more of the genotyped SNPs compared to most *F. graminearum* isolates and appeared to belong to different *Fusarium* species were removed from further analysis.

2.3.2 PCA and Structure analysis

Initially there were 40,169 biallelic SNPs segregating in the set of 454 clone-corrected isolates. We dropped all singleton SNPs and sites with genotype calls missing in >50% (>227) of the isolates. The 16,903 remaining SNPs were then filtered to remove nearby SNPs in high linkage disequilibrium using PLINK v1.07 with the command --indep-pairwise 50 5 0.5. This left us with 10,806 biallelic SNPs. The population structure of 454 *F. graminearum* isolates was inferred by a principal component analysis using the function “snpgdsPCA” in the R package SNPrelate (Zheng et al., 2012). Output was plotted using the R library ggplot2 (Wickham, 2016).

Population structure analysis of the 454 clone-corrected isolates also was conducted by using Bayesian clustering implemented in STRUCTURE 2.3.4 (Pritchard et al., 2000). Support for up to 10 different populations (K=1 to K=10) was tested using 10 replicates for each choice

of K (100 total runs). The burn-in period and reps were both set at 100,000. Output files from STRUCTURE were summarized using STRUCTURE harvester (Earl and VonHoldt, 2012). Output from STRUCTURE harvester was used as the input for CLUMPP (Jakobsson and Rosenberg, 2007). The best K was selected using the Evanno method (Evanno et al., 2005). The method looks for the change in log probability (ΔK) for each increment in K. For the best K, the value of ΔK is maximized as we move from a lower value of K to a higher value suggesting that the higher value is a better fit. A second STRUCTURE analysis was performed using a smaller set of 230 isolates. Isolates were chosen based on the results of the first run such that we had a balanced sample size among the inferred populations. Figures were created using DISTRUCT 1.1 (Rosenberg, 2004). Populations from STRUCTURE analysis were labeled as NA1, NA2, SLA (Southern Louisiana), and GC (Gulf Coast) based on the frequency of different *TRI* loci haplotypes among isolates assigned to these populations. For subsequent analyses, isolates were assigned to populations as follows. Isolates were assigned to the NA1 and NA2 populations when the population membership coefficient was >= 0.9 for the green and red populations, respectively, from the K=3 runs of 454 isolates. The 61 isolates with population membership coefficient >= 0.9 for the blue population from the K=3 runs were subsequently assigned to the SLA population when the population membership coefficient for the blue population from the K=4 runs for 454 isolates was >= 0.9 . Similarly, the 19 remaining isolates from the blue K=3 population that were not assigned to the SLA population but had population assignment coefficients = 1 for the yellow population in the K=4 runs of the sample of 230 isolates are assigned to the GC population. Isolates with population assignment coefficient less than 0.9 in the K=3 runs were defined as unassigned to any population.

2.3.3 Analysis of linkage disequilibrium decay

Decay of linkage disequilibrium with distance was inferred for three *F. graminearum* populations, NA1, NA2, and SLA. A total of 368 isolates, 211, 115, and 42 from the NA1, NA2 and SLA populations, respectively, were included in the analysis. The GC population was not included because of its small sample size (19 isolates). Monomorphic sites and sites with less than 50% of the genotype calls in each of the three populations were excluded. Filtering based on polymorphisms and missing data left us with 11,150, 5,759, and 4,161 SNPs for analysis for the NA1, NA2, and SLA populations, respectively. A custom perl script was used to make .ped and .map input files for PLINK 1.9 (Purcell et al., 2007). LD decay (measured as r^2) was calculated for each pair of SNPs and a custom perl script was used to get the average r^2 for SNPS at different distance intervals (*e.g.*, 10 kb to 1000 kb with 10 kb increments, and 1 kb to 10 kb with 1 kb increments). Mean r^2 computed for SNPs on different chromosomes served as background LD and was used to assess residual structure within the populations. The figures were plotted using ggplot2.

2.3.4 Inferring demographic history

The demographic history of the three *F. graminearum* populations, NA1, NA2 and SLA, was inferred by using the joint allele frequency spectrum derived using δaδi 2.1.1 (Gutenkunst et al., 2009). The ancestral state for each SNP was inferred as described by (Fumero et al., 2021). In brief, *Fusarium culmorum* and *Fusarium pseudograminearum* were used as outgroups for identification of the ancestral state. The site frequency spectrum was generated using the function “dadi.Misc.make_data_dict”. During this process, the sample size was projected down to 100, 50, and 20 individuals for NA1, NA2, and SLA, respectively. Since the demographic analysis requires a matrix of samples and sites with no missing data, projection allows us to keep

a greater number of sites (compared to only using sites with no missing genotypes in the entire population) for computing the frequency spectrum. However, some sites with very low genotyping rates are still eliminated. Separate site frequency spectra were generated for 3D, 2D and 1D demographic modeling; one for 3D (all 3 populations), three for 2D (one for each of the 3 population pairs) and three (one for each of the three populations) for 1D modeling. For the 3D analysis, an initial 12,823 sites were available, but this number was reduced to 6,612 sites after projection.

Four 3D models were tested: split_nomig, split_symmig_all, split_sym_mig_adjacent, and sim_split_sym_mig_all. The split_nomig model involves splitting of the ancestral population into SLA and population A where A further splits into populations NA1 and NA2. No migration is allowed. The split_symmig_all model assumes symmetric migration between all population pairs following the splits as in split_nomig. The split_symmig_adjacent model involves symmetric migration between the adjacent population pairs. Since the order of populations determines both the order of splitting and the adjacent populations for this model, three different scenarios were tested (split_symmig_adjacent A, split_symmig_adjacent B, and split_symmig_adjacent C). The population order for split_symmig_adjacent A is SLA, NA2 and NA1, thus migration is allowed between SLA - NA2 and NA2 - NA1. Similarly, the order is SLA, NA1, and NA2 in split_symmig_adjacent B and NA2, NA1 and SLA in split_symmig_adjacent C. Migration between SLA-NA1 and NA1-NA2 is allowed in split_symmig_adjacent B and between NA1-NA2 and NA1-SLA in split_symmig_adjacent C. The sim_split_sym_mig_all model is different from the other models as this model considers simultaneous splitting of the populations from the common ancestor and symmetric migration between each population pair. Dadi_pipeline (Portik et al., 2017) was used for parameter

optimization. Schematic diagrams of the models can be found at https://github.com/dportik/dadi_pipeline/blob/master/Three_Population_Pipeline/Models_3D.pdf . A single modification in the pipeline allowed us to account for the proportion of ancestral state misidentification in the data. The ancestral state for a site can be misidentified if a back mutation occurred in the site in the species used for obtaining the ancestral state. The modification included the addition of one additional parameter (p_misid) to the model using the function "Numerics.make_anc_stat_misid_func". This parameter was added to the model and optimized with the other parameters. At least five independent optimizations were performed for each model, each with six rounds of optimization with up to 50 replications per round [replications= 20,20,30,40,40,50]. Random parameters were used as input for the first round. The parameters corresponding to the highest log-likelihood from the first round were used as input for the second round, and so on. The Inference.optimize.log_fmin function was used for parameter optimization. The parameters were considered to have converged when the same log-likelihoods with similar parameters were obtained in the final round of each of five independent optimization runs. When parameters did not converge for 5 independent runs, up to 10 independent runs were performed regardless of whether the parameters converged or not. Uncertainty around the optimized parameters was estimated using Godambe Information Matrix (GIM) implemented in δaδi. Six 2D models, three testing symmetric migration (sym_mig) and three more testing non-symmetric migration (asym_mig) between the population pairs were fitted to investigate the patterns of migration at a finer scale. Schematic diagrams of the 2D models and associated parameters can be found at https://github.com/dportik/dadi_pipeline/blob/master/Two_Population_Pipeline/Models_2D.pdf. We do not present results from 3D models with asymmetric migration, as incorporating

asymmetric migration in these models made them very complex and significantly increased the computational time and resources. Parameters for 2D models were optimized as described above. In addition to the 3D and 2D models, five 1D models were also fitted for each population to check for change in population size following population splits (Figure 2.4). The standard neutral model (Snm) assumes no size change after the split. The growth model involves instantaneous growth at some time in the past. The bottle growth model considers an instantaneous bottleneck followed by growth. The three-epoch model involves two different size change events in the past whereas the two-epoch model involves a single size change event. The model fitting and optimization were performed as described earlier. The parameters were optimized in 3 independent runs.

2.3.5 SweepFinder selection scan

The software SweepFinder2 (Degiorgio et al., 2016) was used to identify the signatures of recent positive selective sweeps in three *F. graminearum* populations. Genomic regions affected by recent selective sweeps are characterized by reduced nucleotide diversity and a locally distorted spectrum of allele frequencies. SweepFinder (Nielsen et al., 2005) calculates a composite log-likelihood ratio (CLR) based on the log-likelihood for the genome wide frequency spectrum, which is presumed to reflect demography but not selective sweeps, and the log-likelihood for a region potentially affected by recent selection (Derbyshire, 2020). Sweepfinder2 input files for three populations were created by including polymorphisms (derived allele/minor allele found at least once in the populations) and substitutions (alleles different from the inferred ancestral state for the position and no longer polymorphic, or ‘fixed’ in the population). The ancestral state was determined as described in the demographic modeling and was included for each SNP where it was available. Positions with ancestral state unavailable were coded as folded,

otherwise the position was coded as unfolded. Sites for which the derived allele was absent in the populations (ancestral state fixed) were excluded from analysis. To reduce the sites with missing data, only sites with genotype calls for at least 100 (out of 211), 50 (out of 115), and 20 (out of 42) individuals were included for populations NA1, NA2, and SLA, respectively. 12,117, 13,866, and 5,846 sites were retained for analysis for population NA1, population NA2 and population SLA, respectively. The CLR computed for a window of approximately 5 kb was used to identify regions affected by selective sweeps. A simple genetic map, which was prepared by using the information published in (Laurent et al., 2018) and summarizes variation in recombination rates across the genome, was incorporated into the selection scans. Regions falling under the top 5th percentile of genome wide CLR values were identified as outlier regions (regions under the influence of selection). Adjacent outlier windows were clustered into outlier regions manually. Genes falling within the outlier regions were extracted from the reference genome PH-1 annotation using a custom script. Graphs were plotted using R library "ggplot2". Gene ontology (GO) enrichment analyses were conducted using the GO enrichment tool in FungiDB. A *p*-value of 0.05 (after Bonferroni correction) was considered significant in enrichment analyses. The GO terms and curated phenotypes for the genes were extracted from FungiDB.

2.3.6 Cross population selection scan

Signatures of recent positive selection were also identified using cross-population haplotype homozygosity, or XP-EHH, scans (Sabeti et al., 2007) in the R package rehh 2.0 (Gautier et al., 2017). The XP-EHH test compares the patterns of extended haplotype homozygosity (EHH) in a pair of populations and outlier sites in a population are identified relative to the other population. The outlier region in a population is characterized by a single

long haplotype widespread in the population whereas the reference population has multiple short haplotypes for the region in question. Cross population scans like XP-EHH have better power to detect sweeps of alleles that are close to fixation or have already been fixed. Before performing selection scans using rehh 2.0, missing data was imputed for 368 total isolates (211 from NA1, 115 from NA2, and 42 from SLA) using Beagle 4.0 (Browning and Browning, 2007). Sites with genotype calls in less than 25% of the isolates (92 out of 368) were excluded before imputation. The 18,162 sites that were polymorphic among the 368 isolates after imputation were used in the scans. The rate of imputation error was calculated to be ~ 0.36%. Ancestral state was not available for all SNPs used for rehh 2.0, thus the folded frequency spectrum was used. A haplotype homozygosity scan was performed separately for each chromosome for each population by using the function "scah_hh". The haplotype homozygosity scans for four chromosomes in a population were combined and the cross-population scans were performed using the function "ies2xpehh". "discard_integration_at_border" was set to false and "limehhs" was set at 0.1. SNPs exceeding the 95th percentile of the negative log p-value were identified as outlier SNPs. Information on linkage disequilibrium decay was used to group the outlier SNPs into regions. For population NA1, SNPs were grouped into a region if the distance between two successive outlier SNPs was less than 3 kb, for population NA2, if the distance was less than 2 kb and, for population SLA, if the distance was less than 7 kb. To define the size of the affected region, 1.5 kb, 1 kb, and 3.5 kb was added to both ends of each region corresponding to populations NA1, NA2, and SLA, respectively. Genes falling into the outlier regions were extracted for the four chromosomes using a custom script. Gene ontology enrichment (GO) analysis was conducted using the GO enrichment tool in FungiDB. A p-value of 0.05 (after Bonferroni correction) was considered significant in enrichment analysis.

2.4 Results

2.4.1 Read alignment and variant identification

GBS reads were produced from an initial set of 570 isolates from the genus *Fusarium***.** Four isolates in the initial set of 570 were known to be species other than *F. graminearum* and were included as controls. Trimmed reads were aligned to the *F. graminearum* PH-1 reference genome (NCBI BioProject: PRJEB5475). The GATK tool HaplotypeCaller called 150,299, 104,303, and 104,194 SNPs from the resulting BAM files after the first, second and third round of base quality score recalibration (BQSR), respectively. The 87,385 SNPs common to these three files were used as the truth set for variant quality score recalibration (VQSR) and 90,491 SNPs were obtained by hard filtering SNPs from the 3rd round of BQSR for training the VQSR model. A total of 92,382 SNPs were retained after VQSR, and these SNPs, or subsets of them, were used for subsequent analysis. On average, isolates were successfully genotyped at 61% of the total SNPs. The six isolates with the fewest genotype calls (fewer than 20,000), were removed, leaving the fewest SNPs called among the remaining isolates at 29,949. Four isolates previously identified as *F. asiaticum*, *F. boothi*, *F. acacia-mearnsii*, and *F. mesoamericanum* differed from most of the other isolates at over 14% of the called SNPs. Five additional isolates also differed from the remaining isolates at over 13% of called SNPs and were assumed to be different species. These nine isolates were excluded from subsequent analyses. After identifying a clear gap in the distribution of the percentage of SNPs differing between all pairs of isolates at ~0.2%, pairs or groups of isolates more similar than this were labeled as clonal, while those less similar were considered distinct genotypes. Seventy-two clonal groups were identified using pairwise comparison, each containing 2 to 5 isolates. Specifically, we found 53 pairs, 13 groups of three, two groups of four, and four groups of five. Thus, 101 isolates from putative clonal

groups were removed from the analysis, leaving 454 isolates for further analysis. The 92,382 SNPs retained after VQSR was filtered to remove 40 SNPs with more than two alleles or those that no longer segregated in the set of 454 isolates. This left 40,169 SNPs for subsequent analyses.

2.4.2 Population structure analysis

Based on PCA and STRUCTURE analyses, most isolates were assigned to one of four different populations, while several isolates could not be assigned clearly to a single population. For PCA, two large clusters sharing negative values for PC2 but defined by having either positive or negative values for PC1 were identified (Figure 2.1). Negative PC1 and PC2 values generally correspond to 3ADON isolates from the Upper Midwest and New York and 15ADON isolates from South America (Uruguay), though six 15ADON isolates from New York and one 15ADON isolate from the Upper Midwest tightly cluster with the 3ADON isolates in this PC space. Positive PC1 values and negative PC2 values generally correspond to 15ADON isolates from the Upper Midwest and New York, though one 15ADON isolate from Louisiana, one 3ADON isolate, and two NX-2 isolates tightly cluster with the northern 15ADON isolates in this PC space. These two large clusters do not have sharp boundaries, and many isolates have negative PC2 values and intermediate PC1 values and fall between them, suggesting admixture (hybrid isolates with ancestry from multiple populations).

The former cluster corresponds to the NA2 population, and the latter to the NA1 population, from (Gale et al., 2007; Kelly and Ward, 2018). NX-2 isolates did not form a discrete cluster, and on average were closer to the NA1 population than the NA2 population. Most Louisiana isolates, with higher positive PC2 values, group away from the other isolates in the third and fourth clusters (Figure 2.1), and the two Louisiana clusters are separated from one

another by the top three PCs. These third and fourth clusters were identified as the SLA and GC populations by matching the frequency of different *TRI* loci haplotypes among isolates found in the clusters with those of these populations reported by (Gale et al., 2011; Starkey et al., 2007). Most of the NIV isolates from Louisiana grouped into the SLA population, whereas six NIV, one 15ADON and all 3ADON isolates from Louisiana formed the GC population, which, with the highest PC2 values, clusters away from all other isolates. (Gale et al., 2011; Starkey et al., 2007). In addition to the four main clusters, two small clusters have PC1 values near zero. One consists of one 3ADON isolate from Louisiana and a few 15ADON isolates from the Midwest and New York, while the other is composed of one NIV isolate from Louisiana and a NIV, an NX-2, and three 15ADON isolates from Minnesota. The latter cluster may correspond to the NA3 population (Kelly and Ward, 2018), while the other does not correspond to known *F. graminearum* populations. Thus, across the Americas, population structure appears due to a combination of both *TRI* locus genotype and geographic origin.

The results from the PCA analyses were supported by STRUCTURE analyses. The runs assuming three populations (K=3) were inferred as the best fit in the STRUCTURE analyses of the 454 isolates (Figure 2.2). One population, mostly composed of US 3ADON and South American 15ADON isolates, corresponds to NA2. Another population is composed of NIV, 3ADON and 15ADON isolates from Louisiana, and the population made primarily of US 15ADON and two NX-2 isolates corresponds to NA1. On increasing K from three to four, Louisiana isolates separated into two groups, suggesting further structure within our Louisiana sample. The STRUCTURE analysis of a smaller, more balanced set of 230 isolates indicated K=4 (four populations) as the best fit, confirming the substructure within the Louisiana population (Figure 2.2). In K=3, 211, 115, and 61 isolates were assigned into NA1, NA2 and

Louisiana (SLA+GC) populations; 67 isolates remained unassigned to any populations (supplemental file 2.2). For K=4, 42 out of the 61 Louisiana population isolates were assigned to the SLA population (supplemental file 2.3).

2.4.3 Analysis of LD decay

A negative correlation between LD and distance between the SNPs was observed as the LD decayed rapidly with distance for all three populations. For the NA2 population, the LD decayed to half of the highest value (0.32 for pairs within 1 kb, with average spacing of 164 bp) within 2 kb (Figure 2.3). For the NA1 population, LD decayed to half of the highest value (0.16 for pairs within 1 kb, with average spacing of 185 bp) within 3 kb (Figure 2.3). The LD decays to half of the highest value (0.37 for pairs within 1 kb, with average spacing 176 bp) within 7 kb for the Louisiana population (Figure 2.3). LD reaches approximate background levels within 300 kb for all three populations.

2.4.4 Demographic modeling

The demographic history of the three *F. graminearum* populations, NA1, NA2 and SLA was inferred by using their individual and joint allele frequency spectra using δaδi 2.1.1 (Gutenkunst et al., 2009). The questions of interest were: How long ago did the populations diverge? Did population size change following the split from the common ancestral population? Did the populations evolve in isolation following the splits or with migration between them? We used a combination of 1D, 2D and 3D models, modeling one, two, or three populations at a time, respectively, to answer those questions. Four distinct 3D models were tested: one where the ancestral population simultaneously splits into three, and subsequent migration occurs between all possible population pairs (sim_split_sym_mig_all). In all other 3D models, a single population splits from the ancestral population, followed later by a subsequent split of the

ancestral population into two more populations. In one, migration was allowed between all three populations, and between the first population to split and the intermediate ancestral population (split_symmig_all). In another, no migration was allowed (split_nomig). In the final model, after the second split event, migration is blocked between a 'non-adjacent' pair of populations (split_sym_mig_adjacent). Three different versions of this final model were tested: two where SLA splits first and is either adjacent to NA1 or NA2, and a third where NA2 splits first and is adjacent to NA1. Our preliminary distance-based phylogenetic analysis indicated that SLA had the earliest divergence from the other populations, so that in both split_symmig_all and split_nomig SLA was designated as the first population to split off. In all 3D models we tested with migration, migration rates were forced to be symmetric, meaning the rate from population 1 to 2 was identical to that from 2 to 1.

When 3D models were fitted to the observed joint frequency spectrum of the populations, the three models allowing migration were a better fit than the model with no migration (Table 2.1). The log-likelihood was maximized for the sim_split_sym_mig_all (-5270) model. This suggests that the three populations separated from their common ancestor at roughly the same time. Based on parameters from the sim_split_sym_mig_all, the size of the SLA and NA2 populations are smaller relative to the ancestral population size (Table 2.1). However, the relative size of the NA1 population is 35% larger than the ancestral population (Table 2.1). The model supports a moderate level of gene flow between the populations. The rate of gene flow (expressed as 2*Na*m, where Na is ancestral population size and m is effective migration rate) between NA1 and NA2 is 0.10 and between NA1 and SLA is 0.25. The rate of gene flow between SLA and NA2 was the lowest (0.04).

Two distinct 2D models for each population pair, one with symmetric migration between populations and another that allows for asymmetric migration rates between each population pair, were fitted to evaluate migration patterns at a finer scale. The 2D models also provided information on relative population size and population divergence times that could be compared with parameter values from the 3D models. The model with asymmetric migration rates between the NA1 and NA2 populations fit better than the model with symmetric migration rates, with more individuals migrating from NA1 into NA2 (Supplemental file 2.5). Symmetric migration was a slightly better fit than asymmetric migration for the NA1-SLA pair (Supplemental file 2.5). Symmetric and asymmetric migration had similar support in the NA2-SLA pair (Supplemental file 2.5).

Based on results from the 1D demographic models, all three populations have changed in size in the past. Since the five models we tested are not nested, one cannot rigorously test if one model fits the data better than another. Based on the log-likelihood ratio, the standard neutral model (Snm) fit poorly for all three populations, and the three-epoch model seemed to fit better than the remaining three (Supplemental file 2.6). For SLA and NA1, other than the poorly fitting standard neutral model, all other models were similar in terms of log-likelihood. The three-epoch and bottle growth models have similar fits for NA2. Assuming (for the purpose of comparison between populations) that the three-epoch model holds for all populations, NA2 and SLA underwent a bottleneck and then expanded in size. In contrast, NA1 expanded in size twice and grew considerably (over 9-fold) following the split from its ancestor.

2.4.5 Scans of selection

The software SweepFinder2 (Degiorgio et al., 2016) was used to identify signatures of recent positive selective sweeps in the NA1, NA2, and SLA populations. The ~5 kb windows

corresponding to composite log-likelihood ratio (CLR) scores equal to or above the 95th percentile were deemed outlier regions. The 95th percentiles of CLR were 7.79, 6.74 and 14.47 for NA1, NA2 and SLA, respectively (Figure 2.5). The highest CLR score for SLA was 34.01 on chromosome 4 for the window between 1.403 to 1.408 Mb. Similarly, the highest CLR peak for NA2 was 24.04 between 5.333 and 5.338 Mb on chromosome 1. Likewise, the highest CLR score for NA1 was 15.61 on chromosome 3 in the window between 5.545 Mb to 5.550 Mb. The signatures of selection were pronounced on all chromosomes for NA1 and on chromosomes 1, 3 and 4 for population NA2. There were no outlier regions on chromosome 2 or 3 for the SLA population.

We define candidate genes as all those found in selection scan outlier regions, genes that potentially are or recently were under the influence of positive natural selection. There are 741, 707, and 670 candidate genes for populations NA1, NA2, and SLA corresponding to 104, 73, and 23 outlier windows, respectively. Five candidate genes were common among three populations, FGRAMPH1_01T00639, FGRAMPH1_01T00641, FGRAMPH1_01T00643, and FGRAMPH1_01T07669 on chromosome 1, and FGRAMPH1_01G23777 on chromosome 4. FGRAMPH1_01T07669 codes for sarcolemmal membrane associated proteins. FGRAMPH1_01T00643 is an unnamed protein, and the remaining three are hypothetical proteins.

There are 74 candidate genes in the outlier regions common between NA1 and SLA, 119 common between NA2 and SLA, and 77 common between NA1 and NA2 populations. Gene ontology enrichment analysis of outlier genes showed enrichment of the GO molecular function term “binding” in NA2 candidates (1.2-fold enrichment, p=0.03). Similarly, the GO biological process term “cellular component organization and biogenesis” was enriched (1.7-fold, p=0.04)

in the NA1 candidates. Several cellular component GO terms were enriched in all three populations (Figure 2.6). In all three populations, sweeps were overrepresented in the regions of low recombination, with less than 10% of the sweeps falling in the high recombination regions for each of the three populations. In population NA1 only 4 out of 104 sweep windows were in high recombination regions. In population SLA, only 1 sweep window was found in the high recombination regions compared to 22 sweep windows in the rest of the genome. Similarly, in population NA2, 6 sweep windows were in high recombination regions compared to 67 sweep windows for the rest of the genome.

Cross-population scans using rehh also were used to scan for signatures of recent positive natural selection in all three populations. A total of 256, 197, and 141 regions had characteristics of selective sweeps and fell within the tail of the distribution beyond the 95th percentile of the negative log p-value for populations NA1, NA2, and SLA, respectively. The size of the affected regions spanned from 3 kb to 13 kb for NA1, 2 kb to 6 kb for NA2, and 7 kb to 27 kb for SLA. The largest affected regions were found on chromosome 3 between 3,449,184-3,462,542, chromosome 2 between 462,022-468,247, and chromosome 3 between 27,226,672-27,254,438 for populations NA1, NA2, and SLA, respectively. A total of 463, 303 and 602 genes were identified to fall in the outlier regions of populations NA1, NA2, and SLA, respectively. There are 62 common genes between NA1 and NA2, 68 between NA2 and SLA, and 67 between NA1 and SLA. No genes identified as outliers by the XP-EHH scan were common to all three populations.

Gene ontology enrichment analysis was conducted with the genes in the outlier regions in the three populations to identify the enriched terms. The biological process term cellular iron ion

homeostasis was enriched (p= 0.04) 11-fold in population NA1. There was no enrichment of terms in any categories for populations NA2 and SLA.

In our cross-population scans, some genes were identified in outlier regions twice for the same population, in each of its comparisons against the other two populations. Fifty-seven genes were identified as outliers in NA1 for both NA1-NA2 and NA1-SLA scans. Ten of these genes are predicted to be membrane components (general membrane and membranes of intracellular anatomical structures). Similarly, 25 genes identified in population NA2 were common in NA2-NA1 and NA2-SLA scans, while 67 genes identified in SLA were common in NA1-SLA and NA2-SLA scans.

2.5 Discussion

For this study we used the SNPs obtained from 454 *F. graminearum* isolates to study population structure and diversity, model demographic history of major populations identified and scan genomes to identify genomic regions under recent positive natural selection. PCA and STRUCTURE analysis supported four populations, NA1, NA2, SLA and GC, and several admixed isolates in our collection of *F. graminearum* isolates from the Upper Midwest, New York, Louisiana, and Uruguay (Supplemental file 2.2). These populations have been reported previously to occur in different parts of the US (Gale et al., 2011; Kelly and Ward, 2018; Starkey et al., 2007). Multiple isolates of the NX-2 genotype did not resolve into a separate population as reported by Kelly and Ward (2018), likely due to the small sample size (13 NX-2 isolates) that lacked the power to group the isolates into a separate population. Two NX-2 isolates were assigned to the NA1 population while the rest remained unassigned but most grouped most closely with the NA1 population. NX-2 isolates on average had NA1 assignment coefficients in K=4 of 79%, with only one NX-2 isolate with an NA1 coefficient of less than 50%. NX-2

isolates on average had NA2 assignment coefficients of 11% and SLA and GC coefficients of less than 10% each. Kelly and Ward (2018) also report some admixed NX-2 isolates that share a greater proportion of ancestry with the NA1 population. *TRI12* and *TRI13* haplotypes of NX-2 isolates are similar to those of 3ADON producers, but NX-2 isolates have a greater proportion of ancestry derived from the NA1 population (Liang et al., 2014). This pattern suggests either the individuals resulted from recombination between isolates in the NA1 and NA2 populations or that these NX-2 isolates trace their origin to some more ancestral population and have evolved independently from the NA1 and NA2 populations. *Fusarium graminearum* is known to survive in wild grasses as endophytes, and Kelly and Ward (2018) favor the latter explanation, speculating a potential host jump of the NA3 population from wild grasses to wheat. 15ADON isolates from Uruguay group closely with 3ADON isolates from the US in the NA2 population. Isolates of the US 3ADON chemotype also are similar to Italian *F. graminearum* isolates (15ADON, 3ADON and NIV chemotype) and group into a single cluster in a previous STRUCTURE analysis (Gale et al., 2007). The genetic similarity between US 3ADON isolates and 15ADON isolates from outside the US (*e.g.*, Italy and Uruguay) is consistent with an ancient age of the 15ADON/3ADON polymorphism that is argued to be maintained by balancing selection (Ward et al., 2002). There also are seven North American isolates with a 15ADON *TRI* haplotype in the NA2 population (Supplemental file 2.2), which likely result from interpopulation hybridization. Sixty-seven of 454 isolates (K=3) could not be clearly assigned to any single population and have ancestry derived from two or more populations (Supplemental file 2.3). *Fusarium graminearum* populations coexist in multiple regions of the US (Gale et al., 2007; Liang et al., 2014; Starkey et al., 2007) and the possibility of them hybridizing via sexual reproduction exists. Admixed isolates with no clear population assignment are not unique to our

study and have been reported previously (Gale et al., 2007; Kelly and Ward, 2018; Starkey et al., 2007). Isolates generally group into populations based on *TRI* locus genotype and geographic origin, but there are exceptions. A fair proportion (24%) of admixed isolates and the presence of multiple isolates not conforming to the majority chemotype in their population make chemotype an unreliable predictor of population assignment. Thus, it is important to keep sampling isolates periodically and monitor pathogen populations in the field, genotyping samples more deeply than just relying on *TRI* locus haplotypes. We expect the rate of mismatch between chemotype and population to increase in the future as gene flow between the populations continues. SLA is exclusively composed of isolates of the NIV genotype except for two 15ADON isolates. More isolates from Southern Louisiana (SLA) have some proportion of ancestry shared with the NA1 and GC populations than with the NA2 population. This pattern is expected if the NA1, SLA and GC populations have coexisted together longer than NA2 has coexisted with SLA. This conclusion is consistent with the hypothesis that the NA2 population was introduced into the USA from another location relatively recently (Kelly and Ward, 2018; Ward et al., 2008). 3ADON isolates comprising the NA2 population are known from collections made during the mid- to late 1990s and later (Gale et al., 2007; Liang et al., 2015; Liang et al., 2014). Despite evidence of some gene flow between the populations, they are still quite differentiated. F_{ST} values between populations NA2-SLA, NA1-NA2 and NA1-SLA are 0.551, 0.385, and 0.398, respectively. Insufficient recombination between populations and the recent introduction of new populations are both potential explanations for the continued existence of genetically differentiated populations (Gale et al., 2007). The Gulf Coast (GC) population grouped away from most of the isolates in the PCA analysis. The GC population was originally reported as a genetically distant *F. graminearum* clade composed of isolates with NIV and 3ADON genotypes

(Gale et al., 2011; Starkey et al., 2007). Overall, the *F. graminearum* populations from the US are highly differentiated (Kelly and Ward, 2018).

We analyzed LD decay to infer patterns of recombination within the NA1, NA2 and SLA populations. *Fusarium graminearum* is a homothallic fungus, and a single isolate can reproduce sexually by itself. However, field populations are suggested to outcross frequently in nature (Gale et al., 2011; Talas and McDonald, 2015b; Zeller et al., 2004). The rapid LD decay for our three populations, NA1, NA2 and SLA suggests frequent outcrossing within populations. Talas and McDonald (2015b) report rapid decay of LD over 1000 bp in German field populations of *F. graminearum*. Similar patterns of LD decay also have been reported in Chinese populations (Zhang et al., 2012). The rate of LD decay was fastest in the NA2 population (Figure 2.3). It is surprising that this population, which reportedly has rapidly increased in frequency in the US, has the fastest LD decay. Rapid decay of LD in NA2 could be due to frequent outcrossing of the individuals within the population. In fact, Liang et al. (2014) report a recent change in the mode of reproduction from non-random towards more random mating in population NA2. LD decayed somewhat slowly in the SLA population (decayed to half of the highest value of r^2 within 7 kb) (Figure 2.3). This result could be due either to a relatively small sample size (42 isolates) or to a greater frequency of clonal reproduction relative to sexual reproduction. Gale et al. (2011) reported a mixed mode of reproduction in the SLA population. The background LD (LD between pairs of SNPs on different chromosomes) was very low for all three populations (r^2 =0.02 for NA2, 0.034 for SLA and 0.007 for NA1) suggesting negligible residual structure within the populations (Figure 2.3). *Fusarium graminearum* populations from the US and elsewhere have high evolutionary and adaptive potential given their high genetic diversity coupled with high rates of recombination (Talas and Mcdonald, 2015b). Beneficial mutations contributing to

pathogen fitness, *e.g*., fungicide resistance, could spread rapidly within these populations. Thus, the populations can rapidly adapt to environmental changes, to overcome resistant cultivars and to develop resistance against the fungicides used to manage FHB (Talas and McDonald, 2015a). Recombination itself is a source of multi-locus variation in natural pathogen populations.

The diffusion-based approach implemented in δaδi was used to test the demographic history of the three US *F. graminearum* populations. Change of population size over time was tested by using 1D models. 1D models are unrealistic since they assume populations evolve in isolation with no migration from other populations, but they still provide an approximation of relative population growth. The standard neutral model allows no size change and fit poorly for all three populations, indicating that the populations have undergone size changes (Supplemental file 2.6). Based on the three-epoch model (best fit for all populations), the NA2 and SLA populations experienced population bottlenecks. Bottlenecks following divergence are common in natural populations. The bottleneck in population NA2 can be explained by its introduction into the US from elsewhere. The bottleneck for NA2 was more severe than that for SLA, however NA2 quickly overcame the bottleneck and expand more rapidly than did SLA. In contrast to the NA2 and SLA populations, NA1 did not experience a bottleneck and continued to increase in size following the split. NA1 is believed to be endemic to the northern US and is the largest and most diverse population of *F. graminearum* in the country (Gale et al., 2007; Kelly and Ward, 2018).

In 3D modeling, the addition of a migration component to the models improved the model fit considerably compared to one that lacked migration (Table 2.1). When the populations were modeled to split non-simultaneously, the model parameters, especially time (T1) and migration rates (mA) did not converge well. This result is expected if the divergence time

between the populations is low. Low divergence time means that there are not enough new mutations in the diverging populations nor enough migration events happening between them to accurately estimate the relevant split time and migration parameters. The optimized time T1 (time between two divergence events) is very low for the best replicates for all non-simultaneous splitting models regardless of the order of splitting. Thus, the sim_split_symmig_all model best represents our data (Table 2.1). The optimized parameters for this model converged well and the estimate of uncertainty around the parameters was low (Table 2.1). Model sim_split_symmig_all involves a simultaneous split of all three populations from an ancestral population and symmetric migration between all population pairs. Even though the simultaneous split model was the overall best fit, it is likely that the splitting was not strictly simultaneous, and a series of splits in rapid succession could have happened. The SLA population likely diverged first. Distance based phylogenetic analysis (results not shown, computed using the genetic distance between individuals) supports that SLA diverged earlier than NA1 and NA2. In the three-epoch 1D model, the time since the start of population size changes (the length of the bottleneck, TB, plus the time since bottleneck, TF) is similar for NA1 and NA2 but much higher for SLA, further suggesting SLA has existed longer than the NA1 and NA2 populations. The current sizes of populations NA2 and SLA are each about half of the size of the ancestral population (Table 2.1). Population NA1, however, has expanded in size and is about 1.3 times the size of the ancestral population (Table 2.1). NA1 is the largest population followed by NA2 and SLA based on the sim_split_symmig_all model (Table 2.1). Migration between NA1-SLA is highest (0.25) followed by migration between NA1-NA2 (0.1) and SLA-NA2 (0.04) (Table 2.1). This finding is supported by the clustering patterns in the principal component analysis and population differentiation (F_{ST}) values (Figure2.1). Population NA1 is believed to be endemic to the US

(Gale et al., 2007). The greater gene flow between the NA1 and SLA populations can be explained by greater historic gene flow if SLA were also endemic and co-existed with NA1 longer than either the NA1-NA2 or NA2-SLA pairs. Population NA2 has existed together with the NA1 population in many parts of the US following its recent introduction (Liang et al., 2014; Lofgren et al., 2018), which could explain the intermediate gene flow between these populations. The SLA population, which is almost exclusively NIV genotype isolates, is distributed in the Southern US where the NA2 population so far has been rare (Gale et al., 2011; Starkey et al., 2007). This difference in geographic distribution could limit the gene flow between these populations.

We also fitted 2D models to evaluate finer patterns of migration, *i.e*., symmetric versus asymmetric. Asymmetric migration was a better fit for the NA1-NA2 pair with more gene flow occurring from NA1 to NA2 (Supplemental file 2.5). Similar patterns of biased gene flow were reported by Ward et al. (2008) for these two populations. NA1 is larger than NA2 and we expect more gene flow from the larger population to the smaller one. Both symmetric and asymmetric migration models fit similarly well for the SLA-NA2 pair. A symmetric migration model fit slightly better for SLA-NA1.

The population mutation parameter theta ($4N_e\mu$) for the best fit 3D model sim_split_symmig_all is 581. The 12,823 SNPs used in the analysis spanned over an average length of 403 kb per isolate. This gives a per base pair theta of 0.00144 (581/403,000). If we assume the mutation rate (μ) to be $7.82x10^{-9}$ (previously estimated for *Neurospora*), the ancestral effective population size is ~184,000. This translates to current effective population sizes of NA1, NA2 and SLA of 248k, 121k and 86k, respectively. The estimated divergence time ($2*N_e*T$) is 345,878 years in the past assuming a generation time of 1-year.

We scanned the genomes of isolates from populations NA1, NA2 and SLA with the objective of identifying regions under the influence of natural selection that are likely providing fitness advantages to the pathogen. In our cross-population scans, the average size of outlier regions and the total number of candidate genes were larger for population SLA than for NA1 and NA2. The average size of the outlier regions and the total number of candidate genes was smallest for population NA2. This result is expected as LD decays faster in NA2 than in the other two populations, providing a better resolution of the outlier regions in NA2.

We observed a similar pattern in Sweepfinder scans as the average size of the outlier region (following the merger of nearby regions) correlated with the pattern of LD decay for our populations. The average size of outlier regions was 17 kb, 27 kb, and 83 kb for populations NA1, NA2, and SLA, respectively. LD decayed slower in SLA than in the other two populations, and a larger number of genes flanking the selected gene are expected to be in LD with the target gene. Four outlier regions in chromosome 1 and three regions in chromosome 4 for population SLA are over 100 kb. The especially broad outlier regions in SLA may indicate more than greater average LD. They also can indicate selection events in the population that were particularly recent and/or strong.

A large proportion of the genes identified as candidates in selection scans using both Sweepfinder and XP-EHH are hypothetical, unnamed or unidentified proteins. Functional annotations are available for ~60% of the candidate genes identified by Sweepfinder in all three populations, and curated mutant phenotypes are available for even fewer. Curated mutant phenotypes for important traits (virulence, growth, vegetative spore production, sexual spore production, DON production and fungicide resistance) are available for 64, 60, and 52 Sweepfinder outlier genes for populations NA1, NA2 and SLA, respectively. Similarly,

phenotypes are available for 36, 18 and 45 XP-EHH candidate genes from NA1, NA2 and SLA, respectively. Upon mutation, deletion, or knocking out of many of these genes, *F. graminearum* isolates exhibit reduced virulence, post-penetration defects, reduced spore germination, reduced production of conidia and ascospores, production of aberrant, deformed spores, reduced/aberrant *in vitro* growth, increased resistance to fungicides and/or reduced mycotoxin production. Our results suggest that outlier regions and candidate genes in these regions play an important role in survival, fitness and adaptation of *F. graminearum*. Outlier regions also contain many membrane proteins and transporters with potential roles in virulence, plant-microbe interaction, and fungicide resistance.

One reason to perform selection scans is to identify loci governing fitness and adaptation. Fungicide resistance is one trait of interest. Some of the outlier regions contain genes with known or predicted roles in fungicide response. Azole fungicides used in the management of FHB act by binding to CYP51 paralogues and thus inhibiting ergosterol biosynthesis. Mutations in CYP51 genes are related to fungicide resistance in several fungi (Hamamoto et al., 2000; Klix et al., 2007; Leroux et al., 2007; Spolti et al., 2014b). One CYP51 paralogue, CYP51B (FGRAMPH1_01G02507) is an outlier gene in population SLA in the Sweepfinder scan. Overexpression of CYP51 and transport of fungicide out of the fungal plasma membrane are two strategies used by fungi to develop resistance/tolerance to azole fungicides. ABC transporters might be part of these strategies in *F. graminearum*, as Ammar et al. (2013) identified ABC transporters FgABC3 and FgABC4 as providing tolerance to azole fungicides, although their specific roles remain unknown. Another ABC transporter is encoded by FGRAMPH1_01G09799, an outlier gene in NA1 in the Sweepfinder scan (Supplemental file 2.7). *Fusarium graminearum* mutants in FGRAMPH1_01G09799 grew more vigorously in the

presence of tebuconazole, suggesting that this gene has a role in tebuconazole sensitivity, possibly by transporting or sequestering the fungicide outside the cells (Ammar et al., 2013). In addition, gene FGRAMPH1_01G01687 is an outlier in the XP-EHH scan of NA1. This gene is annotated as a multidrug transporter, and other genes FGRAMPH1_01G05895, FGRAMPH1_01G08449, and FGRAMPH1_01G11505, identified in the same scan, are annotated as drug resistance genes.

Many of our genes identified as candidate outliers from Sweepfinder and XP-EHH scans are putative membrane components with predicted functions in transmembrane transport. A total of 27, 28, and 20 Sweepfinder outlier genes in populations NA1, NA2, and NA3, respectively, are predicted to be transmembrane transporters. Among the XP-EHH candidate genes in NA1 are 26 transmembrane transporters. Four of these transporters are major facilitator superfamily transporters, with predicted functions in drug response for GRAMPH1_01G11623 and FGRAMPH1_01G19589, transmembrane transport for FGRAMPH1_01G11897, and polyamine transport for FGRAMPH1_01G18909. Genes GRAMPH1_01G11623 (MFS transporter) and FGRAMPH1_01G08449 (brefeldin resistance) are candidate genes in both the NA1 and NA2 populations. Similarly, SLA has 32 candidate genes from cross-population scans that are transmembrane transporters. Six of the 32 genes, FGRAMPH1_01G00221, FGRAMPH1_01G04399, FGRAMPH1_01G05779, FGRAMPH1_01G14121, FGRAMPH1_01G22359, and FGRAMPH1_01G25483, are annotated as major facilitator superfamily transporters. ABC and other transporters are important in sequestering and exporting DON to prevent toxicity to the fungus itself. They also transport DON and effectors into plant cells, and export plant defense compounds out of the fungal cell (Ammar et al., 2013; Chen et al., 2019; Wang et al., 2018). Some of the transporter genes identified in our selection scans may

have similar roles, and the prevalence of these genes in the outlier regions suggests that their evolutionary role in survival, fitness and adaptation of this fungus is worthy of further investigation.

Our results show that positive natural selection is currently operating on genes important in plant-microbe interactions. Among the XP-EHH candidate genes, 10 genes in population NA1, 12 in NA2 and 12 in SLA are predicted to be secreted effectors (Brown et al. 2012) (Supplemental files 2.10-2.12). These genes encode for enzymes with degradation functions and cysteine rich molecules, which often are secreted into the host and act as effectors. Some other candidates encode proteins that are secreted but have no known role in degradation. Among the predicted secreted proteins, gene FGRAMPH1_01G14115, which is common to NA2 and SLA, is a component of the fungal cell wall. This protein is predicted to help with the adhesion of the fungus to the host and in evading the host immune response. Similarly, gene FGRAMPH1_01G13287 is common to NA1 and NA2 and is predicted to have hydrolyzing functions. Gene FGRAMPH1_01G08415 is a member of glycosyl hydrolase family 18-6 and is predicted to have chitinase and chitin binding functions. The gene is common to the outlier regions in NA1 and SLA and was also reported as a candidate gene by Kelly and Ward (2018). Similarly, FGRAMPH1_01G08919, a common outlier gene in NA1 and NA2, is predicted to be expressed in the host cell nucleus and regulate host cell transcription. The gene is also among the candidate genes reported by (Kelly and Ward, 2018). A total of 8 genes from NA1, 1 gene from NA2 and 6 genes from SLA from our XP-EHH scan are common with Kelly and Ward (2018). Common genes have predicted functions in hydrolysis, carbohydrate metabolism, transmembrane transport, oxidation-reduction and transcription regulation and are all strong candidates for detailed functional studies.

Molecules modulating plant microbe interactions are exported not only through conventional pathways coupled with signal peptides, but also out of both fungal and plant cells in extracellular vesicles (Zamith-Miranda et al., 2018). Although this mechanism of molecule transport is fairly recently discovered in fungal pathogens, it is becoming increasingly clear that extracellular vesicles serve import roles in plant-pathogen interactions. Two proteins encoded by XP-EHH scan candidate genes in NA1 (FGRAMPH1_01G13589, FGRAMPH1_01G18909), three in NA2 (FGRAMPH1_01G12741, FGRAMPH1_01G13589, FGRAMPH1_01G15555), and three in SLA (FGRAMPH1_01G04863, FGRAMPH1_01G15555, FGRAMPH1_01G27929) are transported in *F. graminearum* extracellular vesicles (Garcia-Ceron et al., 2021). FGRAMPH1_01G15555, a candidate gene common to NA2 and SLA, is a carboxypeptidase that is highly enriched in extracellular vesicles, and likely plays an important role in establishing the fungus in the host. FGRAMPH1_01G18909, a candidate gene from NA1, is highly enriched in vesicles, and is an MFS gene predicted to have a role in ferrichrome transport.

Sweepfinder also identified some outlier regions containing genes encoding proteins important in plant-microbe interactions. Cell wall degrading enzymes are encoded by genes in outlier regions, including three identified in NA1 (FGRAMPH1_01G06361, FGRAMPH1_01G06375, FGRAMPH1_01G25271), one in SLA (FGRAMPH1_01G02483), and one in NA2 (FGRAMPH1_01G03929). In addition to cell wall degrading enzymes, the outlier regions also include genes encoding predicted secreted proteins; four in SLA (FGRAMPH1_01G00613, FGRAMPH1_01G05873, FGRAMPH1_01G23157, and FGRAMPH1_01G23695) and three in NA2 (FGRAMPH1_01G23157, FGRAMPH1_01G26021, FGRAMPH1_01G26599). While some candidate genes are in outlier regions in more than one population, others are unique to outlier regions in individual

populations. This pattern suggests that the populations are evolving independently and may be subject to different selective forces.

A significant proportion of the Sweepfinder outlier genes with a known role in one or more important traits (virulence, vegetative or sexual spore production, *in vitro* growth, spore germination, mycotoxin production and fungicide resistance) are annotated with one or more of the following cellular component GO terms: cytoplasm, intracellular organelle, organelle, intracellular anatomical structure, intracellular membrane bound organelle, membrane bound organelle, protein containing complex, and respirasome. In the SLA population, 14 candidate genes have a known role in one or more of the above traits and 11 of the 14 are associated with at least one of the cellular components mentioned above. Similarly, 22 and 24 candidate genes from NA1 and NA2, respectively, were previously identified as having roles in the above traits, and 8 and 9 of these are associated with a cellular component. This pattern suggests that genes associated with these cellular components play important roles in pathogen fitness.

A number of these genes are predicted to code for components of the Golgi apparatus/Golgi complex (NA1: 10, NA2: 7, SLA: 8), and the endoplasmic reticulum (ER) (NA1: 14, NA2: 16, SLA: 10). During trichothecene biosynthesis, the ER is remodeled in *F. graminearum* to form toxisomes, sites of trichothecene biosynthesis, and Tri proteins colocalize with toxisomes (Chen et al., 2019; Menke et al., 2012). Following synthesis, DON is believed to be stored in vesicles and then exported out of the fungal cell (Chen et al., 2019).

Gene FGRAMPH1_01G17679 is an integral component of the ER and is a known fluconazole transmembrane transporter. It might also have a role in trichothecene transmembrane transport as it is located in ER/toxisomes. FGRAMPH1_01G17609 is a major facilitator superfamily transporter and is encoded in an outlier region in population NA2. The Golgi

apparatus is an important component of the conventional secretion pathway in filamentous fungi, where it receives cargo from the ER and endosomes, and cargo is sorted and packed into vesicles for endocytosis and exocytosis (Pantazopoulou, 2016). Some of our candidate genes are predicted to encode a component of the Golgi apparatus and to possibly have roles in transporting DON from toxisomes into plant cells or into other fungal compartments for storage, although the role of these genes in other housekeeping functions cannot be ignored. FGRAMPH1_01G01291, a candidate gene in NA1, forms the coat of COPI vesicles. Fungi also use clathrin-coated vesicles for secretion. The plant pathogenic fungus *Botrytis cinerea* uses clathrin-coated vesicles to transport cell wall degrading enzymes out of the fungal cell (Souibgui et al., 2021). Two candidate genes, FGRAMPH1_01G18347 (encoding a component of the clathrin coat) in population NA1 and FGRAMPH1_01G02065 (encoding a component of the clathrin adaptor complex) in population NA2, are predicted to help with vesicle-mediated transport and could have similar functions in *F. graminearum*.

Some of our candidate genes likely affect more than one fitness trait in *F. graminearum*. Among the five Sweepfinder candidate genes common to three populations, gene FGRAMPH1_01G07669 codes for sarcolemmal membrane associated protein (SLMAP). SLMAP is a member of the Striatin-interacting phosphatase and kinase (STRIPAK) complex. SLMAP and members of this complex are known to have roles in sexual development, vegetative growth, virulence, cell fusion, primary and secondary metabolism and conidiation in fungi (Bernhards and Pöggeler, 2011; Islam et al., 2017; Kück and Stein, 2021; Wang et al., 2016). The STRIPAK complex acts as a major signaling hub and controls diverse developmental processes in ascomycete fungi (Kück and Stein, 2021). Similarly, transcription factor FGRAMPH1_01G03329 is an outlier in the Sweepfinder scan in populations NA1 and NA2 and

regulates multiple proteins, including SLMAP. Strains with mutations in this gene form no perithecia, and have reduced virulence, growth and DON production (Son et al., 2011). The transcription factor is conserved across multiple fungal species and could have important roles in survival, development and evolution (Son et al., 2011).

Even though *F. graminearum* is one of the most well-studied plant pathogens, large scale population genomics studies are limited, and studies tend to focus on a single or a few major populations. *Fusarium graminearum* in the United States is composed of multiple diverse populations that undergo frequent sexual outcrossing and recombination in nature. Beneficial alleles have the potential to expand rapidly given the high frequency of recombination within populations and moderate level of gene flow between populations. Regular sampling of field isolates and population genomics analysis as conducted in this study are necessary to monitor levels of diversity and the direction of gene flow between populations. Due to the composition of our sample, we did not find enough support for the NA3 population to include it in the analysis of LD decay, demographic modelling or selection scans. Similar studies that include the NA3 population will help elucidate the extent of gene flow between it and other populations and may elucidate the action of natural selection in this population and identify the reason for the increase in frequency of NX-2 isolates in some parts of the country. In the current study, selection scans identified genomic regions and candidate genes in these regions that potentially contribute to pathogen fitness. Genes with potential roles in fungicide resistance, virulence, and DON production are found among the candidate genes and provide new targets for functional studies. Large scale population genomics studies are limited or absent in most plant pathogenic fungi. Similar studies can be useful to identify the causes of disease outbreaks or fungicide resistance,

monitor the expansion of emerging or resistant isolates, identify the mode of reproduction, and study the genetic basis of pathogen adaptation.

2.6 References

Ammar, G.A., Tryono, R., DoÏl, K., Karlovsky, P., Deising, H.B., Wirsel, S.G.R., 2013. Identification of ABC transporter genes of *Fusarium graminearum* with roles in azole tolerance and/or virulence. PLoS One 8, e79042. https://doi.org/10.1371/journal.pone.0079042

Bernhards, Y., Pöggeler, S., 2011. The phocein homologue SmMOB3 is essential for vegetative cell fusion and sexual development in the filamentous ascomycete *Sordaria macrospora*. Curr. Genet. 57, 133–149. https://doi.org/10.1007/s00294-010-0333-z

Brown, N.A., Antoniw, J., Hammond-Kosack, K.E., 2012. The predicted secretome of the plant pathogenic fungus *Fusarium graminearum*: A refined comparative analysis. PLoS One 7, e33731. https://doi.org/10.1371/journal.pone.0033731

Browning, S.R., Browning, B.L., 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. 81, 1084–1097. https://doi.org/10.1086/521987

Chen, Y., Kistler, H.C., Ma, Z., 2019. *Fusarium graminearum* Trichothecene Mycotoxins: Biosynthesis, Regulation, and Management. Annu. Rev. Phytopathol. 57, 15–39. https://doi.org/10.1146/annurev-phyto-082718-100318

dadi_pipeline. accessed September 13, 2023. https://github.com/dportik/dadi_pipeline/blob/master/Three_Population_Pipeline/Models_3D.pdf

dadi_pipeline. accessed September 13, 2023. https://github.com/dportik/dadi_pipeline/blob/master/Two_Population_Pipeline/Models_2D.pdf

Degiorgio, M., Huber, C.D., Hubisz, M.J., Hellmann, I., Nielsen, R., 2016. SweepFinder2: Increased sensitivity, robustness and flexibility. Bioinformatics 32, 1895–1897. https://doi.org/10.1093/bioinformatics/btw051

Derbyshire, M.C., 2020. Bioinformatic Detection of Positive Selection Pressure in Plant Pathogens: The Neutral Theory of Molecular Sequence Evolution in Action. Front. Microbiol. 11, 644. https://doi.org/10.3389/fmicb.2020.00644

Earl, D.A., VonHoldt, B.M., 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4, 359–361. https://doi.org/10.1007/s12686-011-9548-7

Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol. Ecol. 14, 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Fumero, M.V., Yue, W., Chiotta, M.L., Chulze, S.N., Leslie, J.F., Toomajian, C., 2021. Divergence and gene flow between *Fusarium subglutinans* and *F. temperatum* isolated from maize in Argentina. Phytopathology 111, 170–183. https://doi.org/10.1094/PHYTO-09-20-0434-FI

Gale, L.R., Harrison, S.A., Ward, T.J., O'Donnell, K., Milus, E.A., Gale, S.W., Kistler, H.C., 2011. Nivalenol-type populations of *Fusarium graminearum* and *F. asiaticum* are prevalent on wheat in southern Louisiana. Phytopathology 101, 124–134. https://doi.org/10.1094/PHYTO-03-10-0067

Gale, L.R., Ward, T.J., Balmas, V., Kistler, H.C., 2007. Population subdivision of *Fusarium graminearum sensu stricto* in the upper midwestern United States. Phytopathology 97, 1434–1439. https://doi.org/10.1094/PHYTO-97-11-1434

Garcia-Ceron, D., Lowe, R.G.T., McKenna, J.A., Brain, L.M., Dawson, C.S., Clark, B., Berkowitz, O., Faou, P., Whelan, J., Bleackley, M.R., Anderson, M.A., 2021. Extracellular vesicles from *Fusarium graminearum* contain protein effectors expressed during infection of corn. J. Fungi 7, 977. https://doi.org/10.3390/jof7110977

Gautier, M., Klassmann, A., Vitalis, R., 2017. rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. Mol. Ecol. Resour. 17, 78–90. https://doi.org/10.1111/1755-0998.12634

Goswami, R.S., Kistler, H.C., 2004. Heading for disaster: *Fusarium graminearum* on cereal crops. Mol. Plant Pathol. 5, 515–525. https://doi.org/10.1111/J.1364-3703.2004.00252.X

Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H., Bustamante, C.D., 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 5, e1000695. https://doi.org/10.1371/journal.pgen.1000695

Hamamoto, H., Hasegawa, K., Nakaune, R., Lee, Y.J., Makizumi, Y., Akutsu, K., Hibi, T., 2000. Tandem repeat of a transcriptional enhancer upstream of the sterol 14α-demethylase gene (CYP51) in *Penicillium digitatum*. Appl. Environ. Microbiol. 66, 3421–3426. https://doi.org/10.1128/AEM.66.8.3421-3426.2000

Islam, K.T., Bond, J.P., Fakhoury, A.M., 2017. FvSTR1, a striatin orthologue in *Fusarium virguliforme*, is required for asexual development and virulence. Appl. Microbiol. Biotechnol. 101, 6431–6445. https://doi.org/10.1007/s00253-017-8387-1

Jakobsson, M., Rosenberg, N.A., 2007. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23, 1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Kelly, A.C., Ward, T.J., 2018. Population genomics of *Fusarium graminearum* reveals signatures of divergent evolution within a major cereal pathogen. PLoS One 13, e0194616. https://doi.org/10.1371/journal.pone.0194616

Klix, M.B., Verreet, J.A., Beyer, M., 2007. Comparison of the declining triazole sensitivity of *Gibberella zeae* and increased sensitivity achieved by advances in triazole fungicide development. Crop Prot. 26, 683–690. https://doi.org/10.1016/j.cropro.2006.06.006

Kück, U., Stein, V., 2021. STRIPAK, a key regulator of fungal development, operates as a multifunctional signaling hub. J. Fungi 7, 443. https://doi.org/10.3390/jof7060443

Laurent, B., Palaiokostas, C., Spataro, C., Moinard, M., Zehraoui, E., Houston, R.D., Foulongne-Oriol, M., 2018. High-resolution mapping of the recombination landscape of the phytopathogen *Fusarium graminearum* suggests two-speed genome evolution. Mol. Plant Pathol. 19, 341–354. https://doi.org/10.1111/mpp.12524

Leroux, P., Albertini, C., Gautier, A., Gredt, M., Walker, A.S., 2007. Mutations in the CYP51 gene correlated with changes in sensitivity to sterol 14α-demethylation inhibitors in field isolates of *Mycosphaerella graminicola*. Pest Manag. Sci. 63, 688–698. https://doi.org/10.1002/ps.1390

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Liang, J., Lofgren, L., Ma, Z., Ward, T.J., Kistler, H.C., 2015. Population subdivision of *Fusarium graminearum* from barley and wheat in the upper Midwestern United States at

the turn of the century. Phytopathology 105, 1466–1474. https://doi.org/10.1094/PHYTO-01-15-0021-R

Liang, J.M., Xayamongkhon, H., Broz, K., Dong, Y., McCormick, S.P., Abramova, S., Ward, T.J., Ma, Z.H., Kistler, H.C., 2014. Temporal dynamics and population genetic structure of *Fusarium graminearum* in the upper Midwestern United States. Fungal Genet. Biol. 73, 83–92. https://doi.org/10.1016/j.fgb.2014.10.002

Lofgren, L., Riddle, J., Dong, Y., Kuhnem, P.R., Cummings, J.A., Del Ponte, E.M., Bergstrom, G.C., Kistler, H.C., 2018. A high proportion of NX-2 genotype strains are found among *Fusarium graminearum* isolates from northeastern New York State. Eur. J. Plant Pathol. 150, 791–796. https://doi.org/10.1007/s10658-017-1314-6

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303. https://doi.org/10.1101/gr.107524.110

McMullen, M., Bergstrom, G., De Wolf, E., Dill-Macky, R., Hershman, D., Shaner, G., Van Sanford, D., 2012. A unified effort to fight an enemy of wheat and barley: Fusarium head blight. Plant Dis. 96, 1712–1728. https://doi.org/10.1094/PDIS-03-12-0291-FE

Menke, J., Dong, Y.N., Kistler, H.C., 2012. *Fusarium graminearum* Tri12p influences virulence to wheat and trichothecene accumulation. Mol. Plant-Microbe Interact. 25, 1408–1418. https://doi.org/10.1094/MPMI-04-12-0081-R

Nganje, W.E., Kaitibie, S., Wilson, W.W., Leistritz, L.F., Bangsund, D.A., 2004. Economic impacts of Fusarium head blight in wheat and barley: 1993-2001, Agribusiness & Applied Economics.

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M.J., Clark, A.G., Bustamante, C., 2005. Genomic scans for selective sweeps using SNP data. Genome Res. 15, 1566–1575. https://doi.org/10.1101/gr.4252305

Pantazopoulou, A., 2016. The Golgi apparatus: Insights from filamentous fungi. Mycologia 108, 606–622. https://doi.org/10.3852/15-309

Portik, D.M., Leaché, A.D., Rivera, D., Barej, M.F., Burger, M., Hirschfeld, M., Rödel, M.O., Blackburn, D.C., Fujita, M.K., 2017. Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. Mol. Ecol. 26, 5245–5263. https://doi.org/10.1111/mec.14266

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959. https://doi.org/10.1093/genetics/155.2.945

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., De Bakker, P.I.W., Daly, M.J., Sham, P.C., 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575. https://doi.org/10.1086/519795

Puri, K.D., Zhong, S., 2010. The 3ADON population of *Fusarium graminearum* found in North Dakota is more aggressive and produces a higher level of DON than the prevalent 15ADON population in spring wheat. Phytopathology 100, 1007–1014. https://doi.org/10.1094/PHYTO-12-09-0332

Rosenberg, N.A., 2004. DISTRUCT: A program for the graphical display of population structure. Mol. Ecol. Notes 4, 137–138. https://doi.org/10.1046/j.1471-8286.2003.00566.x

Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., Schaffner, S.F., Lander, E.S., Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., Pasternak, S., Wheeler, D.A., Willis, T.D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, Hongbin, Zhao, Hui, Zhou, J., Gabriel, S.B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R.C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Yan, Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Shen, Yayun, Sun, W., Wang, Haifeng, Wang, Yi, Wang, Ying, Xiong, X., Xu, L., Waye, M.M.Y., Tsui, S.K.W., Xue, H., Wong, J.T.F., Galver, L.M., Fan, J.B., Gunderson, K., Murray, S.S., Oliphant, A.R., Chee, M.S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J.F., Phillips, M.S., Roumy, S., Sallée, C., Verner, A., Hudson, T.J., Kwok, P.Y., Cai, D., Koboldt, D.C., Miller, R.D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L.C., Mak, W., You, Q.S., Tam, P.K.H., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., Tanaka, T., Tsunoda, T., Deloukas, P., Bird, C.P., Delgado, M., Dermitzakis, E.T., Gwilliam, R., Hunt, S., Morrison, J., Powell, D., Stranger, B.E., Whittaker, P., Bentley, D.R., Daly, M.J., De Bakker, P.I.W., Barrett, J., Chretien, Y.R., Maller, J., McCarroll, S., Patterson, N., Pe'Er, I., Price, A., Purcell, S., Richter, D.J., Saxena, R., Sham, P.C., Stein, L.D., Krishnan, L., Smith, A.V., Tello-Ruiz, M.K., Thorisson, G.A., Chakravarti, A., Chen, P.E., Cutler, D.J., Kashuk, C.S., Lin, S., Abecasis, G.R., Guan, W., Li, Y., Munro, H.M., Qin, Z.S., Thomas, D.J., McVean, G.,

Auton, A., Bottolo, L., Cardin, N., Eyheramendy, S., Freeman, C., Marchini, J., Myers, S., Spencer, C., Stephens, M., Donnelly, P., Cardon, L.R., Clarke, G., Evans, D.M., Morris, A.P., Weir, B.S., Johnson, T.A., Mullikin, J.C., Sherry, S.T., Feolo, M., Skol, A., Zhang, H., Matsuda, I., Fukushima, Y., MacEr, D.R., Suda, E., Rotimi, C.N., Adebamowo, C.A., Ajayi, I., Aniagwu, T., Marshall, P.A., Nkwodimmah, C., Royal, C.D.M., Leppert, M.F., Dixon, M., Peiffer, A., Qiu, R., Kent, A., Kato, K., Niikawa, N., Adewole, I.F., Knoppers, B.M., Foster, M.W., Clayton, E.W., Watkin, J., Muzny, D., Nazareth, L., Sodergren, E., Weinstock, G.M., Yakub, I., Birren, B.W., Wilson, R.K., Fulton, L.L., Rogers, J., Burton, J., Carter, N.P., Clee, C.M., Griffiths, M., Jones, M.C., McLay, K., Plumb, R.W., Ross, M.T., Sims, S.K., Willey, D.L., Chen, Z., Han, H., Kang, L., Godbout, M., Wallenburg, J.C., L'Archevêque, P., Bellemare, G., Saeki, K., Wang, Hongguang, An, D., Fu, H., Li, Q., Wang, Z., Wang, R., Holden, A.L., Brooks, L.D., McEwen, J.E., Guyer, M.S., Wang, V.O., Peterson, J.L., Shi, M., Spiegel, J., Sung, L.M., Zacharia, L.F., Collins, F.S., Kennedy, K., Jamieson, R., Stewart, J., 2007. Genome-wide detection and characterization of positive selection in human populations. Nature 449, 913–918. https://doi.org/10.1038/nature06250

Son, H., Seo, Y.S., Min, K., Park, A.R., Lee, J., Jin, J.M., Lin, Y., Cao, P., Hong, S.Y., Kim, E.K., Lee, S.H., Cho, A., Lee, S., Kim, M.G., Kim, Y., Kim, J.E., Kim, J.C., Choi, G.J., Yun, S.H., Lim, J.Y., Kim, M., Lee, Y.H., Choi, Y. Do, Lee, Y.W., 2011. A phenome-based functional analysis of transcription factors in the cereal head blight fungus, *Fusarium graminearum*. PLoS Pathog. 7, e1002310. https://doi.org/10.1371/journal.ppat.1002310

Souibgui, E., Bruel, C., Choquer, M., de Vallée, A., Dieryckx, C., Dupuy, J.W., Latorse, M.P., Rascle, C., Poussereau, N., 2021. Clathrin Is Important for Virulence Factors Delivery in the Necrotrophic Fungus *Botrytis cinerea*. Front. Plant Sci. 12, 668937. https://doi.org/10.3389/fpls.2021.668937

Spolti, P., Del Ponte, E.M., Cummings, J.A., Dong, Y., Bergstrom, G.C., 2014a. Fitness attributes of *Fusarium graminearum* isolates from wheat in New York possessing a 3-ADON or 15-ADON trichothecene genotype. Phytopathology 104, 513–519. https://doi.org/10.1094/PHYTO-07-13-0206-R

Spolti, P., Del Ponte, E.M., Dong, Y., Cummings, J.A., Bergstrom, G.C., 2014b. Triazole sensitivity in a contemporary population of *Fusarium graminearum* from New York wheat and competitiveness of a tebuconazole-resistant isolate. Plant Dis. 98, 607–613. https://doi.org/10.1094/PDIS-10-13-1051-RE

Starkey, D.E., Ward, T.J., Aoki, T., Gale, L.R., Kistler, H.C., Geiser, D.M., Suga, H., Tóth, B., Varga, J., O'Donnell, K., 2007. Global molecular surveillance reveals novel Fusarium head blight species and trichothecene toxin diversity. Fungal Genet. Biol. 44, 1191–1204. https://doi.org/10.1016/j.fgb.2007.03.001

Talas, F., Mcdonald, B.A., 2015a. Significant variation in sensitivity to a DMI fungicide in field populations of *Fusarium graminearum*. Plant Pathol. 64, 664–670. https://doi.org/10.1111/ppa.12280

Talas, F., McDonald, B.A., 2015b. Genome-wide analysis of *Fusarium graminearum* field populations reveals hotspots of recombination. BMC Genomics 16, 1–12. https://doi.org/10.1186/s12864-015-2166-0

Wang, C.L., Shim, W.B., Shaw, B.D., 2016. The *Colletotrichum graminicola* striatin orthologue Str1 is necessary for anastomosis and is a virulence factor. Mol. Plant Pathol. 17, 931–942. https://doi.org/10.1111/mpp.12339

Wang, Q., Chen, D., Wu, M., Zhu, J., Jiang, C., Xu, J.R., Liu, H., 2018. MFS transporters and GABA metabolism are involved in the self-defense against DON in *Fusarium graminearum*. Front. Plant Sci. 9, 438. https://doi.org/10.3389/fpls.2018.00438

Ward, T.J., Bielawski, J.P., Corby Kistler, H., Sullivan, E., O'Donnell, K., 2002. Ancestral polymorphism and adaptive evolution in the trichothecene mycotoxin gene cluster of phytopathogenic *Fusarium*. Proc. Natl. Acad. Sci. U. S. A. 99, 9278–9283. https://doi.org/10.1073/pnas.142307199

Ward, T.J., Clear, R.M., Rooney, A.P., O'Donnell, K., Gaba, D., Patrick, S., Starkey, D.E., Gilbert, J., Geiser, D.M., Nowicki, T.W., 2008. An adaptive evolutionary shift in Fusarium head blight pathogen populations is driving the rapid spread of more toxigenic *Fusarium graminearum* in North America. Fungal Genet. Biol. 45, 473–484. https://doi.org/10.1016/j.fgb.2007.10.003

Wickham, H., 2016. ggplot2 Elegant Graphics for Data Analysis (Use R!), Springer.

Zamith-Miranda, D., Nimrichter, L., Rodrigues, M.L., Nosanchuk, J.D., 2018. Fungal extracellular vesicles: modulating host–pathogen interactions by both the fungus and the host. Microbes Infect. 20, 501–504. https://doi.org/10.1016/j.micinf.2018.01.011

Zeller, K.A., Bowden, R.L., Leslie, J.F., 2004. Population differentiation and recombination in wheat scab populations of *Gibberella zeae* from the United States. Mol. Ecol. 13, 563–571. https://doi.org/10.1046/j.1365-294X.2004.02098.x

Zhang, H., van der Lee, T., Waalwijk, C., Chen, W., Xu, Jin, Xu, Jingsheng, Zhang, Y., Feng, J., 2012. Population analysis of the *Fusarium graminearum* species complex from wheat in China show a shift to more aggressive isolates. PLoS One 7, e31722. https://doi.org/10.1371/journal.pone.0031722

Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C., Weir, B.S., 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28, 3326–3328. https://doi.org/10.1093/bioinformatics/bts606

2.7 Figures

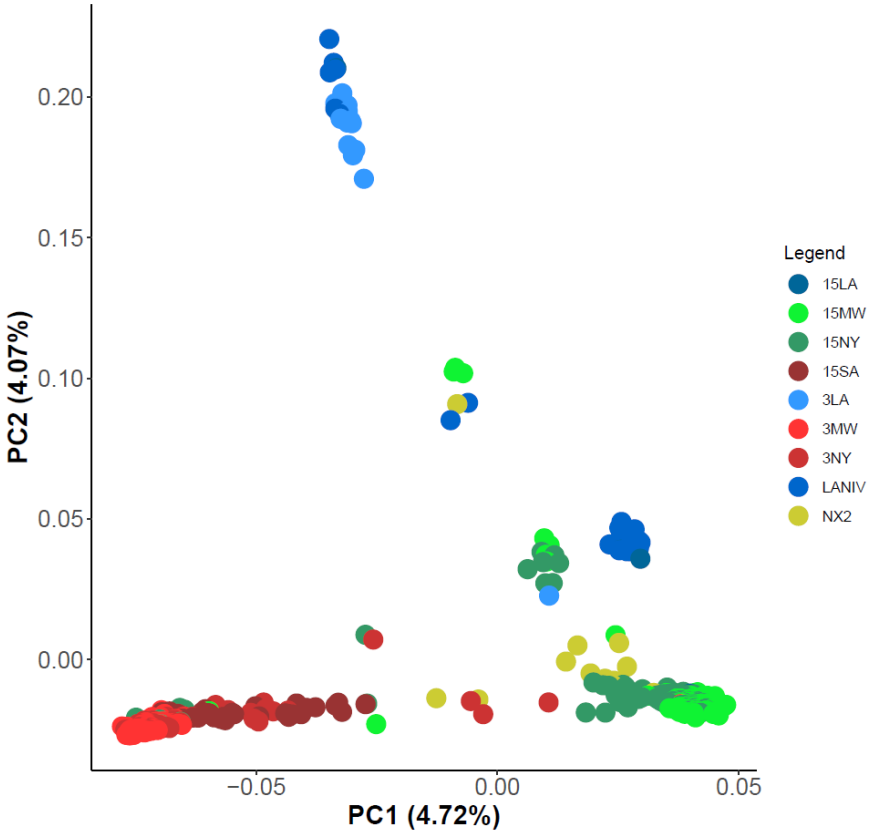


Figure 2.1. Grouping of 454 *Fusarium graminearum* isolates based on the first (PC1) and second (PC2) component axis in principal component analysis.

Percentage in parentheses indicates the variance explained by each PC. Isolates are color-coded based on geographic origin and *TRI* genotype. Genotypes are designated as 3 (3ADON), 15

(15ADON), NX2, and NIV. Geographic origin codes: LA=Louisiana, MW=upper Midwest, NY=New York, SA=South America.

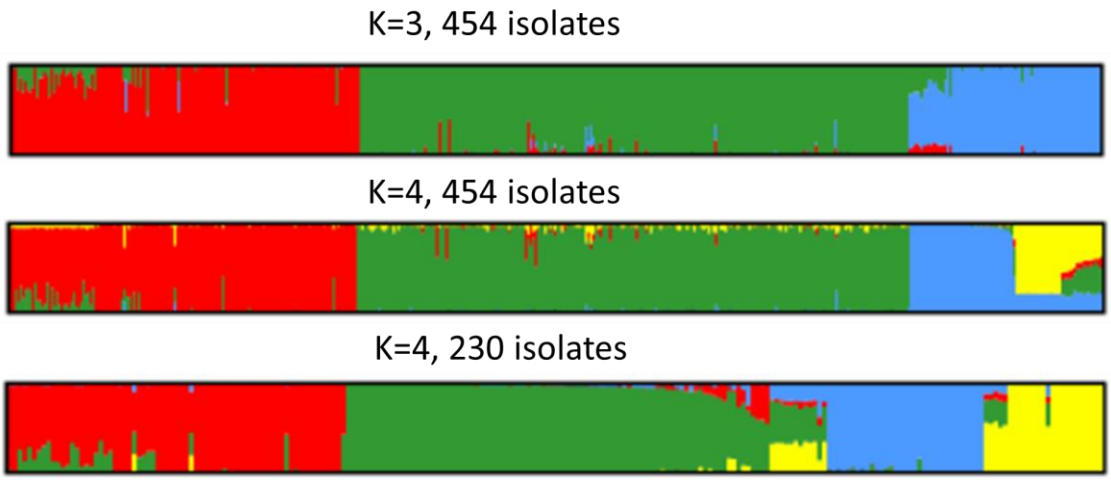


Figure 2.2. Ancestry of *Fusarium graminearum* isolates inferred using STRUCTURE.

K=3 and K=4 were the best fit for a set of 454 and 230 isolates, respectively. Based on the frequency of *TRI* locus genotypes in each STRUCTURE population, we infer the following colors correspond to these previously described populations: red = North America 2 (NA2), green = North America 1 (NA1), blue = Southern Louisiana (SLA), and yellow = Gulf Coast (GC).

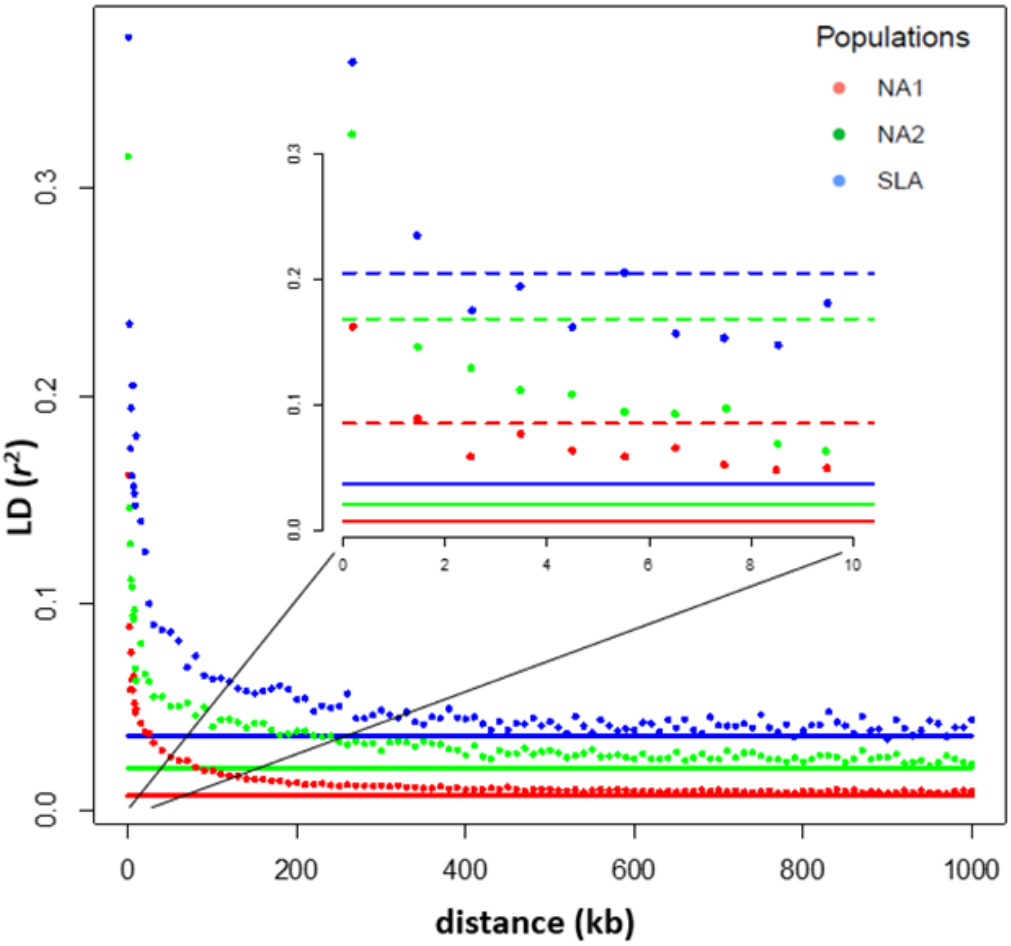


Figure 2.3. Linkage disequilibrium (LD) decay with distance for *Fusarium graminearum* populations.

The x-axes measure the distance between pairs of SNPs, and points represent the average r^2 value for all SNP pairs separated by particular distance intervals. The inset plot uses distance intervals 1 kb wide (e.g., 0-1 kb, 1-2 kb, …, 9-10 kb), while the outer plot duplicates the inset points and also uses four 5 kb wide distance intervals (10-15 kb, 15-20 kb, 20-25 kb, 25-30 kb) and intervals 10 kb wide for distances > 30 kb. Horizontal lines represent r^2 for pairs of unlinked SNPs (SNPs on different chromosomes). Dashed lines on the inset plot represent r^2 values halfway between the maximum and minimum values.

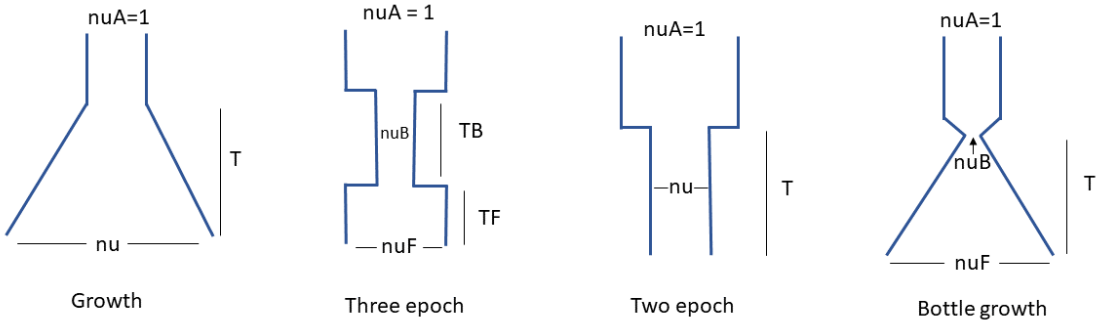


Figure 2.4. Schematic representation of dadi 1D models

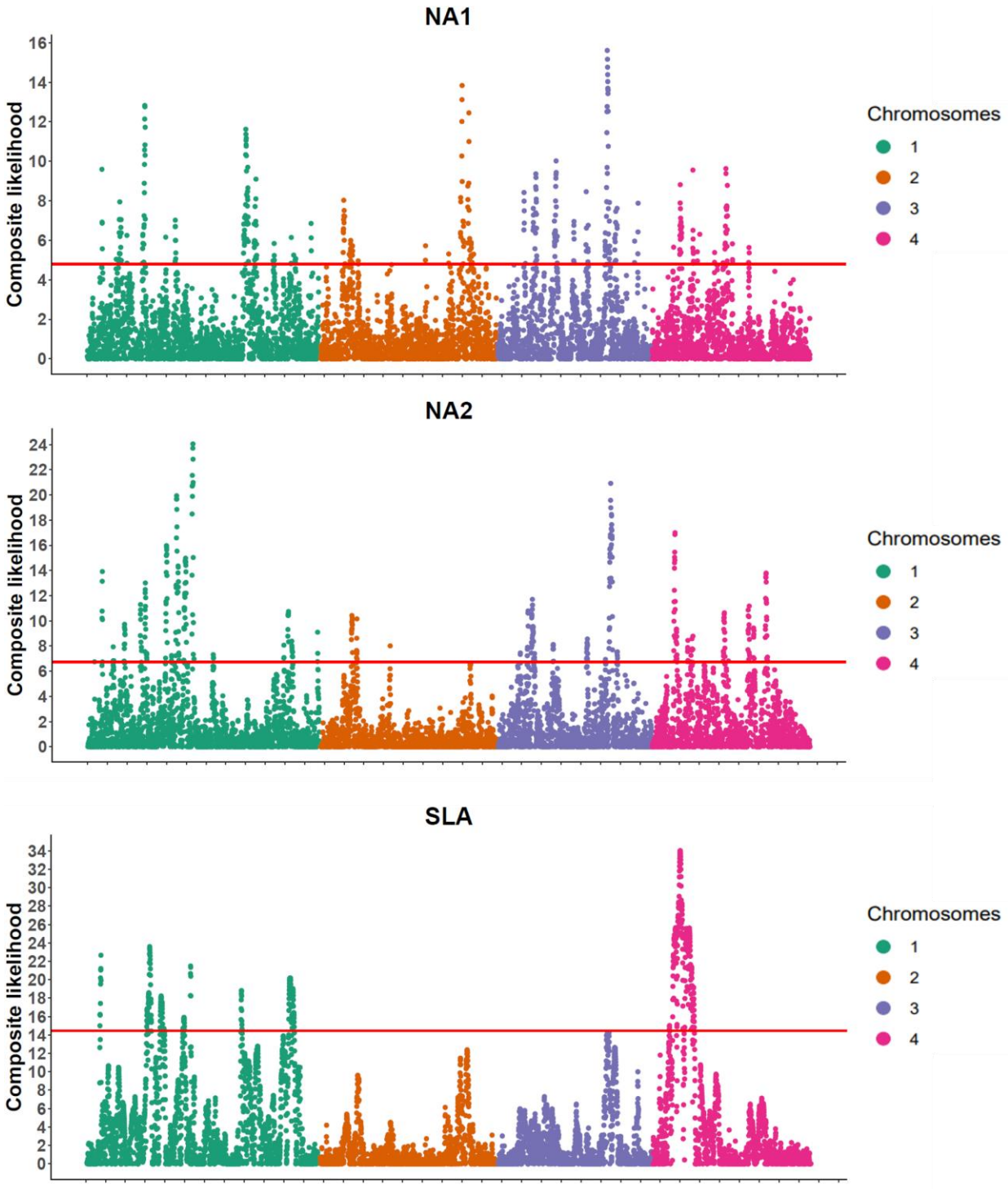


Figure 2.5. Regions showing patterns consistent with recent positive selection in *Fusarium graminearum* populations (NA1, NA2, SLA) as identified by Sweepfinder2.

Selection scans were conducted with each population using Sweepfinder2. Horizontal red lines represent the 95th percentile of composite likelihood values. Tick marks on the x-axis are spaced every 1 Mb.

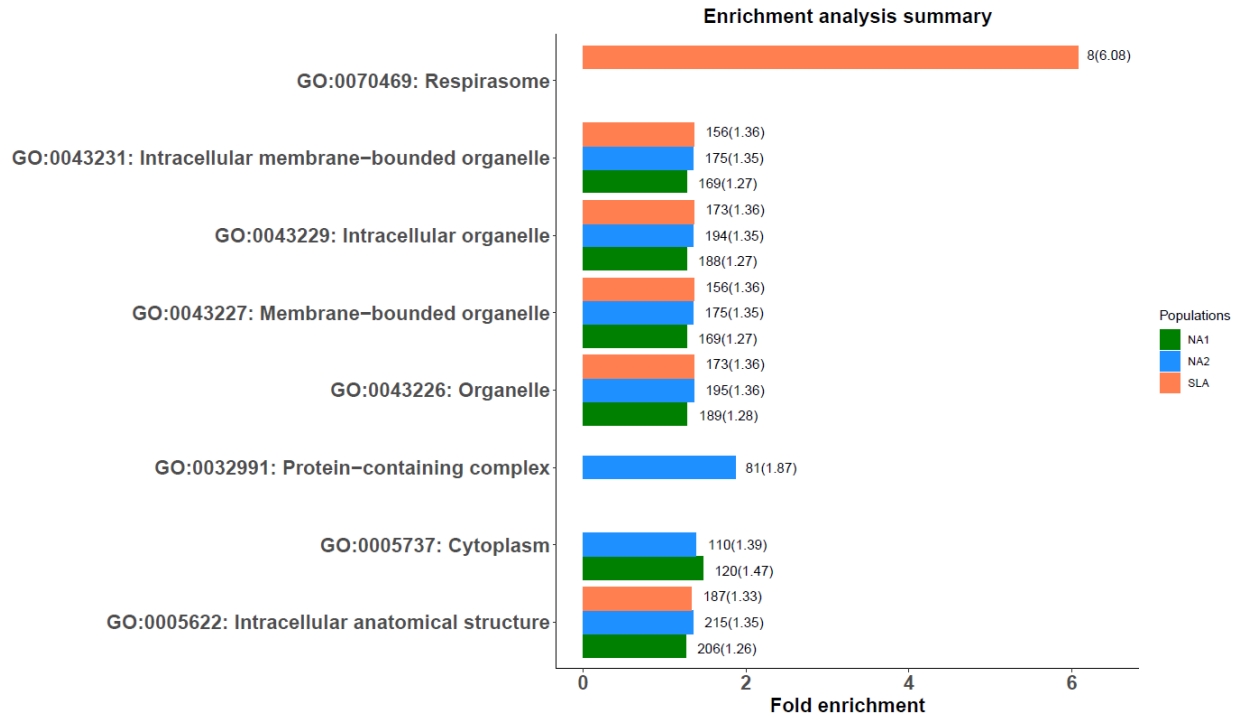


Figure 2.6. Summary of the cellular component GO term enrichment for the genes in the outlier regions identified in the selective sweep scans conducted using Sweepfinder2.

The numbers on top of the bars are the number of genes for a particular GO term and numbers on parenthesis is the fold enrichment of that GO term.

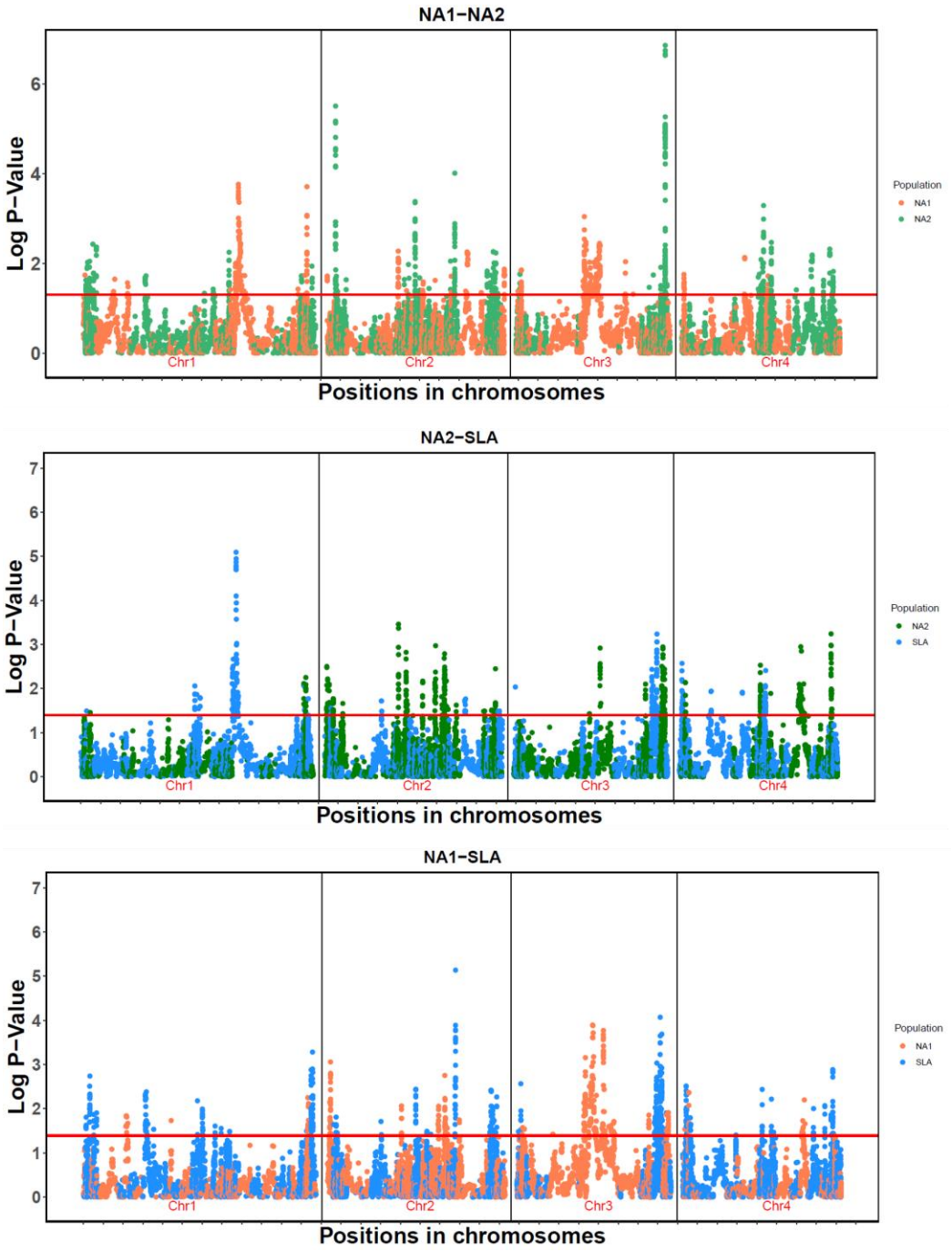


Figure 2.7. Regions under the influence of positive natural selection in *Fusarium graminearum* populations as identified using XP-EHH.

p-values for the cross population extended haplotype homozygosity statistic XP-EHH were obtained by scanning genomes of two populations against each other using R package rehh. The different colors in each panel indicate evidence for selection occurring in each population.

Horizontal red lines represent the 95th percentile of log P-value. Tick marks on the x-axis are spaced every 1 Mb.

2.8 Tables

Table 2.1. Summary of demographic parameter estimates and uncertainty surrounding the parameter estimates for 3D dadi models.

	1: sim_split_symmig_all		2: split_symmig_all		3: split_symmig_adacent A		4: split_symmig_adacent B		5: split_symmig_adjacent C		6: split_no mig	
	Log-likelihood = -5270.49		Log-likelihood = -5423.06		Log-likelihood = -6041.45		Log-likelihood = -5426.62		Log-likelihood = -5377.22		Log-likelihood = -6145.44	
Parameters	Estimates	Uncertainty	Estimates	Uncertainty	Estimates	Uncertainty	Estimates	Uncertainty	Estimates	Uncertainty	Estimates	Uncertainty
$nu1_a$	0.47	0.03	0.42	0.05	0.39	0.02	0.29	0.08	0.58	0.08	0.30	0.05
nuA_b	NA	NA	0.05	0.13	0.20	0.06	0.08	0.01	3.38	0.02	0.86	0.96
$nu2_c$	0.66	0.04	0.56	0.05	0.30	0.02	0.97	0.13	1.12	0.13	0.32	0.08
$nu3_d$	1.35	0.07	1.32	0.45	1.45	0.51	0.49	0.06	0.41	0.06	0.99	0.19
mA_e	NA	NA	4.02	13.64	0.06	2.70	4.12	3.81	0.68	3.73	NA	NA
$m1_f$	0.04	0.01	0.18	0.19	0.16	0.14	0.39	0.14	0.18	0.14	NA	NA
$m2_g$	0.10	0.02	0.17	0.02	0.15	0.02	0.17	0.03	0.35	0.04	NA	NA
$m3_h$	0.25	0.04	0.32	0.03	NA	NA	NA	NA	NA	NA	NA	NA
$T1_i$	0.94	0.07	0.13	0.15	0.05	0.00	0.03	0.01	0.09	0.01	0.01	0.03
$T2_j$	NA	NA	1.23	1.25	0.34	0.01	0.45	0.12	0.76	0.12	0.26	0.07
p_misid_k	0.05	0.01	0.07	0.05	0.04	0.02	0.03	0.01	0.04	0.01	0.03	0.01
$theta_l$	581.34	47.78	624.12	206.57	783.48	84.49	838.07	131.97	651.34	134.92	885.33	129.84

The log-likelihood, parameter estimates and uncertainty estimates correspond to the best replicate for a particular model. The uncertainty was calculated using Godambe information matrix (GIM) implemented in dadi. All migration rates are expressed as 2*Na*m. All times are expressed in units of 2*Na generations. Na is effective ancestral population size. m is effective migration rate and μ is rate of mutation per generation. All population sizes are relative to ancestral population size.

1: **Sim_split_symmig_all** 2: **split_symmig_all** 3: **split_symmig_adjacent A** 4: **split_symmig_adjacent B** 5: **split_symmig_adjacent C** 6: **split_nomig**

a: nu1- size of first population to split. SLA in models 1, 2, 3, 4, and 6; NA2 in model 5.

b: nuA- size of intermediate ancestral population. NA2+NA1 in models 2, 3, 4, and 6; NA1+SLA in model 5.

c: nu2- size of second population after split. NA2 in models 1, 2, 3, and 6; NA1 in models 4 and 5.

d: nu3- size of third population after split. NA1 in models 1, 2, 3, and 6; NA2 in model 4; SLA in model 5.

e: mA- migration between intermediate ancestral population and first population to split. SLA and NA1+NA2 in models 2, 3, and 4; NA3 and NA1+SLA in model 5.

f: m1- migration between populations 1 and 2. SLA and NA2 in models 1, 2, and 3; SLA and NA1 in model 4; NA1 and NA2 in model 5.

g: m2- migration between populations 2 and 3. NA2 and NA1 in models 1, 2, 3, and 4; NA1 and SLA in model 5.
h: m3- migration between non-adjacent populations (1 and 3). SLA and NA1 in models 1 and 2.
i: T1- Time between most ancient split event and either present (model 1) or more recent split (models 2 – 6).
j: T2- Time between present and the more recent split. Between populations NA2 and NA1 in models 2, 3, 4, and 6; NA1 and SLA in model 5.
k: p_misid- population misidentification parameter
l: theta- Population mutation parameter which equals to 4Naμ

2.9 List of supplemental files

Supplemental file 2.1: Eigenvalues from principal component analysis of 454 isolates, and information on origin (collection location and date) of these isolates.

Supplemental file 2.2: Population assignment coefficients obtained using STRUCTURE for 454 clone-corrected *Fusarium graminearum* isolates for runs that assumed 3 (K=3, the best-fit for this sample based on the Evanno method) and 4 (K=4) populations.

Supplemental file 2.3: Population assignment coefficients obtained using STRUCTURE for 230 clone-corrected *Fusarium graminearum* isolates. Four populations (K=4, the best-fit for this sample based on the Evanno method) were assumed.

Supplemental file 2.4: The Delta K (ΔK) statistic from the Evanno method was used to identify the optimum number of populations in a set of 454 and 230 *Fusarium graminearum* isolates.

Supplemental file 2.5: Summary of 2D dadi demographic models for population pairs NA2-SLA, NA1-SLA, and NA2-NA1. Models tested support for symmetric and asymmetric migration between population pairs.

Supplemental file 2.6: Summary of 1D dadi demographic models for populations NA1, NA2 and SLA. Five models were used to test the change in population size.

Supplemental file 2.7: Summary of genes that fall in the Sweepfinder2 outlier regions for population NA1. Gene annotations, gene ontology terms, and known curated phenotypes for genes in the outlier regions were obtained from fungiDB. Information on genes transported and or enriched in vesicles and potentially secreted were obtained from Garcia-Ceron et al. (2021), and Brown et al. (2012).

Supplemental file 2.8: Summary of genes that fall in the Sweepfinder2 outlier regions for population NA2. Gene annotations, gene ontology terms, and known curated phenotypes for genes in the outlier regions were obtained from fungiDB. Information on genes transported and or enriched in vesicles and potentially secreted were obtained from Garcia-Ceron et al. (2021) and Brown et al. (2012)

Supplemental file 2.9: Summary of genes that fall in the Sweepfinder2 outlier regions for population SLA. Gene annotations, gene ontology terms, and known curated phenotypes for genes in the outlier regions were obtained from fungiDB. Information on genes transported and or enriched in vesicles and potentially secreted were obtained from Garcia-Ceron et al. (2021), and Brown et al. (2012)

Supplemental file 2.10: Summary of genes that fall in the outlier regions for population NA1 in cross population haplotype homozygosity scan performed using rehh in R. Gene annotations, gene ontology terms, and known curated phenotypes for genes in the outlier regions were obtained from fungiDB. Information on genes transported and or enriched in vesicles and potentially secreted were obtained from Garcia-Ceron et al. (2021), and Brown et al. (2012)

Supplemental file 2.11: Summary of genes that fall in the outlier regions for population NA2 in cross population haplotype homozygosity scan performed using rehh in R. Gene annotations, gene ontology terms, and known curated phenotypes for genes in the outlier regions were obtained from fungiDB. Information on genes transported and or enriched in vesicles and potentially secreted were obtained from Garcia-Ceron et al. (2021), and Brown et al. (2012)

Supplemental file 2.12: Summary of genes that fall in the outlier regions for population SLA in cross population haplotype homozygosity scan performed using rehh in R. Gene annotations, gene ontology terms, and known curated phenotypes for genes in the outlier regions were obtained from fungiDB. Information on genes transported and or enriched in vesicles and potentially secreted were obtained from Garcia-Ceron et al. (2021), and Brown et al. (2012)

Chapter 3 - Identification of the genetic basis of saprophytic and pathogenic fitness in *Fusarium graminearum*

3.1 Abstract

Fusarium graminearum causes FHB, a devastating disease of wheat and barley worldwide. Many *F. graminearum* traits, including aggressiveness, sexual fertility, *in vitro* growth rate, mycotoxin production *in vitro* and *in planta*, and fungicide resistance, are quantitative in nature. The genetic basis of each of these traits has not been fully elucidated. We collected phenotype measures from over 100 isolates, 135 to 152 in most cases, for each of these traits. Phenotypic measurements were collected in laboratory experiments, except for aggressiveness, which was performed in the greenhouse. Isolates were previously genotyped in the lab using genotyping by sequencing (GBS). Genome-wide association mapping was conducted using the models implemented in R packages GAPIT3 and mrMLM. We identified a total of 13 significant QTNs for growth rate, 30 for sexual fertility, 48 for propiconazole sensitivity, 37 for tebuconazole sensitivity, 8 for DON production *in vitro*, 3 for 15ADON production and 2 for aggressiveness in two wheat genotypes. Some of these QTNs fall on or in close proximity to the genes with previously known roles for the respective traits. However, many of the QTNs are novel and provide new candidates for functional studies.

3.2 Introduction

Fusarium graminearum is a filamentous ascomycete fungus and major causal agent of FHB of wheat and barley. FHB is a major constraint to wheat production throughout the world. *Fusarium graminearum* also causes stalk and ear rot of corn and can also infect other small grain cereals like oats and other grass hosts (Goswami and Kistler, 2004; McMullen et al., 2012).

Besides the direct yield loss related to the disease, farmers also suffer indirect losses caused by price discounts due to the contamination of infected grains with trichothecene mycotoxin (Goswami and Kistler, 2004; McMullen et al., 2012; Wilson et al., 2018). Deoxynivalenol is one major trichothecene mycotoxin produced by *F. graminearum*, and isolates are classified into chemotype based on the different variants of trichothecene mycotoxin produced by the isolates, like 3ADON, 15ADON, NX2 and NIV (Kelly and Ward, 2018; Zeller et al., 2004). In the United States *F. graminearum* isolates exist in populations correlated with the trichothecene mycotoxin chemotypes of the isolates.

In between growing seasons and crop hosts, *F. graminearum* isolates survive as saprophytic mycelia in the residues of small grain crops (Leplat et al., 2013). A faster growth rate helps the fungus to rapidly colonize the residues. The growth rate of *F. graminearum* isolates on PDA and their production of ascospores on corn stalks are significantly correlated (Spolti et al., 2014a). Ascospores are the sexual spores of *F. graminearum* and are the primary source of infection to start a new disease cycle every growing season (Bai and Shaner, 2004; Goswami and Kistler, 2004; McMullen et al., 2012). Ascospores are produced in perithecia, the sexual fruiting bodies, in wheat or corn residues and are forcibly discharged into the air (Trail et al., 2002). Infection starts when an ascospore that lands on a wheat spikelet germinates. Wheat is susceptible to infection from flowering to the soft dough stage, but most susceptible during anthesis when the anthers are fresh and yellow (Bai and Shaner, 2004). Thus, the ability of the population to produce perithecia and ascospores (sexual fertility) is likely one of multiple factors determining disease pressure in the field. The MAT loci are the master regulators of sexual reproduction in *F. graminearum*.

The trichothecene mycotoxin is a known aggressiveness factor in *F. graminearum*. In the absence of trichothecene, *F. graminearum* can infect the wheat spike but the disease does not progress much beyond the point of infection (Proctor et al., 1995). Trichothecene biosynthesis is regulated and carried out by the proteins coded by clusters of genes located in three genomic loci. A total of 12 out of 15 genes identified to have a role in trichothecene biosynthesis are clustered in a single locus called the *TRI* locus, while another two genes cluster in the *TRI1-TRI16* locus and the remaining locus has a single gene, *TRI101* (Alexander et al., 2009; Gale et al., 2005).

In addition to cultural practices like growing resistant cultivars and crop residue management, fungicide application during anthesis is an important, effective, and commonly used component of integrated FHB management. Demethylation inhibitor (DMI) fungicides like propiconazole, metconazole, prothiconazole, and tebuconazole have been used to manage FHB in the US since the mid-1990s (McMullen et al., 2012). DMI fungicides when used in combination (as a mixture) can provide up to 50% protection compared to a control not treated with fungicide (McMullen et al., 2012). They function by binding to 14-a demethylases and inhibiting the synthesis of ergosterol, a major component of fungal cell walls. The fungal 14-a demethylase is encoded by the CYP51 gene, of which *F. graminearum* has three paralogs, CYP51A, CYP51B, and CYP51C. Intensive application of these fungicides has led to the development of fungicide resistance in some *F. graminearum* field isolates/populations collected from Asia, Europe, and the USA (Klix et al., 2007; Spolti et al., 2014b; Yin et al., 2009). Mutation in CYP51 genes, overexpression of CYP51 genes, and transportation of fungicide out of fungal cells are some of the known mechanisms for azole resistance (Hamamoto et al., 2000; Hayashi et al., 2003; Leroux et al., 2007; Ma et al., 2006; Reimann and Deising, 2005).

Previous studies in *F. graminearum* suggest abundant variation in these fitness traits and the quantitative nature of the traits. Isolates of the 3ADON chemotype are reported by some studies to grow faster (Ward et al., 2008), produce more DON (Von der Ohe et al., 2010; Ward et al., 2008) and be more aggressive (Ward et al., 2008) than the 15ADON chemotype. There is no agreement on this topic, as some other studies did not find 3ADON isolates to be any more aggressive than 15ADON isolates (Spolti et al., 2014a; Von der Ohe et al., 2010) or significantly different with respect to growth or ascospore production on corn stalk (Spolti et al., 2014a). Most of the genes identified so far to have a role in fitness come from a candidate gene-based approach. Candidate gene-based approaches are likely to miss important genes in fitness-related pathways, as they are restricted to looking into one or certain types of genes in the genomes. Genome wide association studies (GWAS) scan the genomes free of any hypothesis as to specific causative genes, screen the genome without any inclination to a specific region, type of gene or variant (Kitsios and Zintzaras, 2009) and are likely to find genes and pathways missed by the candidate gene approach. Genome-wide association studies have been successfully used in fungi to identify the genetic determinants of aggressiveness (Atwell et al., 2018; Hartmann et al., 2017; Miedaner et al., 2021) and mycotoxin production (Miedaner et al., 2021). So far, there is one published GWAS study on *F. graminearum* field isolates from Germany that only focuses on aggressiveness, DON production and propiconazole sensitivity.

To better understand the genomic architecture of fitness in *F. graminearum*, we conducted genome wide association studies (GWAS) for five different fitness traits: growth rate, sexual fertility, aggressiveness, mycotoxin production *in vitro* and *in planta* and sensitivity to two azole fungicides.

3.3 Materials and Methods

3.3.1 Growth rate

For long-term storage, *F. graminearum* isolates were stored at -80℃ in 20% glycerol solution. A total of 135 isolates from the NA1 and NA2 populations as well as some admixed isolates were used. Isolates were revived in complete media tubes (Leslie and Summerell, 2007). To measure the growth rate, a single germinated macroconidia was transferred to the center of the Speziller Nahrstoffarmer agar (SNA) (KH_2PO_4: 1g, KNO_3: 1g, $MgSo_4.7H_2O$: 0.5 g, KCL: 0.5 g, glucose: 0.2 g, sucrose 0.2g and agar 20 g in 1 liter water) plate (Leslie and Summerell, 2007) and incubated in the dark for 5 days. Single spores were separated using a micro manipulator. Growth rate was assessed at three temperatures: 15℃, 25℃, and 30℃. For each isolate, measurements were taken from three plates (three replicates). Three incubators set at the three temperatures were used for incubation. Growth measurements were taken in batches of 10 isolates, and 2 common control isolates were included with every batch. At 4 and 5 days of incubation, two perpendicular measurements of the diameter of the growing fungal colony were taken for each plate. Least square means (ls means) for isolates adjusted for temperature and block effect was calculated using an incomplete block design in R using packages lmerTest (Kuznetsova et al., 2017) and ls means (Lenth, 2016). Measurements from day 4 were used, as we observed the colonies reaching the edge of the plates by day 5 for some isolates, so the measured diameter might not reflect the true growth rate. Isolate and temperature effects were used as fixed effects and batch as a random effect.

3.3.2 Ascospore production

The number of ascospores produced by 152 *F. graminearum* isolates was quantified by growing isolates on 60 mm carrot agar plates (Leslie and Summerell, 2007). Selected isolates

included those from the NA1 and NA2 populations as well as some admixed ones. Previously published protocols (Leslie and Summerell, 2007; Trail et al., 2002) with some modifications were used. Briefly, *F. graminearum* isolates were grown on carrot agar plates for 5 days at 24℃ under bright florescent and black light that were set to be on for 12 hours light and off for 12 hours dark conditions. At 5 days, mycelia on carrot agar plates were removed with a sterile toothpick and self-fertilized with 1 ml of 2.5% Tween 60. Glass rods bent to the shape of hockey sticks were used to work the Tween solution around the plate during self-fertilization. Post fertilization, plates were incubated at 24℃ under bright florescent and black light and 12 hours light/ 12 hours dark conditions. Ascospores were collected at 8, 9, and 10 days post-fertilization. For ascospore collection, carrot agar plates were inverted on top of 60 mm plates with 800 microliters of phosphate buffered saline (PBS) and ascospores were collected overnight (12 - 14 hours period). At the end of the collection period, the ascospore suspension in PBS was transferred to 1.5 ml microcentrifuge tubes and refrigerated. Ascospores from the three days were pooled in 15 ml centrifuge tubes and counted using a hemocytometer. The spore suspension was diluted as necessary. Isolates were grown in batches of ~50 isolates with two common control isolates in each batch. Each isolate was replicated three times. Ls means for isolates adjusted for block effect was calculated using incomplete block design in R using packages lmerTest and ls means. Isolate effects were modeled as a fixed effect and batch effect as a random effect.

3.3.3 Greenhouse inoculation and measurement of aggressiveness

Conidia for inoculation were produced on mung bean broth (MBB). The inoculum for MBB was grown on a complete media tube for five days. On the 5^{th} day, the spore suspension was collected by adding ~1 ml of Tween solution in the tube and working it around with the

sterile glass pipet. The collected spore suspension was used to inoculate 40 ml of MBB in 125 ml conical flasks. MBB was prepared by boiling 40 g of mung beans with 1 liter of water for 10 minutes. The beans were filtered using a cheesecloth. The inoculated MBB culture flasks were grown in a shaker for 8-10 days. The shaker was set at 120 RPM and was exposed to white light for 12 hours and 12 hours of darkness. At the end of the growing cycle, mycelium was filtered out using a double layer of cheesecloth, and the spore suspension was collected in a 50 ml centrifuge tube. The spore suspension was centrifuged at 4000 rpm for 5 minutes to get a pellet of spores. MBB was removed without disturbing the pellet and the spores were resuspended in sterile distilled water (different volumes depending on the pellet size). Spore concentration was determined using a hemocytometer. The spore suspension was frozen in screw cap tubes at -80°C until inoculation. On the inoculation day, the frozen tubes of inoculum were thawed, and the spore suspension was diluted to $1x10^5$ spores per ml in sterile distilled water.

The aggressiveness of *F. graminearum* isolates was assessed by inoculating two wheat varieties, Wheaton and Rollag, in the greenhouse over the 2021/22 and 2022/23 growing seasons. Rollag carries the *Fhb1* QTL and is moderately resistant, whereas Wheaton is susceptible to FHB (Anderson et al., 2015; Li et al., 2015). Wheaton and Rollag cultivars were grown in 6 inch square pots. Nine to twelve seeds were planted in a pot and later thinned to have a maximum of 6-7 plants per pot. Planting was staggered into batches with 10 day intervals between the batches. Isolates were divided into 7 batches during the 2021/22 growing season and 4 batches in the 2022/23 growing season. We tested a total of 136 isolates (for both cultivars) during the first and 114 isolates on Rollag and 120 isolates on Wheaton during the second growing season. Isolate 11157 was included with each batch and served as a control isolate. The plants were fertilized at the 2-3 leaves stage with macro and micronutrients. Marathon granules

(1%) were used to control aphids when observed. Towards the time of heading, all the plants in a pot were tied to a bamboo stake to help keep the plants erect. Wheat heads were inoculated at the flowering stage when the anthers were bright yellow. 10 µl of suspension ($1x10^5$ spores per ml) was pipetted into a central spikelet (Jin et al., 2014). The inoculated spikelet was marked with a sharpie and the spike was covered using a transparent plastic bag for 48 hours. Up to 6 spikes per *F. graminearum* isolate were inoculated per replicate, and the average of disease scores across these spikes was used for each isolate. Disease scores were taken by counting the number of symptomatic spikelets and total spikelets in the inoculated head to get the percentage of symptomatic spikelets. Disease scores were taken at 14 days and 21 days post-inoculation. Wheat heads were harvested upon ripening. An automated thresher (Precision Machine, Lincoln, NE) was used for threshing. To retain the shriveled infected grains that were often blown away because of their light weight, chaff for each sample was manually inspected and shriveled low weight grains were collected by hand. The grains were ground using a coffee grinder. When the weight of a grain sample was less than 1 g, unprocessed grains were used for mycotoxin analysis. Measurement of mycotoxins in grain samples (ground or whole) was performed in the mycotoxin testing lab at Virginia Tech using gas chromatography-mass spectrometry (GC-MS). Samples were tested for DON, 15ADON, 3ADON and NIV content.

The percentage of diseased spikelets from two seasons and two cultivars were used to obtain the ls means adjusted for cultivars, growing seasons and batch effect. Only isolates (92) common across two seasons and two cultivars and included in same batch for both Wheaton and Rollag during the respective years were used. Entries where no disease was observed on the 21st day were not used. A split plot design, with cultivars as whole plot and isolates as subplot, was

used to obtain the ls means. The model treats cultivar and isolate effects as fixed effects and year and year x cultivar as random effects.

3.3.4 *In vitro* DON production

The quantity of DON, 3ADON, 15ADON and NIV mycotoxins produced by *F. graminearum* isolates was assessed *in vitro* using the rice culture technique (Burlakoti et al., 2011, 2008; Puri and Zhong, 2010). 20 g of parboiled rice (Ben's) was soaked in 25 ml water in a 250 ml conical flask two nights before inoculation. A day before inoculation, flasks with water-soaked rice were covered with a double layered coffee filter that was secured around the neck with a rubber band. Flasks were autoclaved twice (20-minute liquid cycle). Flasks were cooled in between the autoclave cycles. Inoculation was performed by taking 3 plugs (5mm in diameter) from 4 day old *F. graminearum* cultures on mung bean agar plates. To allow for proper mixing of conidia with rice, conical flasks were shaken every other day for up to 1 week after inoculation. Conical flasks with rice cultures were incubated in the dark at room temperature for 21 days. At 21 days, rice cultures were transferred to a 50 ml centrifuge tube and frozen at -80°C. Rice cultures were freeze-dried at -40°C (Thermo Savant, ModulyoD benchtop dryer) for four days followed by grinding (Cuisinart, grind central coffee grinder). Powdered rice cultures were used for mycotoxin analysis with gas chromatography-mass spectrometry (GC-MS). Analysis for DON, 15ADON, 3ADON, and NIV was conducted at the mycotoxin testing lab at Virginia Tech.

3.3.5 Fungicide sensitivity

Sensitivity of 152 *F. graminearum* isolates towards two DMI fungicides, propiconazole and tebuconazole, was tested using a flat bottom 96 well plate assay (Talas and Mcdonald, 2015). Propiconazole and tebuconazole were tested in 8 different concentrations: 240 ppm, 80 ppm, 30 ppm, 10 ppm, 4.5 ppm, 1.5 ppm, 0.5 ppm and 0 ppm for propiconazole and 216 ppm, 72

ppm, 24 ppm, 8 ppm, 3 ppm, 1 ppm, 0.45 ppm and 0 ppm for tebuconazole. Fungicide solutions were prepared by mixing the fungicide with synthetic nutrient broth. The concentration of DMSO was kept constant at 1.95 % (V/V) for all fungicide concentrations. Streptomycin sulphate was added to the liquid media at the rate of 0.33 mg/ ml to prevent bacterial contamination. 150 µl of fungicide solution and 50 µl spore suspension (adjusted to $1x10^5$ spores/ml) were added to the wells of PR1MA flat bottom 96 well plates (MIDSCI) using a liquid handling robot (Beckman Coulter, BiomeK NX^p). All of the reported concentrations (fungicide, DMSO, and streptomycin) are for the final solution (150 µl fungicide solution + 50 µl spore suspension). Control wells had water rather than a spore suspension added to the fungicide and media solution. Spores were collected by growing *F. graminearum* isolates in 40 ml mung bean broth for about 8 days in 100 ml conical flasks. Mycelia was filtered out with double layered muslin cloth, and the liquid culture was centrifuged at 4000 rpm for 5 minutes to get a pellet of macroconidia in 50 ml centrifuge tubes. Liquid media was carefully removed without disturbing the pellet and spores were resuspended in sterile double distilled water, counted using a hemocytometer and frozen at -80°C. Tubes with concentrated spore solutions were thawed overnight and adjusted to the required concentration on the day of the inoculation. The plates were covered with lids, sealed with parafilm and incubated in the dark at room temperature for seven days. In order to provide humid conditions, plates were incubated over a mat of moist paper towels or incubated in trays filled with water. Absorbance readings were taken before incubation, and at 5 and 7 days at 405 ŋm using a plate reader (FLUOstar Omega SNP, BMG LABTECH). Each isolate was tested in three replicates. Effective concentration 50 (EC50) values for propiconazole and tebuconazole were computed using R package dr4pl (An et

al., 2019). Ls means adjusted for three replicates were calculated using a complete block design where isolate effect is modeled as a fixed effect and replicate as a random effect.

3.3.6 Genotyping

Isolates were sequenced using genotyping by sequencing (GBS) (Yue, 2017). Polymorphisms were extracted using GATK 4.1.8.1 (McKenna et al., 2010). Sites with genotype calls in over 25% of a set of 555 isolates were used for imputation. Missing data was imputed using Beagle 4.0 (Browning and Browning, 2007). Different GWAS analyses include different numbers of isolates, either because a different number was used for different traits, or because additional analyses were performed with a subset of the total sample for a given trait. The subsets of sites that are polymorphic for each smaller set of isolates were obtained and filtered to exclude sites with minor allele frequency less than 0.05.

3.3.7 GWAS analysis

Genome wide association analyses were conducted using single and multi-locus mixed linear models implemented in R packages GAPIT3 (Wang and Zhang, 2021) and mrMLM 5.0.1 (Zhang et al., 2020). GAPIT models include mixed linear model (MLM), multiple loci mixed linear model (MLMM), fixed and random circulation probability unification (FarmCPU), and Bayesian information and linkage disequilibrium iteratively nested keyway (BLINK). Mixed linear models incorporate population structure (Q) and/or kinship (K) to control false positives. For complex traits, the use of the Q or K matrix can confound the signal of association (Liu et al., 2016). To reduce the confounding effect of population stratification, MLMM uses multiple associated markers (pseudo QTNs) as covariates with a testing marker (Segura et al., 2012). FarmCPU is a two-step model (Liu et al., 2016). The fixed effect and random effect models are used iteratively. First, a random effect model is used to select pseudo QTNs using maximum

likelihood. This prevents model overfitting. Markers are then tested using a fixed effect model where multiple pseudo QTNs are used as covariates to remove the confounding effect of kinship. This reduces false positives and increases power. Unlike FarmCPU, BLINK used two fixed effect models iteratively and is computationally more efficient (because it does not use a random effect model) (Huang et al., 2019). The random effect model is replaced by a fixed effect model that used Bayesian information criteria (BIC) to identify pseudo QTNs. BLINK uses the LD pattern between the markers to select the markers to be tested, unlike the binning approach used in FarmCPU. MLM is a single locus model and MLMM, FarmCPU and BLINK are multi-locus models. For GAPIT models, the optimum number of principal component axes to include was determined using Model.selection as true. Principal components (PCs) and kinship matrices were calculated by the software. Markers with FDR corrected P-values less than 0.05 were considered significant.

Models implemented in mrMLM 5.0.1 consider SNPs as random effects. Five random SNP effect models, namely mrMLM, FASTmrMLM (FAST mrMLM multi-locus mixed linear model), FASTmrEMMA (fast multi-locus random-SNP-effect EMMA), pKWmEB (Integrates Kruskal_Wallis test with empirical Bayes), and ISIS EM BLASSO (iterative modified-Sure Independence Screening EM-Bayesian Lasso) were used. pKWmEB is a non-parametric model. mrMLM models operate in two steps. In the first step, marker effect is treated as random, and SNPs, one by one, are tested for association. A less stringent Bonferroni threshold is used to select potentially associated markers for the next step. The second step uses a multi-locus mixed linear model to estimate marker effect using an EM empirical Bayes approach. The relatedness of the individuals was adjusted by supplying the kinship matrix. Population structure adjustment when needed was performed by providing principal component weightings. The optimal number

of components to use was decided based on model selection results in GAPIT3 and diagnostic Q-Q plots. Markers with LOD scores greater than or equal to 3 were considered significant.

3.3.8 Candidate gene analysis

Genes surrounding the significant QTNs were obtained using a custom script. The windows for each QTN were obtained by manually inspecting the LD decay pattern around the QTN using the genotype file for the respective trait and sample in question. The curated phenotypes and gene ontology terms for the candidate genes were obtained from FungiDB.

3.4 Results

3.4.1 Distribution of phenotypic trait measurements

Growth rate of *F. graminearum* isolates is affected by both temperature and isolate genotype. Temperature and isolate effects were highly significant (P-value < 0.0001) in an analysis of variance (ANOVA). Isolates grew faster at 25°C compared to 15°C and 30°C. On average, after 4 days the isolates grew 2.15 cm, 6.24 cm, and 3.28 cm at 15°C, 25°C, and 30°C, respectively. There is no significant difference in growth rate between isolates of the NA1 and NA2 populations. We used the ls means adjusted for the three temperatures as the input phenotype for GWAS as the temperature by isolate interaction was not significant in the analysis of variance. The average, minimum and maximum adjusted growth rate across three temperatures is 3.89, 1.92 and 4.77 centimeters (Figure 3.1).

The distribution of total ascospores produced across 152 isolates did not follow a normal distribution. So, the total ascospores produced was natural log transformed. The distribution of the log transformed ascospore count was somewhat normal. The log of ascospore count ranged from 0.854 to 8.27 and average log count is 6.50 (Figure 3.2).

Our isolates displayed variation in their sensitivity to propiconazole and tebuconazole. Yet none of our isolates were resistant to propiconazole and tebuconazole; growth of all isolates tested was suppressed for both fungicides at the four highest concentrations. Some isolates were relatively more sensitive, and some were relatively more tolerant. The EC50 values for the two fungicides follow a normal distribution with a long right tail. The EC50 values ranged from 0.0086 mM to 0.1 mM for propiconazole and 0.004 mM to 0.049 mM for propiconazole (Figure 3.3, Figure 3.4). Isolate 23310 was the most tolerant and isolate 24315 was most sensitive to both fungicides. Tebuconazole is more effective in suppressing fungal growth as the average EC50 was lower for tebuconazole (0.01733 mM) than for propiconazole (0.0349 mM).

A total of 151 isolates were grown *in vitro* to assess the production of mycotoxins, DON, 15ADON, 3ADON and NIV. All isolates but three (23310, 24309, 23511) produced detectable amount of DON and 15ADON in one or more of the three replicates. NIV was not detected from any isolates and 3ADON was detected in one or more of the three replicates for 126 isolates. Only 40 out of 151 isolates produced detectable 3ADON in a second replicate. 3ADON was not considered for GWAS, as most of our isolates were 15ADON producers and not all the isolates showing any 3ADON production produced 3ADON in all three replicates. The results from three replicates were used to obtain ls means for DON and 15ADON using a completely randomized design where isolate was treated as fixed effect and replicate was treated as random effect. The amount of DON produced by different isolates ranged between 0.7 ppm by isolate 19137 to 340 ppm produced by isolate 23356. The average amount of DON produced was 15.33 ppm (Figure 3.5). Similarly, the amount of 15ADON produced by isolates ranged between 0.167 ppm by isolate 23365 to 174.5 ppm by isolate 23448 (Figure 3.6). The isolates on average produced 7.52

ppm 15ADON. Isolate 23344, a 3ADON isolate by genotype produced the largest amount of 3ADON (17 ppm averaged across three replicates).

The count of total and symptomatic spikelets on an inoculated wheat head was used to obtain the percentage of symptomatic spikelets. This was used as a measure of aggressiveness. There was no isolate by cultivar interaction. Thus, data from two growing seasons for two cultivars was used to obtain overall ls means. The distribution of aggressiveness followed a normal distribution (Figure 3.7). The aggressiveness adjusted across the two cultivars and two seasons ranged from 13.92 to 88.92 with an average of 53.71 (Figure 3.7).

The DON content from greenhouse inoculated grains is only available from the first growing season (2021-2022). 3ADON and NIV were below the threshold of detection (0.2 ppm) in all Rollag samples. Isolates produced varying amounts of DON. The amount of DON produced ranged from 0.2 ppm by isolate 23415 to 110.56 ppm by isolate 24185. The average amount of DON produced in Rollag was 18.23 ppm (Figure 3.8). 15ADON was detected in 101 out of 133 samples and ranged between 0.2 ppm and 3.48 ppm by isolate 24185. The average amount of 15ADON produced was 0.36 ppm (Figure 3.9). The average amount of DON and 15ADON produced in Wheaton samples was higher compared to Rollag. The average amount of DON produced is 118.79 ppm and average 15ADON was 3.3 ppm (Figure 3.10 and 11). DON was detected in all Wheaton samples and 15ADON was detected in all but three (23344, 23310 and 11157 from batch 8). NIV was below the detection threshold in all Wheaton samples. DON production in Wheaton samples ranged from 1.52 ppm in 23415 to 369 ppm in 23355. Similarly, 15ADON production ranged from 0.12 ppm in 23415 to 14.60 ppm in 19105.

3.4.2 Correlation between traits

We calculated Pearson correlation coefficients between all the traits we measured. The ls means or natural log of ls means were used for most traits. We did not have replicated measures for DON and 15ADON produced on grain, so the raw measure of DON and 15ADON from the first greenhouse season was used for these traits. Propiconazole and tebuconazole sensitivities were highly positively correlated with a correlation coefficient of (0.896) (Table 3.1). We also obtained a strong positive correlation (0.665) between DON and 15ADON production in Rollag. We obtained a positive but moderate correlation between ascospore produced in carrot agar and aggressiveness in the greenhouse, as well as between 15ADON in Wheaton and 15ADON in Rollag. Similarly, we obtained low positive correlations between aggressiveness measured in the greenhouse and mycotoxins (DON and 15ADON) produced in grains in both cultivars and aggressiveness and 15ADON measured *in vitro* (Table 3.1). Likewise, low correlation was also observed between fungicide sensitivity (propiconazole and tebuconazole) and growth. In addition, we obtained low positive correlations between ascospore production and 15ADON *in vitro*, and between ascospore production and DON in Rollag (Table 3.1). Growth measured in the laboratory and aggressiveness in wheat are moderately negatively correlated (Table 3.1). We observed a trend of negative correlation between mycotoxin production *in vitro* or *in planta* and fungicide sensitivity, although the correlation is very low.

3.4.3 GWAS scan approach using different models

Several GWAS models, including multi-locus models, are available and have been used here to screen for significant associations (Tamba et al., 2017; Tamba and Zhang, 2018; Wang and Zhang, 2021). These models are based on different genetic or statistical assumptions, and thus not all QTNs can be identified by all models. The best practice is to employ multiple such

models and compare their results in screens for candidate QTNs and genes. In the following, for each trait we report all distinct QTNs exceeding a significance threshold across the set of GWAS models, and highlight QTNs that are identified by multiple models since presumably these are more likely to be robust, true positives. However, we cannot take the significance testing results of each GWAS model at face value when assessing support for candidate associations, since applying multiple GWAS models introduces its own multiple testing problem (beyond the standard multiple testing problem for each model of thousands of SNPs each tested for association, which is dealt with using false discovery rate (FDR) adjustments). The rate of false positives may be higher than the p-values of individual GWAS model tests may indicate, but our aim is to identify multiple candidate QTNs for each trait that can be pursued using follow-up functional studies. For all of the sample-trait combinations, five models (GLM, MLM, MLMM, BLINK and FarmCPU) in GAPIT3 and five models (mrMLM, FASTmrMLM, FASTmrEMMA, ISIS EM-BLASSO, pKWmEB) in mrMLM were used for GWAS analysis. For each model used, only significant SNPs are reported. The GLM and MLM models from GAPIT3 did not identify any significant associations and are not discussed further. For all traits studied, any model implemented that did not produce any significant associations is not mentioned in the results for that trait.

3.4.4 GWAS for growth rate

We conducted three separate GWAS, using genotype and phenotype for: 1) a larger set composed of 135 isolates from the NA1 and NA2 populations as well as some admixed ones; 2) NA1 isolates only (n=55); and 3) NA2 isolates only (n=29). Kinship matrices were used to account for the relatedness among individuals. The Q-Q plot showed inflation of P-values in mrMLM models when values from the first principal component were used to adjust for

population structure. So, the results presented are adjusted for relatedness using only kinship matrices and not adjusted for population structure. We identified three [mrMLM: 3, FASTmrMLM: 1, FASTmrEMMA: 1, ISIS EM-BLASSO: 2, pKWmEB: 0] significant QTNs in GWAS conducted using all isolates, 5 [mrMLM: 3, FASTmrMLM: 0, FASTmrEMMA: 0, ISIS EM-BLASSO: 4, pKWmEB: 0] QTNs using only NA1 isolates and 2 [mrMLM: 0, FASTmrMLM: 2, FASTmrEMMA: 1, ISIS EM-BLASSO: 1, pKWmEB: 1] QTNs using only NA2 isolates (Figures 3.12-3.14; Tables 3.2-3.4). GAPIT models detected no significant SNPs for the set of 135 and NA1 isolates. In the GWAS with all isolates, one QTN (7892085 on chromosome 1) was detected by four methods, mrMLM, FASTmrMLM, FASTmrEMMA, and ISIS EM-BLASSO and another QTN (9701935 on chromosome 1) was detected by mrMLM and ISIS EM-BLASSO. The QTNs explained 3.15% to 15.0% of the total phenotypic variation (Table 3.2). There are no common QTNs detected across analyses of different samples. For the set of NA2 isolates, BLINK and FarmCPU from the R package GAPIT detected 4 common QTNs. One of these QTNs, 2946771 on chromosome 4, was also detected by four different models implemented in the R package mrMLM (Figure 3.15, Table 3.5). In addition, the two GAPIT models also detected 3 other SNPs: 400745 on chromosome 2, and 85478 and 7509544 on chromosome 3 for the set of NA2 isolates (Figure 3.15, Table 3.5).

3.4.5 GWAS for sexual fertility

As for growth rate, three separate GWAS were conducted to include a larger set of 152 NA1, NA2 and admixed isolates, NA1 isolates only (n=58), and NA2 isolates only (n=49). The kinship matrix was enough to account for the cryptic relatedness based on model selection implemented in GAPIT3, and we did not adjust for the population structure. GWAS performed using models implemented in the R package mrMLM with all 152 isolates identified 9 [mrMLM:

5, FASTmrMLM: 6, FASTmrEMMA: 2, ISIS EM-BLASSO: 3, pKWmEB: 4] significant QTNs (Figure 3.16, Table 3.6). Three SNPs (54034 on chromosome 1, 160544 on chromosome 2, and 399628 on chromosome 4) were detected by four methods and two SNPs were detected by two methods. Four SNPs were uniquely detected by only a single model. The QTNs explained from 1.35 to 15.4% of the total phenotypic variation in different models of the mrMLM package (Table 3.6). Three SNPs significant in mrMLM models, 540304 and 10661762 on chromosome 1, and 160544 on chromosome 2, were also significant ion the GAPIT model BLINK (Table 3.9) and explained up to 7.5%, 6.38 and 15.4% of the phenotypic variation, respectively (mrMLM models). SNPs 160544 on chromosome 2 and 10661762 on chromosome 1 were also significant in the GAPIT model FarmCPU (Figure 3.19). In addition, FarmCPU also detected 4489803 on chromosome 2, 3453361 and 7239571 on chromosome 3, and 7290074 and 7700303 on chromosome 4 (Table 3.9). All SNPs detected by BLINK, except 540304 on chromosome 1, and all SNPs detected by FarmCPU, except 10661762 on chromosome 1, have -log10(p-values) above the Bonferroni threshold and are highly significant (Figure 3.19).

Similarly, GWAS conducted using NA1 isolates detected 11 [mrMLM: 3, FASTmrMLM: 4, FASTmrEMMA: 3, ISIS EM-BLASSO: 6, pKWmEB: 0] significant QTNs (Figure 3.17, Table 3.7). Three QTNs were detected by more than one method while eight were detected by only a single model implemented in the mrMLM package, and they explained 4.2% to 22.09% of the phenotypic variation (Table 3.7). One of these 11 significant SNPs, 540228 on chromosome 1, was also detected by the model BLINK. SNP 4489803 on chromosome 2 was detected by the models BLINK and FarmCPU in GAPIT (Figure 3.20) as well as the model mrMLM. The FarmCPU model in GAPIT3 also detected 3 SNPs not detected in mrMLM models: 6679636 and 7079729 on chromosome 3 and 7596245 on chromosome 4 (Table 3.10).

All the SNPs detected by BLINK and FarmCPU are highly significant, with –log10(p-values) above the Bonferroni threshold. There is one common SNP (399628 on chromosome 4) significant for both NA1 only isolates and all 152 isolates. Likewise, association mapping with NA2 isolates only gave 5 [mrMLM: 2, FASTmrMLM: 0, FASTmrEMMA: 0, ISIS EM-BLASSO: 1, pKWmEB: 2] significant QTNs, each of which was detected by only a single model implemented in the package mrMLM and explained 9.3% to 25.5% of the total phenotypic variation (Figure 3.18, Table 3.8). Models implemented in GAPIT3 detected no significant associations. There are no common QTNs for the set of NA2 only isolates and all 152 isolates as well as between the sets of only NA1 and NA2 isolates.

3.4.6 GWAS for fungicide sensitivity

GWAS analysis for the traits propiconazole and tebuconazole sensitivity was performed by using EC50 values estimated for 152 isolates from the NA1 population and 4,857 SNPs distributed across the genome using models implemented in two software packages, GAPIT3 and mrMLM. A kinship matrix was used to account for the relatedness among the individuals and no measure was taken to adjust for population structure as no population structure was detected. Models implemented in GAPIT3 detected no significant associations for either fungicide, while models in the mrMLM package identified 48 [mrMLM: 2, FASTmrMLM: 1, FASTmrEMMA: 0, ISIS EM-BLASSO: 30, pKWmEB: 22] significant SNPs for propiconazole sensitivity with LOD scores ≥ 3 (Figure 3.21, Table 3.11). Six SNPs were detected by more than one method. Even though many (48) QTNs were detected, 38 of them explain less than 1% of the total phenotypic variation in all models that identified them as significant (Table 3.11). QTN 6640556 on chromosome 3 explained the highest percent (10%) of the total phenotypic variation (Table 3.11). Similarly, models implemented in the package mrMLM detected 37 [mrMLM: 0,

FASTmrMLM: 0, FASTmrEMMA: 0, ISIS EM-BLASSO: 22, pKWmEB: 17] significant SNPs for the fungicide tebuconazole (Figure 3.22, Table 3.12). Two of these SNPs were detected by both pkWmEB and ISIS EM-BLASSO. There are 12 common SNPs between propiconazole and tebuconazole (Tables 3.11 and 3.12). Similar to propiconazole, 24 of the QTNs explained less than 1% of the total variation in all models that identified them as significant, and the highest phenotypic variation (3.06%) was explained by QTN 8322036 on chromosome 2.

3.4.7 GWAS for mycotoxin production in rice cultures

Genome wide association (GWAS) performed using models implemented in GAPIT3 did not identify any significant associations for the traits DON and 15ADON production. Models implemented in the R package mrMLM detected 8 [mrMLM: 1, FASTmrMLM: 2, FASTmrEMMA: 0, ISIS EM-BLASSO: 5, pKWmEB: 4] and 3 [mrMLM: 1, FASTmrMLM: 1, FASTmrEMMA: 1, ISIS EM-BLASSO: 3, pKWmEB: 2] significant QTNs for DON and 15ADON, respectively (Figures 3.23 and 3.24; Tables 3.13 and 3.14). Multiple QTNs were detected by more than one model implemented in the package. For the trait DON production, four QTNs were detected by two of the implemented models. QTNs explained up to 10% of the phenotypic variation. Likewise, for the trait 15DON production, one QTN was detected by two different models and one detected by all five models implemented. QTNs explained up to 12% of the phenotypic variation for 15ADON production. There is one common QTN for DON and 15ADON production.

3.4.8 GWAS for aggressiveness

The count of total and symptomatic spikelets on an inoculated wheat head was used to obtain the percentage of symptomatic spikelets. This was used as the measure of

aggressiveness/disease severity. There was no isolate by cultivar interaction. Thus, data from two growing seasons for two cultivars was used to obtain overall ls means.

Genome wide association (GWAS) performed using models implemented in GAPIT3 did not identify any significant associations. Models implemented in the R package mrMLM detected two [mrMLM: 2, FASTmrMLM: 2, FASTmrEMMA: 2, ISIS EM-BLASSO: 2, pKWmEB: 0] significant QTNs, one each on chromosome one and two (Figure 3.25, Table 3.15). QTN 6675327 on chromosome 2 explained 19% and QTN 7371490 on chromosome 1 explained 13% of the total phenotypic variation for the model mrMLM.

3.5 Discussion

We observed significant among-isolate variation in trait measurements in the laboratory and greenhouse experiments. The measures of aggressiveness collected in the growing seasons over the two years on two varieties were highly variable. We obtained a poor correlation in the disease severity values collected during the first and second greenhouse seasons for both cultivars. This is likely due to the different temperature conditions in the greenhouse during the two growing seasons and because we divided our isolates into multiple batches within a single growing season which again were exposed to variable conditions. The broad sense heritability for the traits growth rate, sexual fertility, propiconazole sensitivity, tebuconazole sensitivity, DON production *in vitro* (rice cultures), and 15ADON production *in vitro* are 0.81, 0.85, 0.84, 0.8, 0.37, and 0.45, respectively. Talas and Mcdonald (2015) also reported high heritability of propiconazole sensitivity (0.97) in laboratory experiments conducted using a similar protocol as ours. A moderate heritability (0.67) for DON production was found in a field experiment conducted in Germany (Talas et al., 2016).

We identified a total of 13 significant QTNs for growth rate, a total of 30 for sexual fertility, 48 for propiconazole sensitivity, 37 for tebuconazole sensitivity, 8 for DON production *in vitro*, 3 for 15ADON production and 2 for aggressiveness in two wheat genotypes. We obtained the highest number of significant QTNs for propiconazole sensitivity, but 38 and 24 of significant QTNs explain less than one percent of the phenotypic variation observed, suggesting that azole sensitivity in *F. graminearum* is governed by many QTNs with small effect. Talas et al. (2016) reported 74 significant QTNs for propiconazole sensitivity in *F. graminearum*, where a single QTN explained up to 17% of the total variation. We identified three candidate genes (one for tebuconazole and two for propiconazole) common to the ones reported by Talas et al., (2016). FGRAMPH1_01G13811 (FGSG_16193, FGSG _3830) is a quinate transporter, gene FGRAMPH1_01G19355 (FGSG _06044) is a transcription initiation factor TFIID subunit 12 and FGRAMPH1_01G22503 (FGSG _12935) is a hypothetical protein (Supplemental file 3.2). As in Talas et al. (2016), we did not identify any genes in common with those significantly altered in expression upon treatment with tebuconazole by (Becher et al., 2011). One of our candidate genes, FGRAMPH1_01G09621 (FGSG _17058), is a multidrug resistant type ABC transporter and has been previously identified to specifically play a role in tolerance to SBI class I triazoles (tebuconazole, triticonazole and epoxiconazole) but not to the SBI type II triazoles. The gene was identified as a candidate gene in tebuconazole but not propiconazole sensitivity in our analyses. Our results further support that different ABC transporters might be involved in resistance/tolerance to different fungicides. In addition to the multidrug resistant protein FGRAMPH1_01G09621, we identified five other transport related proteins (FGRAMPH1_01G13811, FGRAMPH1_01G16125, FGRAMPH1_01G19393, FGRAMPH1_01G22013, and FGRAMPH1_01G25701) among our candidate genes. Transport

of fungicide out of the cells is known to be one of the molecular mechanisms for fungicide tolerance/resistance in fungi (Hayashi et al., 2003; Reimann and Deising, 2005), and *F. graminearum* too seems to have those mechanism in place for fungicide tolerance. We did not identify any resistant isolates in our study. Mutations in CYP51 genes have been related to azole resistance in many fungi, including *Mycosphaerella graminicola* (Leroux et al., 2007). However, we did not identify any significant association on or in close proximity to CYP51 genes in *F. graminearum*. This could be due to different reasons: 1. Field isolates of *F. graminearum* from our sample do not have mutations in CYP51 genes that lead to resistance. 2. Genes outside of CYP51 loci contribute the largest portion to variation to fungicide tolerance. Talas and Mcdonald (2015) considered the latter as one of the explanations for no observed correlation between mutations in CYP51 genes and propiconazole resistance in *F. graminearum*. Our results show that genes outside CYP51 loci are responsible for azole tolerance, although functional validation remains to be done. Surprisingly, genes in the fusarin C biosynthesis cluster were among the candidate genes for propiconazole and tebuconazole sensitivity. Sublethal doses of fungicides are known to induce production of secondary metabolites (D'Mello et al., 1998). We speculate that transporter gene of the fusarin C biosynthesis cluster might be associated with enhanced propiconazole and tebuconazole tolerance. *Penicillium expansum* isolates resistant to multiple fungicides produce a higher concentration of patulin and show enhanced expression of efflux transporters associated to patulin biosynthesis (Ntasiou et al., 2023). Similarly, infection of grapes by *Botrytis cinerea* induces production of plant secondary metabolites and ABC transporter overexpression, which in turn induces multi drug resistance in the pathogen. In addition to the fusarin C cluster genes, other secondary metabolite synthesis genes like nonribosomal peptide synthetases (FGRAMPH1_01G08375, FGRAMPH1_01G20955) and

polyketide synthase (FGRAMPH1_01G08383) are also among our candidate genes for azole sensitivity.

In our analysis, we identified seven transcription factors as candidate genes for the traits propiconazole and tebuconazole sensitivity, and this includes the TFIID transcription factor previously discussed (Supplemental file 3.2). Overexpression of CYP51 genes is another known mechanism for tebuconazole resistance in fungi (Hamamoto et al., 2000; Ma et al., 2006). We can speculate that overexpression of other genes like the transporters can lead to fungicide tolerance/resistance, and these transcription factors could be playing a role in the overexpression of genes regulating azole tolerance/resistance.

Heat shock proteins (HSPs) help fungi respond to heat and other types of environmental stressors. HSP90 and HSP70 are also responsible for antifungal resistance in *Candida albicans* and other fungi. HSP70 (FGRAMPH1_01G21231), a candidate gene for both propiconazole and tebuconazole sensitivity, is a novel candidate gene identified in this study. The role of HSP70 in azole resistance should be investigated further (Supplemental file 3.2).

To our knowledge, all the QTNs identified in our study for DON and 15ADON production are novel, as we did not identify any of the candidate genes in the vicinity of the QTNs that have clearly established roles in DON or 15ADON production (Supplemental file 3.2). However, QTN 3285338 on chromosome 3 for DON production has a helicase (FGRAMPH1_01G18675) in its LD block, and this gene is predicted to function as a member of histone H2A.Z, mutations in which lead to reduced DON production (Chen et al., 2020). Similarly, gene FGRAMPH1_01G13371, an MFS transporter gene in the LD block of QTN 5687139 on chromosome 2, is significantly upregulated during FHB infection and could be important in DON production and aggressiveness (Stephens et al., 2008). Other candidate genes

include nonribosomal peptide synthetase (FGRAMPH1_01G08375), NADPH dehydrogenase 2 (FGRAMPH1_01G09217), mate efflux family subfamily (FGRAMPH1_01G12015), MFS monosaccharide (FGRAMPH1_01G12019), transporter SEO1 (FGRAMPH1_01G13345), glucarate dehydratase (FGRAMPH1_01G13369), polyketide synthase (FGRAMPH1_01G14139), amidohydrolase 2 (FGRAMPH1_01G14141) and hypothetical proteins (Supplemental file 3.2). We identified no QTNs related to DON production common between our study and (Talas et al., 2016). This could be because of different phenotyping approaches used in the two studies. Our phenotypes were collected *in vitro* by inoculating rice cultures, while Talas et al. (2016) measured the DON levels from grains infected in the field in Germany by isolates from a population that is presumably different and evolving independently relative to the US populations included in our samples.

The production of some secondary metabolites, such as linoleic acid, zearalenone, melanin, and patulin, and sexual reproduction in fungi are related processes, as these secondary metabolites act as sporogenic factors in fungi (Calvo et al., 2002). For example, linolenic acid compounds induce sexual reproduction in *Aspergillus* (Calvo et al., 2001; Champe and El-Zayat, 1989). Genes from five different biosynthesis gene clusters, C2, C37, C38, C51 and C60, are among the candidate genes for ascospore production (Supplemental files 3.2 and 3.3). Secondary metabolite zearalenone induces sexual reproduction and promotes perithecia formation in *F. graminearum* (Wolf and Mirocha, 1973). Our results indicate that additional secondary metabolites might be involved in sexual reproduction and ascospore production and provide interesting candidates for functional studies. QTN 10661762 on chromosome 1 was significantly associated with ascospore production in the sample including isolates from the NA1 and NA2 populations. The association is detected by the model FASTmrMLM from package mrMLM and

by the model BLINK implemented in GAPIT3. The gene FGRAMPH1_01G07919, coding for peroxisomal-coenzyme A synthetase, falls in the LD block of this QTN. Peroxisomes are important for sexual development in fungi as they make the nutrients available to the developing sexual reproductive structures (Peraza-Reyes and Berteaux-Lecellier, 2013).

Among the QTNs significantly associated with growth rate in the full set of isolates, QTN 9701935 on chromosome 1 falls in the gene FGRAMPH1_01G07179, which is a transcription factor regulating growth and aggressiveness (Dufresne et al., 2008) (Supplemental files 3.2 and 3.3). Similarly, QTN 3635425 on chromosome 2 was associated with growth rate in NA2 isolates and falls in the secondary metabolite synthesis cluster C29 (Sieber et al., 2014). Although the role of the metabolite synthesis cluster in growth remains unknown, secondary metabolites are involved in a wide range of functions in fungi, and cluster C29 can be speculated to have some role in growth rate. QTN 5712361 on chromosome 2 associated with growth rate is near genes FGRAMPH1_01G13381 and FGRAMPH1_01G13389 which are uniquely expressed during saprophytic growth and could be important for growth (Boedi et al., 2016). Most of the QTNs we identified are novel and serve as important candidates for functional studies. Similar to most traits, most of the QTNs and the candidate genes we identified for aggressiveness are novel. Candidate genes included aldehyde dehydrogenase (FGRAMPH1_01G05485), glycerate-and formate-dehydrogenase (FGRAMPH1_01G14281), phosphatase (FGRAMPH1_01G14285) among others.

3.5.1 Overlap of regions between selection scans and GWAS

We used two different approaches to identify the genetic basis of some fitness traits in *F. graminearum*. The traits we are interested in are growth rate, sexual fertility, mycotoxin production, aggressiveness and fungicide sensitivity. The direct approach involved associating

phenotypic variation in *F. graminearum* isolates to the genotypic variation (single nucleotide polymorphisms) using genome wide association (GWAS) mapping. In the indirect population genomics approach, we scanned the genomes of isolates of three *F. graminearum* populations to identify signatures of recent positive natural selection with the assumption that the regions under the influence of natural positive selection likely provide a fitness advantage to the fungus.

Upon comparing candidate genes in the outlier regions from the selection scans to the candidate genes from GWAS analysis, we found some overlap between the candidate genes from the two approaches. Most of the isolates we used for GWAS are from the NA1 or NA2 populations. Analyses of growth rate and sexual fertility used a mixture of isolates from the NA1 and NA2 populations and some admixed isolates. Aggressiveness, DON production and fungicide sensitivity were measured exclusively on isolates from the NA1 population. A total of 43 candidate genes from cross population scans that belong to 23 outlier windows in population NA1 were common to candidate genes from GWAS performed using models implemented in the R package mrMLM (Supplemental file 3.4). Among the candidate genes for growth rate are *TRI15*- transcription factor (FGRAMPH1_01G05879) and Isopenicillin-N N-acyltransferases (FGRAMPH1_01G05883, FGRAMPH1_01G05885). Genes associated with ascospore production include n-acylsphingosine amidohydrolase (FGRAMPH1_01G18813) and genes coding for hypothetical proteins. The rest of the genes common between NA1 cross population scans and GWAS with mrMLM are associated with the various traits we studied (2 with DON production *in vitro*, one with ascospore production, 8 with growth rate, 14 with propiconazole sensitivity, 3 with tebuconazole sensitivity and 5 with aggressiveness). Some of the candidate genes are associated with more than one trait. For example, gene FGRAMPH1_01G16147 is associated with DON and 15ADON production *in vitro*. Likewise, genes

FGRAMPH1_01G19395, FGRAMPH1_01G19399, and FGRAMPH1_01G19401 are common candidates for propiconazole sensitivity and ascospore production (152 isolates) (Supplemental file 3.4). Gene FGRAMPH1_01G11623, an MFS transporter, is a candidate gene for growth rate and is predicted to have a role in filamentous growth. None of these candidates have been previously identified to play a role in our traits of interest through functional studies and provide novel candidates for functional studies. These genes serve as strong candidates as they were identified by multiple approaches.

We used some isolates from the NA2 population for GWAS for the traits growth rate and sexual fertility. We also observed some common candidates in NA2 cross population selection scans and GWAS on these traits. We have four genes associated with each of the two traits that are also candidate genes in selection scans in the NA2 population (Supplemental file 3.5). This includes genes FGRAMPH1_01G11619, FGRAMPH1_01G11621, FGRAMPH1_01G11623 and FGRAMPH1_01G11625 for growth rate (135 isolates) and genes FGRAMPH1_01G00143, FGRAMPH1_01G12291, FGRAMPH1_01G12295, and FGRAMPH1_01G12297 for ascospore production (Supplemental file 3.5). These four genes associated with growth were also identified in cross population selection scans for the NA1 population. Gene FGRAMPH1_01G11623 is predicted to help with filamentous growth as discussed earlier. Gene FGRAMPH1_01G00143, a non-ribosomal peptide synthetase (NRPS8) candidate gene for ascospore production (NA1) was previously identified to be an aggressiveness factor for maize infection by *F. graminearum* (Bahadoor et al., 2018). The gene falls in the biosynthesis gene cluster C02 (Sieber et al., 2014) that codes for secondary metabolites gramillin A and B (Bahadoor et al., 2018). The role of gramillin A and B in sexual reproduction has not been studied. Many secondary metabolites are essential for sexual reproduction and the role of gramillin in sexual reproduction cannot be

ignored. Even though NA2 isolates were not used for the traits aggressiveness and fungicide sensitivity, some of the candidate genes common between NA1 selection scans and GWAS for these traits were also identified to be common with NA2 selection scans. This includes genes FGRAMPH1_01G14273, FGRAMPH1_01G14275, FGRAMPH1_01G14277, and FGRAMPH1_01G14279 for aggressiveness and FGRAMPH1_01G22501 and FGRAMPH1_01G22503 for propiconazole sensitivity (Supplemental file 3.5). Likewise, there are also seven common candidates between SLA cross population selection scans and candidate genes for GWAS on growth rate (identified by both mrMLM and GAPIT3 models for GWAS performed using the set of 135 isolates). These candidate genes also fall in outlier regions in NA1 cross population selection scans. This provides evidence of natural selection acting on common genomic regions in different *F. graminearum* populations and indicates the potential role of variation in some of the same genes underlying variation in the same traits across populations.

3.6 References

Alexander, N.J., Proctor, R.H., McCormick, S.P., 2009. Genes, gene clusters, and biosynthesis of trichothecenes and fumonisins in *Fusarium*. Toxin Rev. 28, 198–215. https://doi.org/10.1080/15569540903092142

An, H., Landis, J.T., Bailey, A.G., Marron, J.S., Dittmer, D.P., 2019. dr4pl: A stable convergence algorithm for the 4 parameter logistic model. R J. 11, 171. https://doi.org/10.32614/rj-2019-003

Anderson, J.A., Wiersma, J.J., Linkert, G.L., Reynolds, S., Kolmer, J.A., Jin, Y., Dill-Macky, R., Hareland, G.A., 2015. Registration of 'Rollag' Spring Wheat. J. Plant Regist. 9, 201–207. https://doi.org/10.3198/jpr2014.07.0048crc

Atwell, S., Corwin, J.A., 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.

Bahadoor, A., Brauer, E.K., Bosnich, W., Schneiderman, D., Johnston, A., Aubin, Y., Blackwell, B., Melanson, J.E., Harris, L.J., 2018. Gramillin A and B: Cyclic Lipopeptides Identified as the Nonribosomal Biosynthetic Products of *Fusarium graminearum*. J. Am. Chem. Soc. 140, 16783–16791. https://doi.org/10.1021/jacs.8b10017

Bai, G., Shaner, G., 2004. Management and resistance in wheat and barley to Fusarium head blight. Annu. Rev. Phytopathol. 42, 135–161. https://doi.org/10.1146/annurev.phyto.42.040803.140340

Becher, R., Weihmann, F., Deising, H.B., Wirsel, S.G.R., 2011. Development of a novel multiplex DNA microarray for *Fusarium graminearum* and analysis of azole fungicide responses. BMC Genomics 12, 1–7. https://doi.org/10.1186/1471-2164-12-52

Boedi, S., Berger, H., Sieber, C., Münsterkötter, M., Maloku, I., Warth, B., Sulyok, M., Lemmens, M., Schuhmacher, R., Güldener, U., Strauss, J., 2016. Comparison of *Fusarium graminearum* transcriptomes on living or dead wheat differentiates substrate-responsive and defense-responsive genes. Front. Microbiol. 7, 1113. https://doi.org/10.3389/fmicb.2016.01113

Burlakoti, R.R., Ali, S., Secor, G.A., Neate, S.M., McMullen, M.P., Adhikari, T.B., 2008. Comparative mycotoxin profiles of *Gibberella zeae* populations from barley, wheat, potatoes, and sugar beets. Appl. Environ. Microbiol. 74, 6513–6520. https://doi.org/10.1128/AEM.01580-08

Burlakoti, R.R., Neate, S.M., Adhikari, T.B., Gyawali, S., Salas, B., Steffenson, B.J., Schwarz, P.B., 2011. Trichothecene profiling and population genetic analysis of *Gibberella zeae* from barley in North Dakota and Minnesota. Phytopathology 101, 687–695. https://doi.org/10.1094/PHYTO-04-10-0101

Calvo, A.M., Gardner, H.W., Keller, N.P., 2001. Genetic Connection between Fatty Acid Metabolism and Sporulation in *Aspergillus nidulans*. J. Biol. Chem. 276, 25766–25774. https://doi.org/10.1074/jbc.M100732200

Calvo, A.M., Wilson, R.A., Bok, J.W., Keller, N.P., 2002. Relationship between Secondary Metabolism and Fungal Development. Microbiol. Mol. Biol. Rev. 66, 447–459. https://doi.org/10.1128/mmbr.66.3.447-459.2002

Champe, S.P., El-Zayat, A.A.E., 1989. Isolation of a sexual sporulation hormone from *Aspergillus nidulans*. J. Bacteriol. 171, 3892–3988. https://doi.org/10.1128/jb.171.7.3982-3988.1989

Chen, Z., Zehraoui, E., Atanasoff-Kardjalieff, A.K., Strauss, J., Studt, L., Ponts, N., 2020. Effect of H2A.Z deletion is rescued by compensatory mutations in *Fusarium graminearum*. PLoS Genet. 16, e1009125. https://doi.org/10.1371/journal.pgen.1009125

D'Mello, J.P.F., Macdonald, A.M.C., Postel, D., Dijksma, W.T.P., Dujardin, A., Placinta, C.M., 1998. Pesticide use and mycotoxin production in *Fusarium* and *Aspergillus* phytopathogens. Eur. J. Plant Pathol. 104, 741–751. https://doi.org/10.1023/A:1008621505708

Dufresne, M., Lee, T. van der, M'Barek, S. Ben, Xu, X., Zhang, X., Liu, T., Waalwijk, C., Zhang, W., Kema, G.H.J., Daboussi, M.J., 2008. Transposon-tagging identifies novel

pathogenicity genes in *Fusarium graminearum*. Fungal Genet. Biol. 45, 1552–1561. https://doi.org/10.1016/j.fgb.2008.09.004

Gale, L.R., Bryant, J.D., Calvo, S., Giese, H., Katan, T., O'Donnell, K., Suga, H., Taga, M., Usgaard, T.R., Ward, T.J., Kistler, H.C., 2005. Chromosome complement of the fungal plant pathogen *Fusarium graminearum* based on genetic and physical mapping and cytological observations. Genetics 171, 985–1001. https://doi.org/10.1534/genetics.105.044842

Goswami, R.S., Kistler, H.C., 2004. Heading for disaster: *Fusarium graminearum* on cereal crops. Mol. Plant Pathol. 5, 515–525. https://doi.org/10.1111/J.1364-3703.2004.00252.X

Hamamoto, H., Hasegawa, K., Nakaune, R., Lee, Y.J., Makizumi, Y., Akutsu, K., Hibi, T., 2000. Tandem repeat of a transcriptional enhancer upstream of the sterol 14α-demethylase gene (CYP51) in *Penicillium digitatum*. Appl. Environ. Microbiol. 66, 3421–3426. https://doi.org/10.1128/AEM.66.8.3421-3426.2000

Hartmann, F.E., Sánchez-Vallet, A., McDonald, B.A., Croll, D., 2017. A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. ISME J. 11, 1189–1204. https://doi.org/10.1038/ismej.2016.196

Hayashi, K., Schoonbeek, H. jan, De Waard, M.A., 2003. Modulators of membrane drug transporters potentiate the activity of the DMI fungicide oxpoconazole against *Botrytis cinerea*. Pest Manag. Sci. 59, 294–302. https://doi.org/10.1002/ps.637

Huang, M., Liu, X., Zhou, Y., Summers, R.M., Zhang, Z., 2019. BLINK: A package for the next level of genome-wide association studies with both individuals and markers in the millions. Gigascience 8. https://doi.org/10.1093/gigascience/giy154

Kelly, A.C., Ward, T.J., 2018. Population genomics of *Fusarium graminearum* reveals signatures of divergent evolution within a major cereal pathogen. PLoS One 13, e0194616. https://doi.org/10.1371/journal.pone.0194616

Kitsios, G.D., Zintzaras, E., 2009. Genome-wide association studies: hypothesis-"free" or "engaged"? Transl. Res. 154, 161–164. https://doi.org/10.1016/j.trsl.2009.07.001

Klix, M.B., Verreet, J.A., Beyer, M., 2007. Comparison of the declining triazole sensitivity of *Gibberella zeae* and increased sensitivity achieved by advances in triazole fungicide development. Crop Prot. 26, 683–690. https://doi.org/10.1016/j.cropro.2006.06.006

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. lmerTest Package: Tests in Linear Mixed Effects Models. J. Stat. Softw. 82, 1–26. https://doi.org/10.18637/JSS.V082.I13

Lenth, R. V., 2016. Least-squares means: The R package lsmeans. J. Stat. Softw. 69, 1–33. https://doi.org/10.18637/jss.v069.i01

Leplat, J., Friberg, H., Abid, M., Steinberg, C., 2013. Survival of *Fusarium graminearum*, the causal agent of Fusarium head blight. A review. Agron. Sustain. Dev. 33, 97–111. https://doi.org/10.1007/s13593-012-0098-5

Leroux, P., Albertini, C., Gautier, A., Gredt, M., Walker, A.S., 2007. Mutations in the CYP51 gene correlated with changes in sensitivity to sterol 14α-demethylation inhibitors in field isolates of *Mycosphaerella graminicola*. Pest Manag. Sci. 63, 688–698. https://doi.org/10.1002/ps.1390

Leslie, J.F., Summerell, B.A., 2007. The Fusarium Laboratory Manual, The Fusarium Laboratory Manual. https://doi.org/10.1002/9780470278376

Li, X., Shin, S., Heinen, S., Dill-Macky, R., Berthiller, F., Nersesian, N., Clemente, T., McCormick, S., Muehlbauer, G.J., 2015. Transgenic wheat expressing a barley UDP-

glucosyltransferase detoxifies deoxynivalenol and provides high levels of resistance to *Fusarium graminearum*. Mol. Plant-Microbe Interact. 28, 1237–1246. https://doi.org/10.1094/MPMI-03-15-0062-R

Liu, X., Huang, M., Fan, B., Buckler, E.S., Zhang, Z., 2016. Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. PLoS Genet. 12. https://doi.org/10.1371/journal.pgen.1005767

Ma, Z., Proffer, T.J., Jacobs, J.L., Sundin, G.W., 2006. Overexpression of the 14α-demethylase target gene (CYP51) mediates fungicide resistance in *Blumeriella jaapii*. Appl. Environ. Microbiol. 72, 2581–2585. https://doi.org/10.1128/AEM.72.4.2581-2585.2006

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303. https://doi.org/10.1101/gr.107524.110

McMullen, M., Bergstrom, G., De Wolf, E., Dill-Macky, R., Hershman, D., Shaner, G., Van Sanford, D., 2012. A unified effort to fight an enemy of wheat and barley: Fusarium head blight. Plant Dis. 96, 1712–1728. https://doi.org/10.1094/PDIS-03-12-0291-FE

Miedaner, T., Vasquez, A., Castiblanco, V., Castillo, H.E., Foroud, N., Würschum, T., Leiser, W., 2021. Genome-wide association study for deoxynivalenol production and aggressiveness in wheat and rye head blight by resequencing 92 isolates of *Fusarium culmorum*. BMC Genomics 22, 1–16. https://doi.org/10.1186/s12864-021-07931-5

Ntasiou, P., Samaras, A., Papadakis, E.N., Menkissoglu-Spiroudi, U., Karaoglanidis, G.S., 2023. Aggressiveness and Patulin Production in *Penicillium expansum* Multidrug Resistant

Strains with Different Expression Levels of MFS and ABC Transporters, in the Presence or Absence of Fludioxonil. Plants 12, 1398. https://doi.org/10.3390/plants12061398

Peraza-Reyes, L., Berteaux-Lecellier, V., 2013. Peroxisomes and sexual development in fungi. Front. Physiol. 4, 244. https://doi.org/10.3389/fphys.2013.00244

Proctor, R.H., Hohn, T.M., McCormick, S.P., 1995. Reduced virulence of *Gibberella zeae* caused by disruption of a trichothecene toxin biosynthetic gene. Mol. Plant-Microbe Interact. 8, 593–601. https://doi.org/10.1094/MPMI-8-0593

Puri, K.D., Zhong, S., 2010. The 3ADON population of *Fusarium graminearum* found in North Dakota is more aggressive and produces a higher level of DON than the prevalent 15ADON population in spring wheat. Phytopathology 100, 1007–1014. https://doi.org/10.1094/PHYTO-12-09-0332

Reimann, S., Deising, H.B., 2005. Inhibition of efflux transporter-mediated fungicide resistance in *Pyrenophora tritici-repentis* by a derivative of 4′-hydroxyflavone and enhancement of fungicide activity. Appl. Environ. Microbiol. 71, 3269–3275. https://doi.org/10.1128/AEM.71.6.3269-3275.2005

Segura, V., Vilhjálmsson, B.J., Platt, A., Korte, A., Seren, Ü., Long, Q., Nordborg, M., 2012. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat. Genet. 44, 825–830. https://doi.org/10.1038/ng.2314

Sieber, C.M.K., Lee, W., Wong, P., Münsterkötter, M., Mewes, H.W., Schmeitzl, C., Varga, E., Berthiller, F., Adam, G., Güldener, U., 2014. The *Fusarium graminearum* genome reveals more secondary metabolite gene clusters and hints of horizontal gene transfer. PLoS One 9, e1103311. https://doi.org/10.1371/journal.pone.0110311

Spolti, P., Del Ponte, E.M., Cummings, J.A., Dong, Y., Bergstrom, G.C., 2014a. Fitness attributes of *Fusarium graminearum* isolates from wheat in New York possessing a 3-ADON or 15-ADON trichothecene genotype. Phytopathology 104, 513–519. https://doi.org/10.1094/PHYTO-07-13-0206-R

Spolti, P., Del Ponte, E.M., Dong, Y., Cummings, J.A., Bergstrom, G.C., 2014b. Triazole sensitivity in a contemporary population of *Fusarium graminearum* from New York wheat and competitiveness of a tebuconazole-resistant isolate. Plant Dis. 98, 607–613. https://doi.org/10.1094/PDIS-10-13-1051-RE

Stephens, A.E., Gardiner, D.M., White, R.G., Munn, A.L., Manners, J.M., 2008. Phases of infection and gene expression of *Fusarium graminearum* during crown rot disease of wheat. Mol. Plant-Microbe Interact. 21, 1571–1581. https://doi.org/10.1094/MPMI-21-12-1571

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. https://doi.org/10.1094/MPMI-09-15-0218-R

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. https://doi.org/10.1111/ppa.12280

Tamba, C.L., Ni, Y.L., Zhang, Y.M., 2017. Iterative sure independence screening EM-Bayesian LASSO algorithm for multi-locus genome-wide association studies. PLoS Comput. Biol. 13. https://doi.org/10.1371/journal.pcbi.1005357

Tamba, C.L., Zhang, Y.-M., 2018. A fast mrMLM algorithm for multi-locus genome-wide association studies. bioRxiv.

Trail, F., Xu, H., Loranger, R., Gadoury, D., 2002. Physiological and Environmental Aspects of Ascospore Discharge in *Gibberella zeae* (Anamorph *Fusarium graminearum*). Mycologia 94, 181. https://doi.org/10.2307/3761794

von der Ohe, C., Gauthier, V., Tamburic-Ilincic, L., Brule-Babel, A., Fernando, W.G.D., Clear, R., Ward, T.J., Miedaner, T., 2010. A comparison of aggressiveness and deoxynivalenol production between Canadian *Fusarium graminearum* isolates with 3-acetyl and 15-acetyldeoxynivalenol chemotypes in field-grown spring wheat. Eur. J. Plant Pathol. 127, 407–417. https://doi.org/10.1007/s10658-010-9607-z

Wang, J., Zhang, Z., 2021. GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction. Genomics, Proteomics Bioinforma. 19, 629–640. https://doi.org/10.1016/j.gpb.2021.08.005

Ward, T.J., Clear, R.M., Rooney, A.P., O'Donnell, K., Gaba, D., Patrick, S., Starkey, D.E., Gilbert, J., Geiser, D.M., Nowicki, T.W., 2008. An adaptive evolutionary shift in Fusarium head blight pathogen populations is driving the rapid spread of more toxigenic *Fusarium graminearum* in North America. Fungal Genet. Biol. 45, 473–484. https://doi.org/10.1016/j.fgb.2007.10.003

Wilson, W., Dahl, B., Nganje, W., 2018. Economic costs of Fusarium Head Blight, scab and deoxynivalenol. World Mycotoxin J. 11. https://doi.org/10.3920/WMJ2017.2204

Wolf, J.C., Mirocha, C.J., 1973. Regulation of sexual reproduction in *Gibberella zeae* (*Fusarium roxeum* "*graminearum*") by F-2 (Zearalenone). Can. J. Microbiol. 19, 725–734. https://doi.org/10.1139/m73-117

Yin, Y., Liu, X., Li, B., Ma, Z., 2009. Characterization of sterol demethylation inhibitor-resistant isolates of *Fusarium asiaticum* and *F. graminearum* collected from wheat in China. Phytopathology 99, 487–497. https://doi.org/10.1094/PHYTO-99-5-0487

Yue, W., 2017. Characterization of genetic variation in secondary metabolites in *Fusarium*. Ph.D. Dissertation. Kansas State University.

Zeller, K.A., Bowden, R.L., Leslie, J.F., 2004. Population differentiation and recombination in wheat scab populations of *Gibberella zeae* from the United States. Mol. Ecol. 13, 563–571. https://doi.org/10.1046/j.1365-294X.2004.02098.x

Zhang, Y.W., Tamba, C.L., Wen, Y.J., Li, P., Ren, W.L., Ni, Y.L., Gao, J., Zhang, Y.M., 2020. mrMLM v4.0.2: An R Platform for Multi-locus Genome-wide Association Studies. Genomics, Proteomics Bioinforma. 18, 481–487. https://doi.org/10.1016/j.gpb.2020.06.006

3.7 Figures

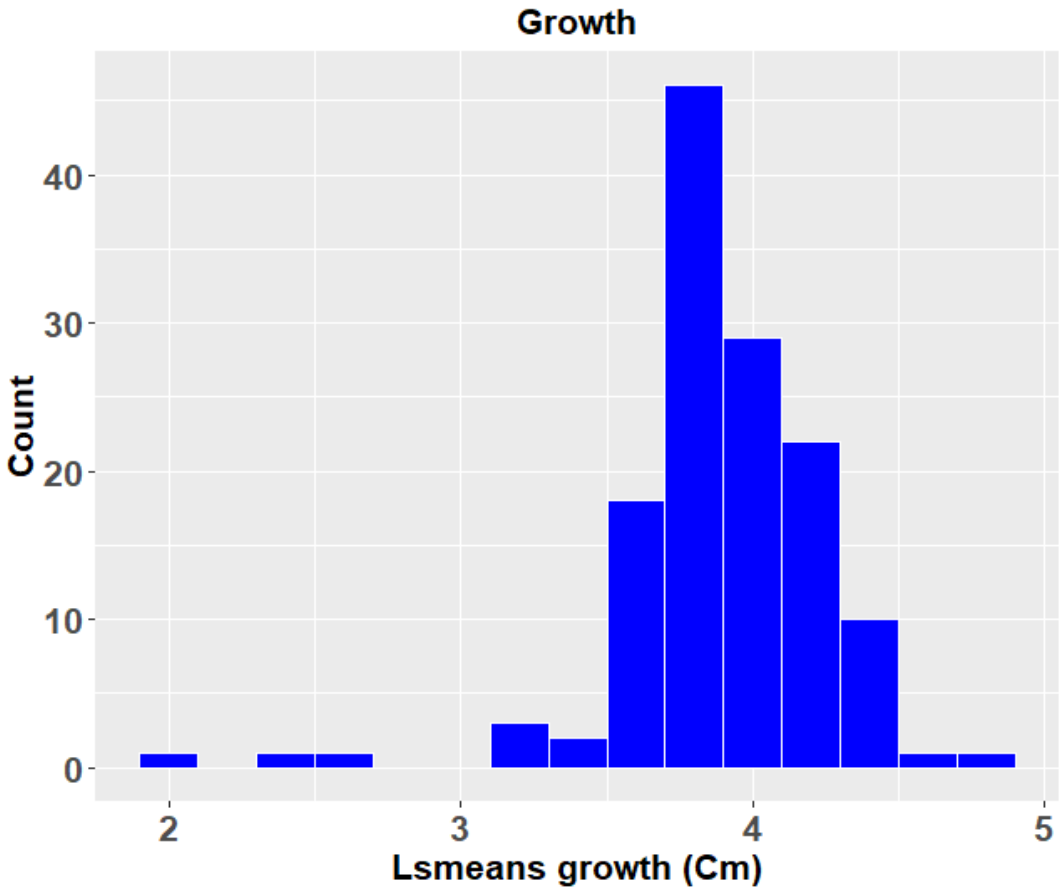


Figure 3.1. Distribution of ls means for the trait growth rate (cm) for the set of 135 isolates.

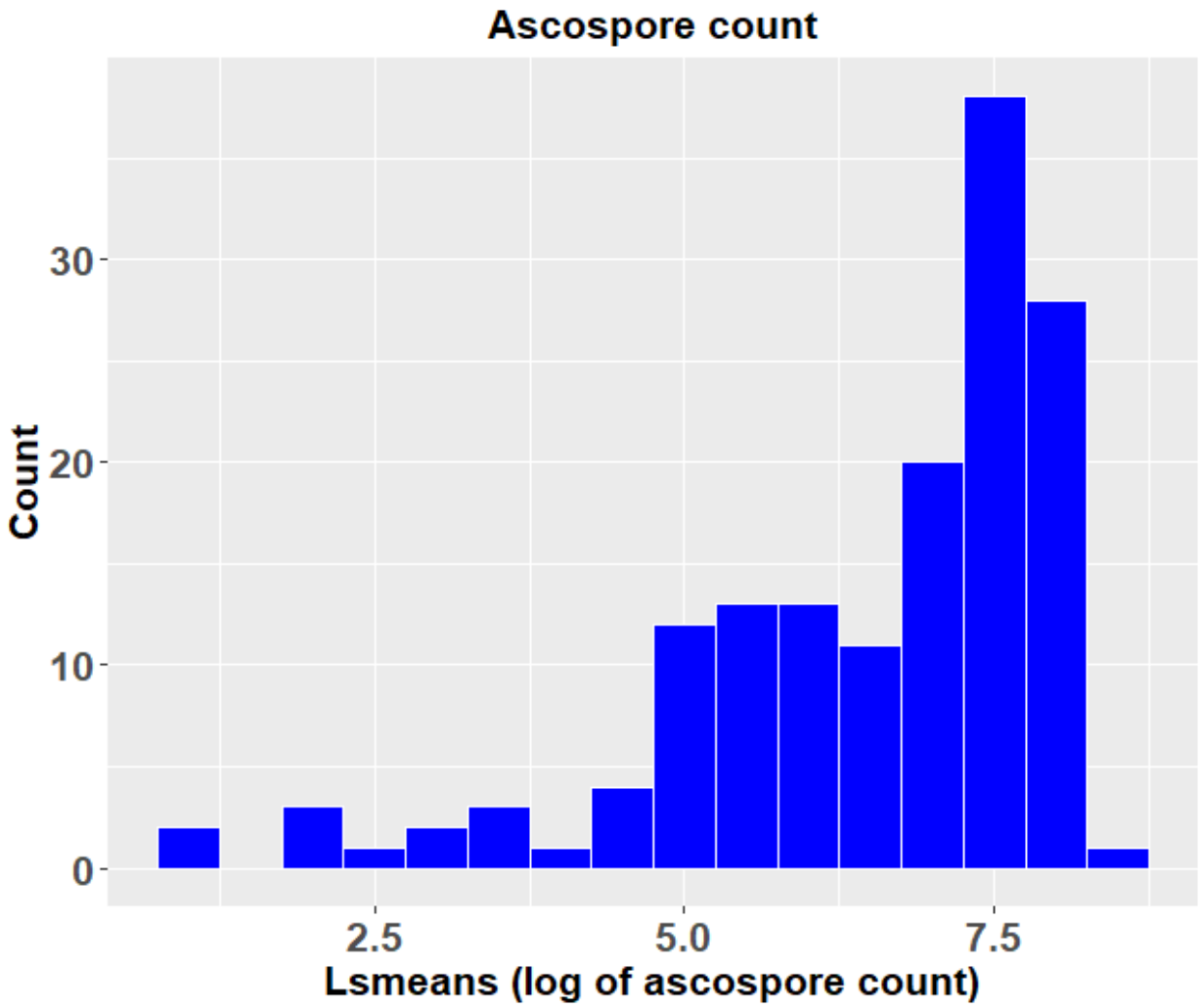


Figure 3.2. Distribution of ls means for the trait sexual fertility (ascospore production) for the set of 152 isolates.

Ls means was obtained using natural log transformed values of total ascospore count.

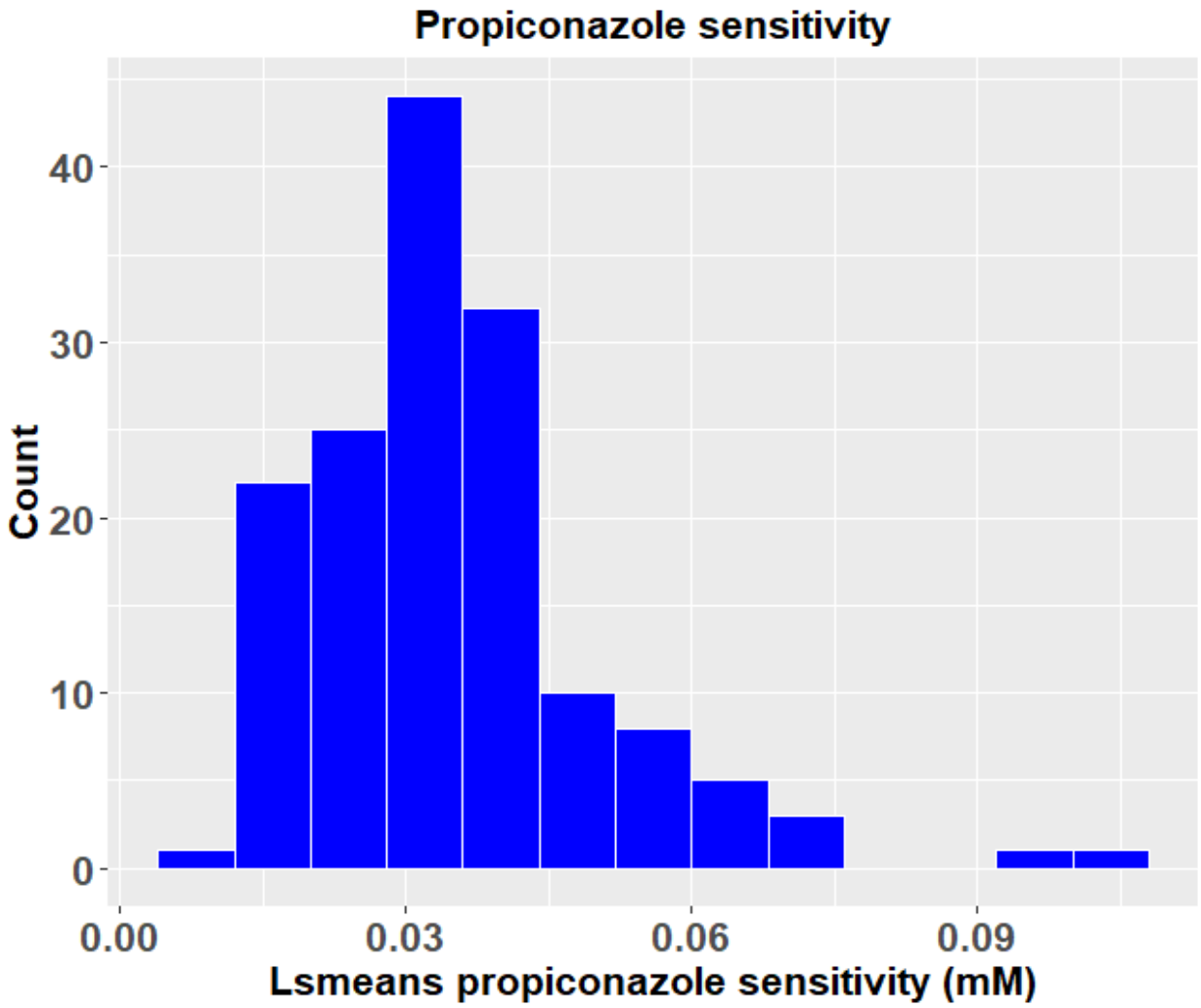


Figure 3.3. Distribution of ls means for the trait propiconazole sensitivity (mM).

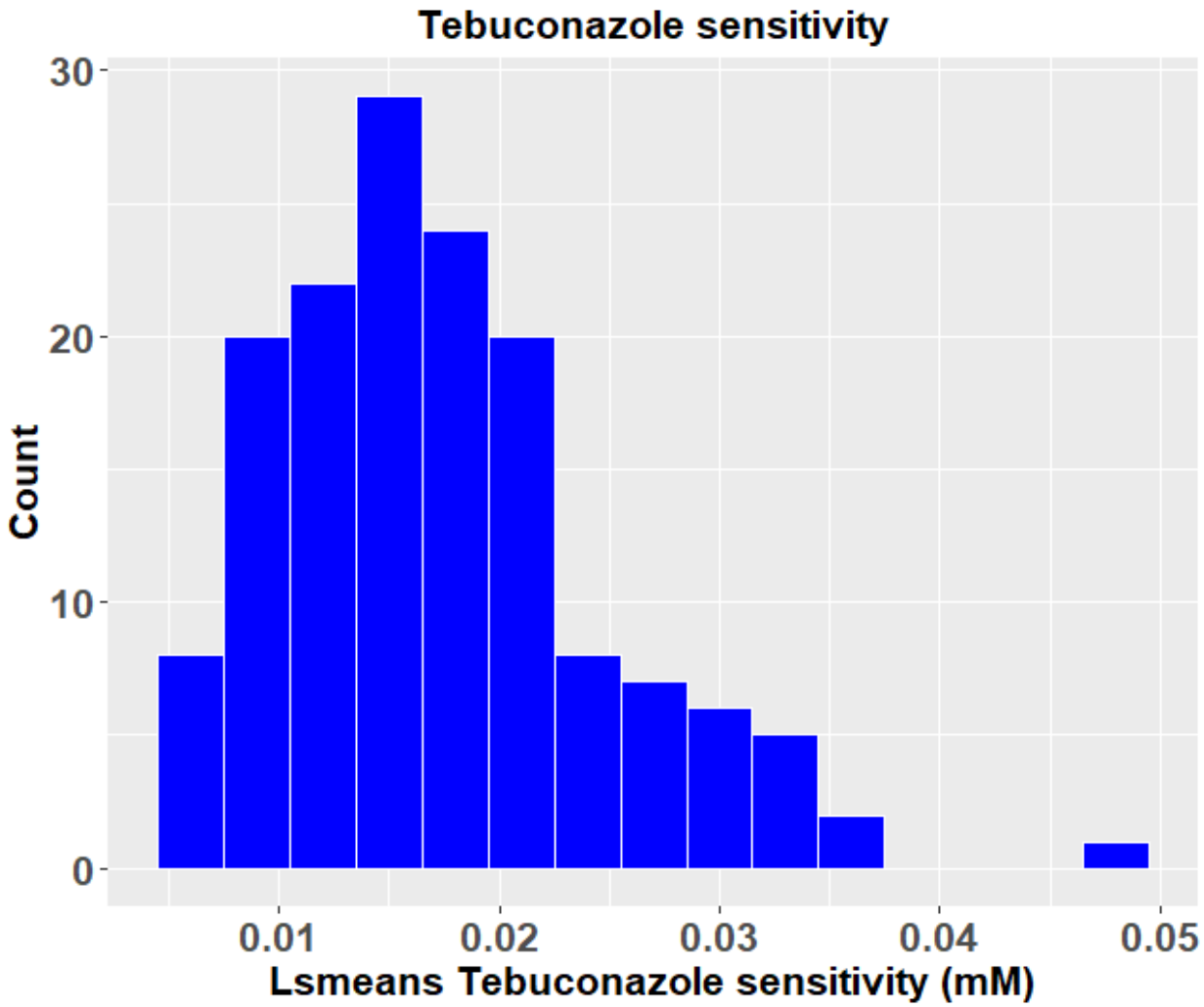


Figure 3.4. Distribution of ls means for the trait tebuconazole sensitivity (mM).

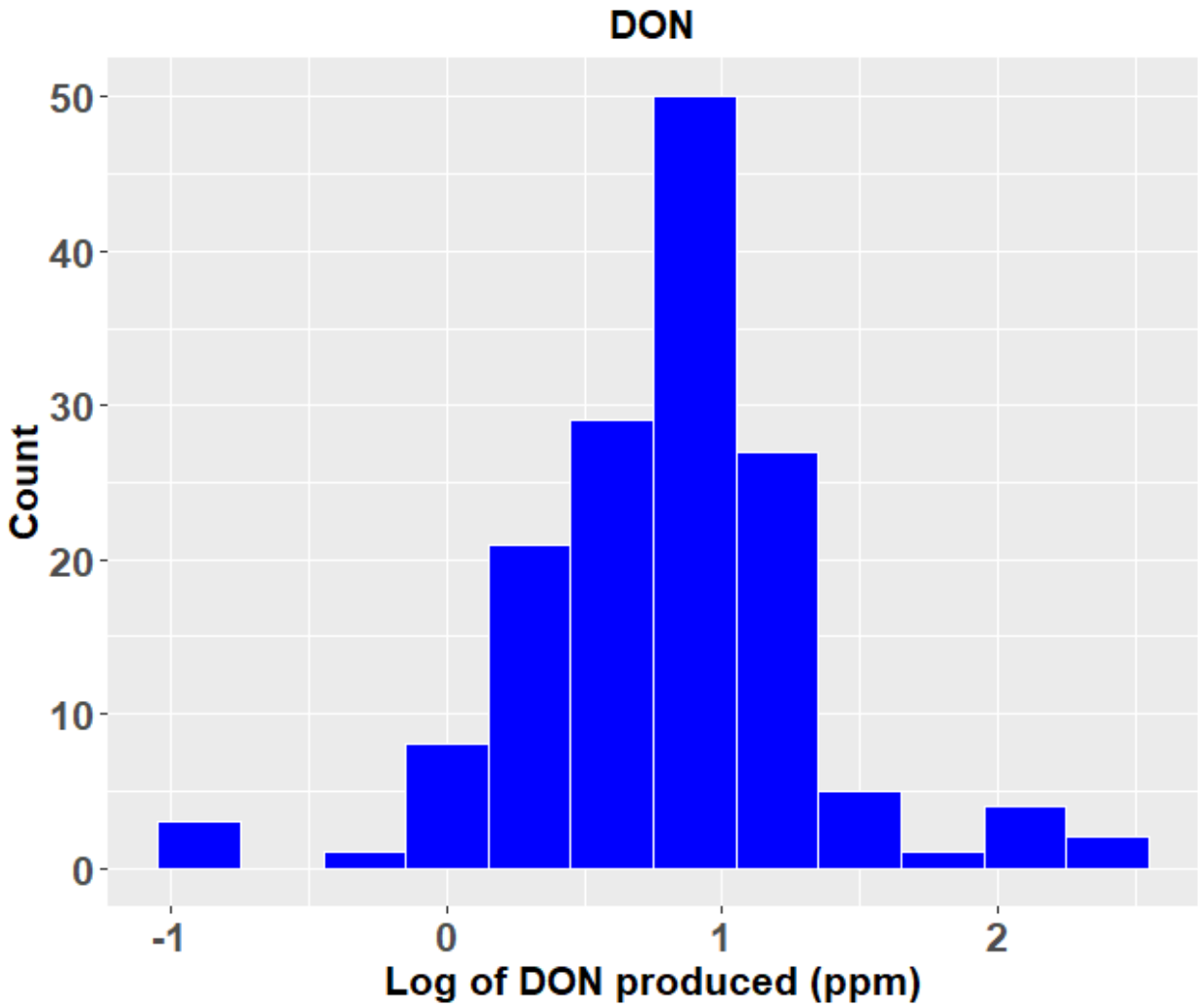


Figure 3.5. Distribution of ls means for the trait DON production *in vitro*.

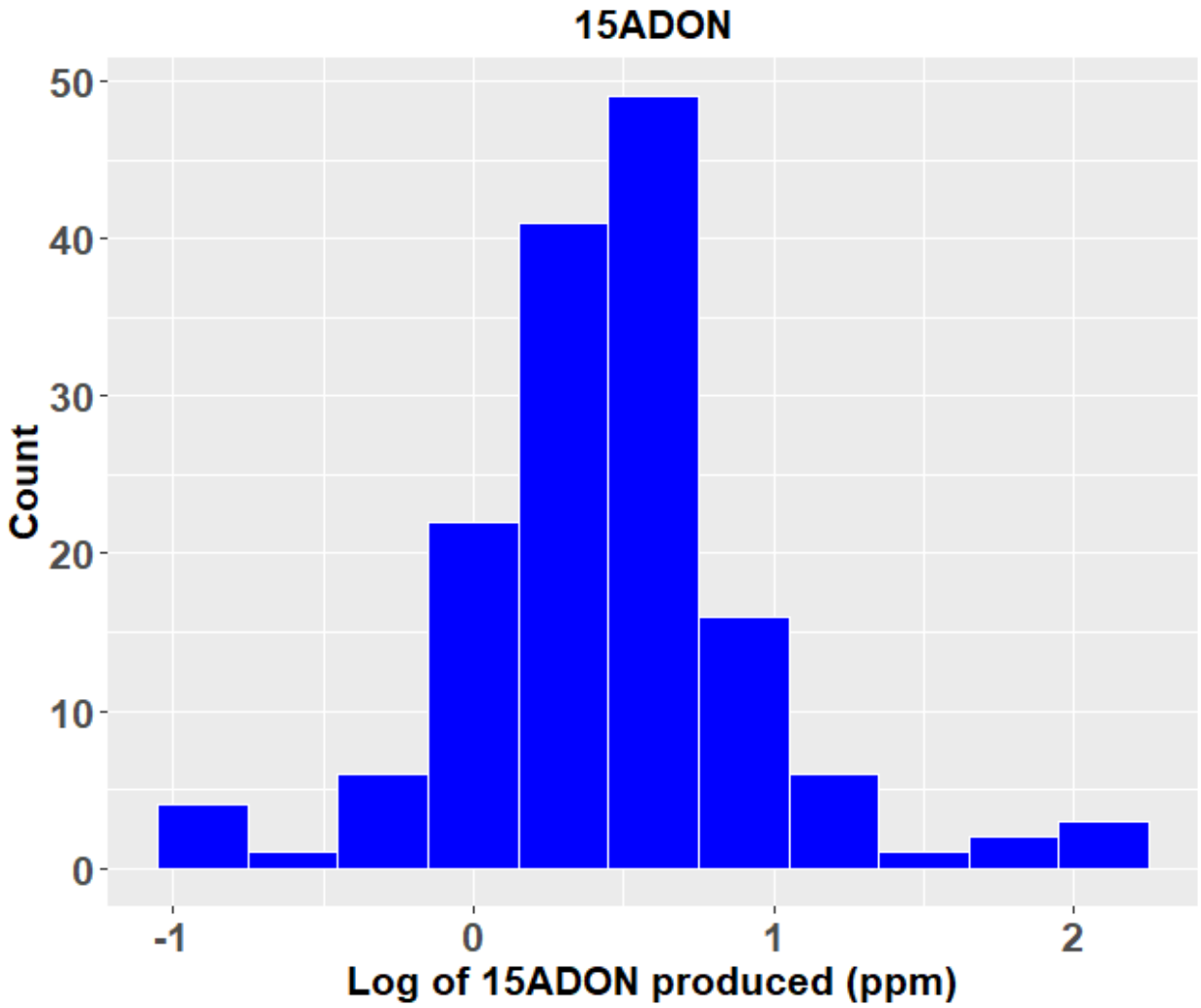


Figure 3.6. Distribution of ls means for the trait 15ADON production *in vitro*.

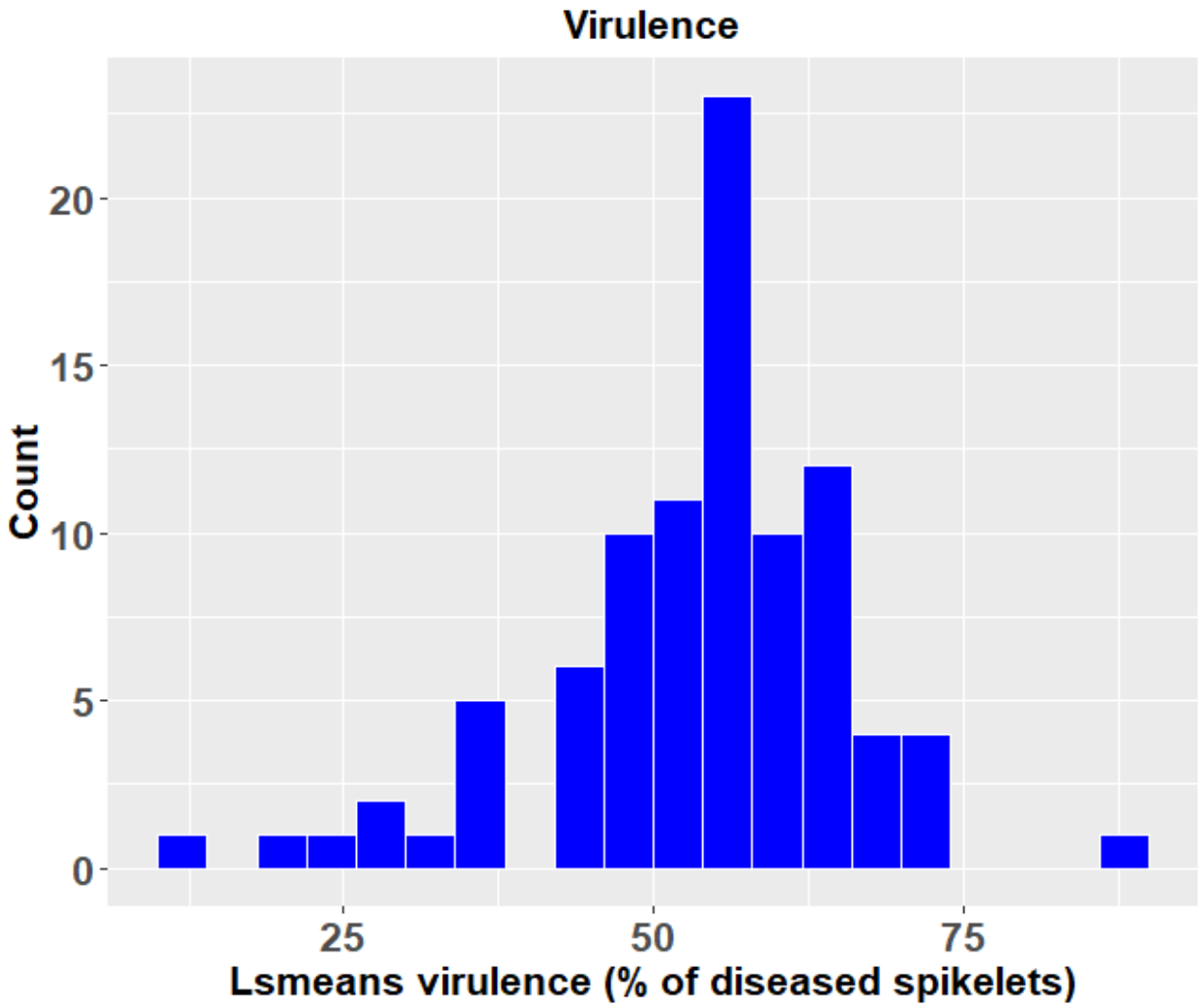


Figure 3.7. Distribution of ls means for the trait aggressiveness (percentage of symptomatic spikelets).

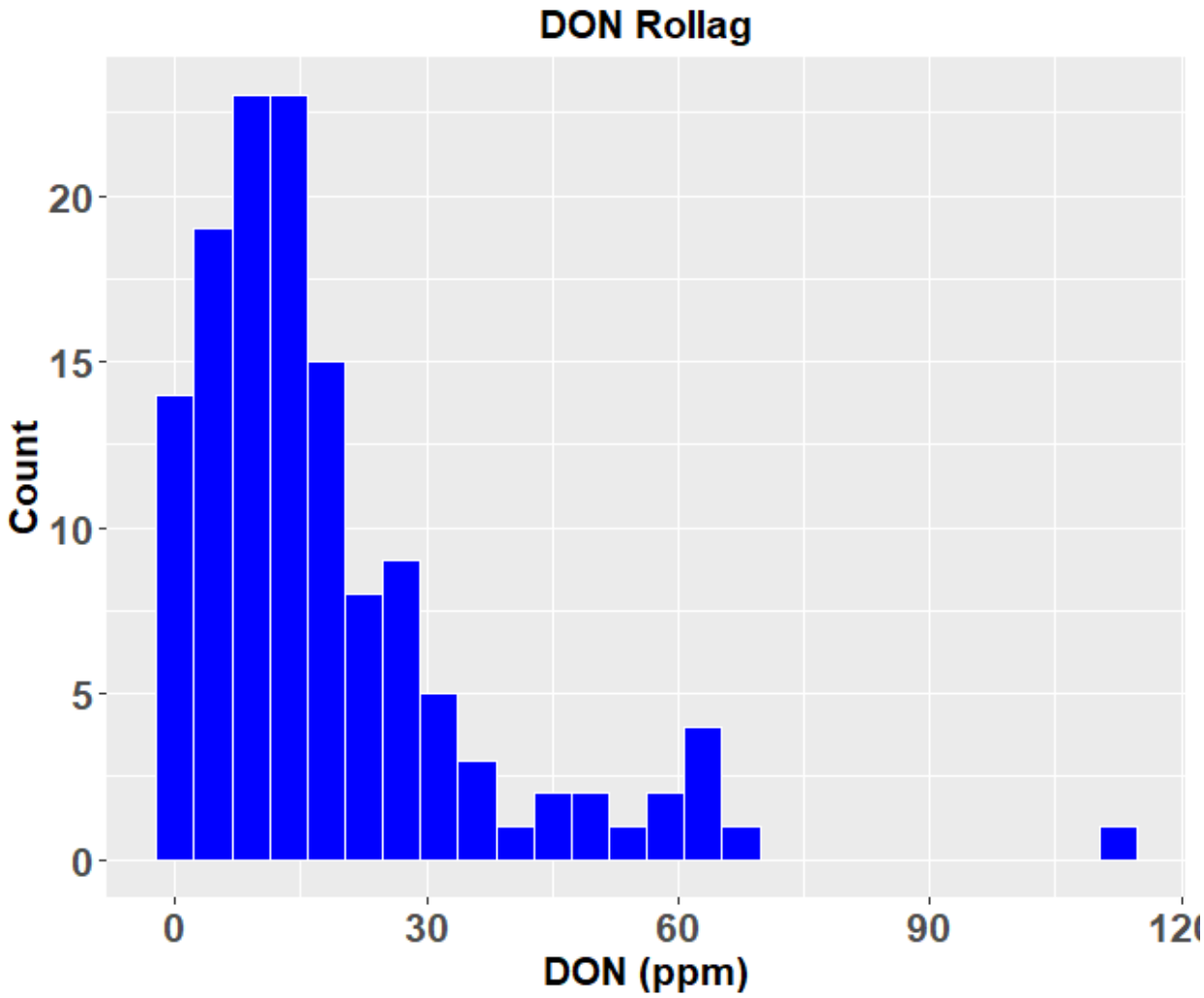


Figure 3.8. Distribution of DON produced by *Fusarium graminearum* in Rollag.

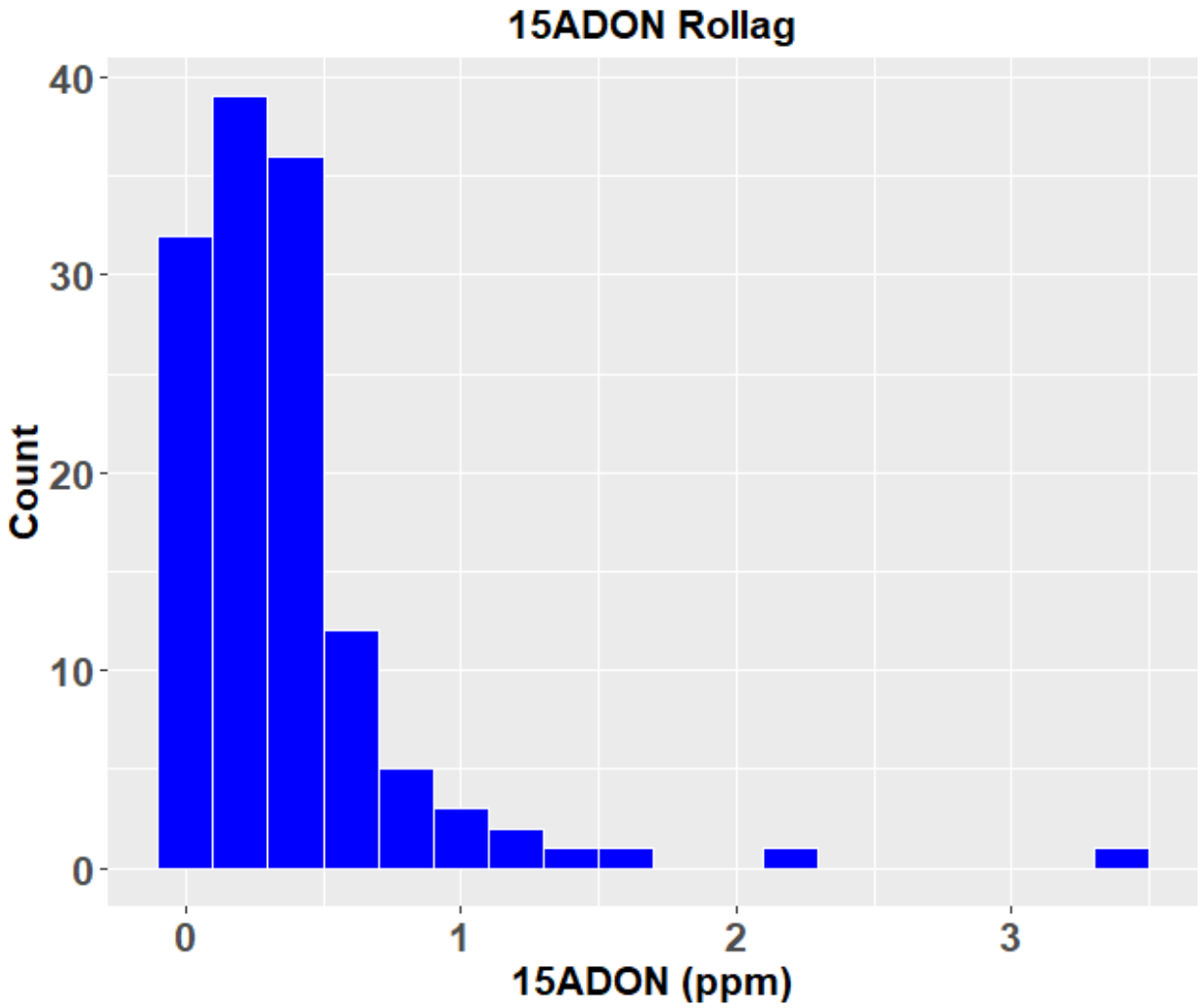


Figure 3.9. Distribution of 15ADON produced by *Fusarium graminearum* in Rollag.

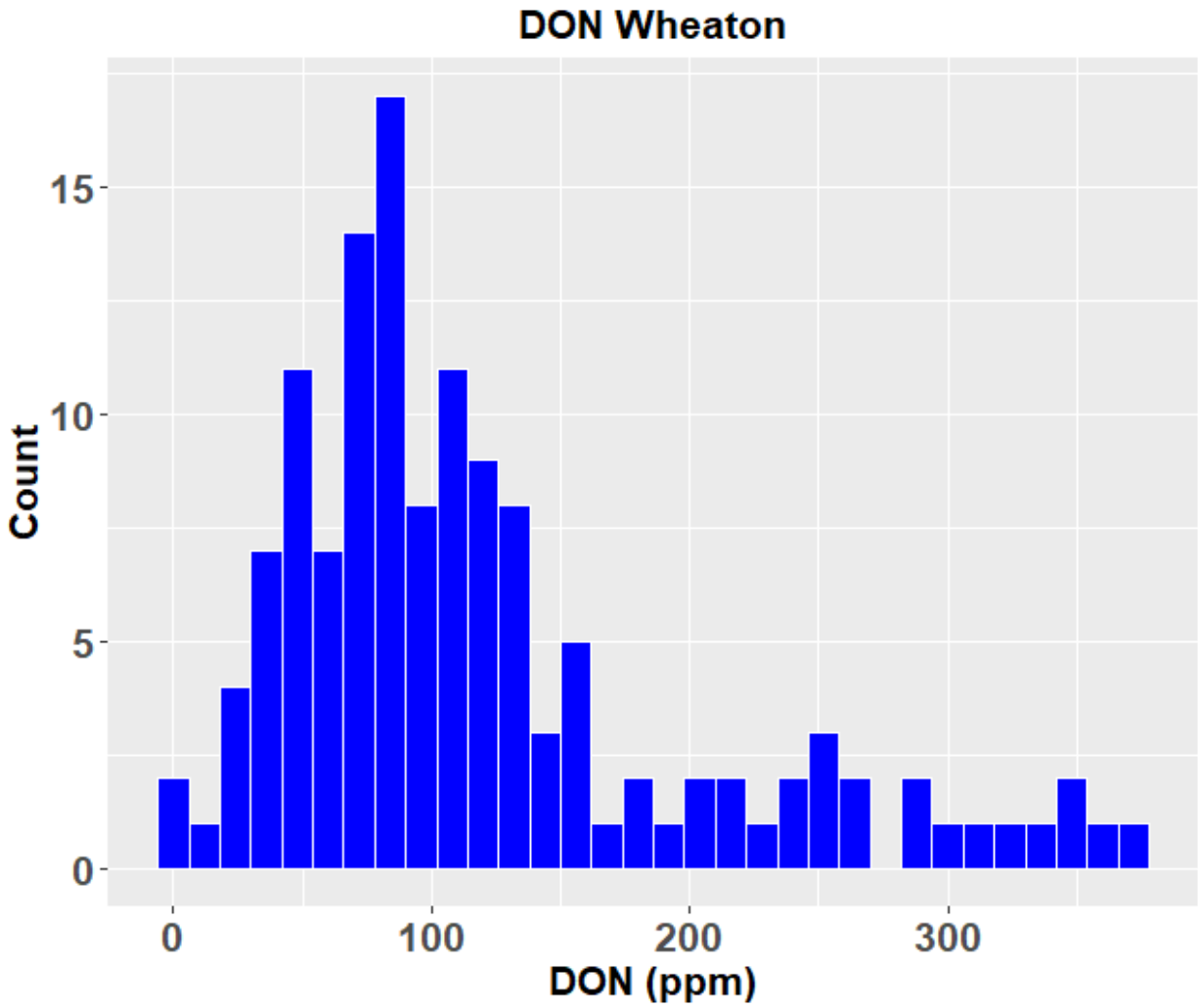


Figure 3.10. Distribution of DON produced by *Fusarium graminearum* in Wheaton.

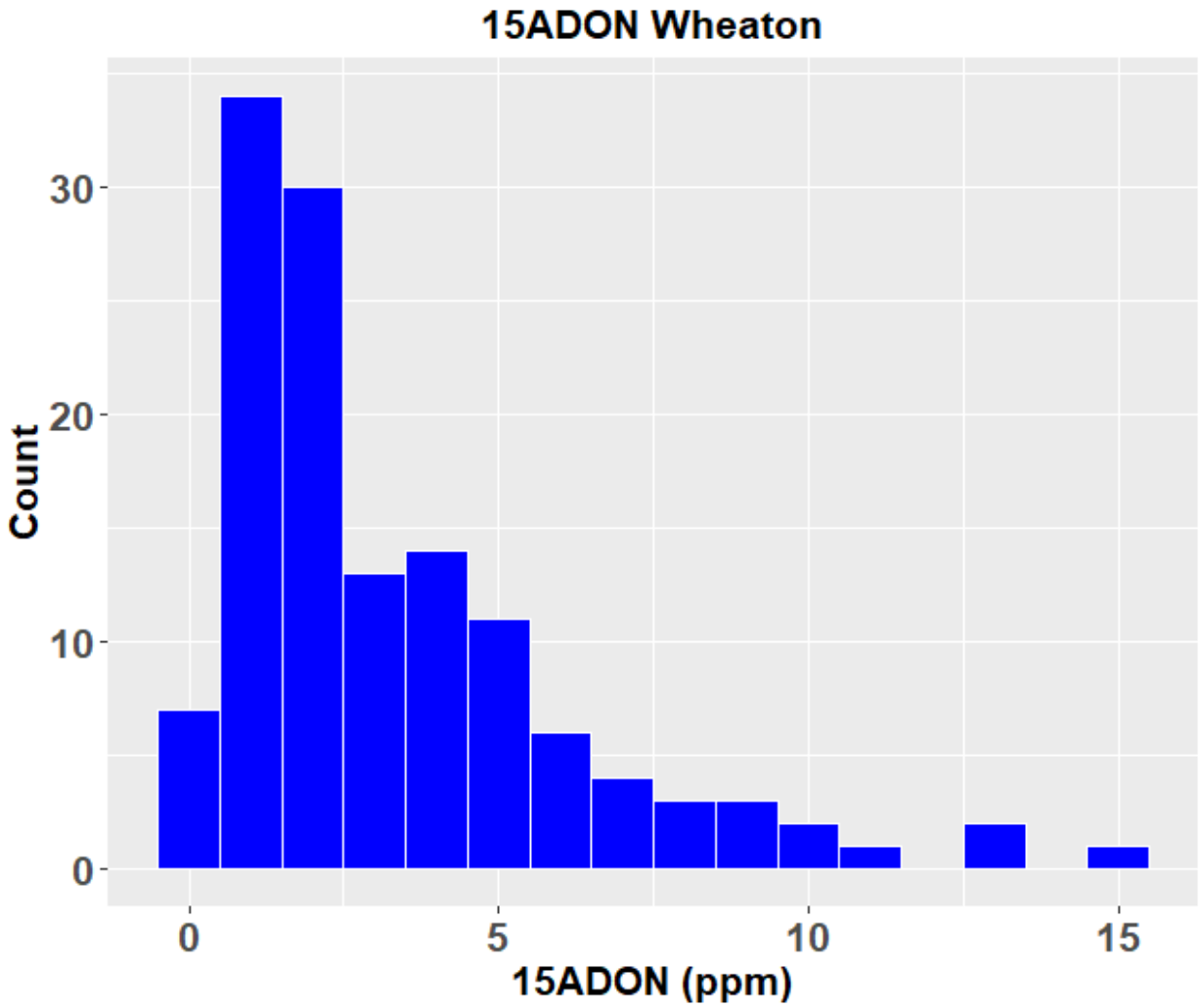


Figure 3.11. Distribution of 15ADON produced by *Fusarium graminearum* in Wheaton.

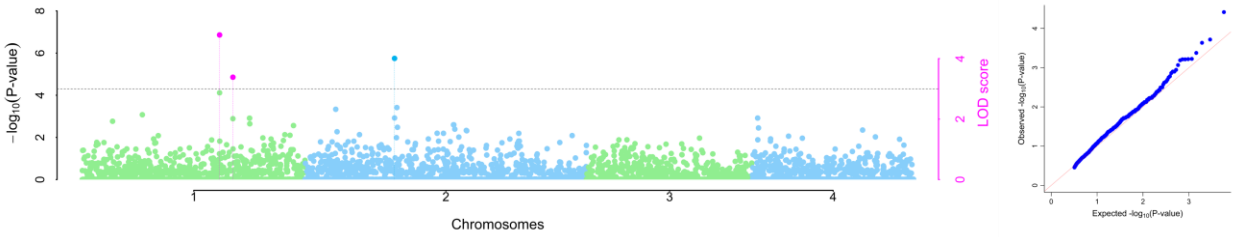


Figure 3.12. Manhattan and Q-Q plots for the trait growth rate (135 isolates).

Manhattan plot shows -log10(P-value) for all SNPs and LOD scores for significant SNPs. SNPs with a LOD score ≥ 3 were considered significant. Significant SNPs are colored in pink or blue and connected to the x-axis by a dotted line. Pink dots are the QTNs detected by a single method and blue dots are QTNs detected by multiple methods.

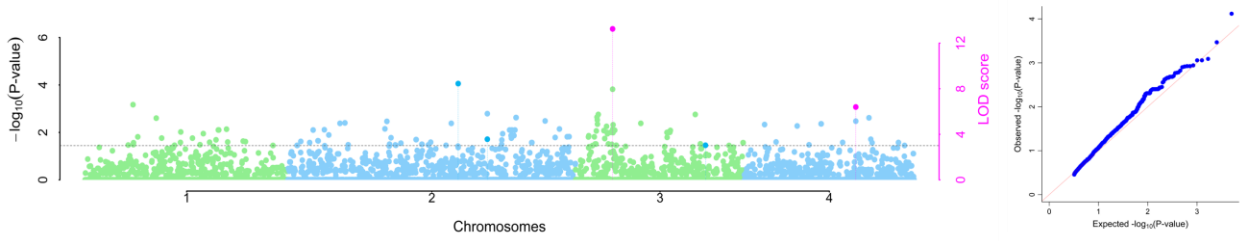


Figure 3.13. Manhattan and Q-Q plots for the trait growth rate (NA1 isolates).

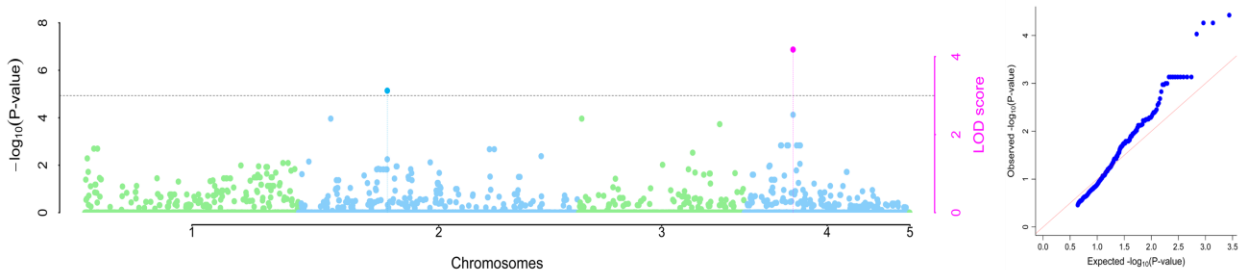


Figure 3.14. Manhattan and Q-Q plots for the trait growth rate (NA2 isolates).

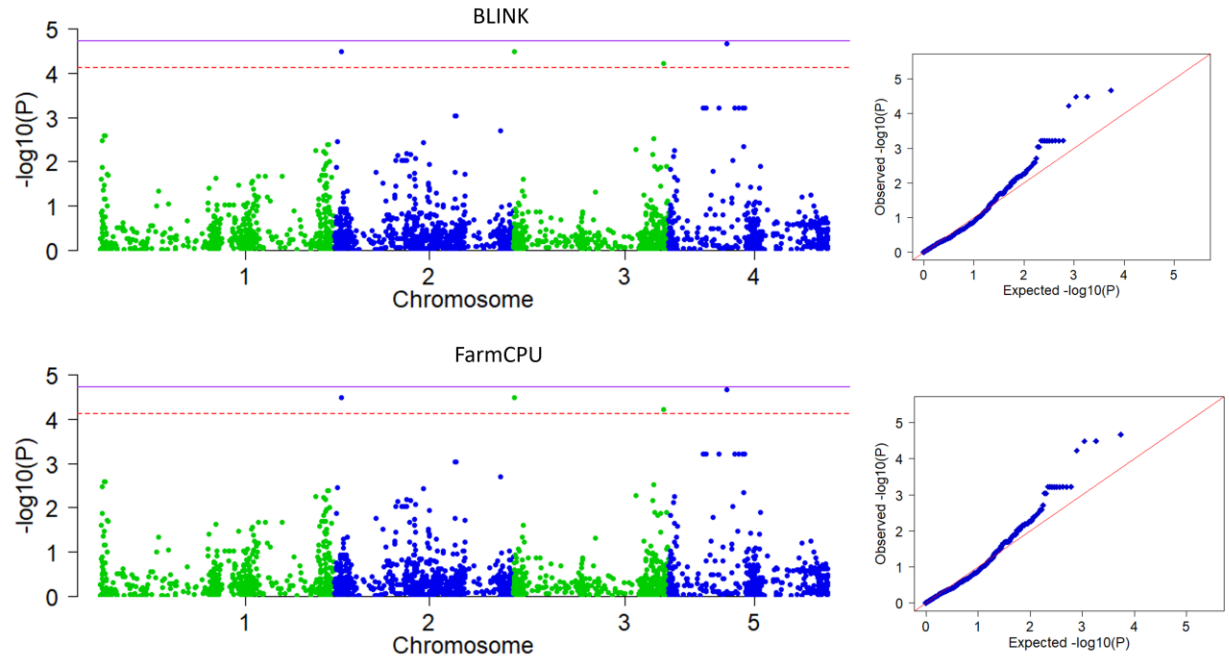


Figure 3.15. Manhattan and Q-Q plots for the trait growth rate (NA2 isolates) using GAPIT3.

Red dotted line represents FDR (Benjamini-Hochberg) cutoff threshold and solid purple line represents Bonferroni threshold. Both thresholds are set at 0.05 p-value.

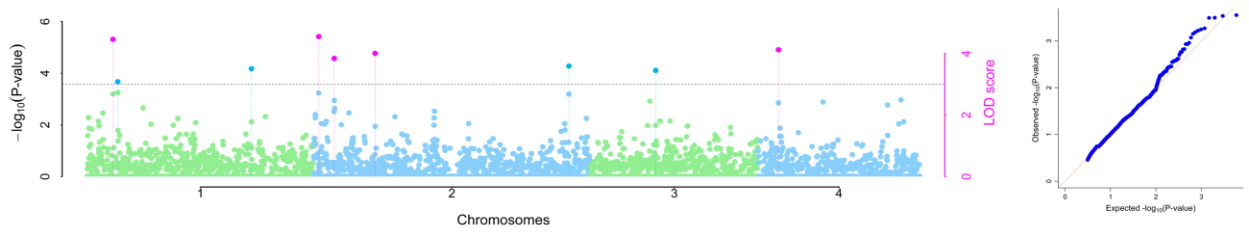


Figure 3.16. Manhattan and Q-Q plots for the trait sexual fertility (152 isolates).

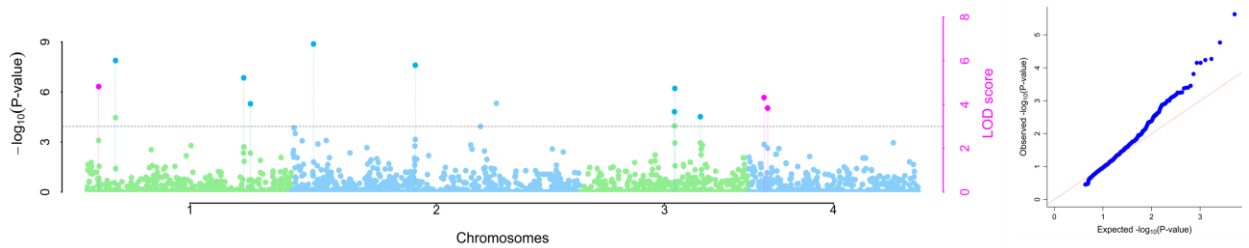


Figure 3.17. Manhattan and Q-Q plots for the trait sexual fertility (NA1 isolates).

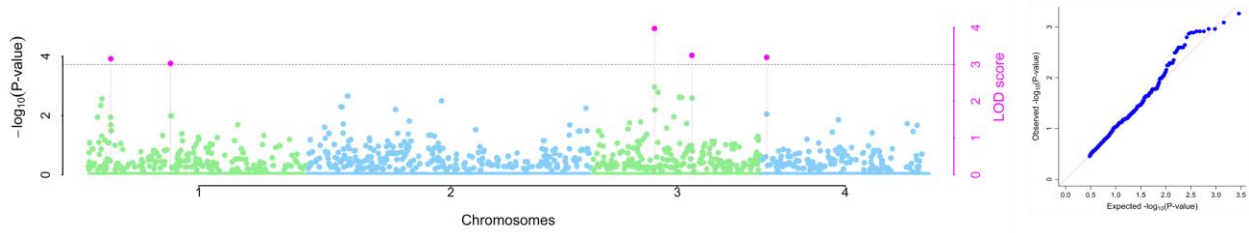


Figure 3.18. Manhattan and Q-Q plots for the trait sexual fertility (NA2 isolates).

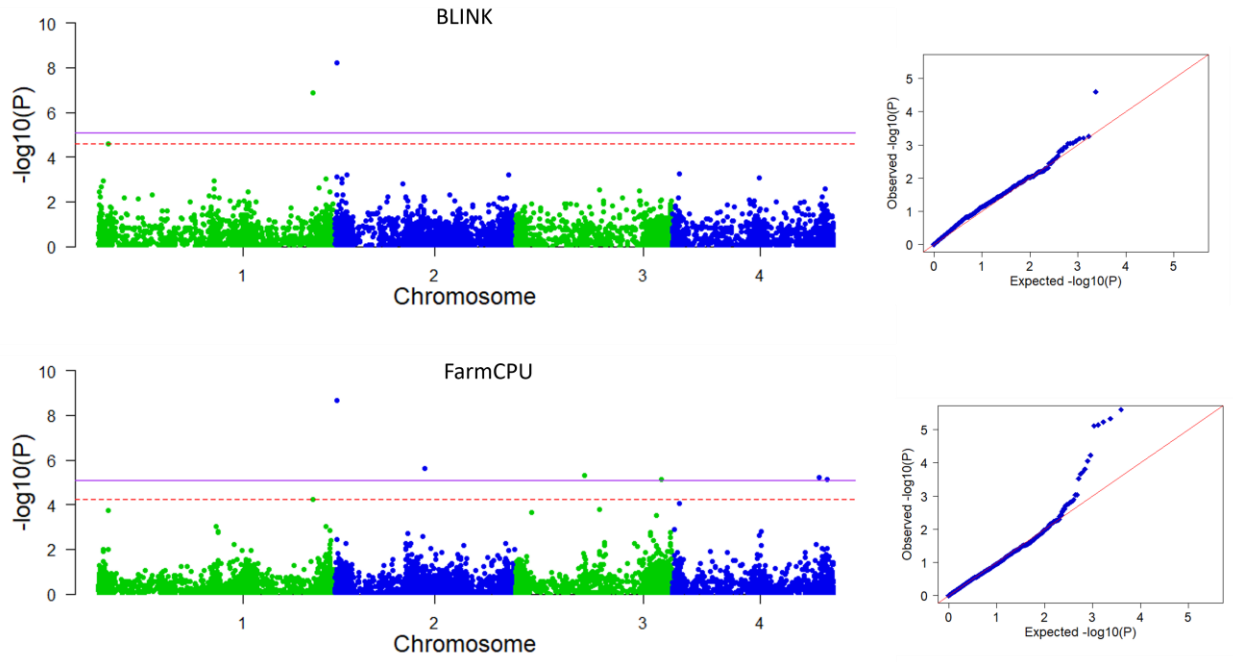


Figure 3.19. Manhattan and Q-Q plots for the trait sexual fertility (all isolates) using GAPIT3.

Red dotted line represents FDR (Benjamini-Hochberg) cutoff threshold and solid purple line represents Bonferroni threshold. Both thresholds are set at 0.05 p-value.

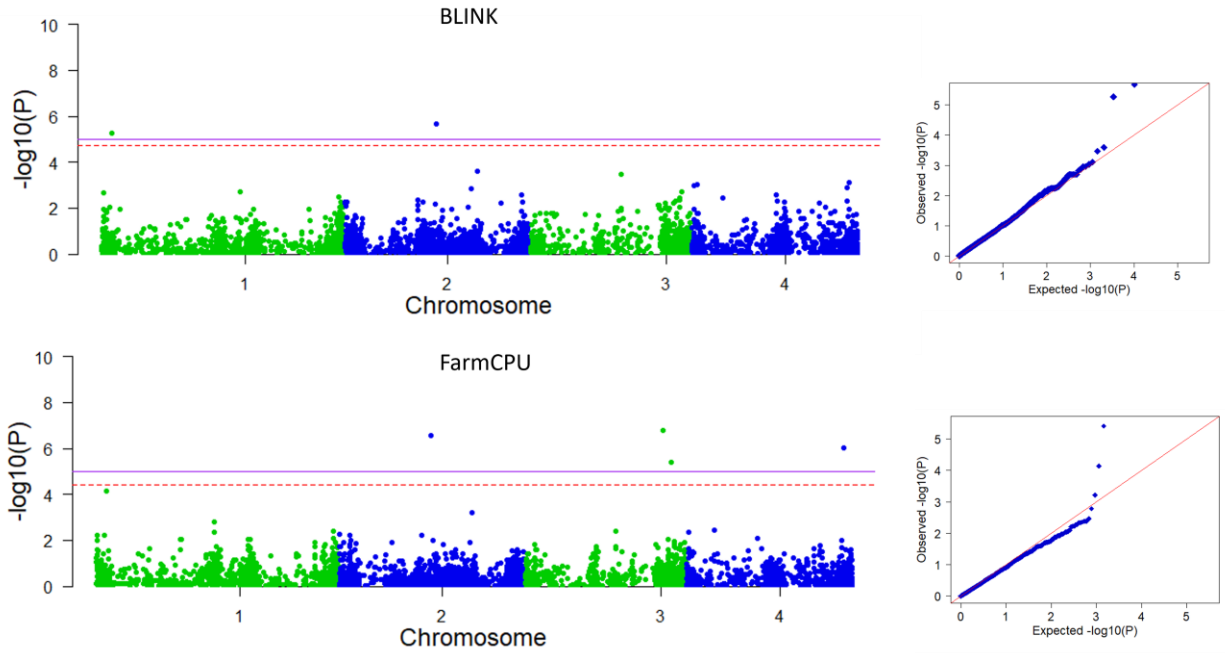


Figure 3.20. Manhattan and Q-Q plots for the trait sexual fertility (NA1 isolates) using GAPIT3.

Red dotted line represents FDR (Benjamini-Hochberg) cutoff threshold and solid purple line represents Bonferroni threshold. Both thresholds are set at 0.05 p-value.

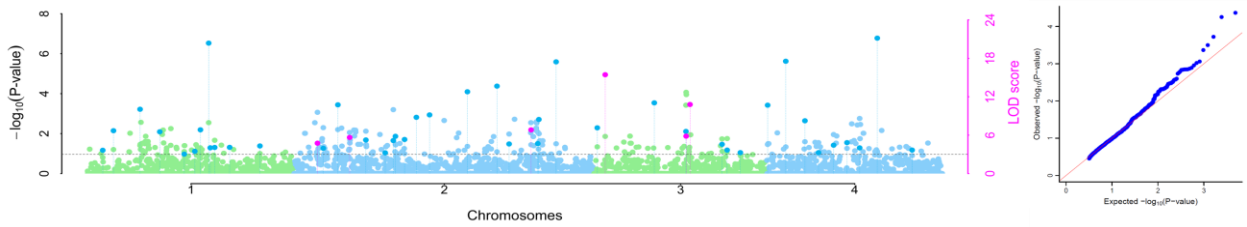


Figure 3.21. Manhattan and Q-Q plots for the trait propiconazole sensitivity.

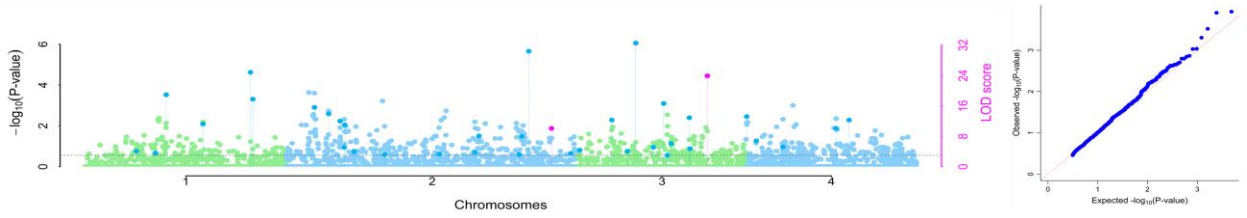


Figure 3.22. Manhattan and Q-Q plots for the trait tebuconazole sensitivity.

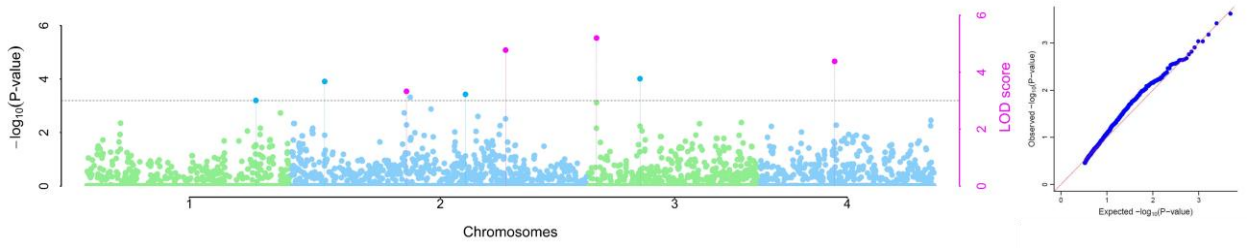


Figure 3.23. Manhattan and Q-Q plots for the trait DON production *in vitro*.

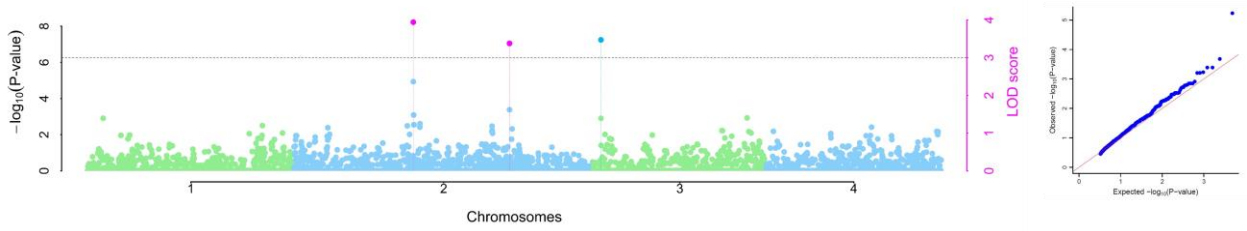


Figure 3.24. Manhattan and Q-Q plots for the trait 15DON production *in vitro*.

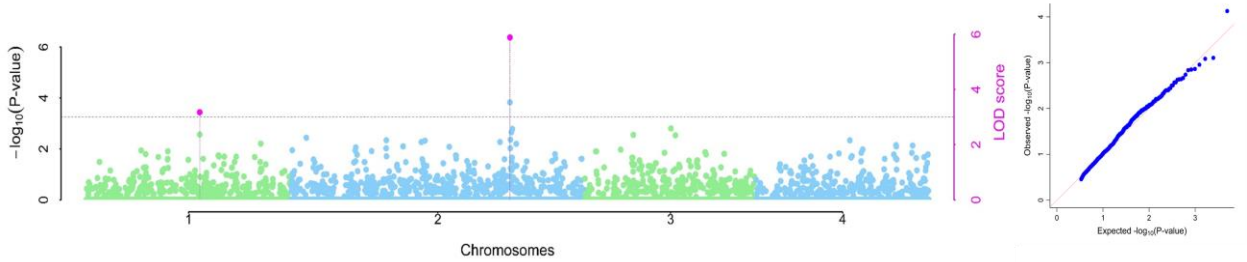


Figure 3.25. Manhattan and Q-Q plots for the trait aggressiveness.

3.8 Tables

Table 3.1. Pearson correlation coefficient between traits studied.

Traits	Growth	Ascospore	Propiconazole	Tebuconazole	Aggressiven	DON in vitro	15ADON in vitro	DON Rollag	15ADON Rollag	DON Wheaton
Growth										
Ascospore	0.112									
Propiconazole	0.236	-0.151								
Tebuconazole	0.249	-0.076	0.896							
Aggressiveness	-0.347	0.537	0.176	0.096						
DON in vitro	-0.119	0.189	-0.173	-0.107	0.192					
15ADON in vitro	-0.189	0.318	-0.138	-0.113	0.214	0.488				
DON Rollag	0.013	0.217	0.008	0.012	0.384	0.119	0.088			
15ADON Rollag	0.009	0.189	-0.065	-0.134	0.399	0.101	0.113	0.665		
DON Wheaton	0.151	0.124	0.051	0.032	0.367	0.081	0.086	0.376	0.107	
15ADON Wheaton	0.036	0.088	-0.010	-0.141	0.393	-0.034	0.063	0.039	0.418	0.297

Table 3.2. List of QTNs significantly associated with growth rate (135 isolates) in models in R package mrMLM.

Method	Chromosome	Marker position (bp)	QTN effect	LOD score	'-log10(P)'	r2 (%)	MAF
FASTmrEMMA	1	7892085	0.27	4.01	4.76	7.54	0.16
FASTmrMLM	1	7892085	0.17	3.63	4.36	12.70	0.16
ISIS EM-BLASSO	1	7892085	0.18	5.55	6.37	12.73	0.16
mrMLM	1	7892085	0.19	7.11	7.98	14.98	0.17
ISIS EM-BLASSO	1	9701935	0.14	3.04	3.73	3.15	0.06
mrMLM	1	9701935	0.21	3.73	4.47	7.49	0.06
mrMLM	2	3764676	0.13	4.01	4.76	9.33	0.25

Table 3.3. List of QTNs significantly associated with growth rate (NA1 isolates) in models in R package mrMLM.

Method	Chromosome	Marker position (bp)	QTN effect	LOD score	'-log10(P)'	r2 (%)	MAF
ISIS EM-BLASSO	2	5712361	-0.07	8.45	9.35	5.42	0.35
mrMLM	2	6320548	0.16	3.57	4.30	11.10	0.09
ISIS EM-BLASSO	3	523887	-0.15	19.20	20.28	28.32	0.35
mrMLM	3	523887	-0.16	7.30	8.17	29.83	0.36
ISIS EM-BLASSO	3	7346499	0.00	3.04	3.73	0.00	0.05
ISIS EM-BLASSO	4	6548217	-0.20	8.21	9.11	11.08	0.05
mrMLM	4	6548217	-0.26	4.58	5.36	18.50	0.06

Table 3.4. List of QTNs significantly associated with growth (NA2 isolates) in models in R package mrMLM.

Method	Chromosome	Marker position (bp)	QTN effect	LOD score	'-log10(P)'	r2 (%)	MAF
FASTmrMLM	2	3635425	0.13	3.12	3.83	11.03	0.31
FASTmrMLM	4	2946771	0.34	4.71	5.49	34.47	0.10
FASTmrEMMA	4	2946771	0.00	3.87	4.62	0.00	0.10
pKWmEB	4	2946771	0.25	4.48	5.26	44.34	0.11
ISIS EM-BLASSO	4	2946771	0.25	3.18	3.88	19.16	0.10

Table 3.5. List of QTNs significantly associated with growth rate (NA2 isolates) in models in R package GAPIT3.

Method	Chromosome	Marker position	Effect	P-value	FDR adjusted P-value	MAF
FarmCPU	2	400745	-0.45	3.27E-05	0.029891454	0.07
Blink	2	400745	-0.45	3.27E-05	0.029891454	0.07
FarmCPU	3	85478	-0.45	3.27E-05	0.029891454	0.07
Blink	3	85478	-0.45	3.27E-05	0.029891454	0.07
FarmCPU	3	7509544	-0.44	5.93E-05	0.040714155	0.07
Blink	3	7509544	-0.44	5.93E-05	0.040714155	0.07
FarmCPU	4	2946771	-0.38	2.16E-05	0.029891454	0.10
Blink	4	2946771	-0.38	2.16E-05	0.029891454	0.10

Table 3.6. List of QTNs significantly associated with sexual fertility (152 isolates) in models in R package mrMLM.

Method	Chromosome	Marker position (bp)	QTN effect	LOD score	'-log10(P)'	r2 (%)	MAF
FASTmrMLM	1	540304	0.68	4.56	5.34	5.61	0.08
ISIS EM-BLASSO	1	540304	0.75	5.07	5.87	6.76	0.08
mrMLM	1	540304	0.79	4.37	5.14	7.51	0.08
pKWmEB	1	540304	0.56	3.09	3.80	6.53	0.08
ISIS EM-BLASSO	1	677363	-0.23	3.09	3.79	1.35	0.19
FASTmrMLM	1	10661762	0.39	3.51	4.23	6.38	0.46
FASTmrEMMA	2	160544	0.87	3.48	4.20	7.77	0.46
ISIS EM-BLASSO	2	160544	0.55	5.12	5.92	12.48	0.46
mrMLM	2	160544	0.60	3.98	4.74	14.82	0.47
pKWmEB	2	160544	0.56	5.24	6.05	15.40	0.47
mrMLM	2	395059	0.85	3.90	4.64	11.68	0.11
pKWmEB	2	395059	0.62	3.79	4.53	9.44	0.11
FASTmrMLM	2	2271329	-0.34	4.55	5.33	4.80	0.46
mrMLM	2	2271329	-0.36	3.46	4.18	5.34	0.46
FASTmrMLM	2	8649328	0.43	3.59	4.32	6.42	0.30
FASTmrMLM	3	4181887	0.41	3.45	4.17	2.93	0.12
FASTmrEMMA	4	399628	1.15	5.91	6.74	9.21	0.21
FASTmrMLM	4	399628	0.43	3.80	4.54	5.15	0.21
mrMLM	4	399628	0.49	4.45	5.22	6.57	0.21
pKWmEB	4	399628	0.48	3.53	4.26	9.00	0.21

Table 3.7. List of QTNs significantly associated with sexual fertility (NA1 isolates) in models in R package mrMLM.

Method	Chromosome	Marker position (bp)	QTN effect	LOD score	'-log10(P)'	r2 (%)	MAF
FASTmrMLM	1	196392	0.57	3.99	4.75	8.93	0.16
FASTmrEMMA	1	196392	1.49	4.82	5.61	15.15	0.16
ISIS EM-BLASSO	1	196392	0.62	7.07	7.94	10.40	0.16
mrMLM	1	540288	1.29	6.01	6.85	22.09	0.07
ISIS EM-BLASSO	1	10977911	-0.35	5.22	6.03	6.27	0.41
ISIS EM-BLASSO	1	11100468	0.33	4.04	4.79	4.21	0.24
ISIS EM-BLASSO	2	455802	0.82	6.77	7.62	8.93	0.07
mrMLM	2	4489803	0.63	5.80	6.62	16.95	0.30
FASTmrMLM	3	6679636	0.68	3.67	4.40	7.64	0.09
ISIS EM-BLASSO	3	6694414	0.71	4.74	5.52	5.14	0.05
mrMLM	3	7230132	-0.61	3.44	4.17	13.21	0.20
FASTmrMLM	4	290113	0.53	4.10	4.86	9.42	0.21
FASTmrEMMA	4	290113	1.19	4.32	5.09	11.95	0.21
ISIS EM-BLASSO	4	290113	0.50	6.49	7.34	8.40	0.21
FASTmrMLM	4	399628	0.40	3.42	4.14	7.09	0.31
FASTmrEMMA	4	399628	0.99	4.25	5.02	10.89	0.31

Table 3.8. List of QTNs significantly associated with sexual fertility (NA2 isolates) in models in R package mrMLM.

Method	Chromosome	Marker position (bp)	QTN effect	LOD score	'-log10(P)'	r2 (%)	MAF
pKWmEB	1	681329	0.000048	3.15	3.86	18.58	0.24
ISIS EM-BLASSO	1	6029757	0.50	3.03	3.73	9.36	0.34
mrMLM	3	3451154	1.44	3.98	4.73	25.56	0.09
mrMLM	3	6652144	1.27	3.25	3.96	20.10	0.07
pKWmEB	4	91835	0.000021	3.19	3.90	21.76	0.33

Table 3.9. List of QTNs significantly associated in sexual fertility (152 isolates) in models in R package GAPIT3.

Method	Chromosome	Marker position	Effect	P-value	FDR adjusted P-value	MAF
Blink	1	540304	0.70	2.50E-05	0.049230237	0.08
FarmCPU	1	10661762	0.32	5.84E-05	0.049313686	0.46
Blink	1	10661762	0.53	1.38E-07	0.000406343	0.46
FarmCPU	2	160544	0.56	2.12E-09	1.25E-05	0.46
Blink	2	160544	0.73	6.13E-09	3.62E-05	0.46
FarmCPU	2	4489803	-0.53	2.49E-06	0.007362405	0.13
FarmCPU	3	3453361	-0.51	4.67E-06	0.007610878	0.16
FarmCPU	3	7239571	-0.61	7.73E-06	0.007610878	0.09
FarmCPU	4	7290074	-0.50	5.87E-06	0.007610878	0.13
FarmCPU	4	7700303	-0.51	7.09E-06	0.007610878	0.13

Table 3.10. List of QTNs significantly associated with sexual fertility (NA1 isolates) in models in R package GAPIT3.

Method	Chromosome	Marker position	Effect	P-value	FDR adjusted P-value	MAF
Blink	1	540288	0.90	5.35E-06	0.013671610	0.07
Blink	2	4489803	-0.54	2.12E-06	0.010855824	0.29
FarmCPU	2	4489803	-0.49	2.62E-07	0.000670941	0.29
FarmCPU	3	6679636	-1.10	1.68E-07	0.000670941	0.09
FarmCPU	3	7079729	-1.03	3.85E-06	0.004925503	0.05
FarmCPU	4	7596245	0.45	9.91E-07	0.001688418	0.34

Table 3.11. List of QTNs significantly associated with propiconazole in models in R package mrMLM.

Method	Chromosome	Marker position (bp)	QTN effect	LOD score	'-log10(P)'	r2 (%)	MAF
pKWmEB	1	237393	-2.00E-04	3.64	4.37	0.02	0.10
ISIS EM-BLASSO	1	489860	-7.00E-04	6.70	7.55	0.11	0.13
pKWmEB	1	3881005	4.00E-04	10.05	10.98	1.49	0.34
ISIS EM-BLASSO	1	5731674	-2.00E-04	6.52	7.37	0.01	0.09
pKWmEB	1	6829683	-2.00E-04	3.04	3.73	0.06	0.23
ISIS EM-BLASSO	1	7218760	-2.89E-05	3.48	4.20	0.00	0.38
ISIS EM-BLASSO	1	7356780	-2.00E-04	6.83	7.69	0.00	0.06
ISIS EM-BLASSO	1	7467655	-0.0075	20.37	21.46	5.26	0.05
ISIS EM-BLASSO	1	7514533	-1.00E-04	4.02	4.78	0.01	0.29
ISIS EM-BLASSO	1	7585853	6.00E-04	4.08	4.83	0.12	0.28
ISIS EM-BLASSO	1	9999192	1.00E-04	4.09	4.85	0.01	0.31
pKWmEB	1	11270238	6.00E-04	4.31	5.08	0.01	0.22
mrMLM	2	459782	-0.0041	3.58	4.30	8.11	0.48
ISIS EM-BLASSO	2	459782	-0.0018	5.91	6.74	1.59	0.47
pKWmEB	2	501355	-5.00E-04	3.95	4.70	0.28	0.10
ISIS EM-BLASSO	2	665406	0.0019	10.74	11.69	1.44	0.30
FASTmrMLM	2	851037	0.0031	3.32	4.04	3.27	0.24
pKWmEB	2	851037	1.60E-03	32.51	33.70	1.86	0.24
ISIS EM-BLASSO	2	851037	0.0027	5.64	6.46	2.48	0.24
pKWmEB	2	2753268	1.40E-03	5.25	6.05	0.54	0.11
ISIS EM-BLASSO	2	3804766	-7.00E-04	3.24	3.95	0.20	0.50
ISIS EM-BLASSO	2	3923133	-0.0025	5.16	5.96	0.83	0.08
ISIS EM-BLASSO	2	3938384	0.0014	5.82	6.64	0.79	0.33
pKWmEB	2	4138058	-4.00E-04	5.33	6.13	0.89	0.13
ISIS EM-BLASSO	2	4358147	1.00E-04	8.79	9.70	0.00	0.05
ISIS EM-BLASSO	2	4679670	-7.83E-05	9.16	10.08	0.00	0.06
ISIS EM-BLASSO	2	5664783	-0.0013	12.77	13.76	0.80	0.44
pKWmEB	2	6243290	-8.06E-05	13.66	14.66	1.01	0.31
pKWmEB	2	6487389	2.00E-04	4.62	5.40	0.40	0.13
pKWmEB	2	8062916	-2.00E-04	8.86	9.77	0.01	0.39
ISIS EM-BLASSO	2	8062916	-2.70E-05	4.78	5.57	0.00	0.38
pKWmEB	2	8232320	1.00E-04	4.68	5.46	0.00	0.31
ISIS EM-BLASSO	2	8244606	3.00E-04	8.45	9.35	0.05	0.47
pKWmEB	2	8502325	-5.00E-04	17.45	18.50	0.58	0.14
ISIS EM-BLASSO	3	72709	8.00E-04	7.14	8.01	0.14	0.14
pKWmEB	3	179032	1.00E-03	6.83	7.69	2.37	0.31
ISIS EM-BLASSO	3	179032	0.0025	24.04	25.16	2.60	0.30
ISIS EM-BLASSO	3	4095333	0.0013	11.05	12.01	0.74	0.33
ISIS EM-BLASSO	3	6640551	-0.006	6.58	7.43	3.39	0.05
mrMLM	3	6640556	-0.0103	5.46	6.28	10.10	0.05
ISIS EM-BLASSO	3	6640556	-0.003	6.23	7.07	0.85	0.05
pKWmEB	3	6679615	-4.00E-04	5.34	6.15	0.92	0.06
ISIS EM-BLASSO	3	6679615	-6.00E-04	16.24	17.28	0.04	0.07
pKWmEB	3	7294460	4.66E-05	4.57	5.35	0.57	0.34
ISIS EM-BLASSO	3	7346278	6.00E-04	3.67	4.40	0.12	0.20
pKWmEB	3	7521430	3.00E-04	3.27	3.98	0.10	0.09
ISIS EM-BLASSO	4	25804	-0.0019	10.68	11.63	1.31	0.25
pKWmEB	4	329152	-1.30E-03	17.54	18.60	3.40	0.05
pKWmEB	4	2203719	-1.00E-04	8.25	9.15	0.02	0.31
pKWmEB	4	3862627	-3.00E-04	3.27	3.98	0.47	0.08
ISIS EM-BLASSO	4	4117472	4.00E-04	4.43	5.20	0.04	0.13
pKWmEB	4	4362316	1.00E-04	4.83	5.62	0.30	0.24
ISIS EM-BLASSO	4	4545038	-0.0018	4.01	4.76	0.60	0.11
pKWmEB	4	5602779	6.59E-05	21.15	22.24	1.87	0.08
ISIS EM-BLASSO	4	7533605	-2.00E-04	3.67	4.40	0.01	0.09

Table 3.12. List of QTNs significantly associated with tebuconazole in models in R package mrMLM.

Method	Chromosome	Marker position (bp)	QTN effect	LOD score	'-log10(P)'	r2 (%)	MAF
pKWmEB	1	3885008	9.54E-05	4.11	4.87	0.26	0.34
ISIS EM-BLASSO	1	5731674	-5.00E-04	3.59	4.32	0.15	0.09
pKWmEB	1	5933539	9.64E-05	18.95	20.02	0.55	0.47
ISIS EM-BLASSO	1	7467655	-0.001	11.23	12.20	0.39	0.05
pKWmEB	1	11258195	3.68E-05	24.88	26.01	2.08	0.09
pKWmEB	1	11270238	1.00E-04	17.78	18.84	0.19	0.22
pKWmEB	2	501355	-7.00E-04	15.62	16.66	0.65	0.10
ISIS EM-BLASSO	2	665406	0.001	13.94	14.95	1.71	0.30
ISIS EM-BLASSO	2	851037	0.0014	12.02	12.99	2.83	0.24
pKWmEB	2	964956	5.68E-05	5.07	5.87	0.16	0.33
pKWmEB	2	985340	8.95E-05	11.02	11.98	2.06	0.17
ISIS EM-BLASSO	2	2645857	-6.00E-04	3.94	4.69	0.46	0.22
ISIS EM-BLASSO	2	3938384	7.00E-04	3.20	3.91	0.80	0.33
ISIS EM-BLASSO	2	5182922	6.00E-04	3.31	4.02	0.26	0.10
ISIS EM-BLASSO	2	6143802	0.0011	3.79	4.53	1.23	0.15
ISIS EM-BLASSO	2	6173074	7.00E-04	8.11	9.01	0.50	0.18
pKWmEB	2	8129795	5.58E-05	3.20	3.91	0.87	0.35
ISIS EM-BLASSO	2	8222086	-0.0013	7.99	8.88	2.26	0.22
pKWmEB	2	8322036	-7.72E-05	30.42	31.59	3.06	0.35
pKWmEB	2	8566172	-3.00E-04	13.45	14.45	1.13	0.06
ISIS EM-BLASSO	2	8566172	-5.00E-04	6.71	7.56	0.11	0.06
pKWmEB	2	8880263	-4.00E-04	3.47	4.19	0.00	0.18
ISIS EM-BLASSO	3	65770	4.00E-04	4.36	5.13	0.28	0.45
pKWmEB	3	503820	6.97E-05	12.29	13.28	0.12	0.34
ISIS EM-BLASSO	3	3183025	4.00E-04	4.07	4.82	0.25	0.33
pKWmEB	3	3813789	3.00E-04	32.56	33.75	0.57	0.07
ISIS EM-BLASSO	3	6299555	6.00E-04	5.12	5.92	0.19	0.09
pKWmEB	3	6562010	7.63E-05	16.65	17.69	1.61	0.15
ISIS EM-BLASSO	3	6640551	-6.00E-04	3.00	3.70	0.12	0.05
ISIS EM-BLASSO	3	6679615	-5.00E-04	6.09	6.93	0.12	0.07
ISIS EM-BLASSO	3	7066756	-8.00E-04	12.90	13.89	1.02	0.35
ISIS EM-BLASSO	3	7067865	-3.03E-05	4.73	5.52	0.00	0.43
pKWmEB	3	7346278	1.00E-03	32.95	34.14	2.53	0.21
ISIS EM-BLASSO	3	7346278	0.0012	14.90	15.92	1.76	0.20
ISIS EM-BLASSO	4	25804	-0.0013	13.20	14.20	2.36	0.25
ISIS EM-BLASSO	4	177976	-5.00E-04	6.82	7.68	0.12	0.08
ISIS EM-BLASSO	4	2203719	-6.00E-04	5.20	6.00	0.54	0.31
pKWmEB	4	4545038	-1.00E-03	10.08	11.02	1.23	0.11
pKWmEB	4	5201402	-2.00E-04	12.27	13.25	0.31	0.05

Table 3.13. List of QTNs significantly associated with DON production *in vitro* in models in R package mrMLM.

Method	Chromosome	Marker position (bp)	QTN effect	LOD score	'-log10(P)'	r2 (%)	MAF
ISIS EM-BLASSO	1	11260778	-0.0001	3.01	3.70	0.00	0.18
FASTmrMLM	2	520988	-0.13	3.67	4.41	3.90	0.19
FASTmrMLM	2	4175663	0.16	3.11	3.82	4.37	0.15
ISIS EM-BLASSO	2	4175663	0.07	3.54	4.27	0.86	0.15
pKWmEB	2	5687139	0.13	3.22	3.93	7.62	0.20
ISIS EM-BLASSO	2	6526109	0.13	5.83	6.66	3.92	0.21
pKWmEB	2	6526109	0.13	3.72	4.46	10.15	0.21
ISIS EM-BLASSO	3	102121	-0.19	5.49	6.31	5.44	0.12
pKWmEB	3	102121	-0.18	4.90	5.70	9.89	0.12
mrMLM	3	3285338	0.20	3.77	4.51	5.87	0.12
ISIS EM-BLASSO	4	4261526	0.12	5.71	6.54	3.97	0.25
pKWmEB	4	4261526	0.10	3.05	3.74	6.53	0.25

Table 3.14. List of QTNs significantly associated with 15ADON production *in vitro* in models in R package mrMLM.

Method	Chromosome	Marker position (bp)	QTN effect	LOD score	'-log10(P)'	r2 (%)	MAF
mrMLM	2	4313942	-0.30	5.00	5.79	12.77	0.11
FASTmrMLM	2	4313942	-0.25	3.79	4.53	8.98	0.11
FASTmrEMMA	2	4313942	-0.44	3.94	4.69	7.00	0.11
pKWmEB	2	4313942	-0.23	3.66	4.40	11.46	0.11
ISIS EM-BLASSO	2	4313942	-0.22	4.25	5.01	7.38	0.11
pKWmEB	2	6526109	0.02	3.13	3.83	6.52	0.21
ISIS EM-BLASSO	2	6526109	0.12	3.63	4.36	4.00	0.21
ISIS EM-BLASSO	3	98785	-0.15	3.47	4.20	2.78	0.09

Table 3.15. List of QTNs significantly associated with aggressiveness in Wheaton and Rollag in models in R package mrMLM.

Method	Chromosome	Marker position (bp)	QTN effect	LOD score	'-log10(P)'	r2 (%)	MAF
mrMLM	1	7371490	-4.51	3.18	3.88	13.14	0.39
FASTmrMLM	1	7371490	-3.04	3.18	3.88	6.15	0.40
FASTmrEMMA	1	7371490	-6.83	3.39	4.11	7.77	0.40
ISIS EM-BLASSO	1	7371490	-3.04	3.18	3.88	6.15	0.40
mrMLM	2	6675327	-5.47	5.89	6.72	19.34	0.41
FASTmrMLM	2	6675327	-4.62	5.89	6.72	14.18	0.40
FASTmrEMMA	2	6675327	-9.18	4.97	5.77	14.00	0.40
ISIS EM-BLASSO	2	6675327	-4.62	5.89	6.72	14.18	0.40

3.9 List of supplemental files

Supplemental file 3.1: Information about isolates used for different phenotypic trait measurement and GWAS

Supplemental file 3.2: List of candidate genes flanking significant QTNs identified using mrMLM

Supplemental file 3.3: List of candidate genes flanking significant QTNs identified using GAPIT3

Supplemental file 3.4: List of genes common in GWAS and cross population selection scans for population NA1.

Supplemental file 3.5: List of genes common in GWAS and cross population selection scans for population NA2.

Chapter 4 - The Landscape and Predicted Roles of Structural Variants in *Fusarium graminearum* Genomes

4.1 Abstract

Structural rearrangements, such as inversions, translocations, duplications, and large insertions and deletions, are large-scale genomic variants that can play an important role in shaping phenotypic variation and in genome adaptation and evolution. We used chromosomal-level assemblies from eight *Fusarium graminearum* isolates to study structural variants and their role in fungal evolution. We generated the assemblies of four of these genomes after Oxford Nanopore sequencing. A total of 87 inversions, 159 translocations, 245 duplications, 58,489 insertions and 34,102 deletions were detected. Regions of high recombination rate are associated with structural rearrangements, and a significant proportion of inversions, translocations, and duplications overlap with the repeat content of the genome, suggesting recombination and repeat elements are major factors in the origin of structural rearrangements in *F. graminearum*. Large insertions and deletions introduce presence-absence polymorphisms for many genes, including secondary metabolite biosynthesis cluster genes and predicted effectors genes. Translocation events were found to be shuffling predicted effector-rich regions of the genomes and are likely contributing to the gain and loss of effectors facilitated by recombination. Breakpoints of some structural rearrangements fall within coding sequences and are likely altering the protein products. Structural rearrangements in *F. graminearum* thus have an important role to play in shaping pathogen-host interactions and broader evolution through genome reorganization, the introduction of presence-absence polymorphisms, and changing protein products and gene regulation.

4.2 Introduction

Phenotypic trait variation arises due to small (SNPs and small indels) and large-scale (structural rearrangements) genomic variants, and these variants are key to genome evolution and speciation (Priest et al., 2020). Meiotic (and parasexual) recombination, DNA repair during meiosis and mitosis, transposon activities, and horizontal gene transfer are some of the mechanisms by which large-scale rearrangements originate in genomes (Mehrabi et al., 2017). Double-stranded DNA breaks generated during these processes are repaired using non-homologous end joining, which is an error prone method of DNA repair.

Studies on large-scale chromosomal rearrangements in fungal genomes have elucidated that structural rearrangements play an important role in fitness, adaptation, fungicide resistance, speciation and evolution (Langner et al., 2021; Massonnet et al., 2018; Tralamazza et al., 2019; Wang et al., 2020). Chromosomal rearrangements have been implicated in presence-absence polymorphisms in secondary metabolite synthesis genes in *Phaeoacremonium minimum* (Massonnet et al., 2018) and the *Fusarium graminearum* species complex (Tralamazza et al., 2019), as well as in the evolution of effector superfamily genes, host-specificity and speciation in the genus *Taphrina* (Wang et al., 2020). Similarly, in the wheat fungal pathogen *Zymoseptoria tritici*, a recent emergence of fungicide resistance is driven by chromosomal rearrangements (Badet et al., 2021). Likewise, in the powdery mildew of grape pathogen *Erysiphe nectar*, copy number variation in EnCYP51 genes is associated with DMI fungicide resistance (Jones et al., 2014). Dispensable mini-chromosomes, a special type of structural variation, most likely have emerged in *Magnaporthe oryzae* by the rearrangement and duplication of core chromosome segments and are implicated in adaptation and host-specificity (Langner et al., 2021).

Despite the important role of structural rearrangements in phenotypic trait variation, adaptation, and evolution, they have not been systematically studied in many fungi because high-quality contiguous assemblies for multiple isolates are lacking. The recent advent of long-read sequencing technologies has facilitated assembly of chromosomal-level assemblies at relatively low cost and should accelerate the study of structural rearrangements in fungal and other genomes. These high-quality genomes will also be useful resources for many other genomics analyses. This study characterizes structural variation in the genomes of the plant pathogenic fungus *F. graminearum*. *Fusarium graminearum* is a fungal pathogen of wheat and barley worldwide and causes FHB disease (Goswami and Kistler, 2004; McMullen et al., 2012). Its genome is composed of four chromosomes and a mitochondrial genome (King et al., 2015), with distinct genomic regions evolving at different rates, fast and slow (Laurent et al., 2018; Wang and Zhang, 2021). The 'fast' genome is enriched in genes with roles in plant-microbe interaction, secondary metabolite synthesis and genes expressed *in planta* during infection, and this region is key to adaptation (Laurent et al., 2018). Previous genome resequencing studies have defined the species' pangenome and delineated core and accessory genes (Alouane et al., 2021; Kelly and Ward, 2018; Machado et al., 2018; Walkowiak et al., 2015, 2016). Accessory genes from this and related FHB species are located in highly diverse, frequently recombining regions near the chromosomal ends and are differentially conserved among pathogen populations (Kelly and Ward, 2018; Walkowiak et al., 2016). In this study, we assembled high-quality genomes of four representative isolates from different *F. graminearum* populations using long reads from Oxford Nanopore technology. We discuss the presence and distribution of these large-scale genomic variants as well as their potential role in survival, adaptation, and pathogen fitness.

4.3 Materials and Methods

4.3.1 Origin of sequenced isolates

The four *F. graminearum* isolates sequenced and assembled here were collected from diseased wheat heads by other researchers and have been included in previous publications (Gale et al., 2007; Kuhnem et al., 2015). Supplemental file 4.1 includes the date and location of collection for each isolate.

4.3.2 DNA extraction, library preparation and sequencing

Mycelia for DNA extraction were collected by growing the isolates on liquid complete media for 2 days and frozen at -80 °C. DNA was extracted using the published protocol by Schwessinger and Mcdonald (2017) with some modifications. Briefly, 500 mg of frozen mycelium was ground in liquid nitrogen with 5 g sand using a mortar and pestle. The powder was transferred to a 50 ml centrifuge tube with 14 ml of lysis buffer and 25 ul RNAse A (20 mg/ml) and mixed gently by inverting the tube. The tube was incubated on a rocking platform at slow speed for 1 hour at room temperature. This was followed by proteinase K treatment (200 µl) and 1.25 hour-long incubation on the rocking platform at room temperature. The tubes were centrifuged at 5,000 x g for 12 minutes at 4°C. The supernatant was treated with phenol: chloroform: isoamyl mixture, incubated for 15 mins (mixed gently by inverting during incubation) and centrifuged at 4000 x g for 10 mins at 4°C. The aqueous layer was transferred to a new tube and phenol: chloroform: isoamyl treatment was repeated once. The aqueous layer was mixed with an equal volume of sodium acetate (NAOAc, 3M, pH 5.2) and DNA was allowed to precipitate at room temperature for 15 mins. The tubes were centrifuged at 8000 x g for 30 mins at 4°C to pellet the precipitated DNA. The pellet was washed twice with 70% ethanol and resuspended in 0.1M Tris HCL (pH 8.5). The DNA was cleaned further by adding 12.5 µl

RNAse A followed by an hour incubation on the rocking platform. This step was followed by phenol: chloroform: isoamyl treatment, precipitation of DNA with sodium acetate and washing with 70% ethanol. In the final step, DNA was resuspended in 50 µl of 0.1M tris-HCL at pH 8.5. High molecular weight DNA (DNA fragments longer than 30 kb) was selected by using Blue Pippin followed by cleanup using AMPure XP beads (Beckman Coulter) before quantifying the DNA using Qubit. Loading solution and electrophoresis buffer provided with the Blue Pippin kit were used.

Each Oxford Nanopore sequencing library was prepared using ligation kit (SQK-LSK109) and about 1000 ng genomic DNA. The SQK-LSK109 protocol was slightly modified. During DNA repair and end prep, incubation at 20°C was increased to 30 minutes from five minutes. Similarly, during clean up with AMPure beads, the DNA sample was incubated for 10 minutes and the first wash was performed with 600 µl of 75% ethanol. Also, during adapter ligation, DNA samples were incubated for 30 minutes following mixing with adapters and ligation reagents and further incubated for 10 mins after mixing the samples with AMPure beads. Libraries for individual isolates were loaded into separate MinION flow cells and sequenced.

4.3.3 Base calling and genome assembly

Following sequencing, DNA bases were called using the Guppy base caller. Genomes were assembled using the Canu-1.9 assembler (Koren et al., 2017). minReadLength and minOverlapLength were set to 5000 and 1000 respectively. corOutCoverage was set to either 40 or 80. Default settings were used for the rest of the parameters. The completeness of each assembly was assessed by BUSCO analysis using the Hypocreales database (Simão et al., 2015).

4.3.4 Error correction

Nanopolish _0.11.3 and Pilon version 1.24 were used to correct the draft assemblies (Loman et al., 2015; Walker et al., 2014). Nanopolish uses the signal of the raw fast5 reads to correct errors. Briefly, the raw fast5 sequences were indexed using nanopolish index and the draft genome assembly was indexed using minimap2 (Li, 2018). The raw fast5 sequences were mapped to the draft genome assembly using minimap2. The resulting BAM file was converted to SAM file format, index and sorted. This was followed by splitting the assembly into 50 kb fragments. Each fragment was corrected by nanopolish and then all fragments were merged to produce the corrected assembly. Two rounds of nanopolishing were performed on each of the four genomes. The corrected assembly from the first round was used as the input for the second round of polishing. The first round of polishing produced about 92,000 changes while changes in the second round averaged about 900 per genome. Insertions accounted for over 90% of the changes made for all genomes.

The four isolates were also sequenced using the Illumina HiSeq platform to obtain short reads for Pilon. For DNA extraction, fungal mycelia was collected after growing the isolates in liquid complete media for 2 days and frozen at -80˚C. The mycelium was freeze-dried for 3 days. About 500 mg of the dried mycelia was ground to powder using a bead beater at 29-30 rpm for 3 minutes. DNA was extracted using a CTAB method. Briefly, 800 µl of warm 2% CTAB buffer and 10 µl of 2-mercaptaethanol was added to each sample and incubated at 65°C for 30 mins with occasional gentle mixing during the incubation. Chloroform: isoamyl alcohol mixture (24:1) was added and gently mixed by inverting the tubes. This was followed by centrifugation at 12,000 rpm for 20 minutes. The aqueous phase was transferred to a fresh tube and DNA was precipitated by adding cold (-20°C) isopropanol. The DNA was pelleted by centrifuging the tube

at 10,000 rpm for 5 mins, air dried and resuspended in 1X TE buffer. The RNA contamination was removed by treating DNA with RNAse A. DNA concentration was quantified using Qubit. An Illumina HiSeq library was prepared with Nextera adapters, and sequenced to obtain 150 bp paired reads. Illumina sequences were trimmed using ILLUMINACLIP in trimmomatic-0.39 (Bolger et al., 2014). Sequences shorter than 60 bp and leading and trailing bases below quality score 3 were removed. Reads were scanned in sliding windows of 4 base pairs and sliding windows with a mean quality score below 13 were dropped. The remaining paired-end reads were used for Pilon. Quality filtered Illumina reads were aligned to the nanopolish-corrected draft assembly using default parameters in BWA-MEM. The output BAM file was changed to a SAM file, indexed, and sorted. Pilon was used to correct errors with parameters --minmq 40, --fix bases, --minqual 15.

In addition to making small insertions, deletions, and substitutions, Pilon flagged potential large assembly errors producing duplications that might require collapsing. Three rounds of Pilon polishing were performed on assemblies obtained after nanopolishing. The number of corrections decreased from round one to round three. Pilon made about 22,000 corrections per genome in the first round. Small insertional corrections were the largest followed by small deletions and substitutions. The second round of pilon polishing made about 100 changes while less than 10 corrections were made in the third round for all four assemblies.

4.3.5 Aligning orphan contigs from conspecific isolates to genome assemblies

The population genomics study of Kelly and Ward (2018) assembled sets of ‘orphan contigs’ from sequencing reads of 60 isolates of *F. graminearum* that did not map to the PH-1 reference genome. The MUMmer tool NUCmer (nucmer --mum) was used to align the orphan contigs sets against each of the four genomes assembled in this study with default parameters,

and the MUMmer tool show-coords was used to parse the alignment output (Marçais et al., 2018). The function count_overlaps from the Python package PyRanges (Stovner and Sætrom, 2020) was used to generate alignment depth information across all 60 isolates along the four nuclear chromosomes from each of the four assemblies.

4.3.6 Identification of structural variants

In addition to the four genomes we assembled, we downloaded chromosomal-level assemblies of three *F. graminearum* isolates, FG-12 (GCA_019343145.1), CML3066 (GCA_900073075.1) and CS3005 (GCA_000599445.1) from the NCBI Genome database. The software SyRI by Goel et al. (2019) was used to detect the structural variants (SVs) in seven *F. graminearum* genomes. Genomes were aligned to reference genome PH-1 (GCA_900044135.1) using the whole genome alignment tool MUMmer (Marçais et al., 2018). The MUMmer program NUCmer was used for the alignment with minimum cluster length (c), breaklength (b) and minimal length of maximum exact match (l) set at 100, 500 and 50, respectively. NUCmer identified all 50-mers, and 50-mers within 500 bp are merged to form a single alignment. Alignments were filtered to remove alignments that were less than 90% identical and shorter than 100 bp using the delta-filter tool in MUMmer. SyRI was then used to identify inversions, translocations, duplications, and indels, as well as SNPs and highly divergent regions (HDRs), using default settings. Only alignments to the four nuclear chromosomes of PH-1 were retained, and mitochondrial genomes were not compared. Chromosomes in the reference and query genomes were labelled to have identical IDs. SyRI first identifies the syntenic regions between the homologous chromosomes. Non-syntenic regions are reversed in one of the chromosomes to identify the inversions (reverse complementing one of the chromosomes makes the inverted regions syntenic). The remaining non-syntenic regions are then screened for translocations or

duplications. A custom script was used to extract the genes (complete and partial) falling in the rearranged segments using the start and end position of the events in the reference PH-1 genome. The rearrangements that partially overlapped genes in PH-1 were manually checked to identify if the chromosomal break points overlapped coding sequence and if a gene falls within the rearranged segment.

Regions present in one or more query genome but absent in PH-1 (insertions relative to PH-1) were searched against a database of 'orphan' protein sequences predicted by Kelly and Ward (2018) (their S2 File) to identify the potential genes present in the inserted segments. The database consists of genes present in one or more of the sixty isolates but absent in PH-1. The functions of the genes in the insertions closely matching the genes in the orphan protein database were identified by performing BLAST searches of the orphan proteins against the GenBank fungal protein database. The presence of secretory signals in genes was assessed using the software SignalP 6.0 (Teufel et al., 2022).

We defined regions with high recombination based on the genetic map by Laurent et al. (2018). Coordinates of the regions were used to determine which structural rearrangements fall within the high recombination regions in *F. graminearum*. We also mapped the PH-1 reference genome to a reference genome of *F. verticillioides* isolate 7600 (GCA_027571605.1) using the NUCmer program in MUMmer to identify ancient chromosomal fusion sites in the PH-1 genome. Overlap of ancient chromosomal fusion sites with regions of high recombination was observed.

Relatedness of the isolates was checked by constructing a neighbor joining tree with the nj function of the ape R package (Paradis and Schliep, 2019) using single nucleotide polymorphisms in the query genomes identified by SyRI.

4.3.7 Detection of repeats and transposable elements

We used the Extensive *de-novo* TE Annotator (EDTA) pipeline to characterize the repeat content and presence of transposable elements (TEs) in the genomes (Ou et al., 2019). EDTA combines the TE annotation programs LTRharvest (Ellinghaus et al., 2008), LTR_retriever (Ou and Jiang, 2018), Generic Repeat Finder (GRF) (Shi and Liang, 2019), TIR-Learner (Su et al., 2019), HelitronScanner (Xiong et al., 2014) and RepeatModeler (Flynn et al., 2020) to identify the repeat elements. The elements are passed through basic and advanced filters of the program to produce a non-redundant TE library for annotation of structurally intact and fragmented TE elements. We supplied the pipeline with the coding sequences of the PH-1 reference genome which is used to purge gene sequences in the TE library.

4.4 Results

4.4.1 Genome assembly

We selected four *Fusarium graminearum* isolates, 23522, 23468, 23389 and 23473, for Oxford Nanopore sequencing. 23468 and 23473 are 15ADON isolates from North Dakota and Minnesota respectively but group distantly in a principal component analysis of a larger set of isolates (Dhakal et al., 2021). 23522 is a 3ADON isolate from Minnesota and 23389 is an NX-2 isolate from New York. Oxford Nanopore sequencing produced a range of sequence data, from 4.68 to 10.14 Gb for the four *Fusarium* isolates (Table 4.1). This amount of raw sequence corresponds to from 123x to 267x expected coverage, based on the 38.04 Mb size of the *F. graminearum* PH-1 reference genome. Assembling the genomes using Canu gave chromosomal-level assemblies for our four isolates. For each of the four assemblies, the four largest contigs represent nearly complete chromosomes, as evidenced by the detection of telomere repeat sequence at the majority of contig ends and the finding that the total length of these four contigs

always exceeded 96.5% of the total length of the four chromosomes from the PH-1 reference assembly (Supplemental file 4.2). The region of ribosomal repeats in the subtelomeric region of chromosome 4 was not fully assembled into contigs in any of the Nanopore assemblies, and the major difference in contig length was due to the absence of most of this region in these assemblies (Figure 4.1, Supplemental file 4.2). We also obtained mitochondrial genomes for all four isolates, though in each case the assembled molecule was substantially larger than expected, and it appears the assembly process generated an artificial duplicated region. Each of the genome assemblies contained at least one contig that did not correspond to a full chromosome or mitochondrial genome, and the assemblies consisted of from 6 to 33 total contigs (Table 4.1). The contigs not corresponding to complete chromosomes from the nuclear genome or the mitochondrial genome generally had poor read support, usually with fewer than 100 reads. For example, three of the smallest contigs for 23468 were assembled with only 11, 1, and 1 reads. The one notable exception was a 4.7 Mb long circular contig with strong read support from the assembly of isolate 23389 that is 88% identical to the bacteria *Stenotrophomonas spp.* The *Stenotrophomonas* contig possibly reflects contamination of this isolate's culture with the bacteria during storage or culturing. The assemblies have an average and minimum BUSCO score of 97.3 and 96.2, respectively (Table 4.1).

4.4.2 An expanded reference genome resource

The high variation in gene content in *F. graminearum* isolates from the same or different populations (Kelly and Ward 2018) highlights the weakness of using only a single reference genome in genomics analyses such as read mapping. In this previous study, reads from sixty isolates that did not map to the PH-1 genome were *de novo* assembled into orphan contigs, which added an average of 705 kb of extra sequence to each isolate's genome. The relative genomic

location of these orphan contigs could not easily be determined using the PH-1 reference, but many contigs could be tentatively placed when compared with new reference genomes, especially ones from the same population. On average, over 486 kb from the orphan contigs could be mapped from each isolate to the 'best match' assembly (the assembly that maximized mapping for an isolate, Supplemental file 4.15). In particular, an average of 769 (and up to 962) kb from isolates assigned to STRUCTURE population NA2 mapped to the 23522 assembly, while an average of 519 (and up to 551) kb from isolates assigned to STRUCTURE population NA3 mapped to assembly 23473 (Supplemental file 4.15). For isolates belonging to the same population as the PH-1 isolate (NA1), that reference genome performed relatively better, but even here on average an additional 267 (and up to 395) kb of contigs could be mapped to another assembly (Supplemental file 4.15). The placement of the orphan contigs on the assemblies is also informative. The majority of them map to subtelomeric regions as predicted in the original study (Kelly and Ward 2018; Figures A.1 – A.10).

4.4.3 Detection of structural variation

In order to maximize our discovery of structural variants (SVs) segregating in *F. graminearum*, we included three publicly available chromosome-level assemblies in addition to the PH-1 reference genome in our analyses. Isolates FG-12, CML3066 and CS3005 have sequence matching the 15ADON genotype and were collected from China, Brazil, and Australia, respectively. Pairwise genome comparisons of the seven query assemblies with the PH-1 reference using the software SyRI identified both SNPs and SVs. Isolate FG-12 is unusual in terms of its high divergence from the other isolates in this study based on a neighbor joining tree constructed using SNPs (Figure A.1). FG-12 is also very different from the rest of the isolates in terms of SVs. This sample was isolated in China from maize seedling roots with root rot

symptoms. Isolate 23473 does not have nearly as many unique SNPs as FG-12, but it nevertheless groups with it in the tree and has one of the highest totals of SVs (in particular, inversions and translocations). In contrast, isolate 23468 is most similar to the reference isolate PH-1 in terms of SNPs and tends to have fewer SV differences from PH-1 relative to the other isolates.

Excluding insertions and deletions smaller than 1 kb in size and all deletions from FG-12 (see below), we identified a total of 568 SVs (Table 4.2, Figure 4.1). Inversions were fewer in number but were generally larger in size (Figure 4.2). Rearrangements were enriched towards the end of the chromosomes. When including SVs of all sizes, 32% of inversions, 58% of translocations, 40% of duplications, 38% of insertions and 35% of deletions originated within 1 Mb of the chromosomal ends in reference isolate PH-1.

Summing across each comparison of the 7 query genomes with PH-1, we identified a total of 87 inverted regions ranging from 156 bp to over 2.6 Mb (Table 4.2). Thirty-six inversions are unique to a single isolate, while another 16 are shared among multiple isolates, leading to 52 inversion-state differences when compared to PH-1. Twenty-five distinct inversions are over 10 kb in size, including half of the shared inversions. FG-12 has the most unique and total inversions, while 23473 has the second-most inversions. Eight inversions are common between 23522 and CML3066 and only one inversion each is shared between 23468 and isolates FG-12 and CS3005. The distribution of inversions per isolate ranged from 4 in 23468 to 29 in FG-12 (Table 4.2). Of all the inversions, 2, 4, 5, 2, 1 and 1 are unique in 23522, 23473, 23389, CML3066, 23468, and CS3005, respectively. Isolate FG-12 has 21 unique inversions, of which 9 are larger than 10 kb, including the largest 2,620,177 bp inversion in chromosome 1 (corresponding to 2,652,877 bp in the PH-1 reference). Inversion rearrangements

are sparse in chromosome 1 and 3 (a total of 20) but common in chromosome 2 and 4 (67 in total). Inversions made up 0.09% of the structural rearrangements but covered 9.38% of the size of the PH-1 genome.

There are a total of 159 translocations in the seven query genomes (Table 4.2). The translocations ranged from 104 bp to 376,520 bp in size, 100 of them being larger than 1 kb. Eighty-five of the translocations are inverted translocations. Sixty-nine translocations are within the same chromosome while 90 are between different chromosomes. Isolate 23473 has the most translocations, while CS3005 has the two largest translocations, 376,520 bp and 152,145 bp in size. The two regions are in chromosome 3 in CS3005 but in chromosomes 1 and 2, respectively, in PH-1. As with inversions, isolates share some translocations in common. However, while translocations identified in different isolates often overlap, the specific positions defining the translocation intervals rarely match, making the identification of translocations inherited from a common ancestor difficult. About a quarter (42) of the translocations originate within 20 kb and about half (93) within 1 Mb from the chromosomal ends in PH-1. Translocations comprised 0.17% of the structural rearrangements. Among the query isolate - PH-1 comparisons, translocations usually cover well under 1% of the size of the PH-1 genome, except for the comparison with CS3005, where translocations cover 1.65% of the genome size (Supplemental file 4.3).

SyRI identified a total of 58,489 insertions and 34,102 deletions (single base pair and larger) in the seven query genomes (Table 4.2). All deletions from FG-12 are excluded from the above count and other analyses as SyRI identified many deletions in FG-12 corresponding exactly to intronic sequences missing in this genome assembly, likely due to some technical error that appears to only affect this one type of SV. Single base pair events made up 57.9% of the

insertions and 52.6% of the total deletions. A total of 26,132 insertions and 16,930 one bp insertions are from a single isolate, FG-12. Insertions and deletions larger than 1 kb are much less common, with each isolate having 23 or fewer of each type (Figure 4.2, Table 4.2). The largest deletion, 22,908 base pairs in size in chromosome 2, is shared by isolates 23389 and 23473. This deletion removes the polyketide synthase (PKS6) gene from the fusaristatin A biosynthesis gene cluster and part of the non-ribosomal polyketide synthase (NRPS7) gene from the same cluster. Similarly, another deletion of 4,658 bp in isolate 23473 affects the PKS2 biosynthesis cluster on chromosome 4. Two genes from this cluster (FGRAMPH1_01G16115, FGRAMPH1_01G16117) are missing and another gene (FGRAMPH1_01G16119) from the same cluster is partially missing. Likewise, the largest insertion identified is 24,017 base pairs in size in chromosome 1 of isolate FG-12. Even though insertions and deletions made up 62% and 36% of the total count of structural rearrangements, they only covered up to 0.53% and 0.34% of the size of the PH-1 genome in any given PH-1-query isolate comparison (Supplemental file 4.3).

The region of ribosomal RNA repeats in the subtelomeric region of chromosome 4 was not completely assembled into contigs in any of the query isolates. Thus, in the SyRI comparisons of PH-1 to each query isolate, the partially assembled repeat regions are misidentified as duplications in PH-1 (Figure 4.1, blue region in the end of chromosome 4). Aside from these artifactual duplications, SyRI identified from 11 to 55 duplications in each query genome (Table 4.2).

As SVs already appeared enriched in subtelomeric regions, we checked the overlap of structural rearrangements with regions of high recombination (Laurent et al., 2018) that include regions at or near each chromosome end. There are a total of 12 such regions, 4 in chromosome

1, 3 each in chromosomes 2 and 4, and 2 in chromosome 3, and these regions cover 29.55% of the genome. Nearly all types of SVs are enriched in these regions of high recombination, as a total of 40.22% of inversions, 44.65% of translocations, 53.46% of insertions (>50 bp), 50.63% of deletions (>50 bp in size, excluding FG-12), and 23.67% of the duplications fall within the regions (Figure 4.1). The proposed ancient chromosomal fusion sites overlap and extend beyond the regions of high recombination rate in this fungus (King et al., 2015; Ma et al., 2010). It is likely that the regions of high recombination extend beyond the current boundaries, as markers used for the construction of the genetic map used to define these regions did not cover the chromosomal ends. If this were true, far more SVs would likely fall in the high recombination regions of the genomes.

4.4.4 Transposable elements and repeat content

Fusarium graminearum isolates differ in terms of the types and total number of transposable elements present in the genome (Table 4.3). The repeat content in the eight *F. graminearum* genomes included in this study ranged from 0.98% in PH-1 to 1.98% in 23473 (Table 4.3). The EDTA pipeline did not identify gypsy retrotransposons in 23473, PH-1 and CS3005. Tc1_mariner transposons are present in all genomes except FG-12, and the Hat type of DNA-transposon is only present in 23522, 23473, 23468 and CS3005. When present, gypsy retrotransposons are present in multiple copies, with 23522 and CML3066 having 114 and 56 copies, respectively. Isolates 23389 and FG-12 have 3 copies and 23468 has 4 copies. A few of the gypsy retrotransposons are intact in the genomes: 3 in 23389, 4 each in 23468 and 23522 and 2 in CML3066 (Supplemental file 4.14). Three classes of DNA transposon: CACTA, mutator and Helitron, are present in multiple copies in all genomes. All of the mutator DNA transposons in all genomes were identified as structurally intact. Most of the Tc1 mariner (except one in PH-

1) and multiple copies of DNA transposons are intact. A total of 7.28% of insertions, 7.11% of deletions (excluding deletions in FG-12) over 50 bp in size, 38.61% of inversions, 63.86% of translocations and 57.69% of the duplications overlap with the repeat content of the genome (Figure 4.3, Figures A.2 - A.6) such that repeat elements are shaping the landscape of SVs in *F. graminearum* genomes.

4.5 Discussion

In this study we sequenced four representative isolates from *Fusarium graminearum* populations using Oxford Nanopore sequencing and obtained chromosomal-level assemblies. These four genomes and four publicly available genomes, including the established PH-1 reference, were used for the analysis of SVs with the objective of identifying their distribution and role in shaping host-pathogen interactions and genome evolution. We detected a total of 87 inversions, 159 translocations, 245 duplications, 58,489 insertions and 34,102 deletions. The general features, distribution across the genomes, predictors and potential roles of SVs in adaptation and evolution are discussed.

Low GC content such as in the regions of short sequence repeats and inactivated transposons, low gene content, recombination rate and the presence of transposable elements are the major predictors of structural rearrangements in *Arabidopsis* and *Zymoseptoria tritici* (Badet et al., 2021). Double-stranded breaks during meiosis leading to the intra-chromosomal or inter-chromosomal rearrangement of segments and the movement of DNA segments by transposons are the major sources of large-scale chromosomal variation (Desjardins, 2006; Mehrabi et al., 2017; Priest et al., 2020; Stukenbrock and Croll, 2014). Other sources include unequal crossover in repeat regions and double-stranded breaks during the mitotic cycle. All of these factors likely serve as predictors of structural rearrangements in *F. graminearum* as well. Analysis of the

genomes of *Z. tritici* progeny from experimental crosses also revealed that meiotic crossovers, and presumably the double-strand chromosomal breaks which can resolve as crossovers, are hotspots of novel SVs (Badet et al., 2021). At a very broad scale along chromosomes, recombination also seems to be a primary driver of structural rearrangements in *F. graminearum* genomes. The *F. graminearum* genome has regions of widely different average recombination rates, a 'two-speed' recombination landscape that is associated with differences in nucleotide diversity and density of host-pathogen interaction genes (Laurent et al., 2018). High recombination regions are characterized by high nucleotide diversity and enriched with genes predicted to code for secreted effectors or have a role in secondary metabolite synthesis. Along chromosomes, meiotic crossovers show a positive correlation with gene density and negative correlation with GC content (Laurent et al., 2018). Genes reported to show host-specific expression are enriched by 2-fold in the high recombination regions, where high recombination likely shuffles the genes with roles in plant-microbe interaction and secondary metabolite synthesis, promoting rapid evolution (Ma et al., 2010). Our results show that recombination is not only associated with high nucleotide diversity and HDRs but also associated with large-scale SVs (Figures 4.3, A.2-A.6). A total of 40.22% of inversions, 44.65% of translocations, 53.46% of large insertions (>50 bp in size), 50.63% of large deletions, and 23.67% of the duplications fall within the regions of high recombination (Talas et al., 2016), which are also sites of ancient chromosomal fusions in the ancestor to *F. graminearum*. The orphan contigs also tend to map to these regions of high recombination (Figures 4.3, A.2-A.6). This makes recombination critically important for adaptation in this fungus. *Fusarium graminearum* is a homothallic fungus, but field isolates undergo frequent outcrossing in nature, and on the hypothesis that recombination is

mutagenic, recombination occurring during both selfing and outcrossing can help drive the evolution of this fungus.

Transposable elements mediate mutations and chromosomal rearrangements, and are key in shaping genome architecture and gene content in many fungal genomes (Stukenbrock and Croll, 2014). The repeat content in *F. graminearum* genomes is very low (0.98% to 1.98%) compared to 24% in *F. oxysporum* (Ma et al., 2010), 44.8% reported in different *Colletotrichum* species (Rao et al., 2018) and 11% in *Magnaporthe oryzae* (Chiapello et al., 2015). This low repeat content is likely related to repeat induced point mutation (RIP), a process in *F. graminearum* and some other fungi that suppresses duplication and transposable element-mediated genome expansion (Cuomo et al., 2007). Both intact and partial remains of transposable elements were identified in this study. While 100% of the mutator, 90% of Tc1-mariner, 59% of CACTA and 90% of helitron transposable elements identified are intact, only a few (7%) of the Gypsy transposons are intact (Supplemental file 4.14). Only four out of 114 Gypsy elements in isolate 23522 and two out of 56 in CML3066 are intact, suggesting RIP is rapidly mutating gypsy transposons in the genome (Supplemental file 4.14). A significant proportion of inversions (38.61%), translocations (63.86%), and duplications (57.69%) overlap with the repeat content of the genome, indicating the role for TEs in structural rearrangements. The finding of both high SNP density and an enrichment of SVs in regions of high recombination and high repeat density suggests that recombination and transposable elements are associated with the origin of small and large-scale genomic variants in *F. graminearum*.

SVs in *F. graminearum* genomes have numerous consequences for gene content, their chromosomal locations, and their coding sequences. A large proportion of the rearranged segments include genic regions annotated in the PH-1 genome. A total of 40.52% of insertions

(>50 bp), 47.58% of deletions (>50 bp), 71.26% of inversions, 15.91% of duplications and 21.38% of the translocations either include at least one gene inside the rearranged segment or have at least one of the breakpoints overlapping a gene (Supplemental files 4.8-4.11). In particular, translocations have helped in shuffling the location of genes coding for effector proteins and have thus helped in the evolution of interactions in this fungus with its host. Effectors are small, often secreted molecules that modulate plant-microbe interactions. Recognition of effectors by plant resistance genes triggers an immune response, so plant pathogens deploy several strategies to avoid recognition. One of the mechanisms is to lose or mutate recognized effector proteins. We identified some effector-rich regions being moved around the genome through translocation events. Two chromosomal regions in isolate CS3005 and one region in FG-12 containing genes coding for predicted secreted effectors are shuffled relative to the reference PH-1 genome by translocation events. A 372 kb region present in chromosome 1 in PH-1 and in chromosome 3 in CS3005 contains 5 predicted effector genes (Supplemental files 4.4 and 4.10). One of the predicted effector genes in this region codes for KP4 killer toxin, which is expressed during wheat seedling infection, and similar proteins are necessary for aggressiveness during head infection (Lu and Faris, 2019). Similarly, another 152 kb region in chromosome 3 in CS3005 is present in chromosome 2 in PH-1 and contains five predicted effector genes (Supplemental files 4.4 and 4.10). A smaller 4.5 kb region in isolate FG-12 is located in a different region within chromosome 1 relative to the reference genome (Supplemental file 4.5). Shuffling of effector-rich regions could change the expression profile of the effector proteins and facilitate effector gain and loss. Recombination of strains having identical effectors in two separate genomic locations can result in progeny having multiple or no copies of the gene.

SVs can also directly create presence-absence or copy number variation in fungal effector proteins and thus have a role in shaping host specialization and plant-microbe interactions in many patho-systems (Hou et al., 2023; Wang et al., 2020). We identified that some predicted effector genes have been lost, which could have been facilitated by translocations as discussed earlier. Deleted segments (>50 bp) in six query isolates partially or completely overlap a total of 229 genes in the PH-1 genome (Supplemental file 4.12). Of these genes, 27 are entirely missing in query isolates (Supplemental file 4.12). Some missing genes are common between isolates. For example, the gene FGRAMPH1_01G08535, which codes for a predicted secreted effector, a hydrolase enzyme, is missing in isolates 23389 and 23473 (Supplemental file 4.12; (Brown et al., 2012). Abundant insertions relative to PH-1 were detected in our analyses, and BLAST searches of these inserted sequences against the orphan protein database from Kelly and Ward (2018) detected predicted effector proteins. For example, a shared insertion common to the isolates 23389, 23473, 23522 and FG-12, likely consists of a cytochrome P450, a predicted secreted protein and a kinase-like protein. Similarly, a 3 kb insertion in isolate 23468 (23468_666) likely is a cluster of secreted proteins as it consists of predicted secreted proteins (identified by SignalP), pectate lyase and a lysin domain protein. Insertions consisting of only lysin motif (lysM) containing proteins were also identified in isolates CML3066 (CML3066_888) and CS3005 (CS3005_889). LysMs are about 50 amino acids long and bind to peptidoglycan, chitin and their derivatives and can help the fungus to evade plant defense responses by binding to the chitin fragments released during enzymatic digestion of plant cell wall during the entry of the fungus into the host cells and/or protecting the fungal cell wall against hydrolytic enzymes (Dubey et al., 2020). Similarly, pectin lyase is an enzyme for degrading pectin in the plant cell wall. In addition, an insertion in isolate 23473 (23473_831) includes a BLAST hit that covers

45% of an orphan protein, and the amino acid sequence of that orphan protein is a 29% match to the stocks4 gene. The stocks4 gene was predicted to produce a secreted protein with high confidence by SignalP. Besides complete loss of genes in some isolates, smaller deletions fall within coding sequences of effectors predicted by Brown et al. (2012) and Rampitsch et al. (2013) and can have important functional consequences as they can make a nonfunctional protein product or help the pathogen evade recognition by its host (Hou et al., 2023). Genes FGRAMPH1_01G00015, FGRAMPH1_01G03877, FGRAMPH1_01G08677 in chromosome 1, FGRAMPH1_01G11565, FGRAMPH1_01G12735, FGRAMPH1_01G14149 in chromosome 2 and FGRAMPH1_01G16399 in chromosome 3 have parts of coding sequences missing compared to PH-1 (Supplemental file 4.12). Our finding of abundant presence-absence variation further supports the view that structural variation is an essential component of plant-microbe evolution.

Secondary metabolite synthesis clusters are highly diverse in fungi, and this diversity is due to the loss of the clusters/genes mediated by complex structural rearrangements or through gain of clusters via horizontal gene transfer (Massonnet et al., 2018; Tralamazza et al., 2019). The fusaristatin A and PKS2 biosynthesis gene clusters are predicted to be nonfunctional in some isolates. The fusaristatin A gene cluster is partially deleted in *F. graminearum* isolates 23389 and 23473 and falls within the largest deletion observed in this study, a 22,098 bp deletion in chromosome 2 (Supplemental files 4.7 and 4.11). Among the two genes in the cluster, the polyketide synthase (PKS6) is absent, and the non-ribosomal peptide synthase (NRPS7) is truncated in these two isolates. Natural populations of *F. pseudograminearum* having similar deletion (PKS6 missing and NRPS7 truncated) in this biosynthetic gene cluster exist in western Australia and are more aggressive pathogens causing crown rot of wheat (Khudhair et al., 2020;

Wollenberg et al., 2019). The experimental deletion of PKS6 and the disruption of NRPS7 in two isolates of *F. pseudograminearum* increased the growth rate and aggressiveness of the mutants in causing crown rot of wheat (Khudhair et al., 2020). The role of PKS6 in the *F. graminearum* reference isolate PH-1 has been previously investigated by the generation of a gene disruption mutant. Disruption of PKS6 in the PH-1 background significantly increased growth rate but did not significantly change pathogenicity towards a susceptible wheat cultivar (Gaffoor et al., 2005). The difference in pathogenicity test results among the two studies could be due to the difference in methods, cultivars used and genetic backgrounds of the isolates. (Gaffoor et al., 2005) only disrupted the PKS6 gene of the cluster while (Khudhair et al., 2020) completely deleted PKS6 and disrupted NRPS7. It is also possible that the cluster is important for crown rot disease and not FHB. The fusaristatin A biosynthesis genes in *F. graminearum* (PH-1) and *F. pseudograminearum* (CS5834) share 96.2% nucleotide identity, and the role of the loss of this gene cluster in the genetic background of more *F. graminearum* isolates should be investigated. Similarly, two of the 18 genes of the PKS2 polyketide synthesis gene cluster were fully deleted and one gene was partially deleted in isolate 23473 (del 4110 in Supplemental file 4.5). The deleted genes, FGRAMPH1_01G16115 and FGRAMPH1_01G16117, are RTA1-like protein and RTA1 domain containing protein, and the truncated gene FGRAMPH1_01T16119 is a c6 transcription factor (Zhao et al., 2014). Deletions in the gene cluster can have important functional consequences, as strains with a disruption in the PKS2 gene (FGRAMPH1_01G16085) of this gene cluster have a reduced growth phenotype (Gaffoor et al., 2005). Our results suggest that structural rearrangements are likely playing a major role in introducing presence-absence variation of secondary metabolite cluster genes in *F. graminearum.*

Gain and loss of chromosomal segments also have introduced presence-absence variation in other types of genes with potential roles in fitness in *F. graminearum,* with important implications. We identified presence-absence variation in a potential chromatin remodeling protein and genes related to trichothecene synthesis, transcription regulation, and transportation in *F. graminearum* isolates (Supplemental file 4.12). Many of these genes may have important roles to play based on their predicted function, but, whether many of these genes play roles in pathogen fitness and host-plant interaction remains unknown. Among the few genes affected by a deletion with a known role in pathogenic fitness is the gene FGRAMPH1_01G07245 (aspartate--tRNA mitochondrial). A deletion falls within the coding sequence of the gene in isolates 23522 and CML3066. Complete gene deletion mutants for the gene exhibit reduced vegetative spore production and an increased aggressiveness phenotype (FungiDB, accessed July 6th 2023). Similarly, a 7,062 bp long deletion in isolate 23473 contains three genes, FGRAMPH1_01G05687, FGRAMPH1_01G05689, and FGRAMPH1_01G05691. Gene FGRAMPH1_01G05687 codes for a trichodiene oxygenase cytochrome P450 and is an important component of plasma membrane, predicted to have iron binding and oxidoreductase functions. FGRAMPH1_01G05687 (FGSG_02367) is predicted to interact directly with *TRI4* and *TRI11* and have a role in trichothecene biosynthesis (Wang et al., 2022). The *TRI4* and *TRI11* genes are involved in early steps of trichothecene synthesis. Likewise, gene FGRAMPH1_01G05689 (FGSG_02368) is a transcription factor regulated by *TRI6*, the main regulator of the expression of *TRI* genes and therefore trichothecene biosynthesis (Shostak et al., 2020). Trichothecene biosynthesis thus could be disrupted in 23473, an isolate with a 15ADON genotype that is otherwise very different from most other 15ADON isolates. It forms a separate clade with FG-12 in the neighbor joining tree (Figure A.1) and does not cluster with most

15ADON isolates in a PCA analysis of a much larger sample of isolates (Dhakal et al., 2021). The loss of transcription factor FGRAMPH1_01G05689 could also affect some of the several processes that *TRI6* is known to regulate, such as cellular metabolism, transport, signal transduction and gene expression regulation.

The identified insertions relative to PH-1 have the potential to code for proteins novel to this species. A total of 46 insertions have translated nucleotide sequences matching one or more sequences in the orphan protein database of Kelly and Ward (2018) with at least 50% coverage and 40% identity. Based on BLAST searches of the orphan proteins the inserted segments matched, insertions likely include coding sequences for proteins like SNF2 helicase-like protein, arylsulphatase, kinesin light chain protein, and ankyrin repeat containing proteins. SNF2 helicase-like proteins have a role in the mechanical remodeling of chromatin structures by moving the nucleosomes around (Ryan and Owen-Hughes, 2011), while arylsulphatases are responsible for sulfur scavenging and influence fungal growth and morphology (Korban et al., 2017). Kinesin is a component of the microtubule cytoskeleton. Although kinesin light chains are absent in some fungi (*e.g*., *Neurospora*) (Seiler et al., 2000), in humans and higher eukaryotes they are known to play an important role in attaching cargo to kinesin-1 protein (Woźniak and Allan, 2006). Kinesin-1 plays a role in moving organelle cargo (endoplasmic reticulum, Golgi apparatus, secretory vesicles, mitochondria and other cellular organelles) in the cellular space (Wozniak et al., 2004). While many accessory genes with presence-absence polymorphisms in *F. graminearum* might not boost fitness in natural environments, some subset may provide a fitness advantage, at least in some environments, and could currently be increasing in frequency towards fixation. These genes that may confer higher fitness could serve important roles in survival, growth, reproduction, fungicide resistance and aggressiveness.

The non-homologous end joining DNA repair pathway, which is used to repair double stranded breaks in the chromosomes, is error prone. When chromosomal rearrangement breakpoints overlap genes, altered and potentially nonfunctional gene products can result in important functional consequences. One end of an inversion in isolate FG-12 falls within coding sequence of the Ecm22 gene, which binds sterol regulatory elements (FGRAMPH-1_01T14981) (Supplemental file 4.9). In the fungus *Saccharomyces cerevisiae*, transcription factor Ecm22p regulates expression of the sterol biosynthesis genes, ERG2 and ERG3 (Vik and Rine, 2001). Ecm22p has overlapping function with the gene UPC2, as they both regulate ERG2 expression (Vik and Rine, 2001). Thus, the disruption of FGRAMPH1_01T14981 in FG-12 is likely affecting sterol biosynthesis unless overcome by the presence of a UPC2 homolog in this strain of *F. graminearum*. Likewise, genes FGRAMPH1_01G00003, FGRAMPH1_01G08549, FGRAMPH1_01G08781, FGRAMPH1_01G13855, FGRAMPH1_01G25033, FGRAMPH1_01G26371, and FGRAMPH1_01G28289 in isolates, FG-12, CS3005, 23389, CS3005, FG-12, CS3005 and CML3066, respectively, have at least one translocation breakpoint within the coding sequence of the gene, and these genes are predicted to function in carbohydrate transport, DNA binding, hydrolase activity, heme binding, pectin esterase activity, nucleic acid binding and protein binding, respectively (Supplemental file 4.11).

SVs have not been systematically studied in many plant pathogens. In this study we identified large SVs in *F. graminearum* genomes, many of which are located towards the chromosome ends and overlap regions of high recombination. Transposable elements and repeat elements are also enriched in the regions of high recombination. We identified presence-absence polymorphisms in biosynthesis gene clusters, predicted effector genes and genes associated with fitness advantages and plant-microbe interaction. We also identified that genomic regions with

roles in plant-microbe interactions (predicted effectors) are shuffled in different genomic regions by structural rearrangements. In addition, overlaps of rearrangement breakpoints with coding sequences were identified. To summarize, SVs in *F. graminearum* have an important role to play in shaping pathogen-host interactions and broader evolution through genome reorganization, the introduction of presence-absence polymorphisms, and changing gene content, protein products and gene regulation.

4.6 Data availability

Data for this study were submitted to NCBI under BioProject ID PRJNA989623. This includes the four chromosome-level genome assemblies, the associated BioSample information, and the raw reads from MinION and Illumina sequencing that were submitted to the NCBI Sequence Read Archive (SRA).

4.7 References

Alouane, T., Rimbert, H., Bormann, J., González-montiel, G.A., Loesgen, S., Schäfer, W., Freitag, M., Langin, T., Bonhomme, L., 2021. Comparative genomics of eight *Fusarium graminearum* strains with contrasting aggressiveness reveals an expanded open pangenome and extended effector content signatures. Int. J. Mol. Sci. 22, 6257. https://doi.org/10.3390/ijms22126257

Badet, T., Fouché, S., Hartmann, F.E., Zala, M., Croll, D., 2021. Machine-learning predicts genomic determinants of meiosis-driven structural variation in a eukaryotic pathogen. Nat. Commun. 12, 3551. https://doi.org/10.1038/s41467-021-23862-x

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Brown, N.A., Antoniw, J., Hammond-Kosack, K.E., 2012. The predicted secretome of the plant pathogenic fungus *Fusarium graminearum*: A refined comparative analysis. PLoS One 7, e33731. https://doi.org/10.1371/journal.pone.0033731

Chiapello, H., Mallet, L., Guérin, C., Aguileta, G., Amselem, J., Kroj, T., Ortega-Abboud, E., Lebrun, M.H., Henrissat, B., Gendrault, A., Rodolphe, F., Tharreau, D., Fournier, E., 2015. Deciphering genome content and evolutionary relationships of isolates from the fungus *Magnaporthe oryzae* attacking different host plants. Genome Biol. Evol. 7, 2796–2912. https://doi.org/10.1093/gbe/evv187

Cuomo, C.A., Güldener, U., Xu, J.R., Trail, F., Turgeon, B.G., Di Pietro, A., Walton, J.D., Ma, L.J., Baker, S.E., Rep, M., Adam, G., Antoniw, J., Baldwin, T., Calvo, S., Chang, Y.L., DeCaprio, D., Gale, L.R., Gnerre, S., Goswami, R.S., Hammond-Kosack, K., Harris, L.J., Hilburn, K., Kennell, J.C., Kroken, S., Magnuson, J.K., Mannhaupt, G., Mauceli, E., Mewes, H.W., Mitterbauer, R., Muehlbauer, G., Münsterkötter, M., Nelson, D., O'Donnell, K., Ouellet, T., Qi, W., Quesneville, H., Roncero, M.I.G., Seong, K.Y., Tetko, I. V., Urban, M., Waalwijk, C., Ward, T.J., Yao, J., Birren, B.W., Kistler, H.C., 2007. The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. Science (80-.). 317, 1400–1402. https://doi.org/10.1126/science.1143708

Desjardins, A.E., 2006. Fusarium mycotoxins: chemistry, genetics, and biology., Fusarium mycotoxins: chemistry, genetics and biology.

Dhakal, U., Leslie, J.F., Toomajian, C., 2021. Analysis of population structure and differentiation, linkage disequilibrium decay, and demographic history of *Fusarium*

graminearum from the US. Phytopathology 111, 85–85. https://doi.org/10.1016/j.tfp.2021.100103

Dubey, M., Vélëz, H., Broberg, M., Jensen, D.F., Karlsson, M., 2020. LysM Proteins Regulate Fungal Development and Contribute to Hyphal Protection and Biocontrol Traits in *Clonostachys rosea*. Front. Microbiol. 11, 679. https://doi.org/10.3389/fmicb.2020.00679

Ellinghaus, D., Kurtz, S., Willhoeft, U., 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics 9, 1–14. https://doi.org/10.1186/1471-2105-9-18

Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., Smit, A.F., 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. U. S. A. 117, 9451–9457. https://doi.org/10.1073/pnas.1921046117

Gaffoor, I., Brown, D.W., Plattner, R., Proctor, R.H., Qi, W., Trail, F., 2005. Functional analysis of the polyketide synthase genes in the filamentous fungus *Gibberella zeae* (Anamorph *Fusarium graminearum*). Eukaryot. Cell 4, 1926–1933. https://doi.org/10.1128/EC.4.11.1926-1933.2005

Gale, L.R., Ward, T.J., Balmas, V., Kistler, H.C., 2007. Population subdivision of *Fusarium graminearum* sensu stricto in the upper midwestern United States. Phytopathology 97, 1434–1439. https://doi.org/10.1094/PHYTO-97-11-1434

Goel, M., Sun, H., Jiao, W.B., Schneeberger, K., 2019. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. Genome Biol. 20, 1–13. https://doi.org/10.1186/s13059-019-1911-0

Goswami, R.S., Kistler, H.C., 2004. Heading for disaster: *Fusarium graminearum* on cereal crops. Mol. Plant Pathol. 5, 515–525. https://doi.org/10.1111/J.1364-3703.2004.00252.X

Hou, X., He, Z., Che, Z., Li, H., Tan, X., Wang, Q., 2023. Molecular mechanisms of *Phytophthora sojae* avirulence effectors escaping host recognition. Front. Microbiol. 13, 1111774. https://doi.org/10.3389/fmicb.2022.1111774

Jones, L., Riaz, S., Morales-Cruz, A., Amrine, K.C.H., McGuire, B., Gubler, W.D., Walker, M.A., Cantu, D., 2014. Adaptive genomic structural variation in the grape powdery mildew pathogen, *Erysiphe necator*. BMC Genomics 15, 1–8. https://doi.org/10.1186/1471-2164-15-1081

Kelly, A.C., Ward, T.J., 2018. Population genomics of *Fusarium graminearum* reveals signatures of divergent evolution within a major cereal pathogen. PLoS One 13, e0194616. https://doi.org/10.1371/journal.pone.0194616

Khudhair, M., Kazan, K., Thatcher, L.F., Obanor, F., Rusu, A., Sørensen, J.L., Wollenberg, R.D., McKay, A., Giblot-Ducray, D., Simpfendorfer, S., Aitken, E., Gardiner, D.M., 2020. Fusaristatin A production negatively affects the growth and aggressiveness of the wheat pathogen *Fusarium pseudograminearum*. Fungal Genet. Biol. 136, 103314. https://doi.org/10.1016/j.fgb.2019.103314

King, R., Urban, M., Hammond-Kosack, M.C.U., Hassani-Pak, K., Hammond-Kosack, K.E., 2015. The completed genome sequence of the pathogenic ascomycete fungus *Fusarium graminearum*. BMC Genomics 16, 1–21. https://doi.org/10.1186/s12864-015-1756-1

Korban, S.A., Bobrov, K.S., Maynskova, M.A., Naryzhny, S.N., Vlasova, O.L., Eneyskaya, E. V., Kulminskaya, A.A., 2017. Heterologous expression in *Pichia pastoris* and biochemical characterization of the unmodified sulfatase from *Fusarium proliferatum* LE1. Protein Eng. Des. Sel. 30, 477–488. https://doi.org/10.1093/protein/gzx033

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., Phillippy, A.M., 2017. Canu: Scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. Genome Res. 27, 722–736. https://doi.org/10.1101/gr.215087.116

Kuhnem, P.R., Spolti, P., Del Ponte, E.M., Cummings, J.A., Bergstrom, G.C., 2015. Trichothecene genotype composition of *Fusarium graminearum* not differentiated among isolates from maize stubble, maize ears, wheat spikes, and the atmosphere in New York. Phytopathology 105, 695–699. https://doi.org/10.1094/PHYTO-10-14-0266-R

Langner, T., Harant, A., Gomez-Luciano, L.B., Shrestha, R.K., Malmgren, A., Latorre, S.M., Burbano, H.A., Win, J., Kamoun, S., 2021. Genomic rearrangements generate hypervariable mini-chromosomes in host-specific isolates of the blast fungus. PLoS Genet. 17, e1009386. https://doi.org/10.1371/JOURNAL.PGEN.1009386

Laurent, B., Palaiokostas, C., Spataro, C., Moinard, M., Zehraoui, E., Houston, R.D., Foulongne-Oriol, M., 2018. High-resolution mapping of the recombination landscape of the phytopathogen *Fusarium graminearum* suggests two-speed genome evolution. Mol. Plant Pathol. 19, 341–354. https://doi.org/10.1111/mpp.12524

Li, H., 2018. Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Loman, N.J., Quick, J., Simpson, J.T., 2015. A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat. Methods 12, 733–735. https://doi.org/10.1038/nmeth.3444

Lu, S., Faris, J.D., 2019. *Fusarium graminearum* KP4-like proteins possess root growth-inhibiting activity against wheat and potentially contribute to fungal virulence in seedling rot. Fungal Genet. Biol. 123, 1–13. https://doi.org/10.1016/j.fgb.2018.11.002

Ma, L.J., Van Der Does, H.C., Borkovich, K.A., Coleman, J.J., Daboussi, M.J., Di Pietro, A., Dufresne, M., Freitag, M., Grabherr, M., Henrissat, B., Houterman, P.M., Kang, S., Shim, W.B., Woloshuk, C., Xie, X., Xu, J.R., Antoniw, J., Baker, S.E., Bluhm, B.H., Breakspear, A., Brown, D.W., Butchko, R.A.E., Chapman, S., Coulson, R., Coutinho, P.M., Danchin, E.G.J., Diener, A., Gale, L.R., Gardiner, D.M., Goff, S., Hammond-Kosack, K.E., Hilburn, K., Hua-Van, A., Jonkers, W., Kazan, K., Kodira, C.D., Koehrsen, M., Kumar, L., Lee, Y.H., Li, L., Manners, J.M., Miranda-Saavedra, D., Mukherjee, M., Park, G., Park, J., Park, S.Y., Proctor, R.H., Regev, A., Ruiz-Roldan, M.C., Sain, D., Sakthikumar, S., Sykes, S., Schwartz, D.C., Turgeon, B.G., Wapinski, I., Yoder, O., Young, S., Zeng, Q., Zhou, S., Galagan, J., Cuomo, C.A., Kistler, H.C., Rep, M., 2010. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. Nature 464, 367–373. https://doi.org/10.1038/nature08850

Machado, A.., King, R., Urban, M., Del Ponte, E.M., 2018. Analysis of *Fusarium graminearum* pangenome, in: Abstracts of Presentations at ICPP2018. p. S1.110.

Marçais, G., Delcher, A.L., Phillippy, A.M., Coston, R., Salzberg, S.L., Zimin, A., 2018. MUMmer4: A fast and versatile genome alignment system. PLoS Comput. Biol. 14, e1005944. https://doi.org/10.1371/journal.pcbi.1005944

Massonnet, M., Morales-Cruz, A., Minio, A., Figueroa-Balderas, R., Lawrence, D.P., Travadon, R., Rolshausen, P.E., Baumgartner, K., Cantu, D., 2018. Whole-genome resequencing and pan-transcriptome reconstruction highlight the impact of genomic structural variation on secondary metabolite gene clusters in the grapevine esca pathogen *Phaeoacremonium minimum*. Front. Microbiol. 9, 1784. https://doi.org/10.3389/fmicb.2018.01784

McMullen, M., Bergstrom, G., De Wolf, E., Dill-Macky, R., Hershman, D., Shaner, G., Van Sanford, D., 2012. A unified effort to fight an enemy of wheat and barley: Fusarium head blight. Plant Dis. 96, 1712–1728. https://doi.org/10.1094/PDIS-03-12-0291-FE

Mehrabi, R., Mirzadi Gohari, A., Kema, G.H.J., 2017. Karyotype Variability in Plant-Pathogenic Fungi. Annu. Rev. Phytopathol. 55, 483–503. https://doi.org/10.1146/annurev-phyto-080615-095928

Ou, S., Jiang, N., 2018. LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. Plant Physiol. 176. https://doi.org/10.1104/pp.17.01310

Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J.R.A., Hellinga, A.J., Lugo, C.S.B., Elliott, T.A., Ware, D., Peterson, T., Jiang, N., Hirsch, C.N., Hufford, M.B., 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. 20, 1–18. https://doi.org/10.1186/s13059-019-1905-y

Paradis, E., Schliep, K., 2019. Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35, 526–528. https://doi.org/10.1093/bioinformatics/bty633

Priest, S.J., Yadav, V., Heitman, J., 2020. Advances in understanding the evolution of fungal genome architecture. F1000Research 9, 776. https://doi.org/10.12688/f1000research.25424.1

Rampitsch, C., Day, J., Subramaniam, R., Walkowiak, S., 2013. Comparative secretome analysis of *Fusarium graminearum* and two of its non-pathogenic mutants upon deoxynivalenol induction in vitro. Proteomics 13, 1913–1921. https://doi.org/10.1002/pmic.201200446

Rao, S., Sharda, S., Oddi, V., Nandineni, M.R., 2018. The landscape of repetitive elements in the refined genome of chilli anthracnose fungus *Colletotrichum truncatum*. Front. Microbiol. 9, 2367. https://doi.org/10.3389/fmicb.2018.02367

Ryan, D.P., Owen-Hughes, T., 2011. Snf2-family proteins: Chromatin remodellers for any occasion. Curr. OpiRyan, D. P., T. Owen-Hughes, 2011 Snf2-family proteins Chromatin Remodel. any Occas. Curr. Opin. Chem. Biol. 15.nion Chem. Biol. https://doi.org/10.1016/j.cbpa.2011.07.022

Schwessinger, B., Mcdonald, M., 2017. High quality DNA from Fungi for long read sequencing e.g. PacBio, Nanopore MinION v4. Protocols.Io 1–6.

Seiler, S., Kirchner, J., Horn, C., Kallipolitou, A., Woehlke, G., Schliwa, M., 2000. Cargo binding and regulatory sites in the tail of fungal conventional kinesin. Nat. Cell Biol. 2, 333–338. https://doi.org/10.1038/35014022

Shi, J., Liang, C., 2019. Generic repeat finder: A high-sensitivity tool for genome-wide de novo repeat detection. Plant Physiol. 180, 1803–1815. https://doi.org/10.1104/pp.19.00386

Shostak, K., Bonner, C., Sproule, A., Thapa, I., Shields, S.W.J., Blackwell, B., Vierula, J., Overy, D., Subramaniam, R., 2020. Activation of biosynthetic gene clusters by the global transcriptional regulator TRI6 in *Fusarium graminearum* 8, 664–680. https://doi.org/10.1111/mmi.14575

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E. V., Zdobnov, E.M., 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Stovner, E.B., Sætrom, P., 2020. PyRanges: Efficient comparison of genomic intervals in Python. Bioinformatics 36, 918–919. https://doi.org/10.1093/bioinformatics/btz615

Stukenbrock, E.H., Croll, D., 2014. The evolving fungal genome. Fungal Biol. Rev. 28, 1–12. https://doi.org/10.1016/j.fbr.2014.02.001

Su, W., Gu, X., Peterson, T., 2019. TIR-Learner, a New Ensemble Method for TIR Transposable Element Annotation, Provides Evidence for Abundant New Transposable Elements in the Maize Genome. Mol. Plant 12, 447–460. https://doi.org/10.1016/j.molp.2019.02.008

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. https://doi.org/10.1094/MPMI-09-15-0218-R

Teufel, F., Almagro Armenteros, J.J., Johansen, A.R., Gíslason, M.H., Pihl, S.I., Tsirigos, K.D., Winther, O., Brunak, S., von Heijne, G., Nielsen, H., 2022. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat. Biotechnol. 40, 1023–1025. https://doi.org/10.1038/s41587-021-01156-3

Tralamazza, S.M., Rocha, L.O., Oggenfuss, U., Corrêa, B., Croll, D., Rose, L., 2019. Complex Evolutionary Origins of Specialized Metabolite Gene Cluster Diversity among the Plant Pathogenic Fungi of the *Fusarium graminearum* Species Complex. Genome Biol. Evol. 11, 3106–3122. https://doi.org/10.1093/gbe/evz225

Vik, Å., Rine, J., 2001. Upc2p and Ecm22p, Dual Regulators of Sterol Biosynthesis in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 21, 6395–6405. https://doi.org/10.1128/mcb.21.19.6395-6405.2001

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., Earl, A.M., 2014. Pilon: An integrated tool for

comprehensive microbial variant detection and genome assembly improvement. PLoS One 9, e112963. https://doi.org/10.1371/journal.pone.0112963

Walkowiak, S., Bonner, C.T., Wang, L., Blackwell, B., Rowland, O., Subramaniam, R., 2015. Intraspecies interaction of *Fusarium graminearum* contributes to reduced toxin production and virulence. Mol. Plant-Microbe Interact. 28, 1256–1267. https://doi.org/10.1094/MPMI-06-15-0120-R

Walkowiak, S., Rowland, O., Rodrigue, N., Subramaniam, R., 2016. Whole genome sequencing and comparative genomics of closely related Fusarium Head Blight fungi: *Fusarium graminearum*, *F. meridionale* and *F. asiaticum*. BMC Genomics 17, 1–15. https://doi.org/10.1186/s12864-016-3371-1

Wang, J., Lin, S., Zhu, G., Zhao, Z., Wang, H., Zhou, M., Zhao, X., Wu, A., 2022. Identification of Candidate Genes Associated with Trichothecene Biosynthesis in *Fusarium graminearum* Species Complex Combined with Transcriptomic and Proteomic Analysis. Microorganisms 10, 1479. https://doi.org/10.3390/microorganisms10081479

Wang, J., Zhang, Z., 2021. GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction. Genomics, Proteomics Bioinforma. 19, 629–640. https://doi.org/10.1016/j.gpb.2021.08.005

Wang, Q., Sun, M., Zhang, Y., Song, Z., Zhang, S., Zhang, Q., Xu, J.R., Liu, H., 2020. Extensive chromosomal rearrangements and rapid evolution of novel effector superfamilies contribute to host adaptation and speciation in the basal ascomycetous fungi. Mol. Plant Pathol. 21, 330–348. https://doi.org/10.1111/mpp.12899

Wollenberg, R.D., Sondergaard, T.E., Nielsen, M.R., Knutsson, S., Pedersen, T.B., Westphal, K.R., Wimmer, R., Gardiner, D.M., Sørensen, J.L., 2019. There it is! *Fusarium*

pseudograminearum did not lose the fusaristatin gene cluster after all. Fungal Biol. 123, 10–17. https://doi.org/10.1016/j.funbio.2018.10.004

Woźniak, M.J., Allan, V.J., 2006. Cargo selection by specific kinesin light chain 1 isoforms. EMBO J. 25, 5457–5468. https://doi.org/10.1038/sj.emboj.7601427

Wozniak, M.J., Milner, R., Allan, V., 2004. N-terminal kinesins: Many and various. Traffic. https://doi.org/10.1111/j.1600-0854.2004.00191.x

Xiong, W., He, L., Lai, J., Dooner, H.K., Du, C., 2014. HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. Proc. Natl. Acad. Sci. U. S. A. 111, 10263–10268. https://doi.org/10.1073/pnas.1410068111

Zhao, C., Waalwijk, C., de Wit, P.J.G.M., Tang, D., Van der Lee, T., 2014. Relocation of genes generates non-conserved chromosomal segments in *Fusarium graminearum* that show distinct and co-regulated gene expression patterns. BMC Genomics 15, 1–17. https://doi.org/10.1186/1471-2164-15-191

4.8 Figures

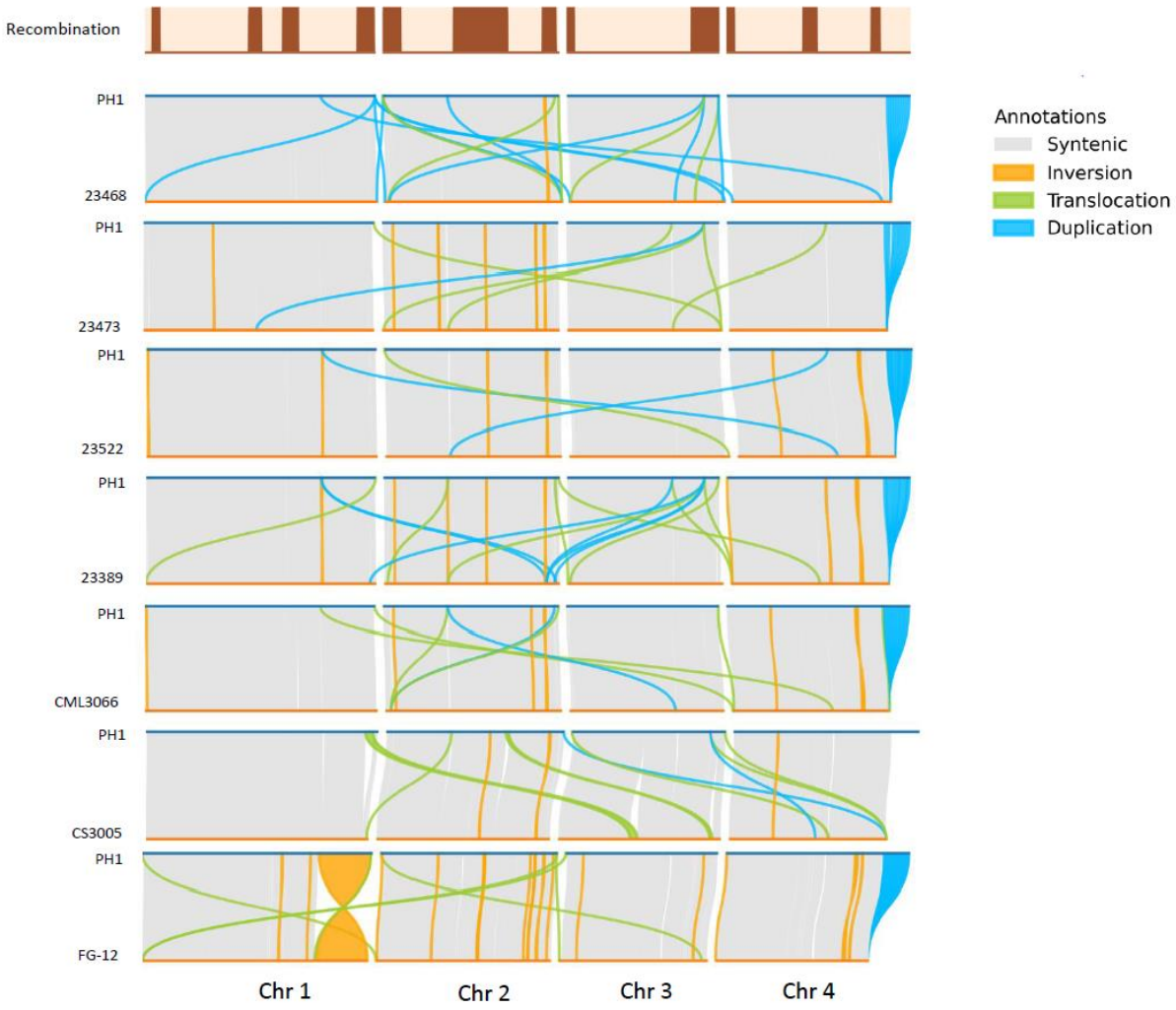


Figure 4.1. Chromosomal rearrangements in *Fusarium graminearum* genomes.
Rearrangements greater than or equal to 1000 bp are plotted. The dark brown bars in the top plot represent regions with high recombination rates.

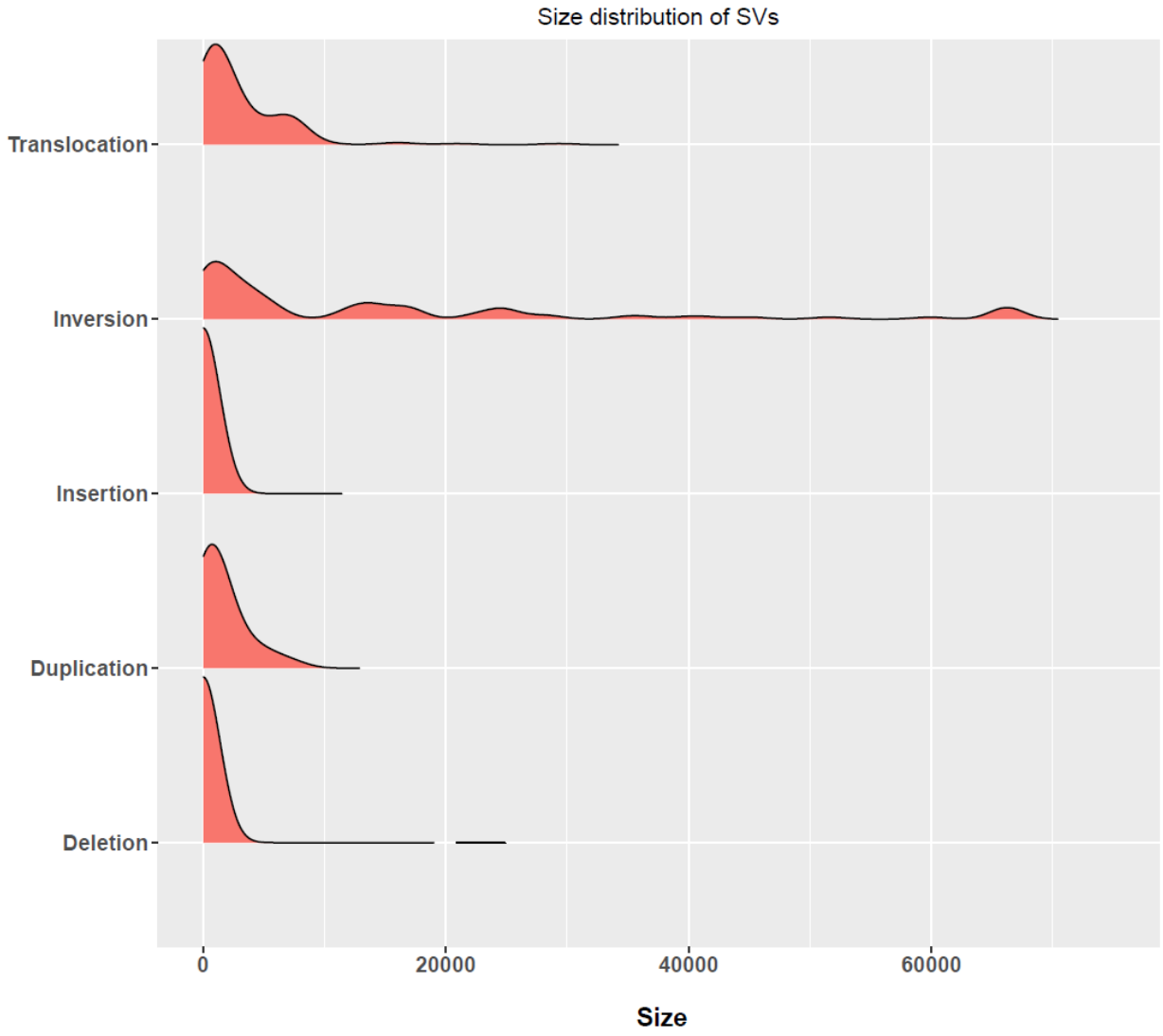


Figure 4.2. Size distribution of structural variants in *Fusarium graminearum* genomes.

The following are not included in the plot: artifactual duplications caused by the incomplete assembly of ribosomal RNA repeats at the C-terminus of chromosomal 4, and because of their size (> 100 kb), the largest inversion at the end of chromosome 1 in FG-12 and the two largest translocations on chromosome 4 in isolate CS3005.

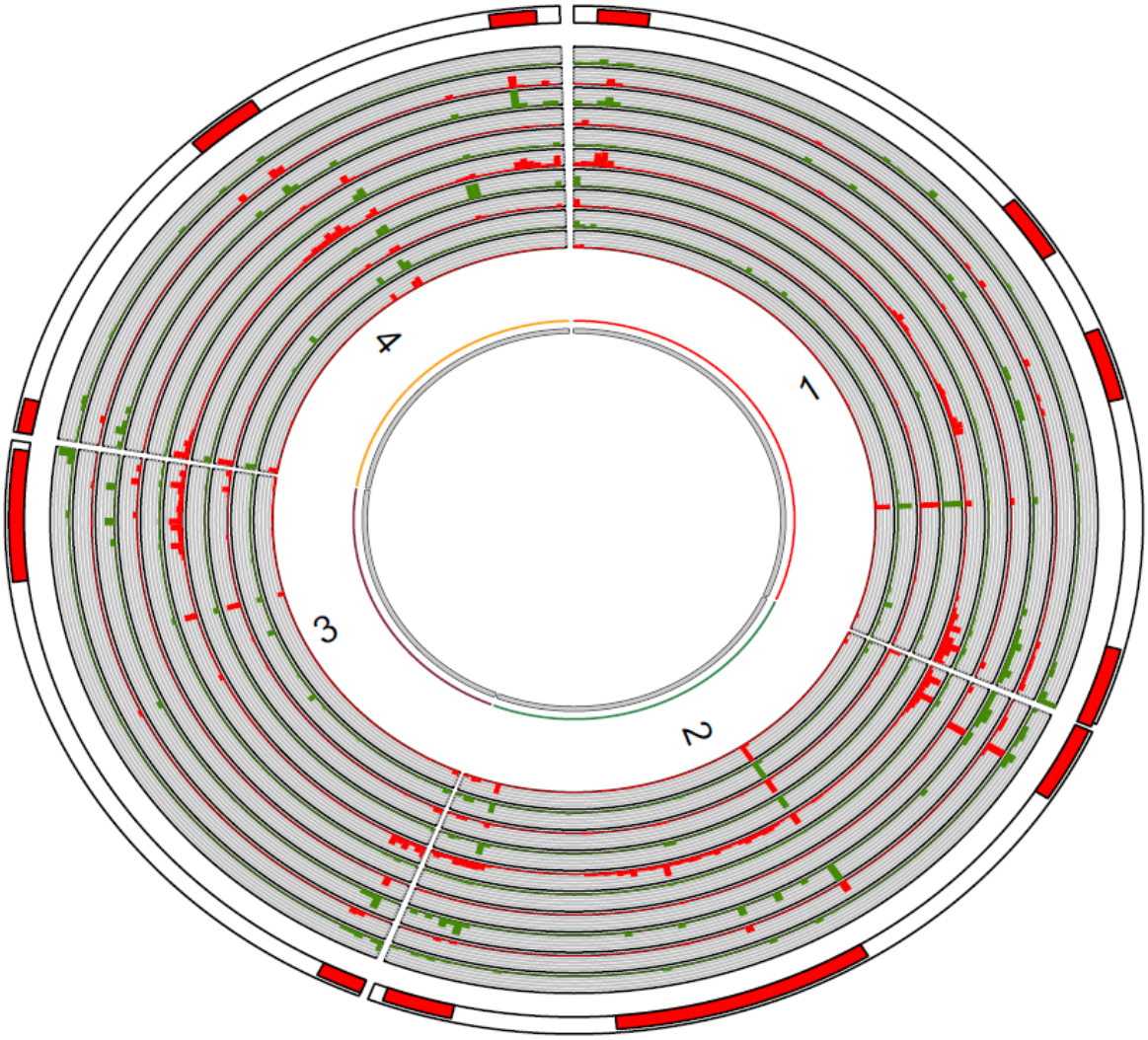


Figure 4.3. Overlap of different features in *Fusarium graminearum* isolate 23389.

Tracks 1 to 10 consist of the total number (including all sizes) of all events or sum of the lengths of events falling within each non-overlapping 50 kb window. Track 1 (from inside): Transposable element/ repeat element count. Track 2: Total Transposable element/ repeat element length. Track 3: Structural rearrangement (Inversion, translocation, duplication) count. Track 4: Structural rearrangement length. Track 5: Single nucleotide polymorphism (difference from PH-1) count. Track 6: Insertion and deletion (indel) count. Track 7: Indel length. Track 8: Highly diverged region (HDR) count. Track 9: HDR length. Track 10: Orphan contig alignment length. Track 11: Regions of high recombination.

4.9 Tables

Table 4.1. Summary statistics of *Fusarium graminearum* genomes used in this study.

Isolates	Chemotype	Raw data (Gb)	Coverage	No. of contigs	Genome size	Nuclear genome size	Busco score
23389	NX2	6.07	160	33	43,661,591	36,906,119	96.2
23468	15ADON	7.3	192	8	37,441,186	37,004,440	97.7
23473	15ADON	10.14	267	10	37,080,489	36,786,251	97.7
23522	3ADON	4.68	123	6	37,374,555	37,126,805	97.7
CML3066	15ADON	NA	NA	5	37,008,806	36,908,675	97.7
CS3005	15ADON	NA	NA	1205	36,667,552	36,313,044	97.5
FG-12	15ADON	NA	NA	6	35,897,492	35,851,270	NA

Genome sizes are in base pairs. The top four assemblies were produced in this study.

Table 4.2. Summary of structural variants in *Fusarium graminearum* genomes.

Isolates	Inversions	Translocations	Insertions	Insertions >1kb	Insertions (1bp)	Deletions	Deletions >1kb	Deletions (1bp)	Duplications
23468	4	18	4,102	12	2,177	5,010	9	2,781	50
23473	14	39	6,822	20	3,503	6,788	13	3,370	42
23522	13	20	5,717	15	3,022	5,470	19	2,667	36
23389	10	28	5,501	14	3,244	7,413	15	4,715	55
CML3066	12	19	5,610	23	2,812	5,168	17	4,292	30
CS3005	5	15	4,605	14	2,181	4,253	10	1,939	21
FG-12	29	20	26,132	18	16,930	0	0	0	11
Total	87	159	58,489	116	33,869	34,102	83	19,764	245

Duplications on the end of chromosome four are not included as the duplications are artifacts and were identified due to incomplete assembly of repeat regions in query genomes. Deletions in FG-12 are not included.

Table 4.3. Summary of transposable elements/repeat content in *Fusarium graminearum* genomes used in this study.

Fg Isolates	Description	LTR		TIR				Non-TIR		Total
		Gypsy	Unknown	CACTA	Mutator	Tc1_mariner	Hat	Helitron	Repeat Regions	
PH1	No. of repeat elements	NA	45	46	20	1	NA	10	260	382
	% of genome masked	NA	0.19	0.17	0.12	~0	NA	0.19	0.31	0.98
23389	No. of repeat elements	3	131	55	23	2	NA	13	361	588
	% of genome masked	0.06	0.48	0.22	0.14	0.01	NA	0.3	0.42	1.63
23468	No. of repeat elements	4	122	51	26	2	32	17	288	542
	% of genome masked	0.08	0.45	0.21	0.16	0.02	0.11	0.31	0.53	1.87
23473	No. of repeat elements	NA	NA	46	26	1	1	16	448	538
	% of genome masked	NA	NA	0.2	0.17	0.01	0.01	0.37	1.22	1.98
23522	No. of repeat elements	114	154	34	30	2	1	14	193	542
	% of genome masked	0.28	0.45	0.2	0.2	0.01	0.01	0.33	0.46	1.94
CML3066	No. of repeat elements	56	45	60	22	1	NA	10	362	556
	% of genome masked	0.08	0.22	0.16	0.14	~0	NA	0.2	0.74	1.54
CS3005	No. of repeat elements	NA	172	53	19	1	1	8	442	696
	% of genome masked	NA	0.33	0.15	0.12	~0	0.01	0.14	0.41	1.16
FG-12	No. of repeat elements	3	340	23	21	NA	NA	9	284	680
	% of genome masked	~0	0.97	0.17	0.15	NA	NA	0.27	0.15	1.71

4.10 List of supplemental files

Supplemental file 4.1: Information on isolates sequenced for this study.

Supplemental file 4.2: Sizes of each of the four chromosomes for isolates assembled during this study.

Supplemental file 4.3: Size, count and proportion of the PH-1 genome affected by different **structural rearrangements for each assembly.**

Supplemental file 4.4: Information (e.g., position in reference and query, size) on inversions identified.

Supplemental file 4.5: Information on translocations identified.

Supplemental file 4.6: Information on duplications identified.

Supplemental file 4.7: Information on insertions greater to or equal to 50 base pairs identified.

Supplemental file 4.8: Information on deletions greater to or equal to 50 base pairs identified.

Supplemental file 4.9: List of genes in PH-1 partially or completely overlapping inversions in query genomes, and phenotypes and GO terms of the genes.

Supplemental file 4.10: List of genes in PH-1 partially or completely overlapping duplications in query genomes, and phenotypes and GO terms of the genes.

Supplemental file 4.11: List of genes in PH-1 partially or completely overlapping translocations in query genomes, and phenotypes and GO terms of the genes.

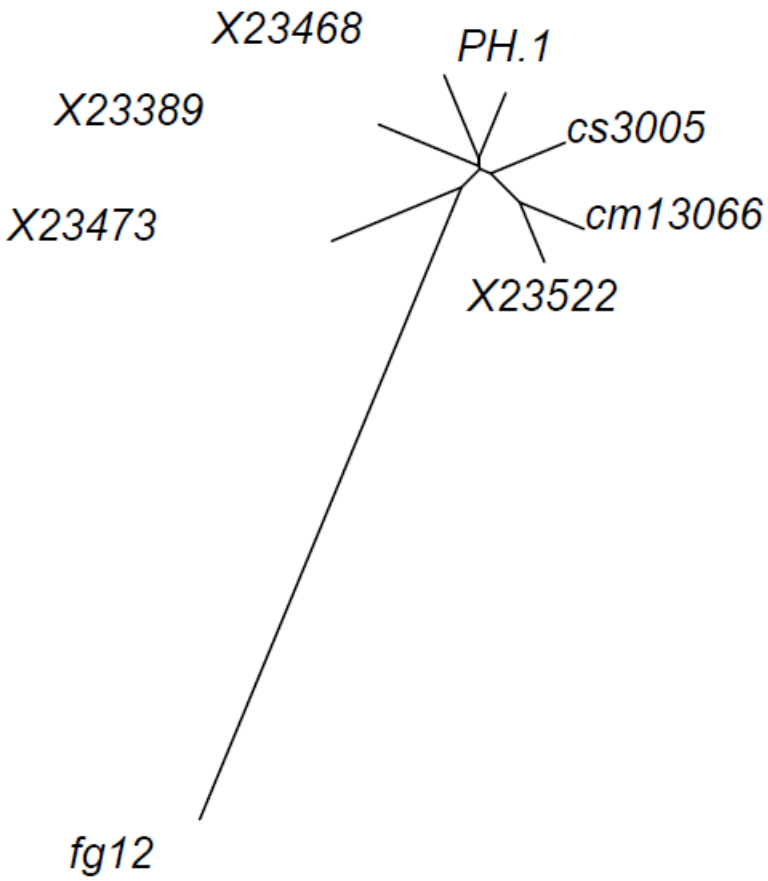
Supplemental file 4.12: List of genes in PH-1 partially or completely overlapping deletions (equal to or greater than 50 base pairs) in query genomes, and phenotypes and GO terms of the genes.

Supplemental file 4.13: Transposable elements/repeat content identified and their class and location in *F. graminearum* genomes.

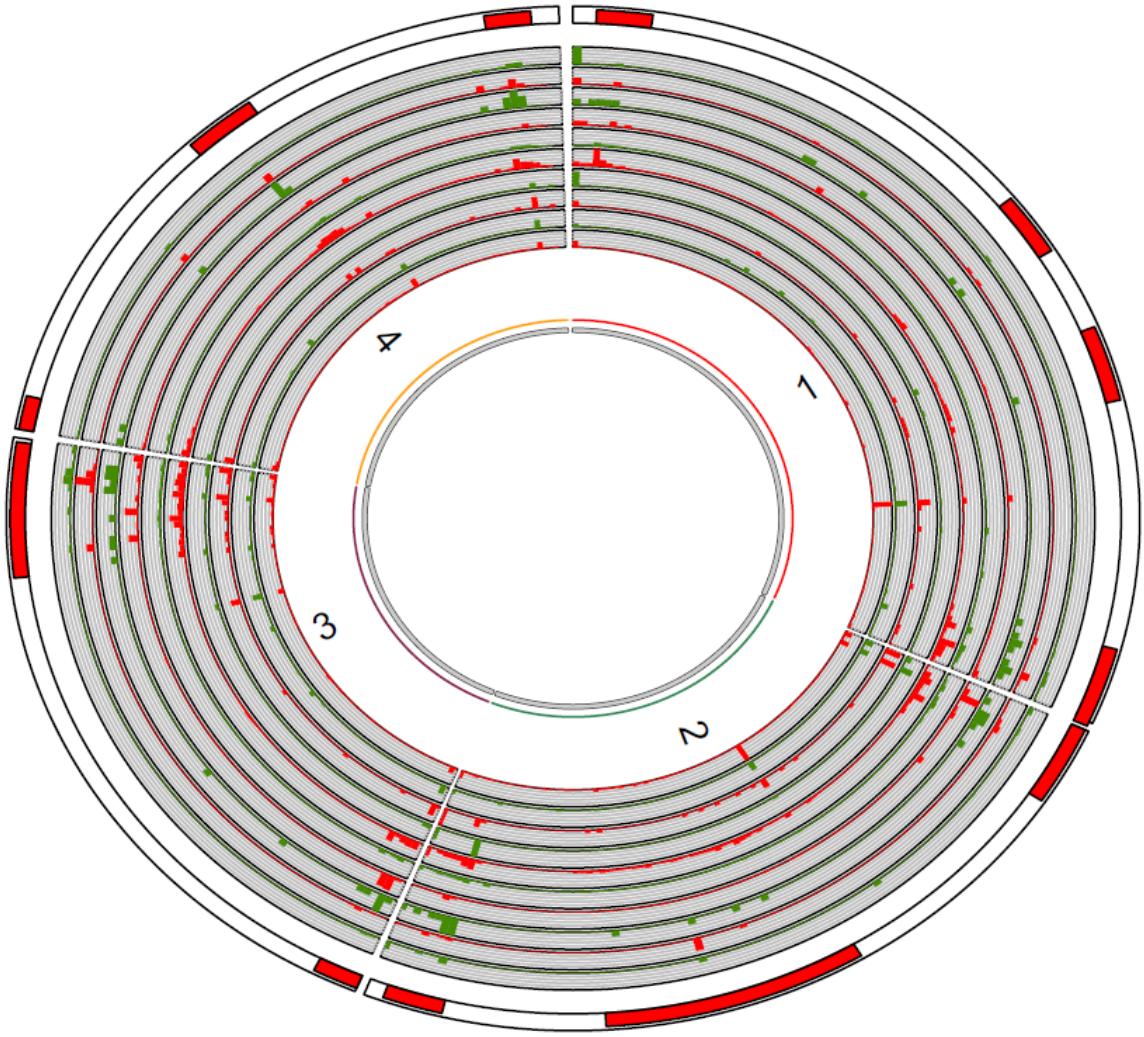
Supplemental file 4.14: Summary of intact transposable elements/repeat content in *F. graminearum* genomes used in this study.

Supplemental file 4.15: Total length of NUCmer alignments of orphan contigs from Kelly and Ward (2018) mapped to *F. graminearum* genomes assembled for this study.

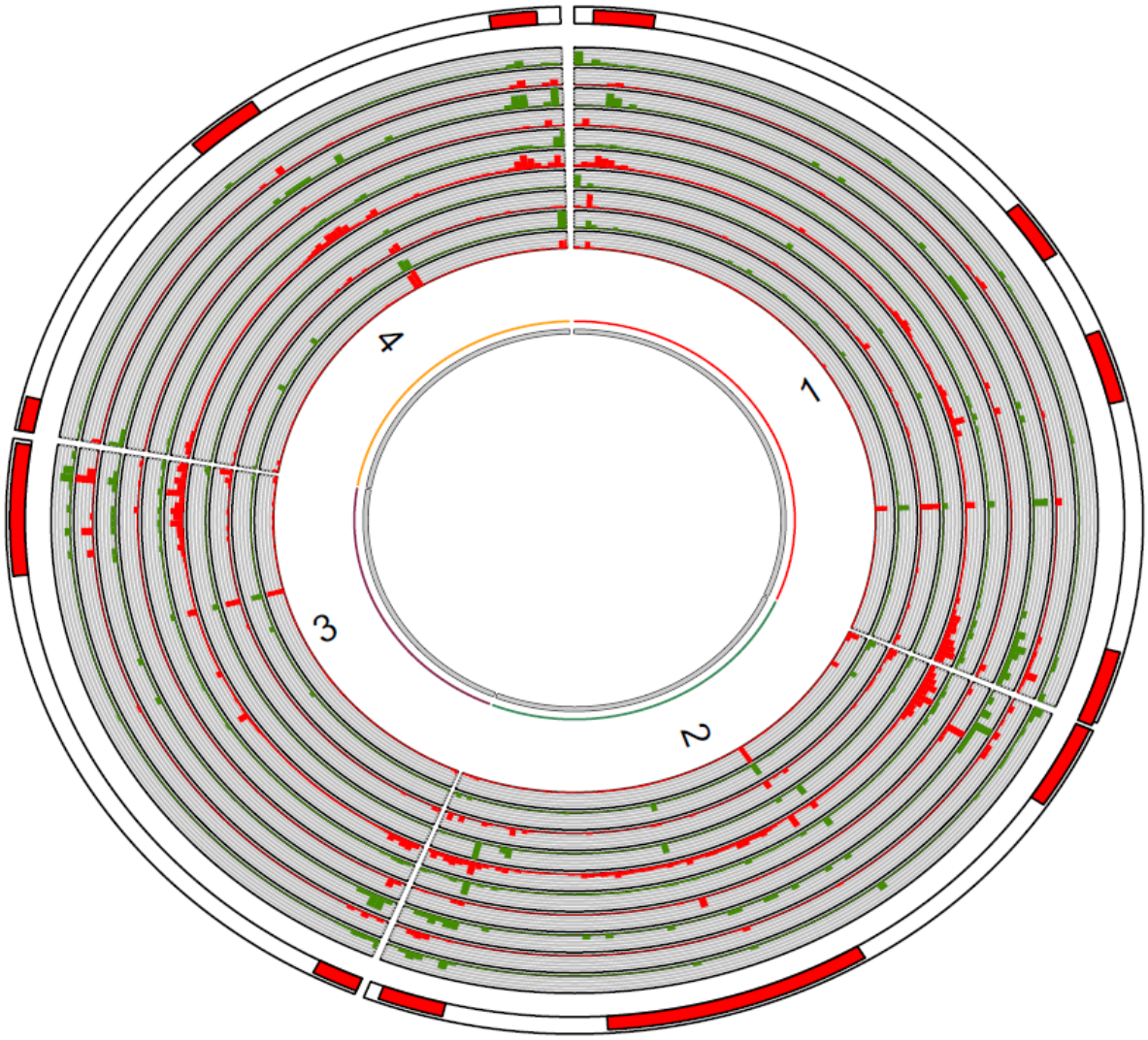
Appendix A Supplemental files chapter 4



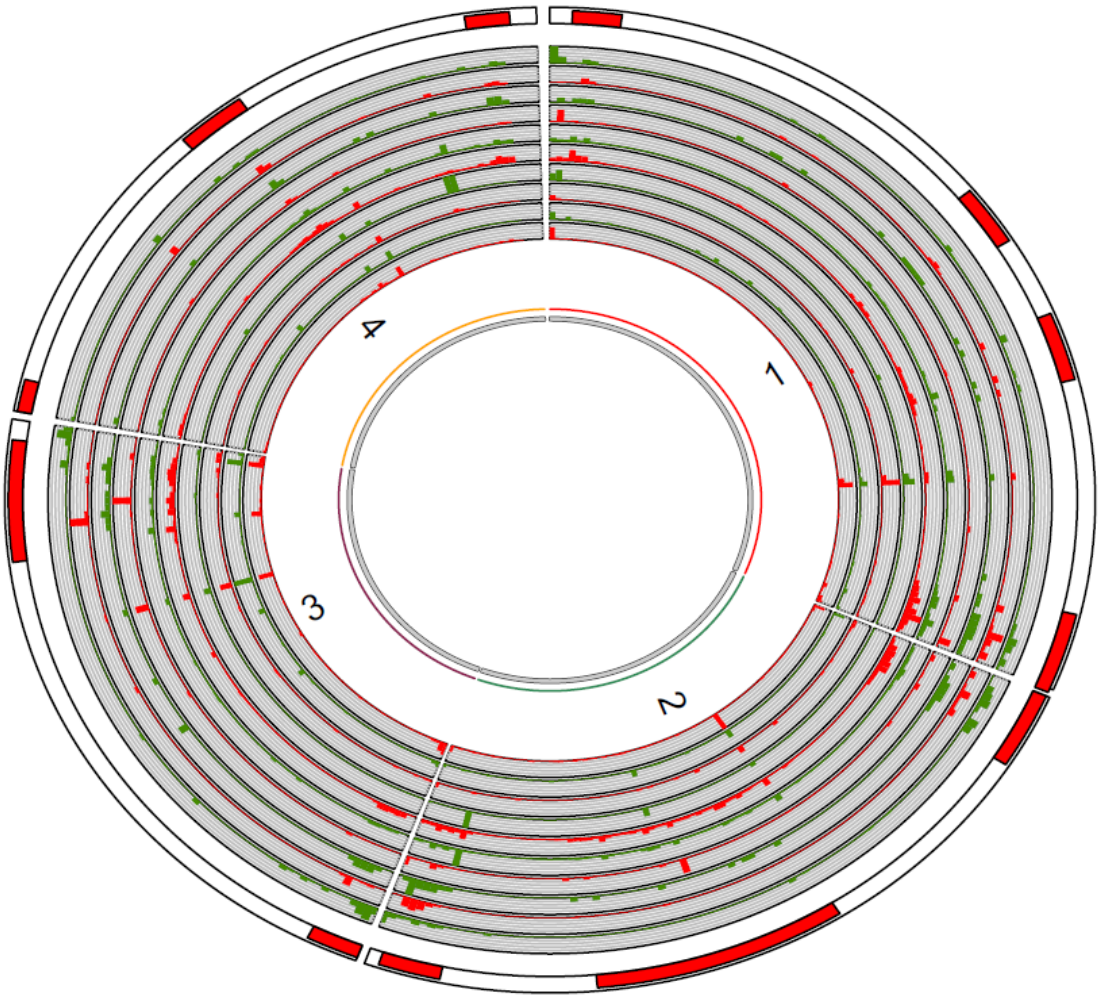
Appendix A Figure A.1 Neighbor-joining tree for isolates used in the study.
Tree was constructed using single nucleotide polymorphisms (SNPs) identified by SyRI in query isolates against PH-1 genome.



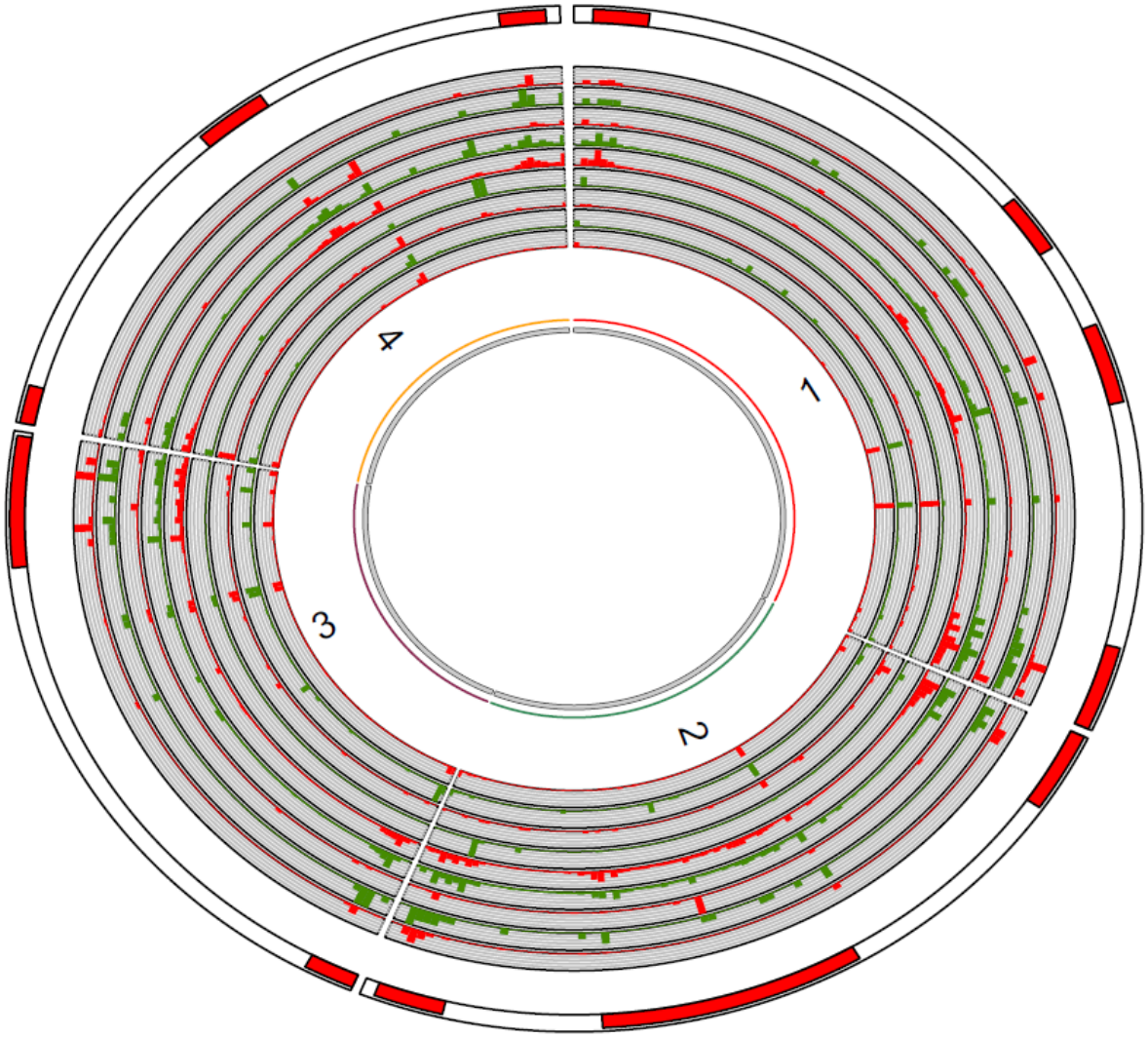
Appendix A Figure A.2 Overlap of different features in *Fusarium graminearum* isolate 23468**.** Tracks 1 to 10 consist of the total number (including all sizes) of all events or sum of the lengths of events falling within each non-overlapping 50 kb window. Track 1 (from inside): Transposable element/ repeat element count. Track 2: Total Transposable element/ repeat element length. Track 3: Structural rearrangement (Inversion, translocation, duplication) count. Track 4: Structural rearrangement length. Track 5: Single nucleotide polymorphism (difference from PH-1) count. Track 6: Insertion and deletion (indel) count. Track 7: Indel length. Track 8: Highly diverged region (HDR) count. Track 9: HDR length. Track 10: Orphan contig alignment length. Track 11: Regions of high recombination.



Appendix A Figure A.3 Overlap of different features in *Fusarium graminearum* isolate 23473. Tracks 1 to 10 consist of the total number (including all sizes) of all events or sum of the lengths of events falling within each non-overlapping 50 kb window. Track 1 (from inside): Transposable element/ repeat element count. Track 2: Total Transposable element/ repeat element length. Track 3: Structural rearrangement (Inversion, translocation, duplication) count. Track 4: Structural rearrangement length. Track 5: Single nucleotide polymorphism (difference from PH-1) count. Track 6: Insertion and deletion (indel) count. Track 7: Indel length. Track 8: Highly diverged region (HDR) count. Track 9: HDR length. Track 10: Orphan contig alignment length. Track 11: Regions of high recombination.

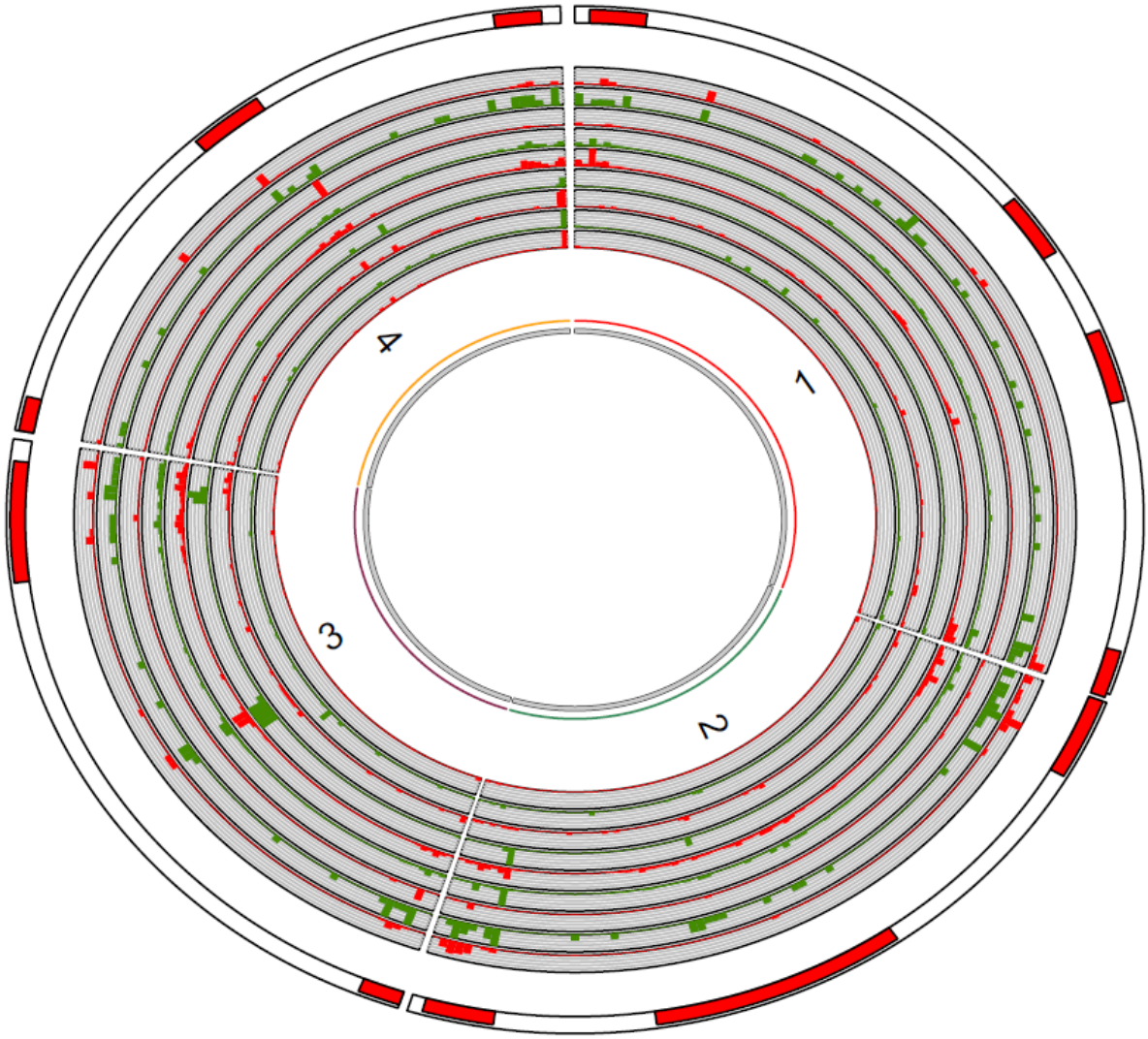


Appendix A Figure A.4 Overlap of different features in *Fusarium graminearum* isolate 23522. Tracks 1 to 10 consist of the total number (including all sizes) of all events or sum of the lengths of events falling within each non-overlapping 50 kb window. Track 1 (from inside): Transposable element/ repeat element count. Track 2: Total Transposable element/ repeat element length. Track 3: Structural rearrangement (Inversion, translocation, duplication) count. Track 4: Structural rearrangement length. Track 5: Single nucleotide polymorphism (difference from PH-1) count. Track 6: Insertion and deletion (indel) count. Track 7: Indel length. Track 8: Highly diverged region (HDR) count. Track 9: HDR length. Track 10: Orphan contig alignment length. Track 11: Regions of high recombination.

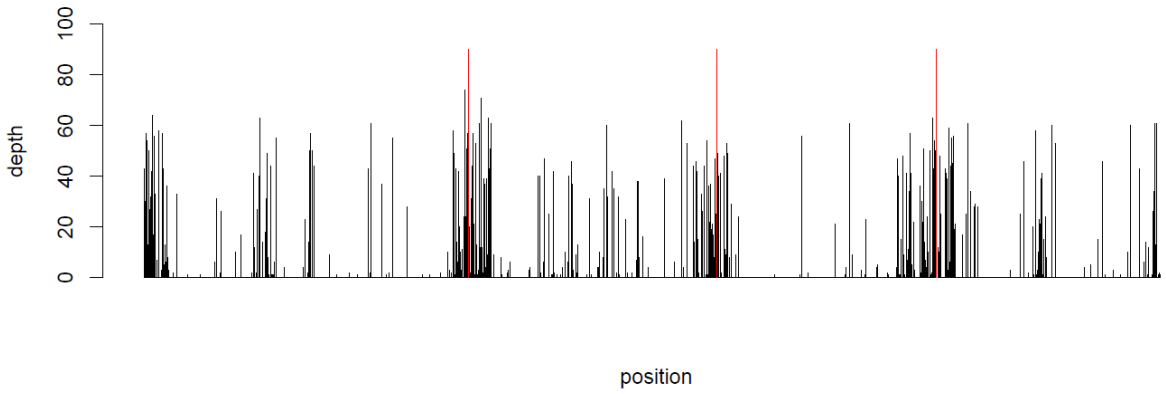


Appendix A Figure A.5 Overlap of different features in *Fusarium graminearum* isolate CML3066.

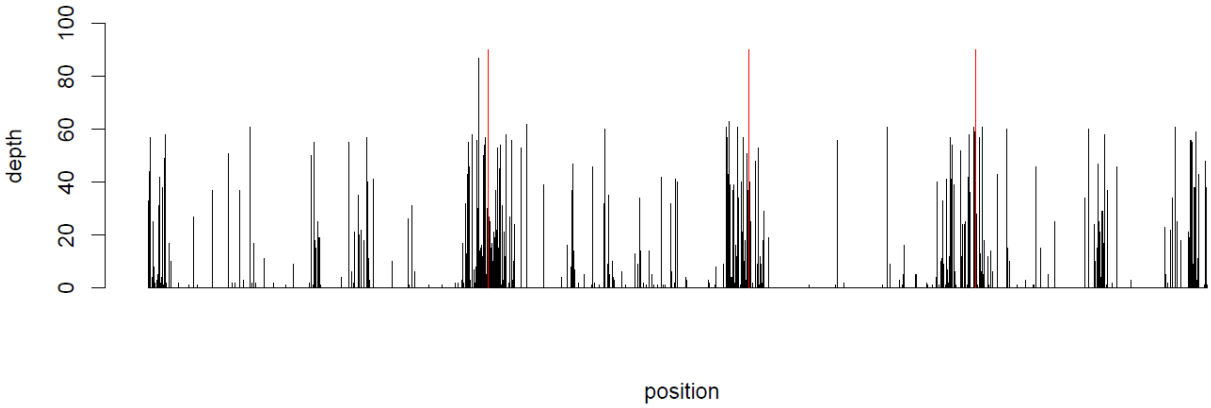
Tracks 1 to 10 consist of the total number (including all sizes) of all events or sum of the lengths of events falling within each non-overlapping 50 kb window. Track 1 (from inside): Transposable element/ repeat element count. Track 2: Total Transposable element/ repeat element length. Track 3: Structural rearrangement (Inversion, translocation, duplication) count. Track 4: Structural rearrangement length. Track 5: Single nucleotide polymorphism (difference from PH-1) count. Track 6: Insertion and deletion (indel) count. Track 7: Indel length. Track 8: Highly diverged region (HDR) count. Track 9: HDR length. Track 10: Orphan contig alignment length. Track 11: Regions of high recombination.



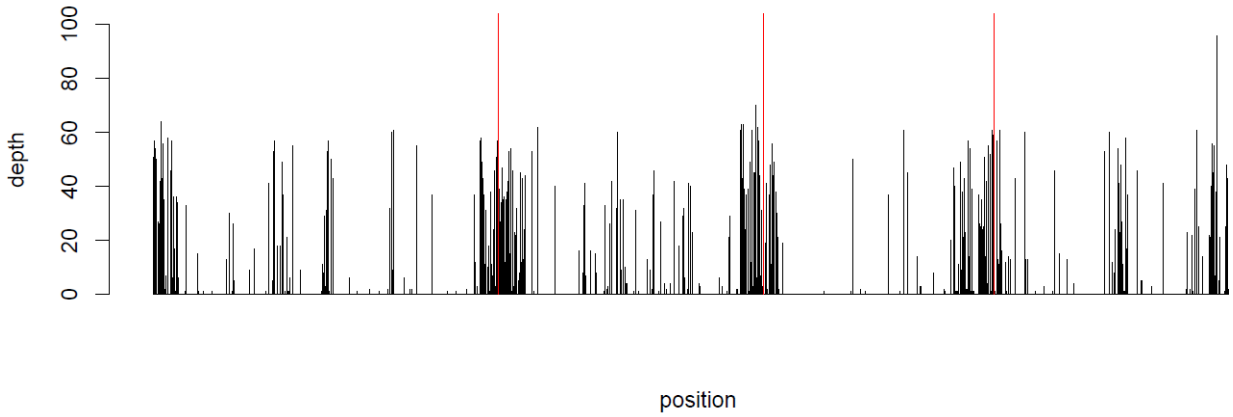
Appendix A Figure A.6 Overlap of different features in *Fusarium graminearum* isolate CS3005. Tracks 1 to 10 consist of the total number (including all sizes) of all events or sum of the lengths of events falling within each non-overlapping 50 kb window. Track 1 (from inside): Transposable element/ repeat element count. Track 2: Total Transposable element/ repeat element length. Track 3: Structural rearrangement (Inversion, translocation, duplication) count. Track 4: Structural rearrangement length. Track 5: Single nucleotide polymorphism (difference from PH-1) count. Track 6: Insertion and deletion (indel) count. Track 7: Indel length. Track 8: Highly diverged region (HDR) count. Track 9: HDR length. Track 10: Orphan contig alignment length. Track 11: Regions of high recombination.



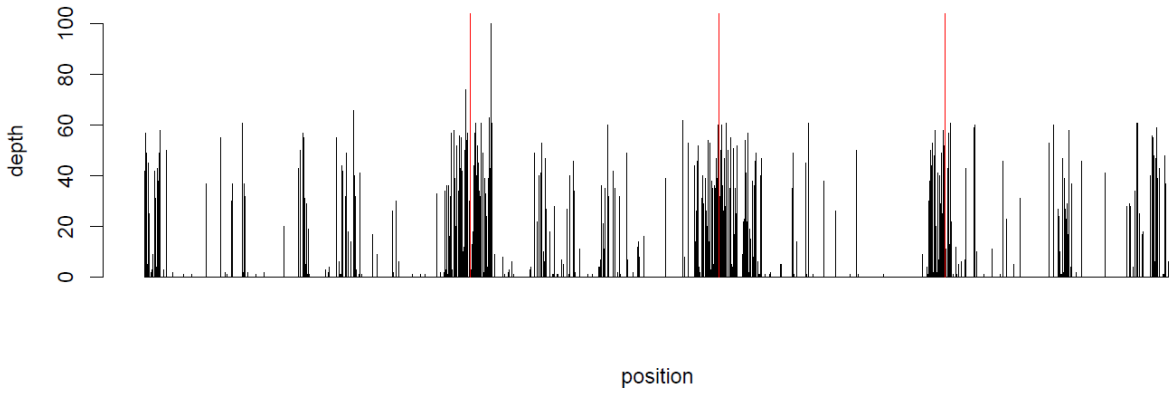
Appendix A Figure A.7 Depth and position along chromosomes of orphan contigs from Kelly and Ward (2018) mapped against the genome of isolate 23389.
Vertical red lines mark the end of chromosomes 1, 2, 3 and 4



Appendix A Figure A.8 Depth and position along chromosomes of orphan contigs from Kelly and Ward (2018) mapped against the genome of isolate 23468.
Vertical red lines mark the end of chromosomes 1, 2, 3 and 4**.**



Appendix A Figure A.9 Depth and position along chromosomes of orphan contigs from Kelly and Ward (2018) mapped against the genome of isolate 23473.
Vertical red lines mark the end of chromosomes 1, 2, 3 and 4**.**



Appendix A Figure A.10 Depth and position along chromosomes of orphan contigs from Kelly and Ward (2018) mapped against the genome of isolate 23522.
Vertical red lines mark the end of chromosomes 1, 2, 3 and 4.