Characterization of genetic variation in secondary metabolites in *Fusarium*

by

Wei Yue

B.S., Shandong Agricultural University, 2001
M.S., Nanjing Agricultural University, 2006

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Interdepartmental Genetics Program
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2017

Abstract

Secondary metabolites (SMs), low molecular weight molecules that are not essential for normal organism growth and development, may confer a selective advantage in some environments. Fungal SMs are structurally and functionally diverse and include mycotoxins, plant regulators and pigments, and the genes that work together in SM biosynthetic pathways are physically clustered in the genome. *Fusarium,* a genus of filamentous fungi, is noted for SM production, especially mycotoxins, which may contribute to plant pathogenesis. *Fusarium* species exhibit differences in their SM profiles, and comparative genomics studies have found corresponding differences in the SM gene clusters in some *Fusarium* species. The investigation of differences in the genomes and SM gene clusters between closely related species, such as *F. proliferatum* and *F. fujikuroi*, may help explain their phenotypic divergence, including differences in SM profiles. In addition, the study of intra-species SM variation may indicate how SM loci affect a pathogen's fitness traits.

My research includes three main projects that address different aspects of *Fusarium* SM variability. To carry out my projects, I established a feasible Genotyping-by-Sequencing (GBS) protocol for *Fusarium*. One project explored the genetic bases underlying phenotypic divergence related to SM profiles and pathogenicity between *F. proliferatum* and *F. fujikuroi* using a quantitative genetics approach. Specifically, I 1) constructed the first high density genetic map based on progeny from an interspecific cross between these two species; and 2) detected a novel regulatory locus for gibberellic acid production and identified a region affecting onion virulence that includes the fumonisin gene cluster. The second project characterized the *F. proliferatum* parent genome from the previous cross and its SM gene clusters using a comparative genomics approach. Specifically, I 1) assembled the *F. proliferatum* genome into 12 chromosomes with a

combined length of ~43 Mb; 2) annotated this assembly and characterized its 50 SM gene clusters; and 3) detected over 100 *F. proliferatum* specific genes that might play roles in this species' host specificity and plant pathogenicity. The third project used a population genomics approach to explore how different *F. graminearum* chemotypes, or isolates classified based on the accumulation of alternate trichothecene toxin types, may differ for fitness traits and whether trichothecene genes are directly responsible for these differences. Specifically, I 1) genotyped over 300 *F. graminearum* strains from New York and the upper Midwest in the U.S. and from South America using our GBS protocol; 2) detected two major subpopulations that were correlated, though imperfectly, to the predicted 3-acetyl deoxynivalenol (3ADON) and 15-acetyl deoxynivalenol (15ADON) chemotypes in the U.S.; 3) identified a rapid linkage disequilibrium decay over a few tens of kb followed by a slower decay to background levels over a distance of 200 kb to 400 kb in selected subpopulations in the U.S.; and 4) found that neither chemotype has a clear fitness advantage in a small set of isolates from New York, but that isolates belonging to one genetic subpopulation may on average have a fitness advantage over isolates from the other subpopulation.

Characterization of genetic variation in secondary metabolites in *Fusarium*

by

Wei Yue

B.S., Shandong Agricultural University, 2001
M.S., Nanjing Agricultural University, 2006

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Interdepartmental Genetics Program
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2017

Approved by:

Major Professor
Christopher Toomajian

Copyright

Abstract

Secondary metabolites (SMs), low molecular weight molecules that are not essential for normal organism growth and development, may confer a selective advantage in some environments. Fungal SMs are structurally and functionally diverse and include mycotoxins, plant regulators and pigments, and the genes that work together in SM biosynthetic pathways are physically clustered in the genome. *Fusarium,* a genus of filamentous fungi, is noted for SM production, especially mycotoxins, which may contribute to plant pathogenesis. *Fusarium* species exhibit differences in their SM profiles, and comparative genomics studies have found corresponding differences in the SM gene clusters in some *Fusarium* species. The investigation of differences in the genomes and SM gene clusters between closely related species, such as *F. proliferatum* and *F. fujikuroi*, may help explain their phenotypic divergence, including differences in SM profiles. In addition, the study of intra-species SM variation may indicate how SM loci affect a pathogen's fitness traits.

My research includes three main projects that address different aspects of *Fusarium* SM variability. To carry out my projects, I established a feasible Genotyping-by-Sequencing (GBS) protocol for *Fusarium*. One project explored the genetic bases underlying phenotypic divergence related to SM profiles and pathogenicity between *F. proliferatum* and *F. fujikuroi* using a quantitative genetics approach. Specifically, I 1) constructed the first high density genetic map based on progeny from an interspecific cross between these two species; and 2) detected a novel regulatory locus for gibberellic acid production and identified a region affecting onion virulence that includes the fumonisin gene cluster. The second project characterized the *F. proliferatum* parent genome from the previous cross and its SM gene clusters using a comparative genomics approach. Specifically, I 1) assembled the *F. proliferatum* genome into 12 chromosomes with a

combined length of ~43 Mb; 2) annotated this assembly and characterized its 50 SM gene clusters; and 3) detected over 100 *F. proliferatum* specific genes that might play roles in this species' host specificity and plant pathogenicity. The third project used a population genomics approach to explore how different *F. graminearum* chemotypes, or isolates classified based on the accumulation of alternate trichothecene toxin types, may differ for fitness traits and whether trichothecene genes are directly responsible for these differences. Specifically, I 1) genotyped over 300 *F. graminearum* strains from New York and the upper Midwest in the U.S. and from South America using our GBS protocol; 2) detected two major subpopulations that were correlated, though imperfectly, to the predicted 3-acetyl deoxynivalenol (3ADON) and 15-acetyl deoxynivalenol (15ADON) chemotypes in the U.S.; 3) identified a rapid linkage disequilibrium decay over a few tens of kb followed by a slower decay to background levels over a distance of 200 kb to 400 kb in selected subpopulations in the U.S.; and 4) found that neither chemotype has a clear fitness advantage in a small set of isolates from New York, but that isolates belonging to one genetic subpopulation may on average have a fitness advantage over isolates from the other subpopulation.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to express my appreciation to my advisor, Dr. Christopher Toomajian for his advice, guidance and support during my Ph.D. study. I thank my committee members, Dr. John Leslie, Dr. Sanzhen Liu, and Dr. Mark Ungerer for their advice in my study. I thank Amy Beyer, Shuangye Wu and the members of Dr. Guihua Bai's laboratory who helped me greatly in my laboratory work. I thank Dr. Barbara Valent for her advice and recommendation in my study and life. I also thank other people in the Plant Pathology department for their assistance and support. Finally, I thank my family for their endless support during my study.

Chapter 1 - Literature review

Fungi, a eukaryotic Kingdom, provides humans with foods like mushrooms, alcohol like beer, antibiotics like penicillin, enzymes like protease, and many other benefits. However, pathogenic fungi also cause problems to humans like valley fever caused by the fungus *Coccidioides* (https://www.cdc.gov/fungal/diseases/) and to crops like rice blast disease caused by the fungus *Magnaporthe* (Talbot 2003). Therefore, we need to understand fungi to win the arms race between pathogenic fungi and their hosts.

Fusarium, a genus of filamentous ascomycete fungi, was first described by Link (1809). It contains many important plant pathogens which can cause significant diseases in major agronomically important plants worldwide such as bakanae disease of rice, *Fusarium* head blight of wheat and barley, ear rot of maize, panama disease of banana, bulb rot of onion, etc. (Leslie and Summerell, 2006). Fusaria are also noted for their production of secondary metabolites (SMs), low molecular weight molecules that are not required for normal growth and development. *Fusarium* SMs include pigments, plant hormones and mycotoxins that contaminate food or feed and cause diseases in humans and animals through eating contaminated food or feed (Leslie and Summerell, 2006). Many of the major plant pathogens of this genus are found in four major phylogenetic groups: the *F. fujikuroi* species complex, the *F. graminearum* species complex, the *F. oxysporum* species complex, and the *F. solani* species complex (Aoki *et al*. 2014).

The *Fusarium fujikuroi* species complex

The *F. fujikuroi* species complex (FFSC) contains more than 50 diverse species that can be grouped into 3 major clades: the African clade, the American clade, and the Asian clade (O'Donnell *et al*. 1998, Kvas *et al*. 2009, Figure 1.1). The best-known pathogen in the African

clade is *F. verticillioides*, which causes maize seedling blight, stalk rot and ear rot throughout the world (Leslie and Summerell, 2006). *F. verticillioides* is noted for its high level of fumonisin production (Desjardins 2006), which is associated with human esophageal cancer (Marasas 2001), and animal diseases like leukoencephalomalacia in horses (Kellerman *et al*. 1990) and porcine pulmonary edema in swine (Harrison *et al*. 1990). Besides that, *F. verticillioides* can produce various other SMs (Table 1.1, Desjardins 2006, Niehaus *et al*. 2016). In the American clade, *F. circinatum* and *F. subglutinans* are two well-known species. The former can cause pitch canker disease of pine trees in the US, South Africa, Spain, Japan, and Chile, and the latter can cause maize stalk and ear rot in cooler maize cultivation areas (Leslie and Summerell, 2006). The profile of SM production is different in *F. circinatum* and *F. subglutinans* (Desjardins 2006).

Two well-studied species in the Asian clade are *F. fujikuroi* and *F. proliferatum*, which are hard to distinguish morphologically (Leslie and Summerell, 2006). The cross between them can produce viable progeny, although the level of cross-fertility is quite low (Desjardins *et al*. 1997, Leslie *et al*. 2004, 2007). The two species also are different in the diseases they cause and the SMs they produce. *F. fujikuroi* causes bakanae disease in rice (elongated seedlings), and *F. proliferatum* can cause diseases in a wide range of plants such as asparagus, date palm, onion, maize, mango, and sorghum. (Leslie and Summerell, 2006). *F. fujikuroi* is known for its gibberellic acid (GA) production (Cerda-Olmedo *et al*. 1994). GA is a plant growth regulator and contributes to the invasion of rice tissue during bakanae disease but does not play a role in initial root colonization (Wiemann *et al*. 2013). *F. fujikuroi* is used to produce GA commercially worldwide (Rademacher 1997). In addition to GA, *F. fujikuroi* produces other SMs, such as fusarin and moniliformin. (Table 1.1, Desjardins 2006, Niehaus *et al*. 2016). *F. proliferatum* can

also produce different SMs like fumonisin (Table 1.1, Desjardins 2006, Niehaus *et al*. 2016). Although no GA production is detected for many *F. proliferatum* isolates, GA production by an orchid-associated strain of *F. proliferatum* has been reported (Tsavkelova *et al*. 2008).

Fusarium genetic maps

A genetic map is the arrangement of genetic loci on a chromosome with estimated genetic distances between them, which serves as fundamental infrastructure for classical genetic analysis and trait mapping (Ott 1991). Several genetic maps are available in the *Fusarium* genus (Table 1.2). Puhalla and Spieth (1985) constructed the first *F. verticillioides* genetic map containing 4 linkage groups with 12 morphological/biochemical markers, which was consistent with the light microscopic observation (Howson *et al*. 1963).

After identifying the electrophoretic karyotype of *F. verticillioides* using pulsed field gel electrophoresis (Xu *et al*. 1995), Xu and Leslie (1996) constructed a new map consisting of 13 linkage groups associated with 12 chromosome-sized DNA fragments with 150 restriction fragment length polymorphisms (RFLPs), random amplified polymorphic DNAs (RAPDs), and morphological/biochemical markers from 121 progeny. The average ratio of physical distance to genetic distance (kb/cM) in this cross was around 32 (Xu and Leslie 1996). To improve this map, Jurgenson *et al*. (2002b) saturated this map with an additional 486 amplified fragment length polymorphisms (AFLPs) markers from the same 121 progeny. The expanded map contains 12 linkage groups corresponding to the detected 12 chromosome-sized DNA fragments. The average kb/cM ratio was reduced to 21 with increased marker density and total map length. In addition, they identified AFLP markers close to several important traits, *e.g*., mating type (*MAT*) and spore killer locus (Jurgenson *et al*. 2002b).

For *F. graminearum*, Jurgenson *et al*. (2002a) created a genetic map containing 9 linkage groups with average kb/cM ratio 28 using 1048 AFLP markers from 99 progeny. Karyotype analysis indicated that *F. graminearum* contains 4 chromosomes rather than 9 (Taga *et al*. 2003). With this map, Cumagun *et al*. (2004) mapped 1 quantitative trait locus (QTL) for pathogenicity and 2 QTLs for aggressiveness of *F. graminearum* toward wheat. However, O'Donnell *et al*. (2004) argue that phylogenetic evidence suggests the parents from the Jurgenson *et al*. (2002a) cross belong to separate phylogenetic species, *F. graminearum* and *F. asiaticum*, making the cross interspecific. As evidence that this cross is not a well-behaved intraspecific cross, they highlight that 5 out of the 9 linkage groups from this map showed prevalent segregation distortion, and 4 linkage groups showed recombination suppression (Jurgenson *et al*. 2002a, Gale *et al*. 2005). Additionally, many cases of gene order disagreement exist in the linkage groups of the genetic map when compared with the *F. graminearum* PH-1 genome sequence assembly (O'Donnell *et al*. 2004). In contrast, Lee *et al*. (2008) argued that some of the segregation distortion observed in the Jurgenson *et al*. (2002a) cross was due to the method used in making the cross, some of the recombination suppression regions were due to inversions, and the marker order of this map matched well with the genome sequence. Later, Gale *et al*. (2005) constructed a genetic map of *F. graminearum* with parents chosen from the same phylogenetic species that had similar properties (9 linkage groups with average kb/cM ratio 29) using 235 sequence-tagged sites (STSs) and AFLP markers from 111 progeny. The 9 linkage groups along with anchored scaffolds from a genome sequence assembly of *F. graminearum* were assembled into 4 chromosomes, confirmed by cytological karyotyping (Gale *et al*. 2005).

Hybridization between closely related *Fusarium* species can occasionally generate viable progeny and can be used for genetic map construction (Desjardins *et al*. 1997, Desjardins *et al*.

2000, Leslie *et al*. 2004, De Vos *et al*. 2007, Mohamed Nor 2014). De Vos *et al*. (2007) constructed a genetic map from a cross between *F. circinatum* and *F. subglutinans,* containing 12 linkage groups with average kb/cM ratio 19.5 using 581 AFLP, mating type and histone loci markers from 94 progeny. The 12 linkage groups were consistent with the PFGE result of *F. subglutinans* (Xu et al. 1995). Using this genetic map, De Vos *et al*. (2011) identified a significant QTL related to mycelial growth and morphology of colony margins.

Another interspecific genetic map was constructed from a cross between *F. fujikuroi* and *F. proliferatum* (Desjardins *et al*. 1997, Leslie *et al*. 2004). Mohamed Nor (2014) created a map of 12 linkage groups by comparing AFLP band sizes to the lengths of AFLP loci predicted from a *F. fujikuroi* genome sequence (Chiara *et al*. 2015) with average kb/cM ratio 34 using 79 AFLP and mating type locus markers from 533 progeny. The 12 linkage groups were consistent with the 12 chromosomes observed in the PFGE results and an independent *F. fujikuroi* genome sequence (Xu *et al*. 1995, Wiemann *et al*. 2013).

Fusarium genomes

Whole genome sequences provide valuable information for genetic research in the genus *Fusarium* for genetic maps and other analyses. Pulsed field gel electrophoresis (PFGE), including contour-clamped homogeneous electric field (CHEF) gel electrophoresis, was used to determine the size of chromosomes and to estimate total genome size before whole genome sequencing became more widespread in fungi (Schwartz and Cantor 1984, Vollrath and Davis 1987). Successful karyotypes for small genomes, such as *Saccharomyces cerevisiae* (Carle and Olson 1985), were constructed by using this method. *Fusarium* species were also evaluated with PFGE methods, including *F. oxysporum* (Boehm *et al*. 1994), *F. fujikuroi* (Xu *et al*. 1995), *F. solani* (Suga et al. 2002), and *F. graminearum* (Gale *et al*. 2005). Chromosome variations in

Fusaria including chromosome length polymorphisms among strains and the presence of conditional dispensable chromosomes (or supernumerary chromosomes) were revealed through analyses of electrophoretic karyotypes (Kistler *et al*. 2013).

Although PFGE analysis offers useful information, whole genome sequencing provides a more powerful method to study *Fusarium* genomes. The first four *Fusarium* species' genomes to be sequenced were sequenced using the traditional Sanger sequencing method at the first decade of the 21st century, including *F. graminearum* (Cuomo *et al*. 2007), *F. solani* (Coleman *et al*. 2009), *F. oxysporum* f. sp. *lycopersici* and *F. verticillioides* (Ma *et al*. 2010) (Table 1.3). Through sequencing and analyzing the first *Fusarium* genome (*F. graminearum*, the smallest genome in the genus sequenced thus far), Cuomo *et al*. (2007) detected low repeat content and few paralogous genes in the genome due to repeat-induced point mutation (RIP), which introduces C to T transitions in duplicated sequences during the sexual cycle (Selker *et al*. 1987). Cuomo *et al*. (2007) also identified that many pathogenicity-related genes are located in highly polymorphic regions (located near telomeres and other discrete regions). Coleman *et al*. (2009) sequenced the low repeat content *F. solani* genome, which contained 17 chromosomes with the total length of 54.4 Mb. Three supernumerary chromosomes contained relatively high repeat content compared to the rest of the genome, and carried genes needed for host colonization. Ma *et al*. (2010) sequenced and annotated two additional *Fusarium* genomes, *F. verticillioides* and *F. oxysporum,* containing 11 and 15 chromosomes with total length of 41.7 Mb and 59.9 Mb, respectively. Notably, *F. oxysporum* had a relatively high level of repeat content (~28%) compared to other published *Fusarium* genomes. Comparative genomics analysis suggested over 25% of the *F. oxysporum* genome corresponds to lineage-specific regions (as opposed to the core genome conserved between this species and other related *Fusarium* species), which carry

horizontally acquired genes potentially related to pathogenicity. Moreover, lineage-specific chromosomes could be experimentally transferred between genetically isolated strains of *F. oxysporum*, and the chromosomal transfer resulted in an accompanying transfer of host-specific pathogenicity (Ma *et al*. 2010).

With the availability of next generation sequencing technology, the cost of whole genome sequencing has decreased rapidly, and genomics research in *Fusarium* has accelerated. Nearly 30 *Fusarium* species have already been sequenced, including multiple species from the FFSC that are the focus of some of this dissertation (https://www.ncbi.nlm.nih.gov/genome) (Table 1.3). Among them, Wiemann *et al*. (2013) first sequenced and comprehensively analyzed *F. fujikuroi*, containing 12 chromosomes with the total length of 43.9 Mb. In total, 44 SM gene clusters were detected in this genome, including two unique to *F. fujikuroi* (PKS19 and NRPS31). Jeong *et al*. (2013) published a draft *F. fujikuroi* genome with the total length of 43.8 Mb, and Chiara *et al*. (2015) sequenced and analyzed three additional *F. fujikuroi* strains with similar genome size (~43 Mb). Comparisons among these five *F. fujikuroi* strains suggest that most of the SM clusters were conserved, and that species- or lineage-specific genes within this species were enriched in subtelomeric regions. Niehaus *et al*. (2016) sequenced and annotated genomes from strains of two additional species in the Asian clade of FFSC: one of *F. mangiferae* and two of *F. proliferatum* (Table 1.3, Figure 1.1). Comparison of sequenced FFSC genomes showed a high level of sequence conservation among species, but the profiles of SM gene clusters were diverse and the expression of these clusters was variable. In addition to GA, FFSC members also could produce two additional phytohormones: auxin and cytokinin. One strain of *F. proliferatum* (ET1) contained two functionally independent GA clusters, which might have different origins (Niehaus *et al*. 2016).

Gibberellin biosynthesis in the *Fusarium fujikuroi* species complex

Gibberellin, a diterpene phytohormone, stimulates plant growth and promotes plant development. Gibberellin was first identified and named from *F. fujikuroi* (teleomorph *Gibberella fujikuroi*), which causes bakanae disease in rice (Phinney 1983). Later, Curtis and Cross (1954) determined the chemical structure of gibberellic acid (GA), which was identical with gibberellin A_3 isolated by Takahashi *et al*. (1955). Since then, over 100 gibberellins have been discovered (http://www.plant-hormones.info/ gibberellin_nomenclature.htm), but only a few (GA_1-GA_4) function as phytohormones.

The GA biosynthetic genes in *F. fujikuroi* are clustered together at the end of chromosome 5, including *DES*, *P450-4*, *P450-1*, *P450-2*, *GGS2*, *CPS/KS*, and *P450-3* (Tudzynski and Hölter 1998, Wiemann *et al*. 2013). Besides *F. fujikuroi*, other members of the FFSC that contain the entire GA gene cluster include *F. sacchari*, *F. proliferatum*, *F. circinatum*, *F. konzum*, and *F. mangiferae*. (Malonek et al. 2005c, Bomke and Tudzynski 2009, Wiemann *et al*. 2013). Although the entire cluster exists, GA production was detected only in *F. fujikuroi*, *F. sacchari*, *F. konzum*, and one orchid-associated *F. proliferatum* strain (Troncoso *et al*. 2010, Tsavkelova *et al*. 2008). The possible reasons for this are: nonfunctional enzymes due to mutations in the coding regions of GA genes, reduced level of transcripts of GA genes, and improper post-transcriptional and/or post-translational modification (Wiemann *et al*. 2013).

F. proliferatum is a close relative of *F. fujikuroi* and contains a similar and intact GA gene cluster. Some strains of *F. proliferatum* have lost the ability to produce GA due to mutations in the coding and 5' noncoding regions of three genes (*ggs2*, *cps/ks*, *P450-4*) of the cluster (Malonek *et al*. 2005a, Malonek *et al*. 2005b). GA production can be restored in *F. proliferatum* by complementation with functional counterparts of these three genes from *F.*

fujikuroi (Malonek *et al*. 2005b). However, Niehaus *et al*. (2016) identified an orchid-associated *F. proliferatum* strain (ET1) that contained two copies of the GA cluster, and both copies were expressed in maize roots. Sequence comparison and phylogenetic analysis suggested that one copy of the GA cluster was similar to that from other Asian clade species, and the other one was more similar to that from American clade species (Niehaus *et al*. 2016). In addition, functional GA gene clusters also have been detected in fungi outside the *Fusarium* genus, *e.g*., *Phaeosphaeria* sp. L497 (Kawaide 2006) and *Sphaceloma manihoticola* (Bomke *et al*. 2008). The simplest explanation for this distribution of the GA gene cluster is horizontal gene transfer (Bomke and Tudzynski 2009, Wiemann *et al*. 2013).

In *F. fujikuroi*, the expression of GA cluster genes is positively regulated by the global nitrogen regulators *AreA* and *AreB* (Tudzynski *et al*. 1999, Mihlan *et al*. 2003, Michielse *et al*. 2014). Other genes identified as regulators of GA cluster genes include nitrogen regulator *nmr*, nitrogen regulator *meaB*, ammonium permease *mepB*, rapamycin kinase *Tor*, and glutamine synthetase (Bomke and Tudzynski 2009, Tudzynski 2014). The *velvet* like genes, *Ffvel1* and *Ffvel2*, and the *LaeA* like gene, *Fflae1*, can also influence the expression of GA cluster genes and GA production (Wiemann *et al*. 2010).

Fusarium graminearum

The *F. graminearum* species complex contains at least 16 phylogenetically distinct species (Aoki *et al*. 2014). The most important pathogen in this group is *F. graminearum* (Fg), which causes Fusarium head blight or scab disease in wheat and barley across the whole world (Leslie and Summerell 2006). This fungus not only reduces the crop yield, but also contaminates grains with mycotoxins such as trichothecenes (Windels 2000). *F. graminearum* can also produce other SMs (Table 1.1, Desjardins 2006).

Trichothecenes are one of the most important families of mycotoxins produced by *F. graminearum*, including type A and type B trichothecenes (McCormick *et al*. 2011, Alexander *et al*. 2011). The difference between type A and type B trichothecenes is the absence or presence of a keto group at carbon atom 8 (C-8) and a hydroxyl group at C-7 (Ueno 1980). The type B trichothecenes have received great attention related to their roles in scab disease in wheat and barley, particularly deoxynivalenol (DON), nivalenol (NIV), and their acetylated derivatives (Alexander *et al*. 2011). The biosynthetic genes of trichothecenes (*TRI* genes) in *F. graminearum* are clustered in three locations: the single-gene *TRI101* locus on chromosome 4, the two-gene *TRI1-TRI16* cluster on chromosome 1, and the 12-gene *TRI* core cluster on chromosome 2 (Rep and Kistler 2010). Due to genetic differences that vary among *F. graminearum* isolates, different amounts of alternate trichothecene toxin types accumulate in different isolates, and isolates can thus be classified into chemotypes based on the predominant trichothecene. In North America, 3-acetyl deoxynivalenol (3ADON) and 15-acetyl deoxynivalenol (15ADON) are common type B trichothecene chemotypes (Ward *et al*. 2008, Schmale *et al*. 2011). The genetic basis for the 3ADON and 15ADON chemotype difference is that the product of the *TRI8* gene in the 12-gene core cluster catalyzes deacetylation of the intermediate 3,15-diacetyl deoxynivalenol at C-3 or C-15, depending on the allele found at *TRI8*, to yield the 15ADON or 3ADON product, respectively (Alexander *et al*. 2011). Due to this simple genetic difference, a PCR assay can be used to predict in which chemotype an *F. graminearum* isolate belongs (Ward et al., 2002). In the work that follows, we will follow the practice of many other researchers by referring to *F. graminearum* chemotypes even though we have only carried out predictive PCR genotyping rather than a full biochemical analysis. Recently, a novel type A trichothecene, NX-2, was identified from a strain collected from Minnesota in the U.S. (Varga *et al*. 2015). Although NX-2

does not have the keto group at C-8, it is very similar to 3ADON structurally (Liang *et al*. 2014, Varga *et al*. 2015).

Several population genetic analyses have been performed with North American Fg isolates. Initially, Zeller *et al*. (2003) reported a single, panmictic population with 253 Fg isolates from Kansas and North Dakota using 94 AFLP markers. However, Gale *et al*. (2007) studied 712 isolates using 10 RFLP markers and reported diverse populations including three trichothecene types (15ADON, 3ADON, and NIV), with 15ADON predominant in the upper Midwest region in the U.S. After that, Ward *et al*. (2008) demonstrated the population subdivision in North America with 130 Fg isolates using 9 variable number of tandem repeat (VNTR) markers. At the same time, they identified evidence for a temporal shift from 15ADON to 3ADON (the frequency of 3ADON increased over 14-fold between 1998 and 2004) in western Canada (Ward *et al*. 2008). Kelly *et al*. (2015) confirmed this temporal shift happened in western Canada but not in eastern Canada with over 4,000 Fg isolates collected from wheat using 8 VNTR markers. Burlakoti *et al*. (2011) indicated a similar shift in the U.S. upper Midwest with 262 Fg isolates from barley using 10 VNTR markers. Liang *et al*. (2014) also detected a similar shift from the 15ADON to 3ADON chemotype in the upper Midwest in the U.S. with 463 Fg isolates collected from wheat using 10 RFLP markers.

Several studies that focused on pathogenic traits found evidence for a 3ADON advantage, which could explain their increase in prevalence in North America. Ward *et al*. (2008) detected higher growth rate, more trichothecene production, and more production of conidia in 3ADON isolates than in 15ADON isolates. Similarly, Puri and Zhong (2010) identified higher disease severity and more DON production in 3ADON isolates than those in 15ADON isolates on both Fusarium head blight susceptible and resistant wheat cultivars. 3ADON isolates also produced

more conidia, but the growth rate between 3ADON and 15 ADON strains did not significantly differ. von der Ohe *et al*. (2010) detected higher DON accumulation in wheat infected with 3ADON isolates, but no significant difference in aggressiveness under field conditions. Gilbert *et al*. (2010) and Foroud *et al*. (2012) identified similar results under greenhouse conditions. Moreover, co-inoculation with 3ADON and 15ADON isolates reduced trichothecene yield and virulence due to intraspecies interactions (Walkowiak *et al*. 2015). In addition, this study compared the genome sequences and transcriptome profiles of isolates from each chemotype.

Genotyping-by-Sequencing (GBS)

To better understand *Fusarium* natural or mapping populations at the level of the whole genome, we used GBS technology. Next-generation sequencing (NGS) technology provides a powerful tool for whole genome sequencing and re-sequencing in genomics research. However, the cost is still too high for large-scale genotyping studies. GBS is a targeted, complexity-reducing genome sequencing method based on restriction enzyme digestion, in which hundreds of DNA barcoded samples are pooled in the same reaction at a relatively low cost compared to whole genome re-sequencing (Elshire *et al*. 2011, Poland *et al*. 2012, Mascher *et al*. 2013, Sonah *et al*. 2013). Poland *et al.* (2012) proposed a two-enzyme GBS approach combined with a Y-adaptor method to reduce the complexity of the genome and to sequence multiple samples (up to 96 samples) in a single Illumina sequencing lane. This process generates high-density single nucleotide polymorphism (SNP) data across the genome with a relatively low per sample cost in wheat and barley. In plants, GBS has been successfully used with maize, barley, wheat, sorghum, soybean, and pea. (Elshire *et al*. 2011, Poland *et al*. 2012, Mascher *et al*. 2013, Morris *et al*. 2013, Sonah *et al*. 2013, Boutet *et al*. 2016).

GBS markers include two types, presence-absence polymorphisms in the restriction cut site between the parents in biparental population, and SNP markers around non-polymorphic restriction cut sites. Since GBS markers are sequence-based, the genetic position of GBS markers used in genetic maps can be easily connected with physical positions by blasting marker sequences against a reference genome, which provides GBS markers with an advantage compared with other markers, *e.g*., AFLPs. Therefore, GBS methods provide valuable resources for genetic and genomic research in *Fusarium*.

In fungi, Milgroom *et al*. (2014) identified strong clonal population structure and detected over 400 recombination events between clonal lineages in *Verticillium dahliae* using GBS markers. Leboldus *et al*. (2015) used a similar GBS procedure with a size selection step and without a Y-adaptor on the Ion Torrent sequencing platform with *Pyrenophora teres* f. *maculata* and *Sphaerulina musiva*. They created a genetic map for a biparental population of *P. teres* f. *teres* with 15 linkage groups based on 1393 SNPs generated from this protocol (Leboldus *et al*. 2015).

Synopsis

F. fujikuroi and *F. proliferatum* are two species within the FFSC. They are closely related, and difficult to distinguish morphologically (Leslie and Summerell, 2006); however, they have different SM profiles and cause different diseases. For example, *F. fujikuroi* is known for GA production and can cause bakanae disease in rice. *F. proliferatum* is known for fumonisin production and can cause onion bulb rot (Leslie and Summerell, 2006). Crosses between some *F. fujikuroi* and *F. proliferatum* isolates can produce a few viable progeny, and parents and progeny from the cross differ in GA production and onion virulence (Desjardins *et al*. 1997, Leslie *et al*. 2004, 2007, Mohamed Nor 2014). The genetic basis for the phenotypic divergence between the

two species is not known. No high-density genetic map existed for either of these species and no *F. proliferatum* genome sequence was publicly available at the start of this study.

In Chapter 2, I describe my construction of a high-density genetic map from a cross (Mohamed Nor 2014) between these two species, an extension of previous work but now using GBS markers generated from an adapted GBS protocol (Poland *et al*. 2012). Based on this high-density map, I performed quantitative trait locus (QTL) analysis on previously scored onion virulence and GA production traits (Mohamed Nor 2014). I detected a QTL for onion virulence, and a novel locus for GA production.

In Chapter 3, I describe my work performing *de novo* assemblies of a draft genome of *F. proliferatum*. Using the constructed high-density genetic map in Chapter 2, I anchored nearly 98% of the total assembly onto 12 linkage groups, corresponding to 12 chromosomes present in each parent species (Xu *et al*., 1995). I annotated this *F. proliferatum* genome and characterized its SM gene cluster profile. Despite the high similarity of the two genomes, certain genomic regions might play important roles in their differential plant pathogenicity.

My dissertation also includes a project that focuses on *F. graminearum*, described in Chapter 4. *F. graminearum* causes Fusarium head blight on wheat and barley worldwide, which leads to huge yield losses and severe grain contamination with toxic SMs like trichothecenes (Desjardins 2006, Leslie and Summerell 2006). The dominant mycotoxin of scab disease in the U.S. is the trichothecene deoxynivalenol (DON), and *F. graminearum* isolates can be categorized based on the relative level of two acetylated DON forms, 3ADON and 15ADON (Alexander *et al*. 2011). Recently, a new type of trichothecene toxin produced by some *F. graminearum* strains, termed NX-2, was discovered (Varga *et al*. 2015). In this project, I detected two major subpopulations of US *F. graminearum* isolates correlated to 3ADON and 15ADON chemotypes

with high quality SNP markers across the genome. In addition, I found that neither trichothecene chemotype has a clear fitness advantage in a small set of isolates from New York, but that isolates originating from one subpopulation may on average have a fitness advantage over isolates from the other subpopulation. This result may help explain the apparent contradictory evidence regarding whether 3ADON chemotypes have a selective advantage and suggests that any selective advantage is not due directly to the difference in which toxin accumulates but to other genetic changes elsewhere in the genome.

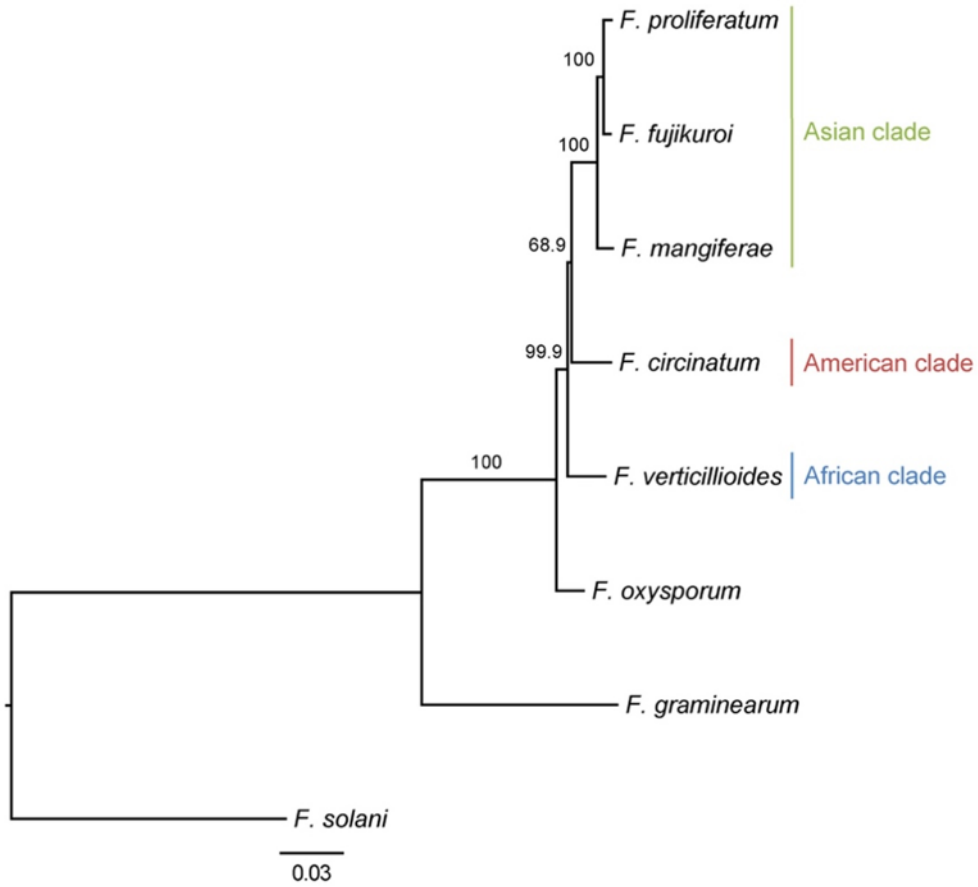

Figure 1.1 Phylogenetic relationships of *Fusarium* species.

Figure from Wiemann *et al*. 2013. Branches show bootstrap values (%), scale bar indicates amino acid substitutions per site.

Table 1.1 Secondary metabolites in *Fusarium*.

Species	Secondary metabolites produced	Reference
F. verticillioides	fumonisin, fusaric acid, fusarin, naphthoquinone, moniliformin, bikaverin, fusarubin, acorenol	Desjardins 2006, Niehaus et al. 2016
F. circinatum	beauvericin, fusaproliferin	Desjardins 2006
F. subglutinans	fusaproliferin, beauvericin, moniliformin, fusaric acid, little or no fumonisin	Desjardins 2006
F. fujikuroi	gibberellic acid, moniliformin, fusarin, fusaric acid, beauvericin, bikaverin, fusarubin, apicidin F, kaurene, acorenol, koraiol	Desjardins 2006, Niehaus et al. 2016
F. proliferatum	fumonisin, moniliformin, fusarin, fusaric acid, beauvericin, fusaproliferin, eniatin, bikaverin, fusarubin, kaurene, acorenol, gibberellic acid (orchid strain)	Desjardins 2006, Tsavkelova et al. 2008, Niehaus et al. 2016
F. graminearum	trichothecene, aurofusarin, butenolide, chlamydosporol, culmorin, cyclonerodiol, fusarins, and zearalenone	Desjardins 2006

Table 1.2 Genetic maps in *Fusarium*.

Cross	Marker type	Marker number	Linkage group	Length (cM)	Physical distance to map distance (kb/cM)	Reference
F. verticillioides x *F. verticillioides*	morphological/biochemical markers	12	4	na	na	Puhalla and Spieth 1985
F. verticillioides x *F. verticillioides*	RFLP, RAPD, morphological/biochemical markers	150	13	1452	32	Xu and Leslie 1996
F. verticillioides x *F. verticillioides*	AFLP, RFLP, RAPD, morphological/biochemical markers	636	12	2188	21	Jurgenson *et al*. 2002b
F. graminearum x *F. asiaticum (F. graminearum* lineage 6*)*	AFLP	1048	9	1286	28	Jurgenson *et al*. 2002a
F. graminearum x *F. graminearum*	STS and AFLP	237	9	1234	29	Gale *et al*. 2005
F. circinatum x *F. subglutinans*	AFLP, Mating type and Histone loci	581	12	2774	19.5	De Vos *et al.* 2007
F. fujikuroi x *F. proliferatum*	AFLP and Mating type locus	79	12	1264	34	Mohamed Nor 2014

RFLP: restriction fragment length polymorphism, RAPD: random amplified polymorphic DNA, AFLP: amplified fragment length polymorphism, STS: sequence-tagged site, na: not available.

Table 1.3 Selected *Fusarium* genomes.

Species	Strains	Size (Mb)	Chromo-some	Repeats (%)	Predicted genes	Reference
F. graminearum	PH-1	36.2	4	0.66	13,332	Cuomo *et al*. 2007
F. oxysporum	4287	59.9	15	28.1	17,735	Ma *et al*. 2010
F. solani	77-13-4	54.4	17	5.1	15,707	Coleman *et al*. 2009
F. circinatum	FSP34	43.97	unknown	1.91	15,060	Wingfield *et al*. 2012
F. verticillioides	7600	41.7	11	0.86	14,179	Ma *et al*. 2010
F. fujikuroi	IMI58289	43.9	12	4.08	14,813	Wiemann *et al*. 2013
F. fujikuroi	B14	43.8	12	unknown	14,017	Jeong *et al*. 2013
F. fujikuroi	FGSC8932	43.1	12	unknown	14,832	Chiara *et al*. 2015
F. fujikuroi	KSUX-10626	43.1	12	unknown	14,801	Chiara *et al*. 2015
F. fujikuroi	KSU3368	43.2	12	unknown	15,188	Chiara *et al*. 2015
F. mangiferae	MRC7560	46.2	11	5.29	15,804	Niehaus *et al*. 2016
F. proliferatum	62905	43.2	12	3.74	15,254	Niehaus *et al*. 2016
F. proliferatum	ET1	45.2	12	4.58	16,143	Niehaus *et al*. 2016

Chapter 2 - High-density genetic map and QTL analysis in *Fusarium*

Abstract

The *Fusarium fujikuroi* species complex contains many agronomically important plant pathogens noted for the production of secondary metabolites such as mycotoxins and gibberellic acids (GAs) that are involved in pathogen virulence. Two members of this species complex, *F. proliferatum* and *F. fujikuroi*, are closely related phylogenetically but differ for pathogen traits. Our long-term aim is to explore the genetic bases behind these differences, and we have begun by conducting QTL analyses of progeny from an interspecific cross between these species. We adapted a Genotyping-by-Sequencing (GBS) protocol to generate high-density genotype data from 253 progeny of a single cross. Illumina reads from our GBS library were mapped to the *F. fujikuroi* reference sequence to produce over 26,000 GBS loci. We constructed a genetic map containing 6,381 segregating GBS markers across the 12 chromosomes from these species by using the R/qtl package. To demonstrate the utility of the map, we performed QTL mapping on GA production and onion virulence using 240 of the 253 progeny. We identified one novel GA regulating locus on chromosome 1 in addition to the known GA biosynthetic gene cluster on chromosome 5. We also detected one significant QTL for onion virulence on chromosome 9. This high-density genetic map provides valuable information for understanding the genetic bases underlying differences between these two closely related *Fusarium* species. The adapted GBS protocol and analysis pipeline provide powerful tools for additional *Fusarium* genetic and genomic studies. The detected QTLs broaden our knowledge of genetics on GA production and onion virulence.

Introduction

Fusarium is a genus of filamentous fungi that includes the *Fusarium fujikuroi* species complex, the *Fusarium graminearum* species complex, the *Fusarium oxysporum* species complex, and the *Fusarium solani* species complex (Aoki *et al*., 2014). The *Fusarium fujikuroi* species complex contains ~50 distinct species in three major clades: Asian, American and African (Leslie and Summerell 2006, Wiemann *et al*. 2013). Species within this complex can cause diseases of many plants, including bakanae disease of rice, stalk and ear rot of maize, and pitch canker of pine. (Leslie and Summerell 2006, Aoki *et al*. 2014).

Some *Fusarium* species can synthesize secondary metabolites (SMs), compounds that are not required for vegetative growth but may be important as toxins or during fungal-host interactions. Fungal SMs are diverse in structure and biological activity, including mycotoxins that are toxic to human and animals, e.g. fumonisin, and plant growth and development regulators, e.g. gibberellins (Desjardins 2006).

F. fujikuroi and *F. proliferatum* are two closely related species within the *Fusarium fujikuroi* species complex. They are difficult to distinguish morphologically, but have different SM profiles and are pathogens of different plants (Leslie and Summerell 2006). For example, *F. proliferatum* is noted for fumonisin production and *F. fujikuroi* is noted for gibberellic acid (GA) production (Desjardins 2006). *F. proliferatum* can cause maize stalk and ear rot, onion bulb rot, and numerous other diseases, while *F. fujikuroi* is best known for causing bakanae disease in rice (Leslie and Summerell 2006).

Although they are distinct species, *F. proliferatum* and *F. fujikuroi* can interbreed and produce viable progeny, usually with *F. proliferatum* serving as the female parent and *F. fujikuroi* serving as the male parent (Desjardins *et al*. 1997, Leslie *et al*. 2004). Leslie *et al*.

(2004) showed that the mating type tester strains of *F. fujikuroi* were cross-fertile with the mating type tester strains of *F. proliferatum*, and also identified a field strain that was cross-fertile with the mating type tester strains of both species. Amplified fragment length polymorphism (AFLP) markers segregated amongst the progeny of these crosses, but not always in a 1:1 Mendelian ratio (Leslie *et al*. 2004).

The mating type tester strains of *F. fujikuroi* and *F. proliferatum* differ in GA production, fusarin production, fumonisin production, fusarubin production and pathogenicity to onion and to rice, with segregation amongst the progeny distorted for all of these traits (Studt *et al*. 2012, Mohamed Nor, 2014). For example, the *F. fujikuroi* parent can produce GA, while the *F. proliferatum* parent does not (Studt, *et al.* 2012, Mohamed Nor, 2014). Of over 400 progeny, 58% produced no GA, with the remaining progeny producing some GA but not a uniform amount (ranging from 0.1 to 3.5 mg/ml) (Mohamed Nor 2014). Progeny containing the GA biosynthetic gene cluster from the *F. fujikuroi* parental strain included producers of both low and high levels of GA, suggesting that a genomic region besides the GA gene cluster influences GA production (Studt *et al*. 2012, Mohamed Nor, 2014).

The *F. fujikuroi* parent is not pathogenic to onion while the *F. proliferatum* parent is. Onion pathogenicity/virulence segregates amongst the progeny with most being intermediate between the two parents. However, transgressive progeny that were more virulent than the *F. proliferatum* parent were observed, as were progeny with an unusual, previously undescribed pathogenicity phenotype (Mohamed Nor 2014).

The genetic basis that underlies the phenotypic divergence between *F. fujikuroi* and *F. proliferatum* is unknown, and no high-density genetic map exists for either of these species. These shortcomings limit the research in this area. To address these shortcomings, I: 1)

established a GBS protocol for *Fusarium*; 2) constructed a genetic map based on progeny from an interspecific cross between *F. fujikuroi* and *F. proliferatum*; and 3) conduct a quantitative trait locus (QTL) analysis of GA production and onion virulence to uncover the genetic difference between these two species. The hypotheses tested were that genes outside the GA gene cluster influence GA production, and that specific genetic loci contribute to onion virulence. This work provides a feasible GBS protocol for *Fusarium* and a high-density genetic map in these two species. This work also provides evidence of particular genes responsible for GA production and onion virulence.

Materials and Methods

Fungal strains

The interspecific cross between *F. fujikuroi* (FGSC8932, MAT-2, C1995) and *F. proliferatum* (FGSC7615, MAT-1, D4854) was made by using FGSC7615 as the female parent and FGSC8932 as the male parent (Leslie *et al*. 2004, Mohamed Nor 2014). Two hundred and fifty-three progeny (KSU ID 22718-23231) previously analyzed for GA production and onion virulence (Mohamed Nor 2014) were used to construct the genetic map, and 240 of these progeny were used for the QTL analysis.

DNA isolation

Each strain, originating from a single ascospore (Mohamed Nor 2014), was cultivated for two days in 30 ml of complete medium (Correll *et al*., 1987) in a 125-ml Erlenmeyer flask on an orbital shaker (150 rpm) at room temperature (22–26°C). The mycelia were harvested by filtration with non-gauze Milk Filter disks (KenAG, Ashland, OH, USA), freeze-dried in a FreeZone Freeze Dryer (Labconco Corporation, Kansas City, MO, USA), and ground to a powder in a Retsch Mixer Mill (Verder Scientific Inc., Germany) with steel beads.

Genomic DNA was isolated with a modified cetyltrimethyl ammonium bromide (CTAB) protocol (Leslie and Summerell 2006). In brief, the ground mycelia was mixed with 800 µl pre-warmed (65°C) CTAB buffer and 10 µl 2-mercaptoethanol and incubated at 65°C for 30 minutes. The resulting solution was homogenized gently with 800 µl chloroform:isoamyl alcohol (24:1) and centrifuged at 13,300 x g for 20 minutes. The upper aqueous phase was removed to a new 1.5 microcentrifuge tube and mixed well with an equal volume of isopropanol (-20°C). After 5 minutes' incubation at room temperature, the sample was centrifuged at 9,200 x g for 5 minutes. The DNA pellet was suspended in 500 µl TE (pH 8.0) (10 mM Tris base, 1 mM Na_2EDTA). To remove RNA, 2 µl RNaseA (7000U/ml, 5 PRIME, Inc., Gaithersburg, MD, USA) was added and the solution incubated at 37°C for 30 min. After incubation, the DNA solution was homogenized gently with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) and centrifuged at 13,300 x g for 10 minutes. The upper aqueous phase was transferred to a new tube, homogenized with an equal volume of isopropanol (-20°C), incubated at room temperature for 5 minutes, and centrifuged at 9,200 x g for 5 minutes. Isopropanol was drained off and the DNA pellet was washed with 70% ethanol (-20°C) and 100% ethanol (-20°C), resuspended in TE (pH 8.0) and stored at -20°C. DNA quality was checked by separation in a 1% agarose gel. DNA concentration was measured in a PicoGreen assay by using the manufacturer's protocol (Life Technologies, Carlsbad, CA, USA) and the assay results were read in a Synergy H1-hybrid Reader (BioTek Instruments, Inc., Winooski, VT, USA).

GBS library construction

The genomic DNAs (FGSC8932, FGSC7615, progeny 10710) from the cross were digested with *PstI* (High-Fidelity, 6-base cutter), *MspI* (4-base cutter) or *MspI*/ *PstI* following the manufacturer's protocol (New England Biolabs, Ipswitch, MA, USA). The amount of adapter

was optimized by using genomic DNAs (FGSC7615 and progeny 170815) with different volumes (2 µl, 4 µl, 6 µl, 8.3 µl) of mixed adapters containing 0.4 µM barcode adapter and 5 µM Y-adapter from the Poland lab. The PCR annealing temperature was optimized on a gradient block from 65.1ºC to 68ºC with DNA from FGSC7615 and using the PCR protocol in Poland *et al*. (2012). The PCR annealing temperature also was tested with the pooled library DNA in a gradient block from 60 ºC to 67.6 ºC.

The library was constructed with a double restriction enzyme protocol (Poland *et al.,* 2012) adapted for *Fusarium*. In brief, 100 ng of high quality genomic DNA was digested with 8 U each of *PstI* (High-Fidelity) and *MspI* in a 20 µl reaction volume of NEB CutSmart® Buffer (New England Biolabs, Ipswitch, MA, USA) at 37ºC for 2 hours followed by enzyme inactivation at 65ºC for 20 minutes. Adapters from the Poland lab with 4 barcodes assigned to each sample were ligated to the digested DNA with 200 U NEB T4 DNA ligase in 36 µl of NEB CutSmart® Buffer at 22ºC for 2 hours followed by 65ºC for 20 minutes to inactive the ligase. A single library in one tube was created by pooling individual libraries from 96 ligated samples, and the pooled library was amplified in 24 cycles of PCR (95ºC 30 sec, 67.6ºC 110 sec). After purification of the amplified library with a QIAquick PCR Purification Kit (Qiagen, Germany), the quality of the library was checked in a BioAnalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA), and the concentration of the library was measured with a Qubit Fluorometer (Life Technologies, Carlsbad, CA, USA). Completed libraries were sequenced on an Illumina MiSeq with single-read 300 cycle kits (Illumina, Inc. San Diego, CA, USA).

Sequence trimming

Raw sequence reads from the Illumina MiSeq were separated by barcode by using Geneious software (Geneious version 7, http://www.geneious.com, Kearse *et al.*, 2012).

Sequences from the parents and each progeny were quality trimmed from the 3' end until quality score of remaining bases were greater than 20, and all sequences shorter than 15 bp in length after trimming were discarded using a custom Galaxy pipeline (Goecks *et al*. 2010, Appendix A). Trimmed sequences were mapped to the *F. fujikuroi* parent sequence (FGSC8932, Chiara *et al.* 2015) using Bowtie2 (Langmead and Salzberg 2012). Using the STACKS software, the BAM files of trimmed reads mapped to the Ff parent were analyzed; distinct GBS loci were defined across all samples and genotypes were called for each locus across all progeny (Catchen *et al*. 2011).

Genetic map construction

The construction of the genetic map had two stages. The preliminary genetic map was constructed from the first 79 progeny genotyped. GBS loci that were monomorphic, produced GBS products that were unique to either parent, or were lacking too many genotype calls were dropped. After computing the recombination fraction with the '*est.rf*' function of the R/qtl package (version 1.35-3) for all remaining pairs of GBS markers and testing for segregation distortion of these markers, markers that showed segregation distorted at the 0.05 significance level after a Bonferroni correction were dropped, and when markers had a recombination fraction of 0 between them, markers were dropped to remove this redundancy (Broman *et al.*, 2003). The remaining markers were used to construct a genetic map with Haldane's mapping function as implemented in the R/qtl package (version 1.35-3, Broman *et al*. 2003). Linkage groups were inferred with pairwise recombination fraction information by using the '*formLinkageGroups*' function. Based on the GBS markers' *F. fujikuroi* parent supercontig mapping information from STACKs, linkage groups were split into multiple parts or merged with other groups as necessary

to yield 12 linkage groups that corresponded to the 12 homologous chromosomes common to *F. fujikuroi*, *F. proliferatum* and *F. verticillioides*.

Markers showing segregation distortion are often clustered in the genome when they represent true distortion and are not due to technical artifacts. To decrease the size of large gaps in linkage groups, markers that had been removed because of segregation distortion but were predicted to belong to the linkage groups with the large gaps based on the contigs on which they were found were re-inserted into these linkage groups. After that, redundant markers were rechecked and markers removed to eliminate redundancy.

Subsequently, 174 more progeny were genotyped using the same procedure. GBS markers that were monomorphic or were missing too many genotype calls were dropped, while the remaining markers were retained for genetic map construction. The GBS marker sequences were blasted against the *F. fujikuroi* IMI58289 genome sequence (Wiemann *et al*. 2013) to determine marker position in this genome. Combining the preliminary linkage groups information, supercontig mapping information, and marker sequence blast results, each marker was assigned to a linkage group. The marker order within each linkage group was determined by the '*orderMarker*' and '*ripple*' functions. The marker order of the linkage map was compared with their order based on the physical map, and the final order was determined by using the '*compareorder*' function based on the LOD score.

QTL analysis

QTL interval mapping was performed on gibberellic acid production and virulence on onion (Mohamed Nor 2014) with a Haley-Knott regression method as implemented in the R/qtl package (Broman and Sen, 2009). For each phenotype, genome scans for single and two interacting QTLs were performed using the functions '*scanone*' and '*scantwo*' respectively. The

LOD scores that corresponded to an $\alpha = 0.05$ statistical significance threshold for possible QTL peaks identified in each phenotype-scan combination were determined by performing permutation tests on the phenotype data by using the '*scanone*' and '*scantwo*' functions with argument '*n.perm=1000'*. All significant QTL identified in either scan for each phenotype were combined into a multiple-QTL model using the function '*fitqtl*' and the best QTLs were determined by using the '*stepwiseqtl*' function which performs forward selection and backward elimination of the QTL included in the model. Interval estimates of QTL locations were determined by the function '*bayesint*', which generates 95% Bayes credible intervals. Genomic sequence analysis of QTL regions was performed using Geneious software (Geneious version 7, http://www.geneious.com, Kearse *et al.*, 2012).

Results

Optimization of GBS library preparation

The quality of the genomic DNA from *F. fujikuroi* FGSC8932, *F. proliferatum* FGSC7615, and the cross progeny was checked in a 1% agarose gel and enzyme digestion assay (Figure 2.1). The DNA was high molecular weight and high quality without contaminations, so the *Fusarium* DNA prepared with our DNA extraction protocol was suitable for GBS analysis.

To optimize the GBS protocol for *Fusarium*, varying amounts of adapter and polymerase chain reaction (PCR) annealing temperatures were evaluated. For the adapters, 2 µl, 4 µl, 6 µl, 8.3 µl of adapter (0.4 µM barcode adapter and 5 µM Y-adapter) were tested in reactions with one parent (*F. proliferatum* FGSC7615) and one progeny (170815). The lowest level of adapters yielded the least adapter dimer (Figure 2.2). PCR annealing temperatures from 60°C to 68°C were tested on parent FGSC7615 DNA and pooled GBS library DNA. The PCR with annealing temperature 67.6ºC produced the least primer dimer and the highest density of library molecules

with higher molecular weight in both assays (Figure 2.3). Therefore, the optimal condition for GBS with DNA from this cross was 2 μl adapters and a 67.6°C PCR annealing temperature.

GBS library construction and sequencing

High molecular weight genomic DNA from 253 progeny and both parents of the cross was used to construct GBS libraries. For the parents, FGSC7615 and FGSC8932, 5 replicates of each were included in a library to maximize the number of parental sequence reads and to identify as many loci as possible. Three GBS libraries were constructed. Almost no primer dimer or adapter dimer was produced in the 1st GBS library, and the unimodal distribution of fragment sizes had an average of 300 bp (Figure 2.4), indicating a high quality GBS library. The average fragment sizes in the other two GBS libraries were 260 bp and 280 bp, respectively. These libraries also were of high quality (data not shown). All three libraries were diluted to 4 nM for sequencing in separate Illumina MiSeq runs.

Processing GBS reads

Three 96-well plates of GBS samples (three total libraries) generated over 60 million raw reads with an average size of 300 bp. Geneious was used to group the raw reads for each sample on the basis of the 4 barcodes specific to each well of the 96 wells per plate that made up each pooled library (Figure 2.5). Reads split by sample served as input for a Galaxy analysis pipeline (Appendix A). In this pipeline, reads were trimmed and filtered for minimal quality thresholds and then mapped to the FGSC8932 genome to create BAM files for each progeny using Bowtie2 (Chiara *et al.* 2015, Langmead and Salzberg 2012). We used Stacks to scan these BAM files and defined more than 26,000 GBS loci that had a minimum of two reads in at least one parent from this interspecific cross (Catchen *et al*. 2011). These loci included those with reads in nearly all progeny but sequence differences between the two parents, and presence/absence loci that were

present in most progeny inheriting the marker region from one parent but absent in the second parent and in all progeny inheriting the region from the second parent. Loci of both of these types with relatively low levels of missing data were chosen for genetic map construction.

Genetic map construction

Based on GBS locus sequences recovered from the parent strains, we defined 26,000 GBS loci. For our preliminary genetic map using 79 progeny, 2,685 GBS markers were retained after excluding markers that were monomorphic, specific to one parent, or were missing too many genotype calls. Of these 2685 markers, 1683 were redundant, *i.e*., occupied the same position on a genetic map and 216 were significantly distorted, and both redundant and distorted markers were excluded. The remaining 786 markers were used to construct the preliminary genetic map using functions from the R/qtl package. Initially, 15 linkage groups composed of multiple GBS markers were identified, but linkage groups were split or merged based on information about to which supercontig of the *F. fujikuroi* parent each marker mapped until we reached 12 linkage groups. Linkage groups 3, 5, and 7 contained large gaps, so 81 markers that had been removed due to segregation distortion but predicted to correspond to these linkage groups were re-inserted into the map. After removing an additional 189 redundant markers, 668 polymorphic markers derived from 79 progeny were included in the preliminary linkage map. Most of the supercontigs from the *F. fujikuroi* parent genome assembly (Chiara *et al*. 2015) could be assigned to one of the 12 linkage groups.

After producing GBS genotypes for an additional 174 progeny, 6,381 markers had genotype data available from at least 127 of the 253 total progeny (> 50%) in the cross, which were selected to construct the genetic map. The final genetic map had a total length of 1900 cM and an average distance between distinct markers of 1.4 cM (Figure 2.6, Table 2.1). The map

contained 12 linkage groups, which is consistent with published pulsed field gel electrophoresis (PFGE) results documenting 12 chromosomes in both *F. fujikuroi* and *F. proliferatum* (Xu *et al.*, 1995). This map is the first high-density genetic map for this cross and for either of the parent species.

Linkage group 1 contained 940 markers, the largest number of any linkage group, while the smallest linkage group, 12, contained only 39 markers. The number of markers on the other 10 linkage groups ranged from 342 to 720 per linkage group, with a genome-wide average of > 3 markers per cM. The map distance between adjacent markers ranged from 1.4 to 76.3 cM with the largest gap on linkage group 3 (Figure 2.6, Table 2.1).

Genetic distances between markers on the map were compared to physical distances inferred by using the *F. fujikuroi* IMI58289 genome sequence (Wiemann *et al*. 2013, Figure 2.7). The 12 linkage groups matched perfectly with the corresponding 12 chromosomes from the *F. fujikuroi* genome. Moreover, the genetic marker order within each linkage group generally matched its physical order on each chromosome. The largest gap (76.3 cM) in the map is on linkage group 3 and corresponds to a physical gap within a single FGSC8932 contig of less than 10 kb, indicating an excessive number of recombination events in this region (Figure 2.7). Some sets of progeny that come from the same perithecium share common recombination events in this and other regions (Toomajian, unpublished). Further studies are needed to understand the high level of crossing over in this interval. Different chromosome lengths and low levels of similarity between chromosome 12 from *F. fujikuroi* (~700 kb, Wiemann *et al*. 2013) and from *F. proliferatum* FGSC7615 (~400 kb, Chapter 4) resulted in relatively few (only 39) markers on the genetic map of this chromosome (Figure 2.7).

Markers that all map at 8.91cM on linkage group 1 (*i.e*., not separated by any recombination events) span a ~49 kb genomic region on chromosome 1 of *F. fujikuroi* IMI58289. We found the sequences from the parents in this region are inverted relative to one another (unpublished *F. proliferatum* genome). Inversions act as crossover suppressors by interfering with chromosome pairing and by rendering progeny that incur an odd number of crossovers within the inverted region inviable (Kirkpatrick 2010). This region contained 19 genes, including a secondary metabolite gene cluster non-ribosomal peptide synthetase gene (NRPS21) (Wiemann *et al*. 2013).

One progeny (200108) was missing a portion of the end of chromosome 1, corresponding to >0.5 Mb of that chromosome. Based on the annotated *F. fujikuroi* IMI58289 genome sequence (Wiemann *et al*. 2013), 193 genes including 3 SM gene clusters are located in this region (the apicidin gene cluster, NRPS31, and two clusters with unknown metabolite products, PKS2 and STC1).

QTL analysis of gibberellic acid (GA) production

We conducted a QTL analysis of GA production with the 240 progeny that were used to construct the genetic map. The *F. fujikuroi* parent produces GA, but the *F. proliferatum* parent does not (Studt *et al*. 2012). GA was extracted with ethyl acetate and was measured by using HPLC (high performance liquid chromatography) from the progeny of the cross after growing in culture (for full details of this work, see Mohamed Nor 2014, Figure 2.8). The *F. proliferatum* parent produced 0 mg/ml of GA, while the *F. fujikuroi* parent produced 0.99 mg/ml of GA. Of the 240 progeny used in this study, 130 produced from 0.1 mg/ml to 2.9 mg/ml of GA. Over half of the GA-producing progeny (78 out of 130) produced ≤ 0.5 mg/ml of GA. Only 3 progeny

produced GA >2 mg/ml, and 49 out 130 produced GA in amounts ranging from 0.6 to1.8 mg/ml. The remaining 110 progeny produced no detectable GA in this study (Mohamed Nor 2014).

Following QTL mapping of GA production levels, one major QTL was detected on chromosome 5, and two minor QTLs were found on chromosome 1 (LOD threshold = 3.03 at 5% significant level, Figure 2.9). GA production is related to a well-known gene cluster at one end of chromosome 5, which contains multiple enzyme encoding genes in this biosynthetic pathway (Tudzynski, 1999, 2005, Wiemann *et al*. 2013). Based on the *F. fujikuroi* IMI58289 genome sequence, 3 genes (*FFUJ_14337*, *FFUJ_14336*, *FFUJ_14335*) of the GA biosynthetic gene cluster were in the 95% Bayes credible interval of this major QTL interval on chromosome 5 near position 178.4 cM (0.8 cM interval with 20.1 kb corresponding genomic region) (Figure 2.10, left panel). These results confirm that the GA gene cluster contributes to the GA phenotype.

We also detected two adjacent minor QTLs on chromosome 1 (Figure 2.9). One QTL was near position 189.8 cM and the other QTL was near position 190.0 cM. Most of the progeny carrying the *F. fujikuroi* main GA cluster on chromosome 5 produced GA. Progeny that also carried the *F. fujikuroi* allele on chromosome 1 QTL had significant higher GA production than did progeny with the *F. proliferatum* allele for this region ($p = 0.0008$, Figure 2.13). Progeny carrying the *F. proliferatum* GA cluster on chromosome 5 produced almost no GA in average. No significant difference in GA production existed when these progeny were split on the basis of which allele from this chromosome 1 QTL region they carried ($p = 0.3019$, Figure 2.13). Therefore, the QTL on chromosome 1 interacts epistatically with the GA cluster on chromosome 5. The physical size of the 0.5 cM 95% Bayes credible interval (combined interval from 189.9 to 190.3 cM on chromosome 1) was 17 kb and contained 5 genes (Figure 2.10, right panel). One

gene (*FFUJ_00242*, 1575 bp, from 5,901,507 bp to 5,903,081 bp) in the *F. fujikuroi* IMI58289 genome was identified as a probable COP9 signalosome complex subunit 2, which is involved in the regulation of protein degradation (Figure 2.11). Comparison of this gene and its homolog in the *F. proliferatum* genome indicated high similarity in gene sequence and predicted coding sequence (CDS) (Figure 2.12 top right). The remaining 4 genes were uncharacterized in the *F. fujikuroi* IMI58289 genome. Blast searching in the Conserved Domain Database (CDD, Marchler-Bauer and Bryant, 2004; Marchler-Bauer *et al*. 2011) indicated that they contained the following conserved domains: Methyltransferase, Peptidase_C65, Major Facilitator Superfamily (MFS), and Metal Ion Transporter (MIT)_CorA-like superfamily (Figure 2.11). A 15 bp (5 aa) indel was found between the gene *FFUJ_00241* and its homologous gene in *F. proliferatum.* Comparison of the remaining 3 genes (*FFUJ_00240*, *FFUJ_00243,* and *FFUJ_00244*) between *F. fujikuroi* and *F. proliferatum* indicated they are highly similar.

QTL analysis of virulence on onion

The onion virulence assay was performed on 240 progeny and both parents from the cross (Mohamed Nor 2014). The *F. proliferatum* parent is pathogenic (2 on the virulence rating scale), and the *F. fujikuroi* parent is not (0 on this scale) (Figure 2.14). The disease rating clearly segregates amongst the progeny. Among 240 progeny, 72 were non-pathogenic like the *F. fujikuroi* parent, and 74 were pathogenic and had similar virulence to the *F. proliferatum* parent. Eighteen transgressive progeny were more virulent than the *F. proliferatum* parent, and 76 progeny were less virulent than the *F. proliferatum* parent (but still more virulent than the *F. fujikuroi* parent).

In a QTL analysis of the 240 progeny for onion virulence, one significant QTL (LOD threshold = 3.21 at 5% significant level) was detected corresponding to a 4.5 cM 95% Bayes

credible interval on chromosome 9 (Figures 2.15 & 2.16). The corresponding genomic region (76.4 kb) located on one end of chromosome 9 contained 30 genes (Figure 2.17 top). Surprisingly, half (15/30) of the genes belonged to the fumonisin biosynthetic gene cluster.

In addition, Mohamed Nor (2014) identified a novel onion pathogenicity phenotype referred to as blister, which appeared as swollen onion tissue on the outer scale around the inoculation site. We did not detect any QTL for this trait in our study.

Discussion

Two closely related *Fusarium* species, *F. proliferatum* and *F. fujikuroi*, are morphologically identical, but have different secondary metabolite and pathogenicity phenotypes. We constructed a high-density genetic map with 6381 GBS markers from an inter-specific cross between strains of these two species. With this high-density genetic map, we performed QTL analyses of GA production and onion virulence with the progeny used in map construction.

By adding the freeze-drying and mill grinding step into the DNA extraction protocol, we isolated large amounts of high molecular weight genomic DNA effectively from nearly 1,000 samples with low cost in our laboratory. This modified protocol not only builds a very confident foundation of our research, but also provides a practical DNA isolation method for NGS applications in *Fusarium* fungi. The high-density genetic map was constructed by adapting the Genotyping-by-Sequencing method, which allows for simultaneously genotyping large numbers of samples (Elshire *et al*. 2011, Poland *et al*. 2012). The GBS protocol was optimized by reducing the adaptor amount and increasing the PCR annealing temperature used in GBS for wheat and barley (Poland *et al*. 2012). These adjustments yield high quality GBS libraries without obvious primer dimers or adaptor dimers (Figure 2.4). Three high quality GBS libraries

yielded over 60 million sequence reads from Illumina MiSeq. To split reads based on barcodes and trim and remove low quality raw sequencing reads, we built a bioinformatics pipeline that combined Geneious and Galaxy software. With this pipeline, raw sequence reads could be trimmed and mapped to a reference genome to get BAM files. Over 26,000 GBS markers were inferred from these BAM files in Stacks.

About one quarter of the 26,000 GBS markers were used to construct the first high-density genetic map of a cross between *F. proliferatum* and *F. fujikuroi* or of either parent species. Since the total genome size of *F. fujikuroi* IMI58289 is 43.8 Mb (Wiemann *et al*. 2013), one cM corresponds to on average ~23 kb. For each chromosome, this value varied from 13.3 kb to 32.7 kb. Jurgenson *et al*. (2002b) reported one cM corresponds to ~21 kb on average in the *F. verticillioides* genome. Gale *et al*. (2005) reported one cM corresponds to ~29 kb on average in the *F. graminearum* genome.

The high-density genetic map was used for QTL analyses of GA production and onion virulence amongst the progeny of the cross. Previously, a seven-gene GA gene cluster (Tudzynski and Holter, 1998) was linked to GA biosynthesis in *F. fujikuroi,* and the global nitrogen regulator *AreA* and *AreB* homologs regulated expression in this gene cluster along with other minor regulators (Mihlan *et al*. 2003, Michielse *et al*. 2014). The *velvet*-like global regulators also can to regulate GA production (Wiemann *et al.* 2010). We detected a major QTL for GA production on chromosome 5. This QTL interval includes the well-known GA gene cluster, which is consistent with this cluster's proven role in GA production (Tudzynski and Holter, 1998). The known GA regulator, *area-GF*, is located on chromosome 6 (Tudzynski *et al*. 1999), and components of the *velvet*-like regulatory complex (*FfVel1, FfVel2, FfLae1*) are located on chromosome 1 (Wiemann *et al*. 2010). More specifically, *FfVel1* is located near 1.3

Mb, *FfVel2* is located near 1.7 Mb, and *FfLae1* is located near 4.8 Mb. Therefore, the QTL region on chromosome 1 detected in this study (located near 5.9 Mb) represents a novel locus. The QTL we detected also has a significant epistatic interaction with the major QTL on chromosome 5 (Figure 2.13). A sequence encoding a putative COP9 signalosome complex subunit 2 (*FFUJ_00242*) exists within the genomic region on chromosome 1 that corresponds to the QTL (Wiemann *et al.* 2013). This complex plays an essential role in protein degradation through ubiquitin-proteasome pathway (Schwechheimer *et al*. 2001). The complex also is involved in regulation of secondary metabolism in *Aspergillus nidulans* (Nahlik *et al*. 2010, Bayram and Braus, 2011). We suggest that the *F. fujikuroi* homolog of this complex regulates GA production, and perhaps that of other secondary metabolites in *Fusarium* as well. These results are consistent with the hypothesis of Studt *et al.* (2012) that other regulators of GA production exist outside the GA gene cluster. The other four genes in this region also could contribute to this QTL, and their functions need to be validated experimentally.

A significant QTL related to onion virulence mapped to chromosome 9 (Figure 2.15, 2.16). The corresponding genomic region in the *F. proliferatum* parent contains the entire fumonisin biosynthetic gene cluster (15 genes total), while the *F. fujikuroi* parent lacks seven of the genes in the cluster (Figure 2.17 bottom, Chiara *et al*. 2015). The *F. proliferatum* parent and at least some of the progeny used in this QTL study can synthesize fumonisin, but not the *F. fujikuroi* parent (Studt *et al*. 2012). *F. proliferatum* can cause bulb rot of onions (du Toit *et al*. 2003), and reports suggest that fumonisins may play a role in this disease and diseases of other crops (Stankovic *et al*. 2007, Arias *et al*. 2012). The location of the QTL for onion virulence supports the hypothesis that fumonisins have an important role in onion pathogenicity or

virulence. The size of this QTL is rather large, however, and it is possible that one or more of the genes outside the cluster but within the QTL also play a role in the pathogenicity process.

F. proliferatum and *F. fujikuroi* are closely related species, with different SM profiles and that can cause different diseases. The genetic bases underlying these phenotypic differences are unclear. The lack of a high-density genetic map limited the research to understand this genetic basis. To address these knowledge gaps, we established a feasible GBS protocol to analyze the genetic variation in *Fusarium.* This protocol provides an effective tool for *Fusarium* genetic and genomic studies. With the GBS markers, we created the first high-density genetic map from an interspecific cross between *F. proliferatum* and *F. fujikuroi.* This map laid a solid foundation for further understanding the genetic basis behind the phenotypic divergence between these two species. To test our hypotheses that other loci besides the GA gene cluster influence GA production and that genetic loci contribute to variability in onion virulence among the progeny, we conducted a QTL analysis of GA production and onion virulence. We detected one novel regulator locus for GA production and one onion virulence locus that includes the fumonisin biosynthetic gene cluster. The detected loci need to be further validated experimentally. These findings enable the identifications of genes responsible for GA production and onion virulence.

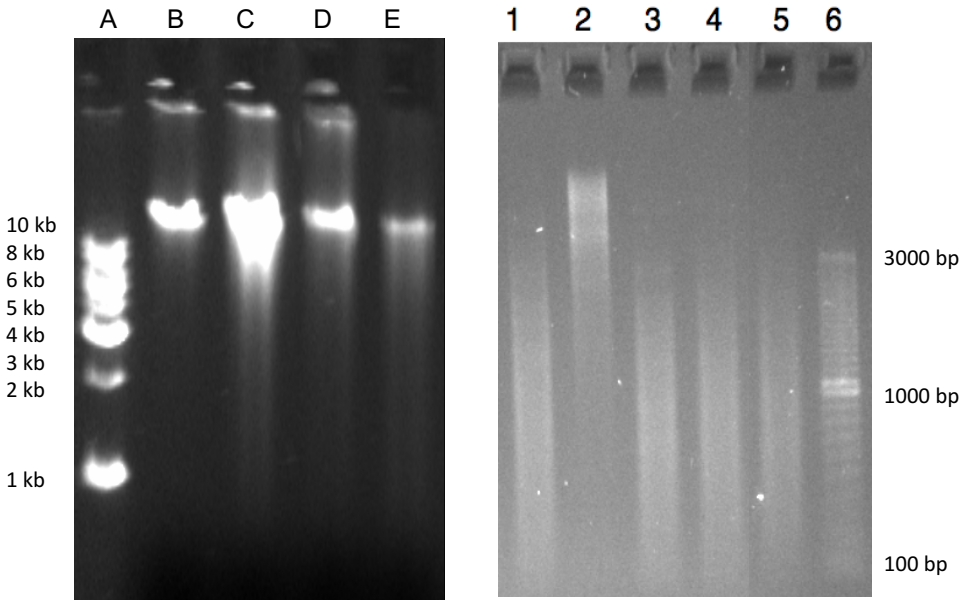

Figure 2.1 High molecular weight genomic DNA (left panel) and enzyme digested product (right panel).

Left panel: A: 1 kb ladder, B: λ DNA, C: FGSC8932, D: FGSC7615, E: progeny 10710; right panel: 1: FGSC7615 digested with *Msp I* (4-base cutter), 2: FGSC7615 digested with *PstI* (6-base cutter), 3: FGSC7615 digested with *Msp I*/ *Pst I*, 4: FGSC8932 digested with *Msp I*/ *Pst I*, 5: progeny 10710 digested with *Msp I*/ *Pst I*, 6: 100 bp DNA ladder.

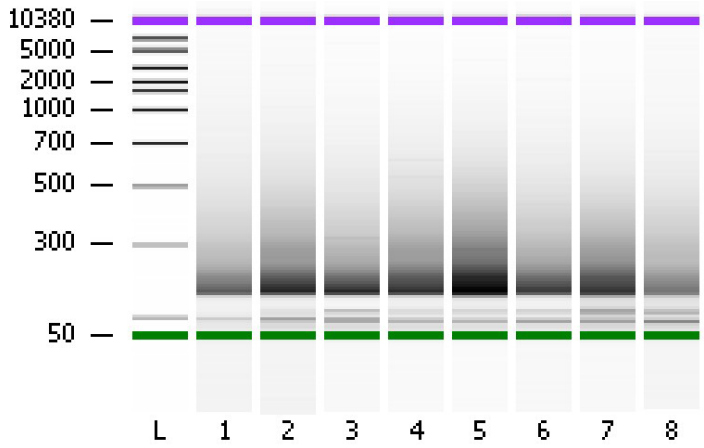

Figure 2.2 BioAnalyzer result of GBS adapter amount assay.

L: DNA ladder; 1-4: 2 µl, 4 µl, 6 µl, and 8.3 µl adapters (0.4 µM barcode adapter and 5 µM Y-adapter) on DNA FGSC7615; 5-8: 2 µl, 4 µl, 6 µl, and 8.3 µl adapters (0.4 µM barcode adapter and 5 µM Y-adapter) on DNA 170815. The first and second band above 50 bp represents the primer dimer and adapter dimer in each lane, respectively.

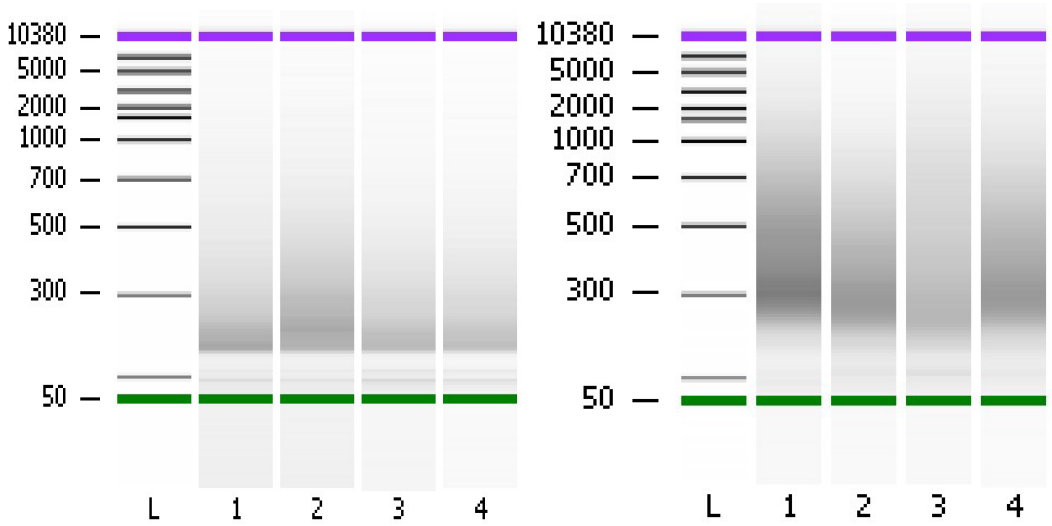

Figure 2.3 BioAnalyzer result of PCR annealing temperature assay.

Left panel: L: DNA ladder; 1-4: 68°C, 67.6°C, 66.7°C, 65.1°C on DNA FGSC7615; right panel: L: DNA ladder; 1-4: 67.6°C, 65.1°C, 61.6°C, 60°C on pooled library DNA. The first band above 50 bp represents the adapter dimer in each lane.

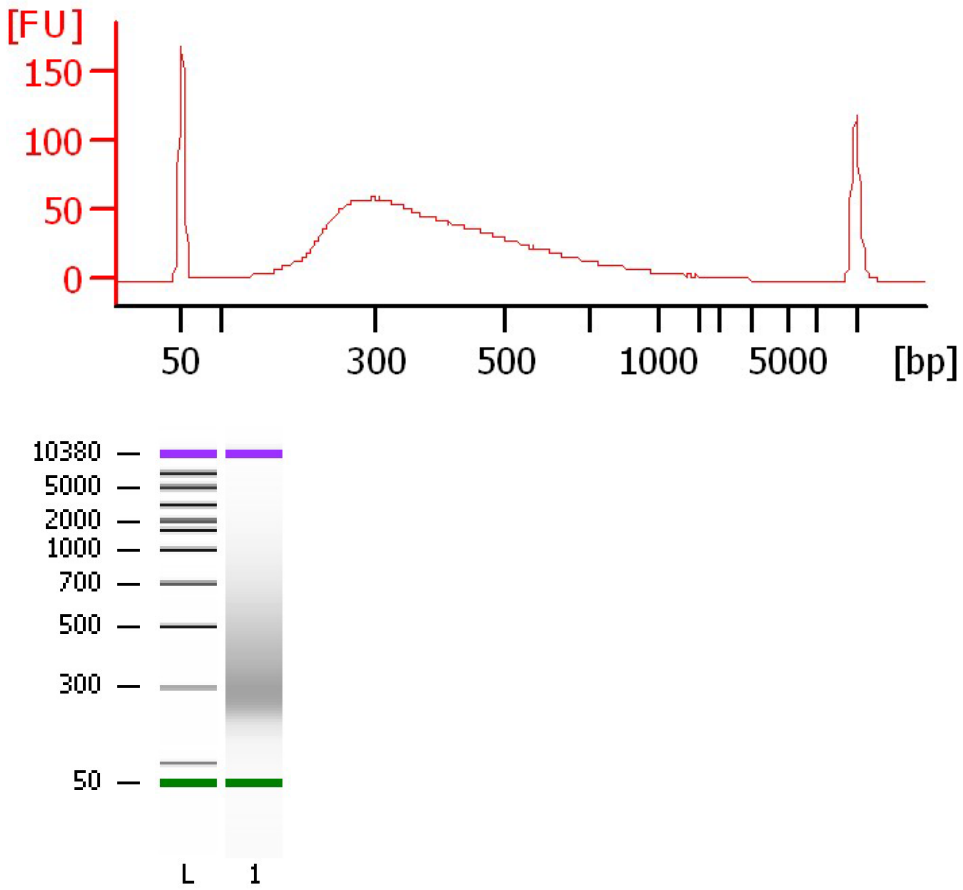

Figure 2.4 BioAnalyzer result of GBS library.

Upper panel: The x-axis represents fragment size (bp), and y-axis represents fluorescence units (FU). The GBS library shows a continuous unimodal fragment distribution with the mode at approximately 300 bp. Two sharp marker peaks, 50 bp and 10380 bp, flank the library distribution but are not part of the library. Lower panel: electrophoresis profile of the same GBS library shown in the upper panel, L: DNA ladder, 1: GBS library.

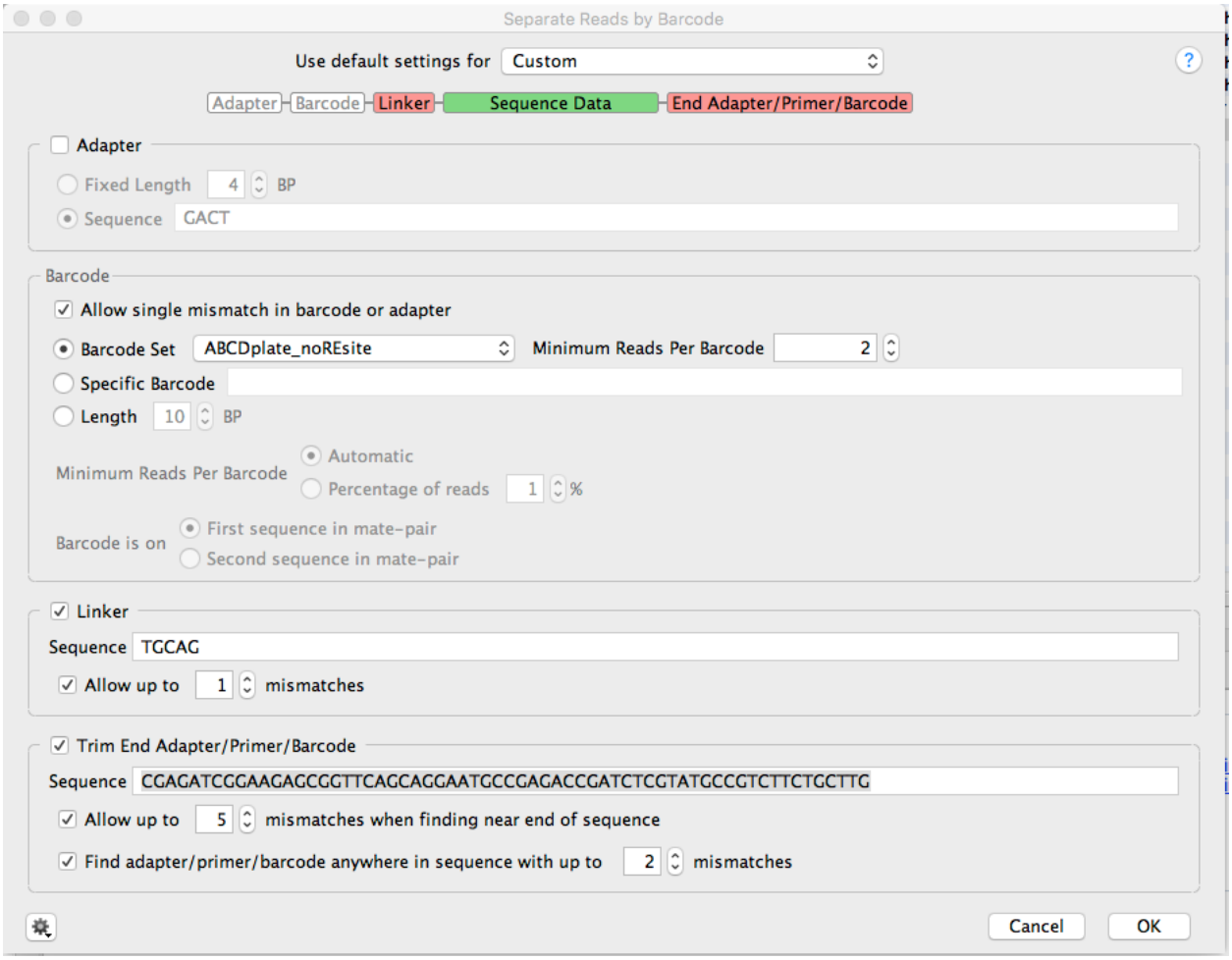

Figure 2.5 Geneious setting for separating reads by barcode and adapter trimming

Only a single mismatch is allowed in both the barcode and linker (which represents the remnant of the *PstI* cut site on the immediate 5-prime flank of the unselected sequence of each molecule in our library), a minimum of 2 reads are required for each barcode, and 5 mismatches are allowed in the 3-prime adapter if found near the end of a sequence (*e.g.*, if the GBS locus is much shorter than the sequencing read length).

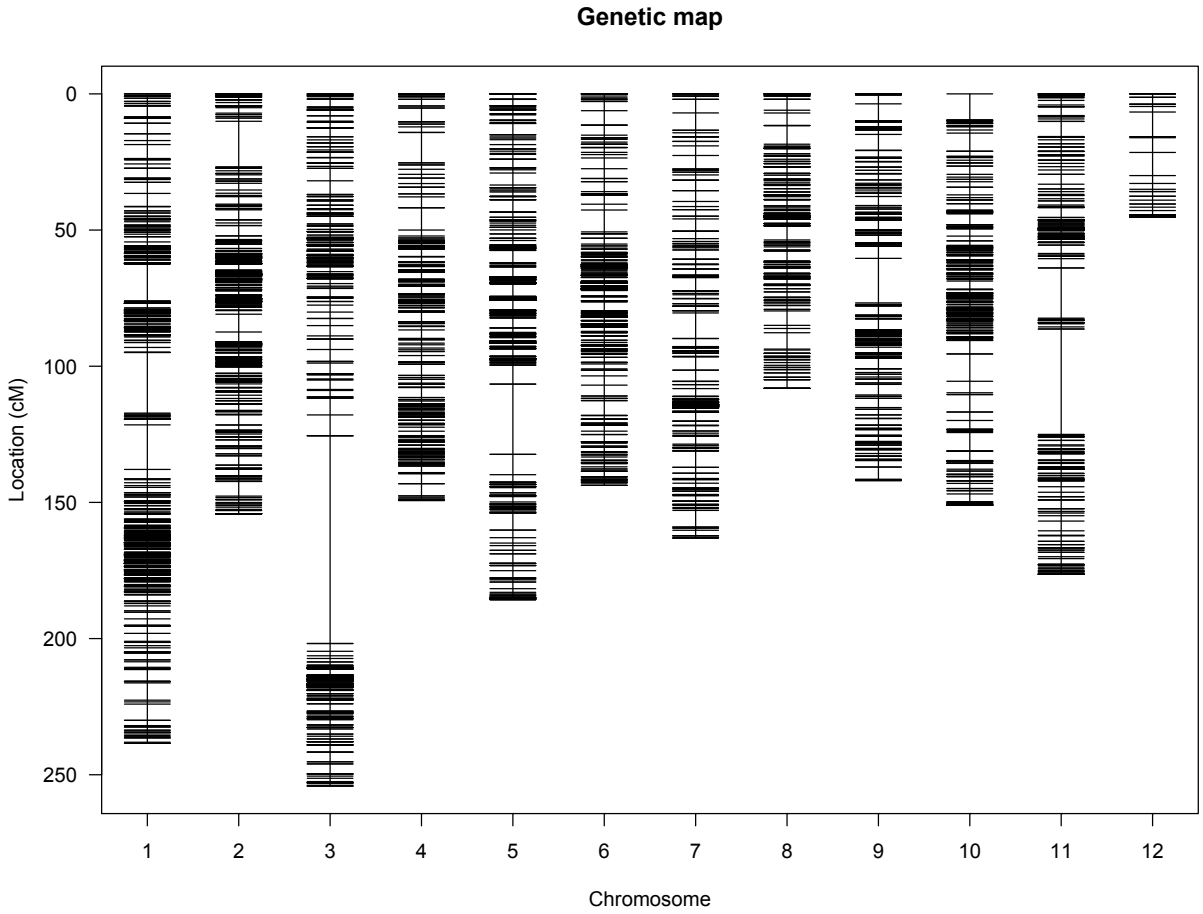

Figure 2.6 Genetic map of cross *F. fujikuroi* and *F. proliferatum*.

The map is constructed with 6381 GBS markers from 253 progeny. The largest gap (76.3 cM) is located on linkage group 3.

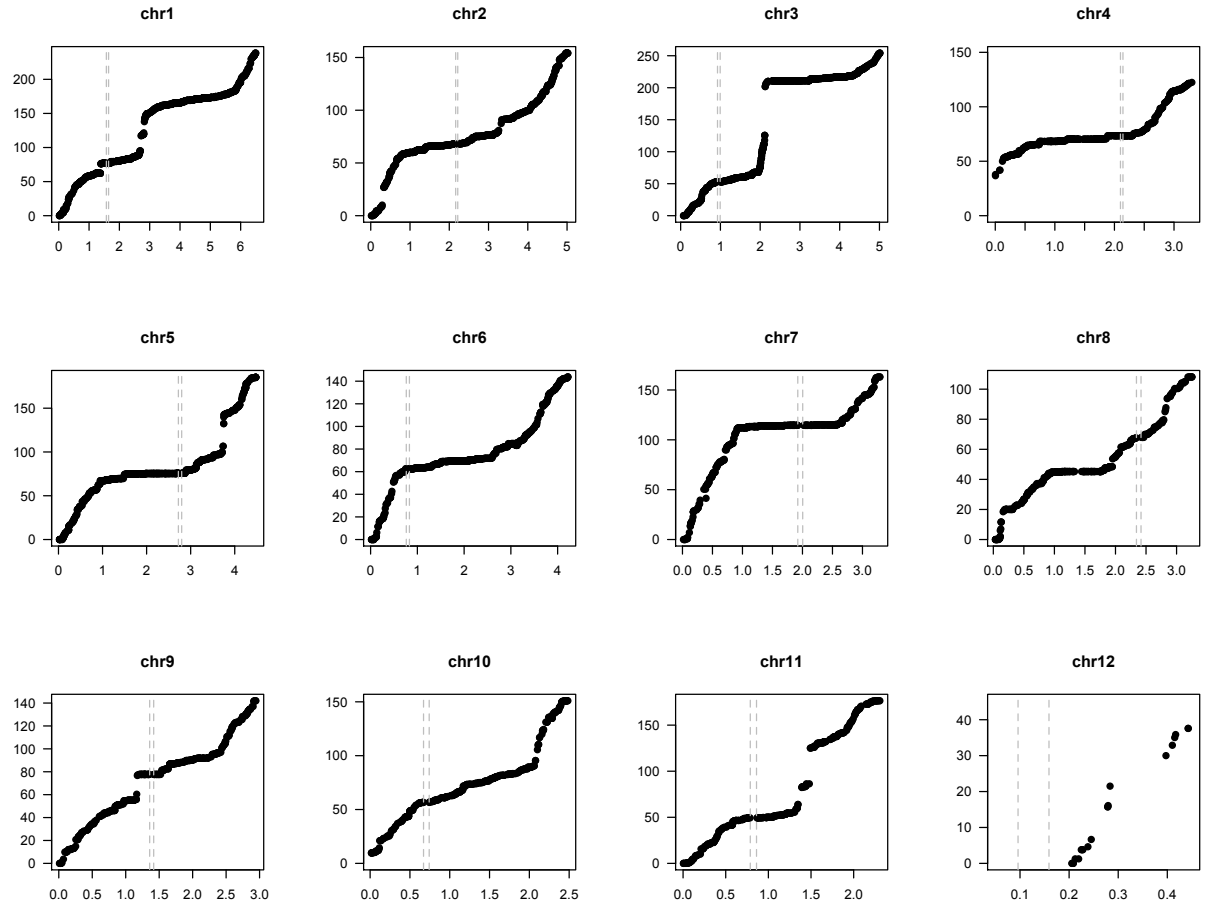

Figure 2.7 Comparison of genetic location and physical location.

In all 12 single plots, the x-axis represents physical position in units of Mb (million base pairs) for each GBS marker, and the y-axis represents their genetic position in units of cM. The interval between each pair of dashed lines is the centromere region for each chromosome (Wiemann *et al*. 2013).

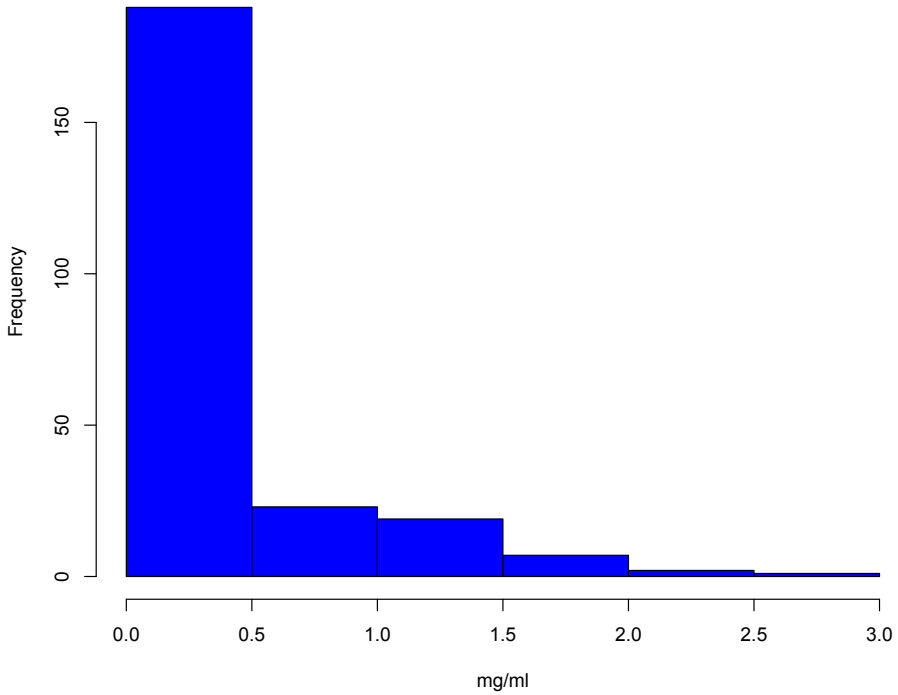

Figure 2.8 Histogram of gibberellic acid production of 240 progeny in the cross between *F. fujikuroi* and *F. proliferatum.*

The x-axis represents the gibberellic acid level, and the y-axis represents the count of progeny falling into each histogram range.

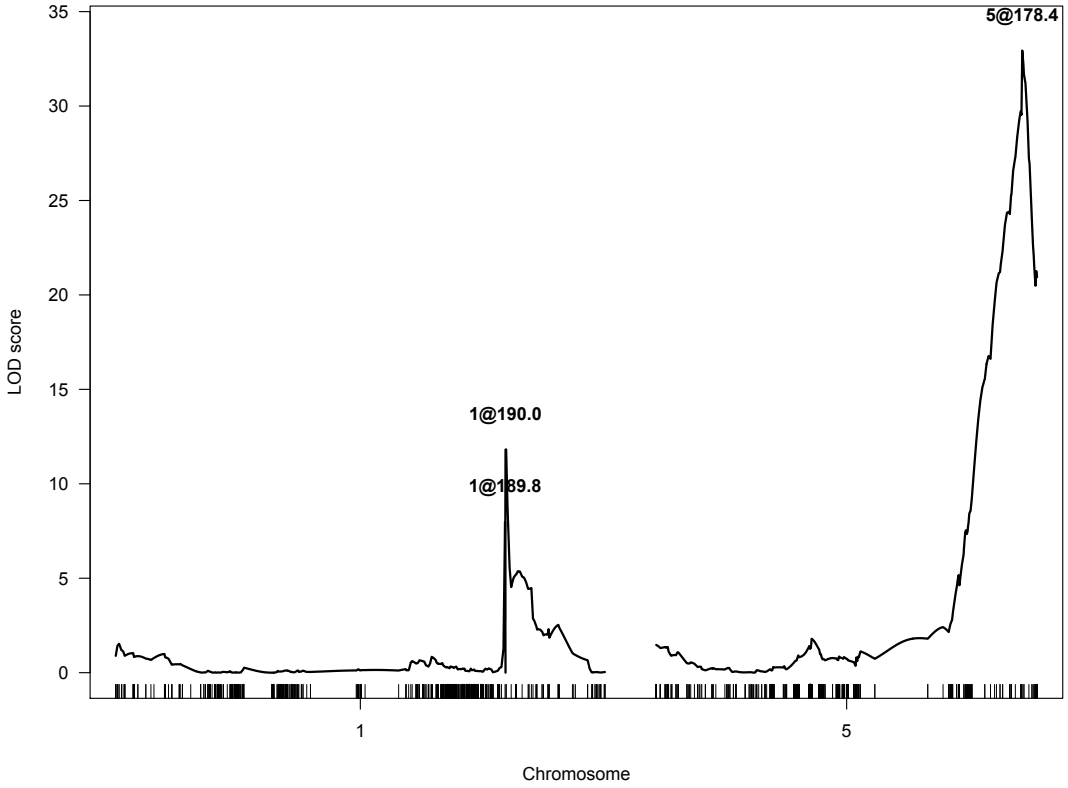

Figure 2.9 Detected QTLs for GA production on chromosome 1 and 5.

One major peak was detected on chromosome 5 at position 178.4 cM. Two minor peaks were found on chromosome 1 at position 189.8 cM and 190.0 cM. LOD threshold is 3.03 at 5% significant level.

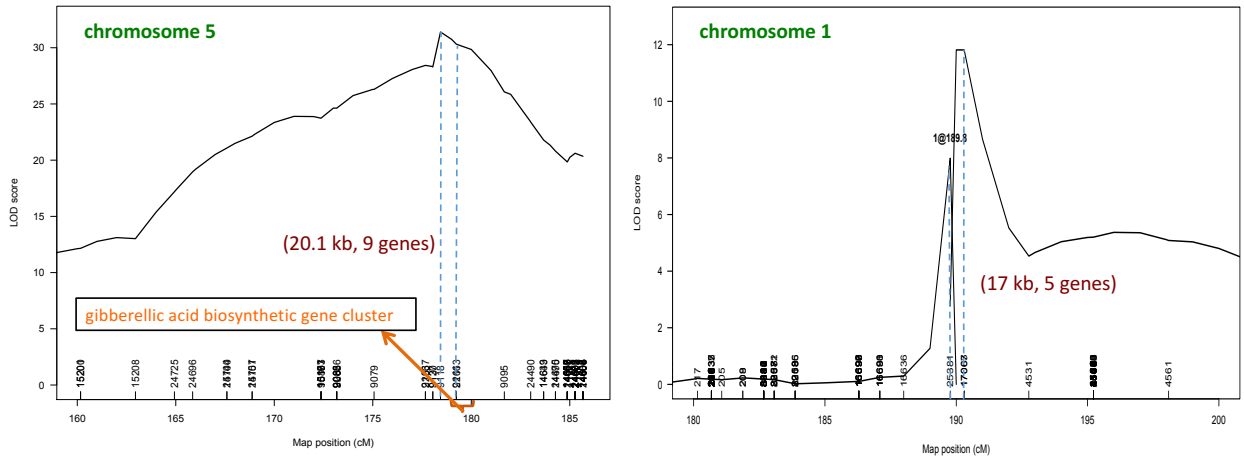

Figure 2.10 Detected QTLs for GA production on chromosome 5 and 1 in detail.

Left panel shows detected QTL on chromosome 5. The corresponding genomic region on *F. fujikuroi* IMI58289 at chromosome 5 is 20.1 kb long, containing 9 genes. This QTL interval covers part of the known gibberellic acid biosynthetic gene cluster. Right panel shows two detected QTLs on chromosome 1. The corresponding genomic region is 17 kb, containing 5 genes.

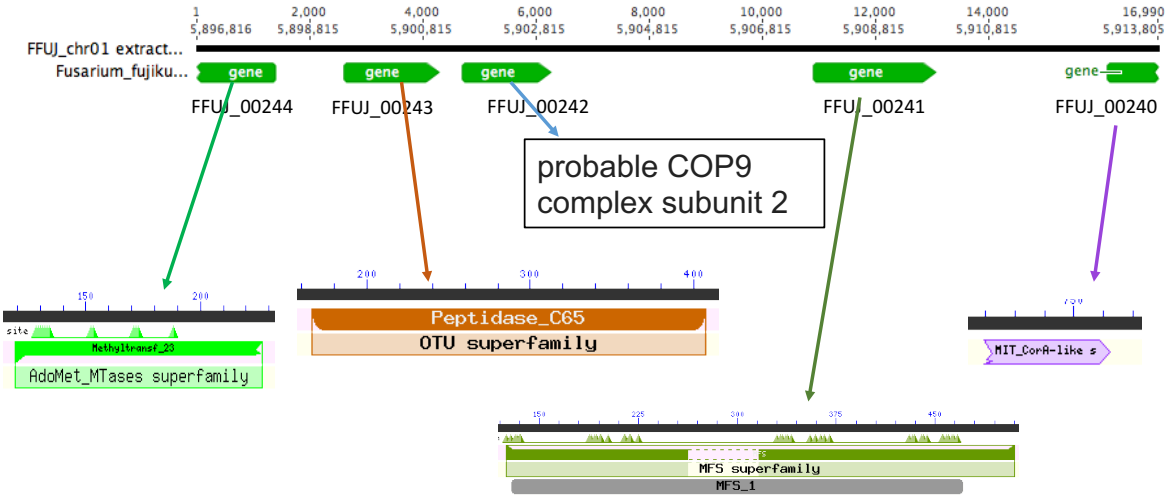

Figure 2.11 Annotation of corresponding genomic region for GA production on chromosome 1 in *F. fujikuroi* IMI58289.

Gene annotation from Wiemann *et al*. (2013). Conserved domain in each protein was detected by blasting against the Conserved Domain Database at NCBI.

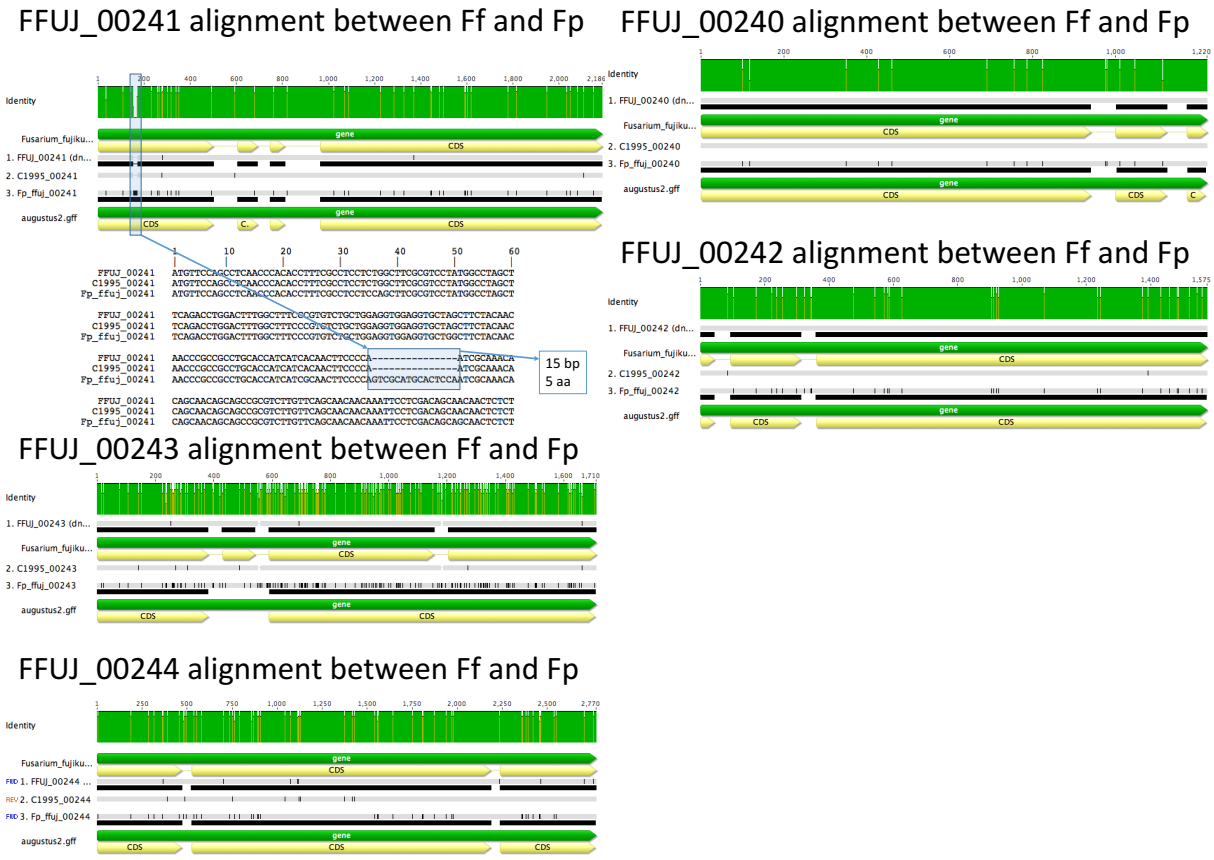

Figure 2.12 Sequence comparison of 5 genes in the detected QTL on chromosome 1 between *F. fujikuroi* (IMI58289 and FGSC8932 (C1995)) and *F. proliferatum* (FGSC7615).

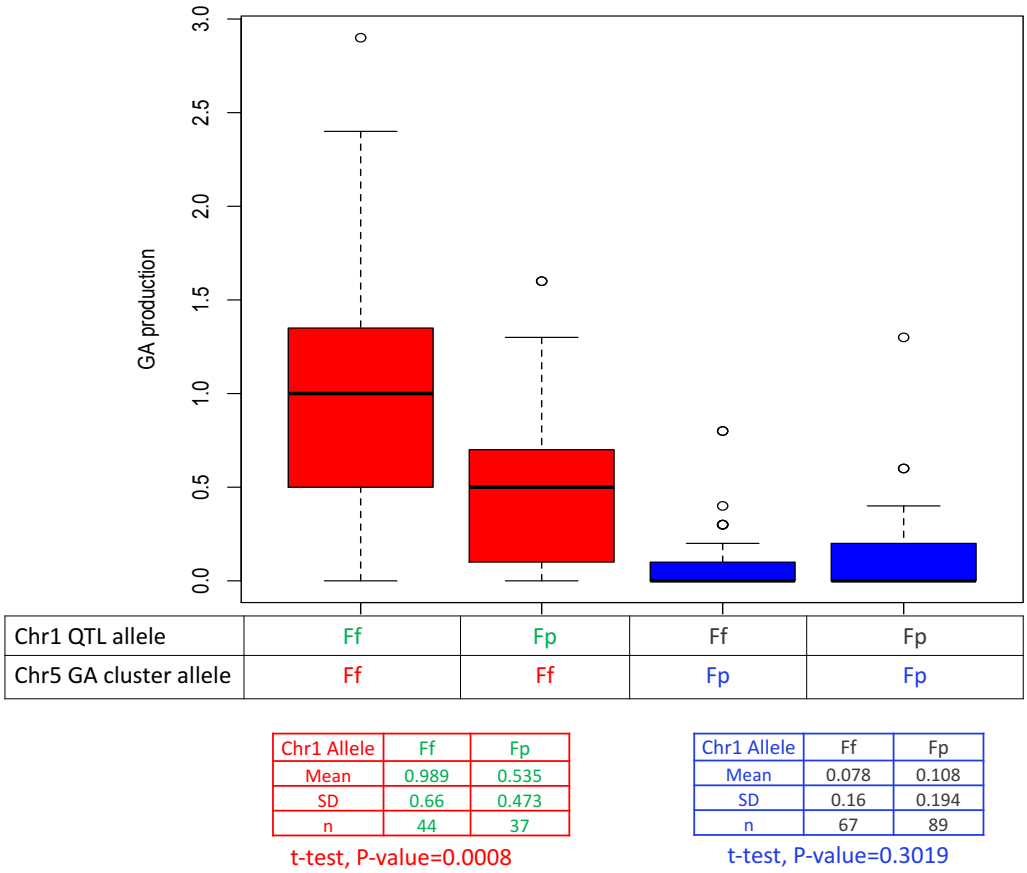

Figure 2.13 Statistical analysis of the GA production (mg/ml) of progeny with or without GA cluster on chromosome 5 (Chr5) or detected GA QTL on chromosome 1 (Chr1).

Some progeny with the GA cluster from the *F. fujikuroi* (Ff) parent (red groups) produced GA, but within this set progeny with the Chr1 QTL allele from the Ff parent had significantly higher GA production ($p < 0.001$) than progeny with the *F. proliferatum* (Fp) parent allele. Progeny with the GA cluster from Fp parent (blue groups) produced almost no GA in average, with no significant difference ($p > 0.1$) between progeny containing the Ff allele versus the Fp allele on the Chr1 QTL.

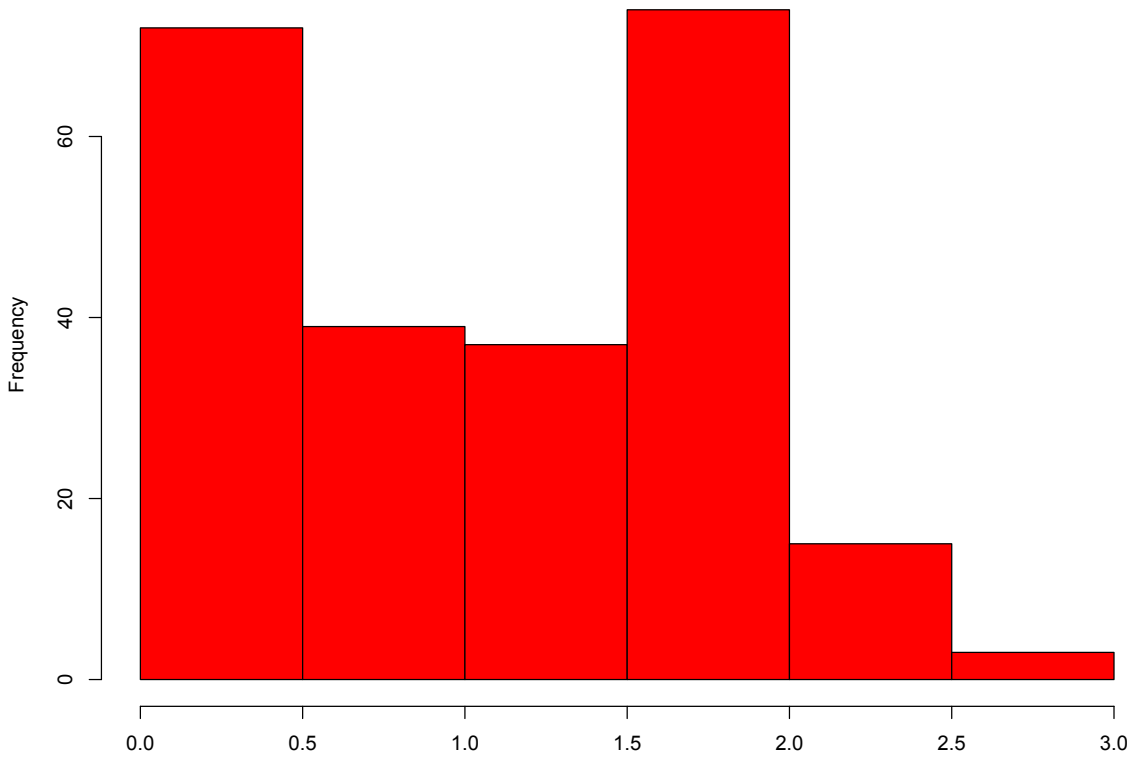

Figure 2.14 Histogram of average virulence to onion of 240 progeny in the cross between *F. fujikuroi* and *F. proliferatum.*

The x-axis represents a pathogenic virulence scale (large number means more virulent), and the y-axis represents the frequency of each category. The *F. fujikuroi* parent has a value of 0 (non-pathogenic), and the *F. proliferatum* parent has the value of 2 (pathogenic). Values below 2 represent progeny that are less virulent than the *F. proliferatum* parent; values above 2 represent progeny that are more virulent than the *F. proliferatum* parent.

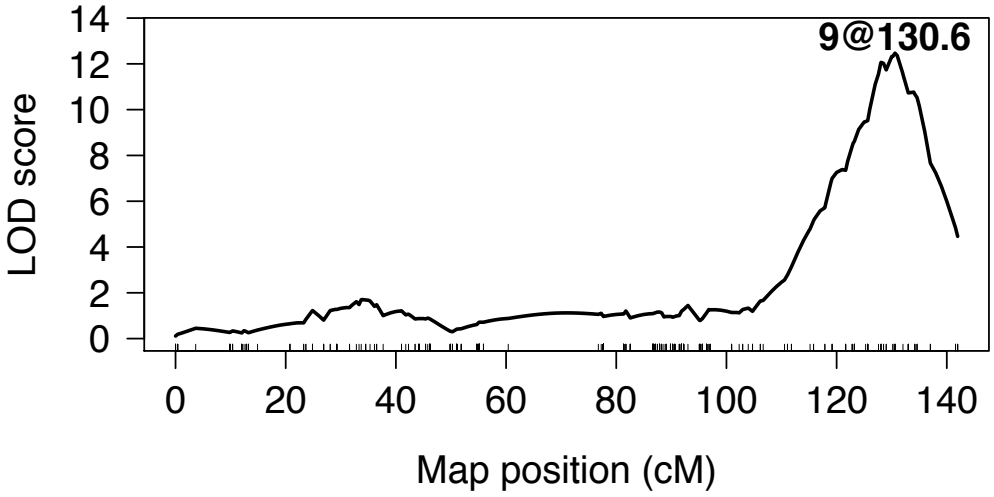

Figure 2.15 Detected QTL for onion virulence on chromosome 9.

One major QTL peak is detected on chromosome 9 at 130.6 cM. LOD threshold is 3.21 at 5% significant level.

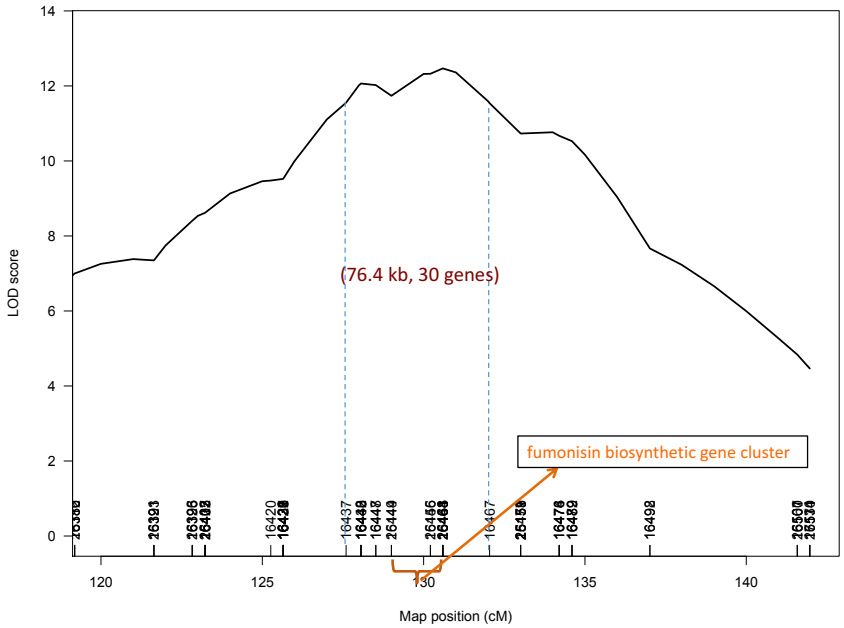

Figure 2.16 Detected QTL for onion virulence on chromosome 9 in detail.

The corresponding chromosome 9 genomic region from the *F. fujikuroi* genome reference IMI58289 is 76.4 kb long and contains 30 genes. This QTL interval covers the known fumonisin biosynthetic gene cluster.

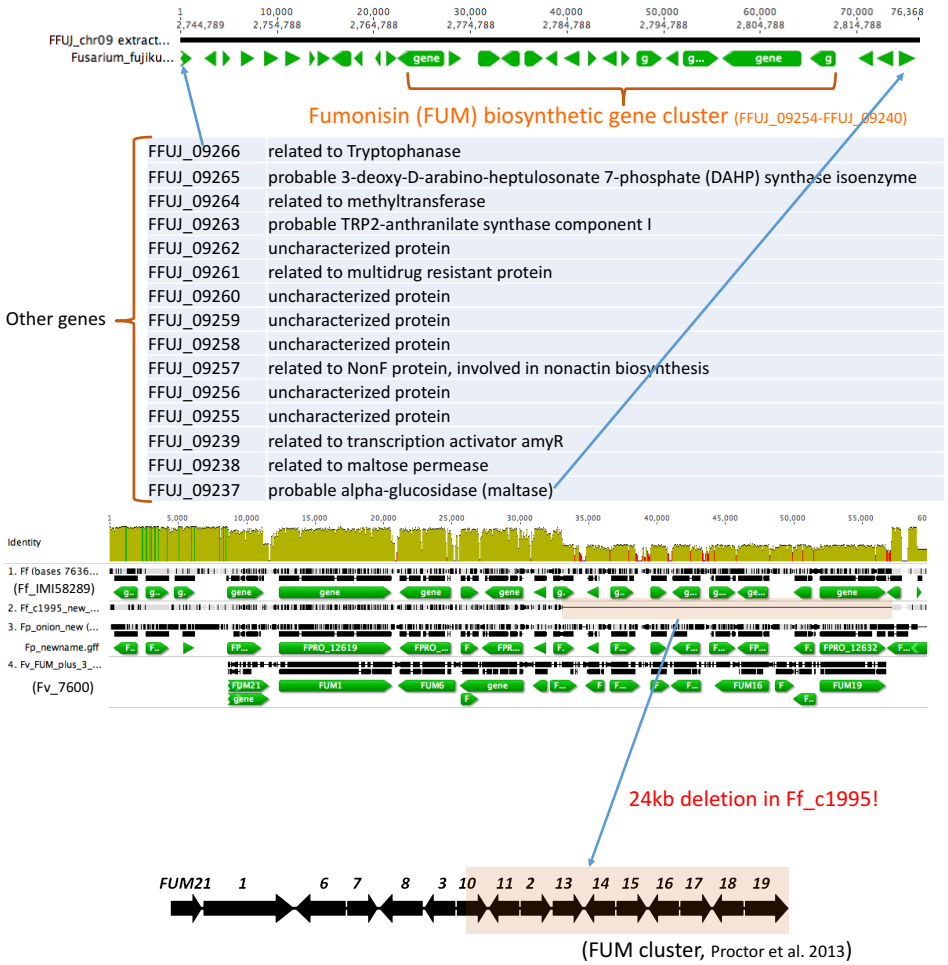

Figure 2.17 Sequence analyses of corresponding genomic region of detected onion QTL on chromosome 9.

Top: annotation of the genomic region in *F. fujikuroi* IMI58289. Bottom: multiple sequence alignments of the corresponding region in *F. fujikuroi* IMI58289, *F. fujikuroi* FGSC8932 (Ff_c1995), *F. proliferatum* FGSC7615 (Fp), and *F. verticillioides* (Fv_7600).

Table 2.1 Genetic map summary.

Chromosome	Number of markers	Number of distinct markers	Length (cM)	Average spacing of distinct markers (cM)	Maximum spacing (cM)
1	940	194	238.4	1.2	22.3
2	720	146	154.1	1.1	16.8
3	713	143	254.1	1.8	76.3
4	610	125	149.3	1.2	11.2
5	689	132	185.7	1.4	25.8
6	603	119	141.4	1.2	8
7	513	99	162.6	1.7	9.4
8	389	86	108.1	1.3	6.9
9	441	100	142	1.4	16.4
10	382	118	148.8	1.3	10
11	342	114	169.8	1.6	38.7
12	39	20	45.3	2.4	9.1
Overall	6381	1396	1899.5	1.4	76.3

Chapter 3 - *Fusarium proliferatum* Genome analysis

Abstract

Fusarium, a genus of ascomycete fungi, contains many important plant pathogens, some of which reduce crop yield and quality by producing contaminating mycotoxins. Among them, *F. proliferatum* (Fp) and *F. fujikuroi* (Ff) are closely related but have different host specificities and produce different mycotoxins. To explore the genetic bases behind these differences, we generated a genetic map from an inter-species cross between them and sequenced the Fp genome to compare with available Ff genomes. We sequenced the Fp parent from this cross to a depth of 50x and assembled a draft genome. This assembly has an N50 of over 824 kb and contains 241 supercontigs, the largest one over 2.1 Mb, with a combined length of 43.1 Mb. With the help of our high-density genetic map that uses sequence-based markers, we anchored 97.8% (42.2 Mb) of the total assembly onto the 12 linkage groups, corresponding to the 12 chromosomes present in each parent species. We annotated this Fp genome, and characterized its SM gene cluster profile, including 50 intact or incomplete SM gene clusters. Despite the high similarity and collinearity of these two species' genomes, we detected 131 Fp specific genes that might play important roles in this species' host specificity and plant pathogenicity. The genome assembly provides valuable resources for further studies on the genomics and biology of *Fusarium*.

Introduction

Fusarium is a genus of filamentous ascomycete fungi that contains many important phytopathogens, which cause severe diseases on all important agronomical plants globally and produce various secondary metabolites (SM) (Leslie and Summerell 2006). The *Fusarium fujikuroi* species complex includes 3 major clades: Asian, African and American (Geiser *et al*. 2013, Wiemann *et al*. 2013, Aoki *et al*. 2014). Within the Asian clade, *F. fujikuroi* (Ff) and *F.*

proliferatum (Fp) are two closely related species with different SM profiles and that cause different plant diseases. Ff is noted for gibberellin (GA) production and can cause bakanae disease in rice, while Fp is noted for fumonisin production and can cause disease in a much broader range of plants (Leslie and Summerell 2006). However, little is known about the genetic basis of Fp's broad host specificity.

In fungi, genes involved in SM biosynthesis usually are clustered together in the genome. SM gene clusters typically contain three different types of genes: SM biosynthetic enzymes, transporter proteins, and pathway-specific transcription factors (Hoffmeister and Keller 2007). Among them, SM biosynthetic enzymes are the core elements of the SM gene cluster, and usually include one core enzyme (sometimes two) from one of the following families: nonribosomal peptide synthetases (NRPS), polyketide synthases (PKS), terpene cyclases (TC), and dimethylallyl tryptophan synthases (DMATS) (Hoffmeister and Keller 2007).

Next-generation sequencing technologies have tremendously increased the power for whole genome sequencing. Currently, more than 20 *Fusarium* species have published genome sequences, including 7 species from the *Fusarium fujikuroi* species complex (https://www.ncbi.nlm.nih.gov/genome). For Ff and Fp, 5 different Ff genomes have been sequenced: IMI58289, B14, FGSC8932, KSU3368, and KSUX-10626 (Wiemann *et al*. 2013, Jeong *et al*. 2013, Chiara *et al*. 2015) and 2 Fp genomes have recently been released: 62905 and ET1 (Niehaus *et al*. 2016). Over 60 SM gene clusters have been detected from these Fp and Ff genomes, which include over 200 genes ranging from 1 to 15 genes per cluster (Wiemann *et al*. 2013, Chiara *et al*. 2015, Niehaus *et al*. 2016). Twenty-three SM products have been assigned to these clusters, including three phytohormones (gibberellin, auxin, and cytokinin), apicidin, fusaric acid, etc (Bomke and Tudzynski 2009, Tsavkelova *et al*. 2012, Wiemann *et al*. 2013,

Niehaus *et al*. 2014b, Niehaus *et al*. 2016). In addition, Niehaus *et al*. (2016) found evidence for copy number variation for some SM clusters. Their results showed that Fp_ET1 contained two copies of the GA gene cluster (both expressed during maize infection) and three copies of the cytokinin (CK) cluster (no expression detected under tested conditions), while Ff_IMI58289 and Fp_62905 only contained one copy of the GA gene cluster and two copies of the CK cluster.

The objectives of this study are: 1) to assemble and annotate a Fp genome; 2) to compare this Fp genome with Ff and other published Fp genomes; 3) to characterize the secondary metabolite gene cluster profile of this Fp genome; and 4) to generate lists of candidate genes that might play roles in host specificity and pathogenicity.

Materials and Methods

Fungal strain and genome sequencing

Fp strain FGSC7615 (MAT-1, D4854) was used for whole genome shotgun sequencing of a 260 bp library with paired-end 250 bp read length using Illumina MiSeq by the Integrated Genomics Facility at Kansas State University, KS, USA.

Genome assembly

The genome was assembled using two independent de Bruijn graph based algorithms. Discovar *de novo* from the Broad Institute was used with the DiscovarExp command (September 2014; http://www.broadinstitute.org/software/discovar/blog/). Assembly By Short Sequences (ABySS, Simpson *et al*. 2009) was also run using the script AssembleG.pl with the following parameters: -n 235, -nodes 8, -mem_per_core 3, -s 47, -l 197, -i 10, -m 61, -c (https://github.com/i5K-KINBRE-script-share/transcriptome-and-genome-assembly/blob/master/KSU_bioinfo_lab/AssembleG/AssembleG_LAB.md).

Assembly annotation

RepeatMasker was used with default parameters to identify the Fp assembly repetitive sequences (http://www.repeatmasker.org, Smit *et al*. 1996). The RepBase database (version 20140131) was used to scan known repetitive elements (Jurka *et al*. 2005). In addition, LTR_finder was used to identify long terminal repeats (LTRs, Xu and Wang, 2007).

The gene annotation of the Fp assembly was generated with the Maker Web Annotation Service (http://weatherby.genetics.utah.edu/cgi-bin/mwas/maker.cgi) by splitting the total assembly into several pieces. The Maker pipeline was trained with Ff_IMI58289 cDNA sequence (ftp://ftp.ensemblgenomes.org/pub/fungi/release-23/cdna/fusarium_fujikuroi) and protein sequences from Ff IMI58289, *F. verticillioides* 7600, *F. graminearum* PH-1, *F. oxysporum lycopersici* 4287, *F. pseudograminearum* CS3096, and *F. solani* FGSC9596 (https://www.ncbi.nlm.nih.gov/assembly/). AUGUSTUS was selected as the gene predictor and otherwise the default settings were used within the Maker pipeline. This online pipeline aligned EST and protein evidence to the masked assembly to generate *ab initio* gene predictions. Then, this pipeline produced the final annotations with evidence-based quality values from these data (http://gmod.org/wiki/MWAS_Tutorial).

Following Ohm *et al*. (2012), gene models from the Maker pipeline were functionally annotated using the Conserved Domain Database (CDD, Marchler-Bauer and Bryant, 2004; Marchler-Bauer *et al*. 2011), InterProScan (Zdobnov and Apweiler, 2001), SignalP (Petersen *et al*. 2011), TMHMM (Melen *et al*. 2003), BLASTp (Altschul *et al*. 1990) against the non-redundant GenBank database at NCBI, SwissProt (http://www.expasy.org/sprot), Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa *et al*. 2008) and eukaryotic orthologous groups of proteins (KOG) (Koonin *et al*. 2004) database. In addition, tRNAscan-SE

(Lowe and Eddy, 1997) was used to predict tRNA-encoding genes in the Fp assembly. Potential small, secreted, cysteine-rich proteins (SSPs) were filtered with a Perl script using criteria described by Arango Isaza *et al*. (2016).

Assembly and annotation completeness

Assembly and annotation completeness were assessed by Benchmarking Universal Single-Copy Orthologs v1 (BUSCO, Simao *et al.* 2015) with 1438 single-copy orthologs found in at least 90% of 125 fungal species from OrthoDB v9 (www.orthodb.org).

Genome comparison

Three Ff genomes (FGSC8932, KSU3368, and KSUX10626) and two Fp genomes (Fp_62905 and Fp_ET1) were retrieved from https://www.ncbi.nlm.nih.gov/assembly/ and https://www.ebi.ac.uk/ena/data/view/. The annotation files for the two Fp genomes were retrieved from http://www.helmholtz-muenchen.de/genomes.jsp?category=fungal. Pair-wise comparison among genomes at the nucleotide level and amino acid level was performed in NCBI BLAST toolkit 2.3.0 (Camacho *et al*. 2008). Whole genome nucleotide alignments between Fp FGSC7615 and Ff_IMI58289 were made using MUMmer (Kurtz *et al*. 2004) and visualized using the mummerplot command within the MUMmer package. The set of proteins from these two genomes were blasted against themselves and against proteins in the OrthoMCL (Li *et al*. 2003) database. Protein groups were assigned using OrthoMCL with a cutoff of e-value < 10^{-5} and at least 50% amino acid identity and match length, and synteny blocks among genomes were defined by using OrthoClusterDB (Ng *et al*. 2009). The collinearity between genomes was estimated by OrthoClusterDB with standard parameters. The ratio of the rate of non-synonymous to synonymous substitution (Ka/Ks) for orthologous pairs from genome Fp_FGSC7615 and Ff_IMI58289 was calculated using the integrated codeml program of the PAML package in

SynMap with default parameter sets (Lyons *et al*. 2008). Additional sequence alignments were performed with Geneious software (Geneious version 7, http://www.geneious.com, Kearse *et al.*, 2012).

Results

Genome assembly and annotation

The Fp_FGSC7615 genome sequence was generated with a whole genome shotgun approach by using Illumina MiSeq, resulting in around 4.5 million paired-end reads. These reads were assembled into 613 contigs with an N50 of over 304 kb with the ABySS pipeline, resulting in a combined length of 43.1 Mb (50x coverage). Forty-five contigs were longer than 304 kb and the largest was over 1 Mb (Table 3.1). The raw reads were assembled into 241 contigs with an N50 of over 824 kb using the *de novo* Discovar pipeline, resulting in a combined length of 43.1 Mb with 17 contigs over 824 kb and the largest one over 2.1 Mb. The Discovar assembly had a larger N50 and smaller number of contigs compared with the ABySS assembly. Therefore, the Discovar assembly was chosen to do further analysis.

The GC content of the assembly was 48.78% with 0.76% of the total assembly masked by RepeatMasker. The masked region contained 139 LTR elements with total length 17.8 kb, and 60 DNA transposons with total length 17.4 kb (Table 3.2). However, no LTR elements were detected by LTR_FINDER. The assembly is predicted to contain a total of 14,393 genes by the Maker Web Annotation Service. To assess the completeness of the Fp_FGSC7615 assembly, we performed a BUSCO analysis on the assembled genome and its annotation. The completeness of the Fp genome assembly was 99% (1,424 of 1,438 core orthologs) with 13 fragmented orthologs and 1 missing ortholog, and the completeness of the annotation was 98% (1,419 of 1,438 core

orthologs) with 11 fragmented orthologs and 8 missing orthologs (Table 3.3). These BUSCO results indicate that our assembly is highly complete.

The average length of annotated genes in the Fp_FGSC7615 assembly was 1.7 kb with mean protein length 503.6 aa (Table 3.4). In addition, tRNAscan-SE showed the Fp_FGSC7615 assembly had slightly more tRNA-encoding genes (298) than did Ff_IMI58289 (269). Using the non-redundant (nr) database in NCBI during the analysis performed, BLASTP showed that 12,711 out of 14,393 total predicted Fp_FGSC7615 proteins had significant matches with e-value < 10^{-5} (Figure 3.1). Among these 12,711 proteins, most of them (99.3%) had their top hit matching sequence from one of 13 different *Fusarium* species. The top 4 *Fusarium* species were *F. fujikuroi* (80.1%), *F. oxysporum* (13.4%), *F. verticillioides* (0.7%), and *F. proliferatum* (0.3%). The few Fp hits was due to no Fp genome annotation in the NCBI database when the analysis was performed. In addition, only 93 (0.7%) proteins had top hits outside the genus *Fusarium*. These genome statistics and BLAST results confirmed that Fp and Ff are closely related at the genome level (Wiemann *et al*. 2013).

Genetic map assisted scaffolding

A high-density genetic map of 12 linkage groups was constructed with genotype data from 6,381 high quality Genotyping-by-Sequencing markers and 253 progeny from the cross between Fp and Ff (previous chapter). By comparing marker sequence with the Fp_FGSC7615 assembly, we anchored 97.9% (42.2 Mb, 109 contigs) of the total assembly onto the 12 linkage groups, corresponding to the 12 chromosomes present in each parent species (Xu and Leslie 1996; Wiemann *et al*. 2013). The remaining unanchored contigs (132 contigs) included 113 contigs with the length of < 10 kb, 16 contigs with the length of 10 – 50 kb, 2 contigs with the length of 50 – 100 kb, and 1 contig with the length of ~190 kb. The lengths of orthologous

chromosomes in Fp_FGSC7615 and Ff_IMI58289 were similar except the former has a longer chromosome 4 (over 0.8 Mb larger) and a shorter chromosome 12 (only 60% of the counterpart in the latter, Figure 3.2).

Synteny between *F. proliferatum* and *F. fujikuroi*

The Ff_IMI58289 genome was used as the primary reference for the Ff species due to its completed assembly and extensive annotation (Wiemann *et al*. 2013). MUMmer analysis showed that Fp_FGSC7615 is highly similar to Ff_IMI58289 at the nucleotide level, 96.5% identity in the 38.8 Mb (89.7% of the Fp_FGSC7615 genome) aligned regions (Figure 3.3), and only ~1.2% of the aligned sequence exhibited insertions or translocations larger than 1 kb. In addition, the two new Fp genomes Fp_62905 and Fp_ET1 had 99.1% and 98.4% identity in 41.9Mb (97% of the Fp_FGSC7615 genome) and 41Mb (95% of the Fp_FGSC7615 genome) aligned regions respectively with the Fp_FGSC7615 genome. OrthoClusterDB (Ng *et al*. 2009) analysis with 9,184 orthologous pairs detected 1,592 synteny blocks between Fp_FGSC7615 and Ff_IMI58289 (Table 3.5). These blocks contained from 2 to 32 genes, but most (over 70%) of the blocks contained 2-5 genes. Supernumerary chromosome 12 only had 2 short synteny blocks (2 genes each), which were conserved in all available Fp genomes (Figure 3.4). For chromosomes 8, 10, and 11, the synteny blocks were distributed relatively evenly across each chromosome. The remaining eight chromosomes showed less similarity at the ends of each chromosome (Figure 3.4). Of these eight, no similarity was detected at either end of chromosome 4, which confirmed that Ff_IMI58289 is missing these two parts of the Fp_ FGSC7615 genome (Wiemann *et al*. 2013). Finally, the synteny conservation calculated with the detected synteny blocks between Fp_FGSC7615 and Ff_IMI58289 is 0.976, similar to Fp_62905 and Fp_ET1 (0.995 and 0.978 respectively, Niehaus *et al*. 2016).

Comparison of protein coding genes among *F. proliferatum* and *F. fujikuroi* genomes

To further estimate the protein conservation between Fp and Ff, pair-wise BLASTP analysis using the NCBI BLAST toolkit was made between Fp_FGSC7615 and 4 different Ff genomes (Ff_IMI58289, FGSC8932, KSU3368, and KSUX10626). With the threshold e-value < 10^{-5}, Fp_FGSC7615 proteins were highly conserved with Ff proteins (proportion of proteins, 98.4% of Ff_IMI58289, 97.8% of FGSC8932, 98% of KSU3368, and 98% of KSUX10626). Together, only 131 (0.9%) Fp_FGSC7615 proteins showed low or no similarity with any of these four Ff genomes, suggesting these proteins are either Fp specific or specific to the Fp_FGSC7615 isolate.

Two additional Fp genomes (Fp_62905 and Fp_ET1) had genome statistics similar to our Fp_FGSC7615 assembly, except the other genomes had shorter average gene length (Table 3.4, Niehaus *et al*. 2016). Pair-wise comparison to our Fp_FGSC7615 genomes indicated that 98.7% (14,205 out of 14,393) proteins were similar to Fp_ET1, and 98.5% (14,175 out of 14,393) proteins were similar to Fp_62905. 97.9% (14,095 out of 14,393) proteins were common to all 3 Fp genomes, suggesting they represent a core set of proteins in the Fp species. Additionally, 110 proteins exist in both Fp_FGSC7615 and Fp_ET1 (but not in Fp_62905), and 80 exist in both Fp_FGSC7615 and Fp_62905 (but not in Fp_ET1).

Of the 131 proteins present in Fp_FGSC7615 but absent from 4 Ff genomes, only 43 are common to Fp_FGSC7615, Fp_ET1 and Fp_62905, suggesting that they belong to the core set of Fp proteins. Annotation results suggested that 39 out of 43 were similar to hypothetical proteins from one of 5 different species (22 from *F. oxysporum*, 13 from *F. verticillioides*, 2 from *F. graminearum*, 1 from *F. poae*, and 1 from *Colletotrichum*), and the remaining 4 had no match within the searched annotation database. In addition, 60 out of these 131 proteins were

Fp_FGSC7615 lineage specific (*i.e*., no match in the 2 Fp and 4 Ff strains compared). Annotation results showed that 44 out of these 60 were similar to hypothetical proteins from 3 different species (33 from *F. oxysporum*, 10 from *F. verticillioides*, 1 from *F. pseudograminearum*), and the remaining 16 had no match within the searched annotation database. Therefore, by comparing with Ff genomes and performing BLAST searches to public databases, these results suggest that Fp specific genes have either been inherited from ancestral *Fusarium* species but lost in Ff, or have been acquired via horizontal gene transfer from more distant species (Chiara *et al*. 2015).

Selective constraint and adaptive evolution

To investigate the selective constraint across the genome, the Ka/Ks ratio was calculated for each detected orthologous pair between Fp_FGSC7615 and Ff_IMI58289. The majority (over 98%) of orthologous genes had a Ka/Ks ratio less than 0.9, indicating negative selection working on these genes to conserve their function. A few (0.68%) had a Ka/Ks ratio between 0.9 and 1.1, a value consistent with a lack of selection acting on these genes (*i.e*., neutrality). Only 75, or 0.84%, of compared orthologous pairs had a Ka/Ks ratio greater than 1.1, suggesting positive selection may be working on these genes. These genes, if under positive selection, might play roles in adaptive evolution in Fp_FGSC7615 and Ff_IMI58289. The average Ka/Ks ratio for orthologous genes on dispensable chromosome 12 was 0.3, while the number on the core genome (including chromosome 1 to 11) was 0.19. Genes on the dispensable chromosome showed less selective constraint than genes in the core genome, suggesting the existence of a 'two-speed genome structure' (Sperschneider *et al*. 2015). The average Ka/Ks ratio for orthologous genes per 100 gene window across the Ff_IMI58289 genome ranged from 0.1 to 0.3. Many of the peaks of average Ka/Ks ratio, and many of the individual gene Ka/Ks outliers (>

1.1), were located near the centromeres for a majority of chromosomes (Figure 3.5), implying less selective constraint near centromere regions.

Pathogenicity related genes

Plant cell wall degrading enzymes (CWDEs) are essential tools for plant pathogens to invade plant cells for colonization (Kubicek *et al*. 2014; Choi *et al*. 2013). For most types of potential CWDE genes, Fp_FGSC7615 had a similar distribution pattern with those in Ff_IMI58289 and the other two available Fp genomes (Table 3.6). Fp genomes (10-12 genes) had more α-rhamnosidases than Ff had (7 genes). However, Fp_FGSC7615 contains many more potential genes responsible for degrading the main-chain of pectin, which is concentrated in the primary cell wall (McCann and Roberts 1991). Therefore, those genes might provide Fp_FGSC7615 an additional advantage during colonization in plants, resulting in a wide host specificity (Leslie and Summerell 2006).

Small, secreted, cysteine-rich proteins (SSPs) in fungal pathogens are the most common type of fungal effectors that serve as pathogenicity or virulence factors during fungal infection (De Wit *et al*. 2009; Arango Isaza *et al*. 2016). Fp_FGSC7615 contains 857 SSPs using the following criteria: smaller than 300 amino acids in length, at least four cysteine residues, no signal anchor motifs, and no transmembrane domains (Arango Isaza *et al*. 2016). Of these 857 SSPs, only two were Fp_FGSC7615 lineage specific genes.

Secondary metabolite (SM) gene clusters

The Fp_FGSC7615 genome contained intact or partial counterparts of 44 SM gene clusters identified in Ff_IMI58289 by Wiemann *et al*. (2013). Thirty-seven out of 44 were intact clusters, while the remaining seven were incomplete by BLAST analysis, including 2 sesquiterpene synthases (STCs, STC7 and STC8), 3 PKSs (PKS10, PKS12 and PKS17), and 2

NRPSs (NRPS17 and NRPS31). For cluster STC7 and STC8, only short residual sequences (13% and 19% in length) of these 1-gene clusters were present in the Fp_FGSC7615 genome.

Like Fp_62905, Fp_FGSC7615 contains less than half of the core gene (*FUS1*, a PKS/NRPS gene) required for fusarin C production in the PKS10 cluster, and does not have the remaining 8 genes in this cluster (Niehaus *et al*. 2016). Niehaus *et al*. (2016) indicated that the *PKS12* cluster was potentially functional in Ff, but pseudogenized in Fp_ET1 and Fp_62905 even though they shared the same flanking genes. However, Fp_FGSC7615 not only shares the genes flanking the PKS cluster, but the cluster itself appears intact, although the cluster is inverted relative to Ff_IMI58289 (Figure 3.6). Upon closer examination, the two genes of the *PKS12* cluster in Fp_FGSC7615 were intact and contained the same core conserved domains that are found in Ff_IMI58289 (Figure 3.7). Therefore, the *PKS12* in Fp_FGSC7615 also is potentially functional. The 7-gene *PKS17* cluster is conserved and expressed in Fp_ET1, Fp_62905 and Ff_IMI58289 (Niehaus *et al*. 2016). The Fp_FGSC7615 assembly has the core polyketide synthase gene and four other related genes from this cluster, but instead of having the remaining two of the cluster genes (corresponding to *FFUJ_12060* and *FFUJ_12061* in Ff_IMI58289), Fp_FGSC7615 has an additional three genes functionally unrelated to the missing genes in Ff_IMI58289 (Figure 3.8). Additionally, Ff_IMI58289 has one additional gene flanking the *PKS17* cluster when compared with the three Fp genomes.

NRPS31 (*FFUJ_00003- FFUJ_00013*), which is responsible for apicidin F production, is present on the tip of chromosome 1 in Ff but not found in any other available genome from the *Fusarium fujikuroi* species complex (Wiemann *et al*. 2013, Niehaus *et al*. 2014a, Chiara *et al*. 2015). This cluster also is missing from the corresponding region of Fp_FGSC7615. However, at the end of chromosome 11 Fp_FGSC7615 does contain sequence that is highly similar (89.6%

identity across 1.9 kb) to *FFUJ_00004* (1.9 kb, which serves as a major facilitator superfamily transporter in apicidin F production) and a sequence with low similarity (~50% identity across 19.4 kb) to *FFUJ_00003* (the core non-ribosomal peptide synthetase). By BLASTing these sequences identified in Fp_FGSC7615 against Ff_IMI58289 we identified a *FFUJ_00003*-like sequence on chromosome 11, which is highly similar (96.4% identity across 14.7 kb) to that in Fp_FGSC7615. The *FFUJ_00004*-like sequence was absent on chromosome 11 of Ff_IMI58289. Additionally, sequences highly similar to the *FFUJ_00003*-like and *FFUJ_00004*-like sequences found in Fp_FGSC7615 were detected in Fp_ET1 (97.6% across 19.5 kb, 98.5% identity across 1.9 kb) and in Fp_62905 (98.1% across 19.4 kb, 94.9% identity across 0.95 kb). This pattern of shared or similar sequences suggests that the common ancestor of Fp and Ff contained an intact *NRPS31* cluster and a partial cluster that was presumably nonfunctional. Subsequently, some Ff isolates have inherited both intact and partial clusters, while Fp only acquired the partial cluster.

The *NRPS17* cluster, probably required for the ferrichrome siderophore production, is intact in Ff_IMI58289 (Niehaus *et al*. 2016) but truncated in Fp_FGSC7615 (half of the 4 genes are deleted). In contrast, the potentially functional fusaric acid biosynthesis cluster (encoding two core enzymes, PKS6 and NRPS34, Niehaus *et al*. 2014b) was also detected in Fp_FGSC7615.

Potentially functional clusters that produce phytohormones (gibberellins, auxins, and cytokinins) and are present in most *F. fujikuroi* species complex members (Niehaus *et al*. 2016) were also detected in Fp_FGSC7615. A single gibberellin producing cluster (GA cluster) similar to that in Ff_IMI58289 (DTC2) is located on chromosome 5, in contrast to Fp_ET1 which was found to have two GA clusters (Niehaus et al. 2016). Both auxin-producing clusters (the IAA mini-cluster and the YUCCA-like gene) are located on chromosome 3. The cytokinin producing

clusters (CK1, CK2, and the tRNA-isopentenyl-transferase, or tRNA-IPT, gene) are located on chromosome 3, chromosome 5, and chromosome 7, respectively. Similar to Ff_IMI58289 and the other two Fp genomes (Fp_62905 and Fp_ET1), the GA cluster and the CK2 cluster in FpFGSC7615 are close to each other (~30 kb apart) with the YUCCA-like gene in between on the tip of chromosome 5, while the IAA cluster and the CK1 cluster were relatively far from each other (~270 kb) on the tip of chromosome 3.

Discussion

Fp and Ff are two closely related *Fusarium* species, but they have different host specificities and produce different mycotoxins (Leslie and Summerell 2006). To better understand the genetic bases behind these differences, we sequenced a Fp genome (FGSC7615) to compare with available Ff genomes.

We *de novo* assembled the Fp_FGSC7615 genome with 50x coverage from 4.5 million paired-end Illumina MiSeq reads using Discovar *de novo*. This assembly had an N50 of over 824 kb and contains 241 supercontigs with a combined length of 43.1 Mb, the largest supercontig over 2.1 Mb. Using the Maker Web Annotation Service, we predicted 14,398 genes in this assembly. Compared to the well-studied Ff_IMI58289 genome, the Fp_FGSC7615 assembly has a similar genome size and total number of predicted genes (Table 3.4). However, two new published Fp assemblies (Fp_62905 and Fp_ET1) have slightly larger genome sizes and more predicted genes that are on average shorter, possibly due to strain differences or to a bias in automated gene annotation that depended largely on *F. fujikuroi* data (Niehaus *et al*. 2016).

Pair-wise protein comparisons between Fp_FGSC7615 and two other Fp genomes (Fp_62905 and Fp_ET1) revealed that 97.9% of 14,393 belong to a core set of Fp proteins (*i.e*., present in all Fp isolates), and the remaining 2.1% belong to a lineage specific set of proteins

(*i.e*., specific to a subset of sequenced Fp isolates or to a single Fp isolate). This finding is consistent with the finding that *Fusarium* genomes can be divided into two portions: the core genome and the adaptive/accessory genome (Ma *et al*. 2013). Annotation data from Fp_FGSC7615 indicated that the majority of the lineage specific proteins were uncharacterized (data not shown).

With the help of a high-density genetic map between Fp and Ff, we anchored ~98% of the Fp_FGSC7615 assembly into 12 groups, corresponding to 12 chromosomes in Ff_IMI58289 (Wiemann *et al*. 2013). Fp_FGSC7615 had similar chromosome lengths compared with Ff_IMI58289 except chromosome 4 and 12 (Figure 3.2). Pair-wise nucleotide comparisons between them showed high similarity (96.5% in aligned regions). In addition, OrthoClusterDB (Ng *et al*. 2009) analysis of the two genomes also indicated high synteny conservation (0.976) based on their orthologous proteins. Together these results confirmed that Fp_FGSC7615 is highly similar to Ff_IMI58289 at the nucleotide and protein level, similar to the comparison between Ff_IMI58289 and Fp_ET1 or Fp_62905 (Niehaus *et al*. 2016). Although highly similar to Ff, Fp does have 131 Fp species-specific proteins (60 unique to Fp_FGSC7615) that were not found in Ff. Thirty-four out of these species-specific proteins are potential SSPs, which might play roles in pathogen virulence (De Wit *et al*. 2009; Arango Isaza *et al*. 2016). Additionally, 7 pairs of SSP orthologous genes common to Fp_FGSC7615 and Ff_IMI58289 have a Ka/Ks >1.1, providing some evidence that natural selection may have altered the function of these genes as a result of adaptive evolution during species divergence. Therefore, these results provide a list of candidate loci for host specificity and pathogenicity.

Although highly similar to each other, Fp_FGSC7615 and Ff_IMI58289 do carry different mating type idiomorphs. As parents of a successful interspecific cross that has produced

viable progeny, this was necessarily true (Desjardins *et al*. 1997, Leslie *et al*. 2004, Mohamed Nor 2014), but carrying opposite idiomorphs is not sufficient to guarantee a successful cross, as the majority of isolates attempted in Fp-Ff interspecies crosses are infertile. BLAST results showed that Fp_FGSC7615 contains the *MAT-1* mating type genes (*MAT1-1-1*, *MAT1-1-2*, *MAT1-1-3*) on chromosome 6, while both Fp_ET1 and Fp_62905 had the *MAT-2* mating type gene (*MAT1-2-1*) similar to Ff_IMI58289 and Ff_FGSC8932 (Chiara *et al*. 2015).

Fp_FGSC7615 varied in the SM gene clusters relative to the Ff genome. Seven out of the 44 clusters (Wiemann *et al*. 2013) changed through deleting genes (STC7, STC8, PKS10, NRPS17, NRPS31, PKS17), adding genes (PKS17), or reordering genes in the cluster (PKS12) (Ma *et al*. 2013). The remaining 37 clusters were intact compared to Ff_IMI58289 and potentially functional (Wiemann *et al*. 2013, Niehaus *et al*. 2014b). Moreover, Fp_FGSC7615 carried homologous sequences for phytohormone production such as the gibberellin related GA cluster (DTC2), auxin related IAA clusters (IAA-mini cluster and YUCCA-like gene) and cytokinin related clusters (CK1, CK2, and tRNA-IPT gene) (Niehaus *et al*. 2016). However, unlike Fp_ET1, which has 2 GA clusters (DTC2-1, DTC2-2) and 2 cytokinin CK2 clusters (CK2-1, CK2-2), Fp_FGSC7615 only had one copy of each cluster. In total, 50 intact or incomplete SM gene clusters were found in Fp_FGSC7615: 2 DMATs, 2 DTCs, 8 STCs (including 2 incomplete clusters), 18 PKSs (including 3 incomplete clusters), 15 NRPSs (including 2 incomplete clusters), 1 IAA, 1 YUCCA-like gene, 2 CKs and 1 tRNA-IPT gene. A total of 202 genes were found in these 50 clusters, ranging from 1 to 15 genes per cluster and averaging 4 genes per cluster.

In summary, we *de novo* assembled and annotated the Fp_FGSC7615 genome into 12 chromosomes. We characterized its SM gene cluster profile, and detected 50 intact or incomplete

SM gene clusters. By comparing with published Ff and Fp genomes, we identified 131 Fp species-specific genes with 60 unique to Fp_FGSC7615, which might play roles in host specificity and pathogenicity. These results provide valuable resources for further studies on the genomics and biology of *Fusarium* fungi.

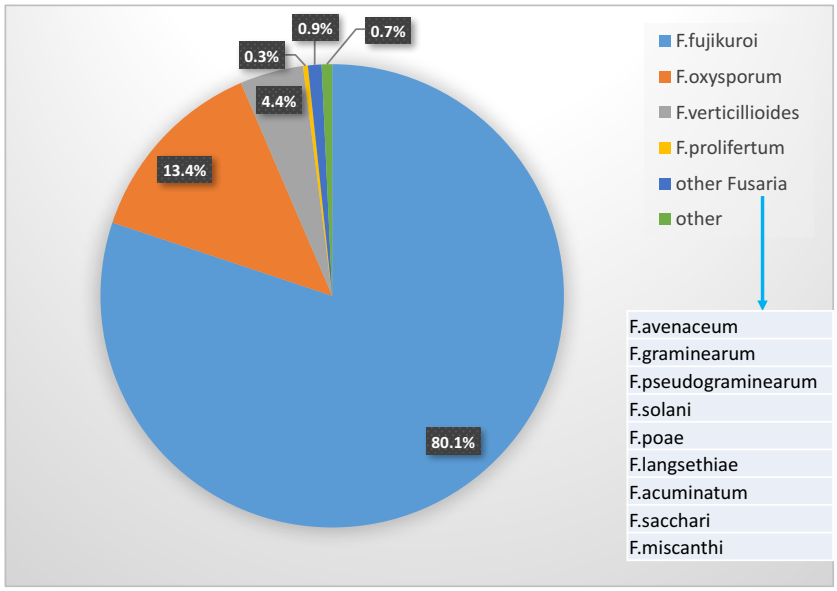

Figure 3.1 BLASTP result with non-redundant database in NCBI.

12,711 out of 14,393 total predicted proteins had matches with e-value < 10^{-5}. Pie chart represents the species of origin of the top-scoring hit via BLASTP of each predicted Fp_FGSC7615 protein.

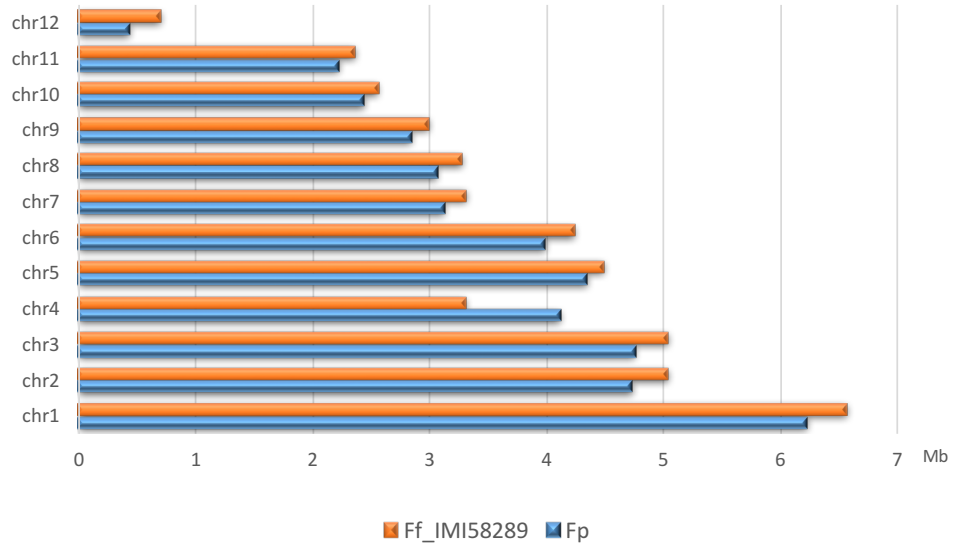

Figure 3.2 Chromosome length statistics of Fp (FGSC7615, scaffolded) and Ff_IMI58289.

x-axis represents the chromosome length in Mb. Fp was over 0.8 Mb longer than Ff on chromosome 4, and had a shorter chromosome 12 than Ff.

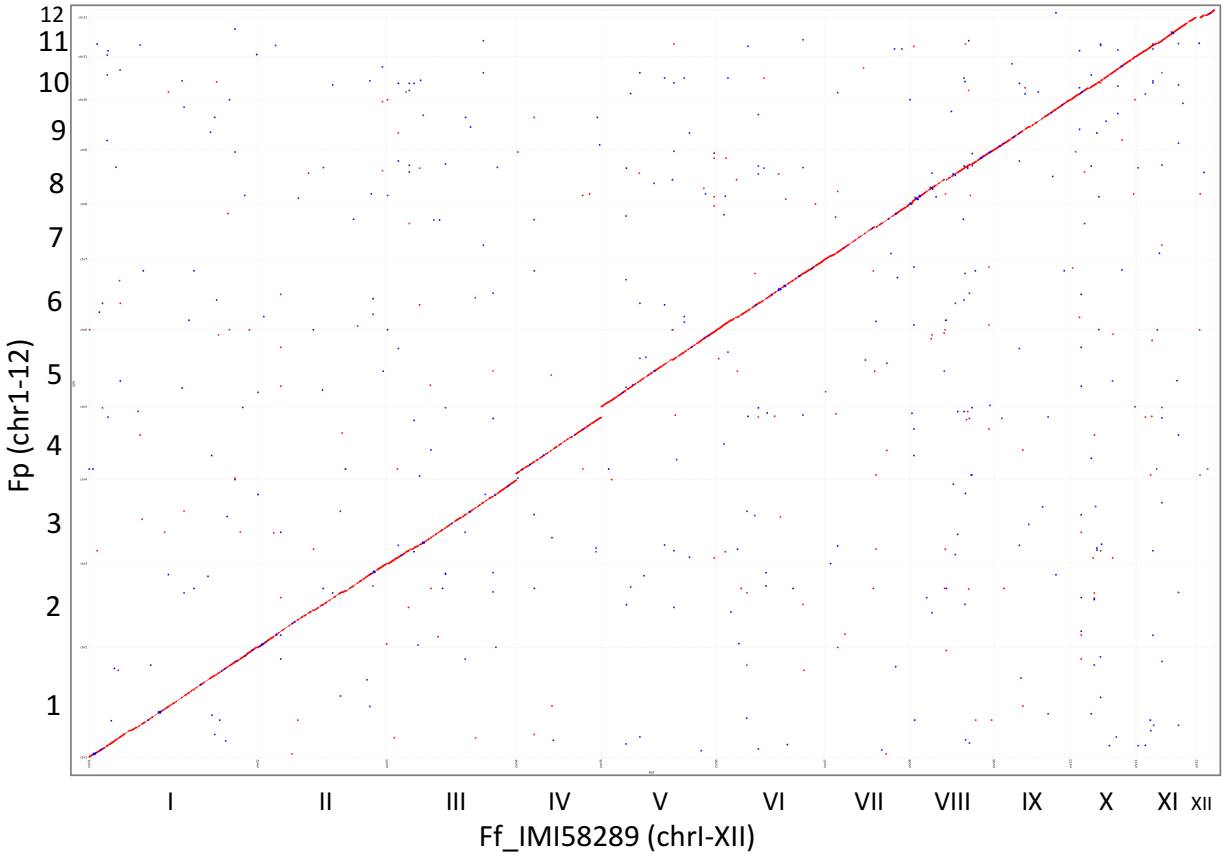

Figure 3.3 Whole genome comparison of Fp (FGSC7615) and Ff_IMI58289.

Dotplot of the comparison calculated using MUMmer 3.0. Red dots represent orthologous DNA, and blue dots represent inverted segments. Unanchored scaffolds of the Fp_FGSC7615 assembly are not shown. High collinearity was detected across the genome, except for chromosomes 4 and 12.

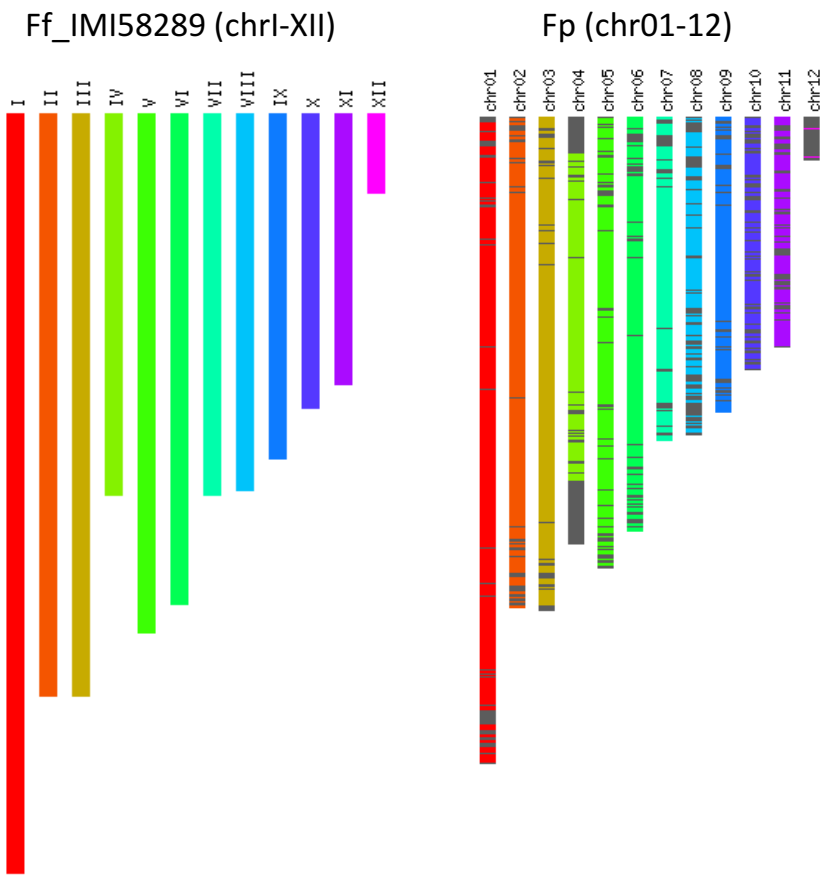

Figure 3.4 Synteny blocks detected by OrthoClusterDB between Fp (FGSC7615, scaffolded) and Ff_IMI58289.

Each color segment in the Fp genome image corresponds to a synteny block detected by OrthoCluster. The color of the synteny block corresponds to the reference chromosome in the Ff_IMI58289 genome. The gray color in the Fp genome indicates no detected synteny blocks in this region.

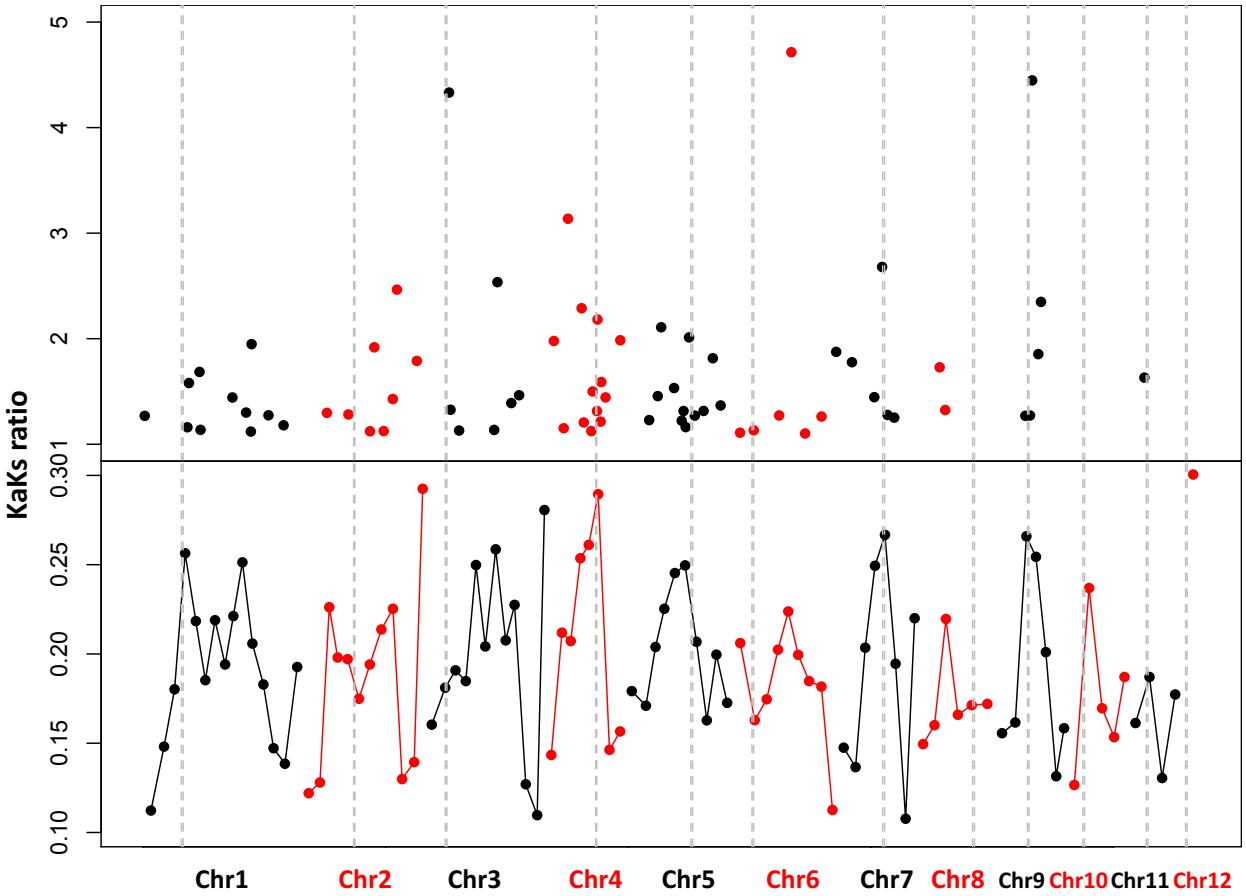

Figure 3.5 Distribution of Fp (FGSC7615) and Ff_IMI58289 orthologous gene pairs with Ka/Ks ratios >1.1 (top) and sliding window average Ka/Ks ratio per 100 genes (bottom) across the Ff_IMI58289 genome.

x-axis shows concatenated chromosomes from 1 to 12. Vertical dashed lines indicate centromere positions (Wiemann *et al*. 2013). The Ka/Ks ratio peaks near the centromere for the majority of the centromeres.

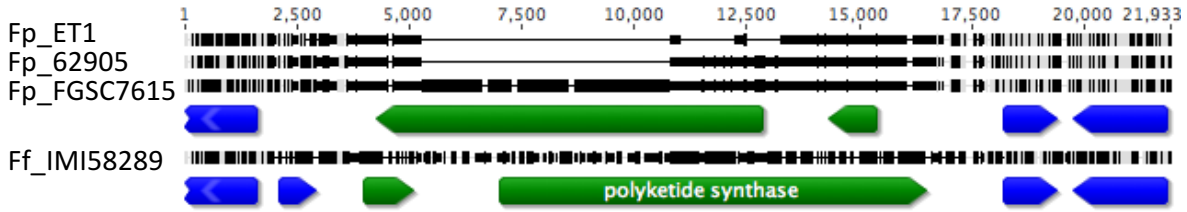

Figure 3.6 Comparison of the PKS12 gene cluster in Fp_ET1, Fp_62905, Fp_FGSC7615, and Ff_IMI58289.

Green arrows represent genes within the PKS12 cluster, while blue arrows represent flanking genes around the cluster. Sequences within Fp species were similar except some deletions in Fp_ET1 and Fp_62905. The clusters in Fp were inverted compared to that in Ff_IMI58289. The straight line in the top 2 Fp sequences indicates a deletion. The thicker black line represents DNA sequence.

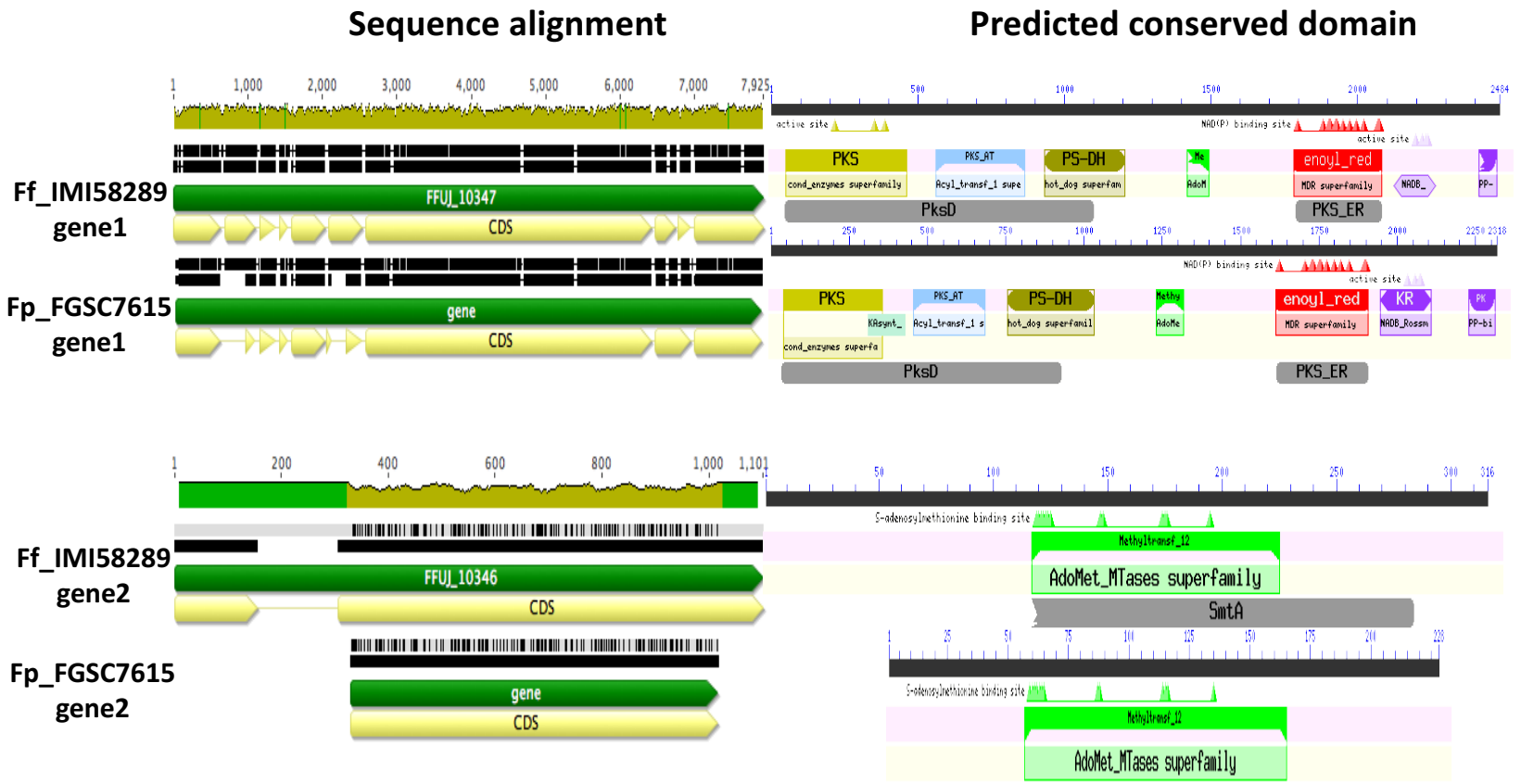

Figure 3.7 Comparison of genes within the PKS12 cluster in Ff_IMI58289 and Fp_FGSC7615.

The sequences of Gene1 in both species were highly similar except that different CDS (coding sequence) splicing is predicted. Gene2 of Fp_FGSC7615 is missing one exon compared to that in Ff_IMI58289. Similar conserved domains were predicted in both gene1 and gene2 from these two species.

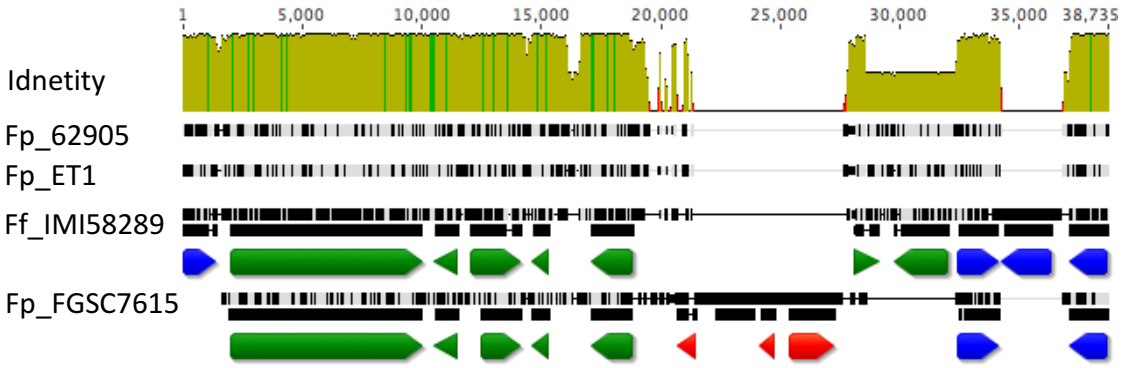

Figure 3.8 Comparison of PKS17 cluster among Ff_IMI58289 and three *F. proliferatum* (62905, ET1, FGSC7615) genomes.

Green arrows represent genes in PKS17 cluster, blue arrows represent flanking genes of the cluster, and red arrows represent additional genes in Fp_FGSC7615.

Table 3.1 Assembly statistics of the *F. proliferatum* (FGSC7615) genome assembly.

Method	number of contigs	n:N50	N50(bp)	Max(bp)	Sum(Mb)
Discovar	241	17	824278	2100024	43.1
ABySS	613	45	304297	1001439	43.1

n:N50: number of contigs with the length over N50

Table 3.2 Result of RepeatMasker.

```
==================================================
file name: Fp_discovar.fasta
sequences:           241
total length:   43151681 bp  (43150181 bp excl N/X-runs)
GC level:         48.78 %
bases masked:     329792 bp ( 0.76 %)
==================================================
               number of      length   percentage
               elements*    occupied  of sequence
--------------------------------------------------
Retroelements          154        20889 bp    0.05 %
   SINEs:                0            0 bp    0.00 %
   Penelope              0            0 bp    0.00 %
   LINEs:               15         3055 bp    0.01 %
    CRE/SLACS            1          588 bp    0.00 %
     L2/CR1/Rex          0            0 bp    0.00 %
     R1/LOA/Jockey       0            0 bp    0.00 %
     R2/R4/NeSL          0            0 bp    0.00 %
     RTE/Bov-B           0            0 bp    0.00 %
     L1/CIN4             0            0 bp    0.00 %
   LTR elements:       139        17834 bp    0.04 %
     BEL/Pao             0            0 bp    0.00 %
     Ty1/Copia          47         5425 bp    0.01 %
     Gypsy/DIRS1        92        12409 bp    0.03 %
       Retroviral        0            0 bp    0.00 %

DNA transposons         60        17357 bp    0.04 %
   hobo-Activator        6          472 bp    0.00 %
   Tc1-IS630-Pogo       26        12898 bp    0.03 %
   En-Spm                0            0 bp    0.00 %
   MuDR-IS905            0            0 bp    0.00 %
   PiggyBac              6         2138 bp    0.00 %
   Tourist/Harbinger     5          242 bp    0.00 %
   Other (Mirage,        0            0 bp    0.00 %
    P-element, Transib)

Rolling-circles          0            0 bp    0.00 %

Unclassified:            4          479 bp    0.00 %

Total interspersed repeats:       38725 bp    0.09 %


Small RNA:              16         6491 bp    0.02 %

Satellites:             60         7444 bp    0.02 %
Simple repeats:       6149       245803 bp    0.57 %
Low complexity:        666        31528 bp    0.07 %
==================================================
```

* Most repeats fragmented by insertions or deletions have been counted as one element.
The query species was assumed to be fungi
RepeatMasker version open-4.0.5, default mode
Run with rmblastn version 2.2.27+
RepBase Update 20140131, RM database version 20140131

Table 3.3 Summary of BUSCO assessment result.

Assessment mode	BUSCO notation
genome assembly	C:99%[D:8.5%],F:0.9%,M:0.0%,n:1438
annotation (gene set)	C:98%[D:12%],F:0.7%,M:0.5%,n:1438

C: complete, D: duplicated, F: fragmented, M: missing, n: total groups searched

Table 3.4 Genome statistics.

	Fp_ FGSC7615	Fp_62905[a]	Fp_ET1[a]	Ff_IMI58289[b]
Genome size (Mb)	43.1	43.2	45.2	43.9
GC-content (%)	48.8	48.7	48.1	47.4
Protein coding genes	14,393	15,254	16,143	14,813
Gene density (# genes per Mb)	334	353	358	338
Average gene length (kb)	1.7	1.5	1.4	1.5
Mean protein length (aa)	503.6	483	475.1	484.7

a: Niehaus *et al*. 2016, b: Wiemann *et al*. 2013

Table 3.5 Synteny block statistics between Fp (FGSC7615) and Ff_IMI58289.

Characters	Numbers
Total number of blocks	1592
Range of block size (in number of genes)	2 – 32
Average block size	5
Median of block size	4
Number of insertions/deletions	33

Table 3.6 Comparison of potential Cell Wall Degrading Enzymes among five *Fusarium* species.

Category	Gene family	Fp_FGSC 7615	Ff_IMI5 8289	Fp_ET1	Fp_62905
Accessary - cellulose	Alpha-glucosidases	23	26	28	28
Accessary - pectin	Beta-D-galactosidases	6	7	9	8
	Pectin methylesterases	1	1	1	1
	Rhamnogalacturonan acetylesterases	2	1	2	2
Accessary - xylan	Alpha-glucuronidases	9	12	14	14
Leaf surface - cutin	Cutinases	30	24	24	25
Main-chain degrading - cellulose	Cellobiohydrolases	1	0	0	0
Main-chain degrading - glacto(gluco)mannan	Alpha-mannosidases	17	9	9	9
	Beta-endo-mannanases	1	1	2	2
Main-chain degrading - pectin	Alpha-rhamnosidase	12	7	11	10
	Pectate lyase	23	13	16	15
	Pectin lyase	44	5	5	4
	Polygalacturonase	4	6	7	7
	Rhamnogalacturonan lyases	2	0	0	0
Main-chain degrading - xylan	Edoxylanases	7	7	11	8

Genes in Fp_FGSC7615, Ff_IMI58289, Fp_ET1, and Fp_62905 were identified from their annotation files. The genes groups were determined based on Choi *et al*. 2013.

Chapter 4 - Population analysis of *Fusarium graminearum* strains in the U.S.

Abstract

Fusarium graminearum (Fg) causes Fusarium head blight (scab disease) in wheat and barley, resulting in huge yield loss and severe food contamination worldwide. Population genomic studies can provide important information on the evolutionary processes acting in this species. Most previous population studies relied on a small set of markers across the genome. To investigate the population composition of Fg strains in the U.S. with genome wide markers, we established a feasible high throughput Genotyping-by-Sequencing (GBS) method for a Fg population study. Through genotyping over 300 Fg strains from New York and the upper Midwest in the U.S. and from South America using our GBS protocol, we generated over 8000 high quality single nucleotide polymorphism (SNP) markers for population genomic analysis with a custom bioinformatics pipeline. Principal component analysis and STRUCTURE analysis revealed two major subpopulations that were correlated, though imperfectly, to the 3-acetyl deoxynivalenol (3ADON) and 15-acetyl deoxynivalenol (15ADON) chemotypes in the U.S. Linkage disequilibrium (LD) analysis revealed a rapid initial decay over a few tens of kb followed by a slower LD decay to background levels over a distance of 200 kb to 400 kb in selected subpopulations in the U.S. In addition, reanalysis of Fg strains collected from New York with our genome-wide genotype supports the hypothesis that neither trichothecene chemotype has a clear fitness advantage, but that isolates originating from one subpopulation may on average have a fitness advantage over isolates from the other subpopulation.

Introduction

Fusarium graminearum (Fg), an important member of the *F. graminearum* species complex, is the causal pathogen of Fusarium head blight (scab disease) on wheat and barley, which leads to huge yield losses and severe grain contamination with toxic secondary metabolites like trichothecenes (Steffenson 2003). This toxin inhibits host defense response during disease spread within the spikes in wheat and barley (Bai *et al*. 2002, Jansen *et al*. 2005).

Although species in the Fg species complex produce different types of trichothecenes, type B trichothecene is predominant in isolates collected in North America, and includes deoxynivalenol (DON), 3-acetyl deoxynivalenol (3ADON), 15-acetyl deoxynivalenol (15ADON), and nivalenol (NIV) (Alexander *et al*. 2011, Liang *et al*. 2014, Varga *et al*. 2015). Recently, a novel type A trichothecene, NX-2, was identified, which is geographically restricted to southern Canada and the northern U.S. and occurs at low frequency (Liang *et al*. 2014, Varga *et al*. 2015, Kelly *et al*. 2015, 2016). Due to genetic differences that vary among Fg isolates, different amounts of alternate trichothecene toxin types accumulate in different isolates, and isolates can thus be classified into chemotypes based on the predominant trichothecene. In North America, 3ADON and 15ADON are two predominant chemotypes (Ward *et al*. 2008, Schmale *et al*. 2011). Recent surveys with a small set of markers show a temporal shift from the 15ADON chemotype to the 3ADON chemotype in the U.S. and Canada (Ward et al. 2008, Burlakoti *et al*. 2011, Liang *et al*. 2014, Kelly *et al*. 2015). With very few exceptions, there have been no population studies of U.S. Fg samples using a large number of genetic markers that are evenly and densely distributed across the genome. Therefore, Fg populations in the U.S. must be investigated using a large set of markers, such as Genotyping-by-Sequencing (GBS) markers (Poland *et al.* 2012) that can be implemented at a relatively low cost.

Several studies suggest that 3ADON strains have a pathogenic advantage compared with 15ADON strains (Ward *et al*. 2008, Puri and Zhong 2010, von der Ohe *et al*. 2010, Gilbert *et al*. 2010, Foroud *et al*. 2012, Walkowiak *et al*. 2015). In contrast, Spolti *et al*. (2014a) tested 14 pathogenic and saprophytic fitness traits using 50 strains (25 strains of each genotype) from New York, but found no significant difference between 3ADON strains and 15ADON strains. No significant 3ADON genotype shift occurred from 2007 to 2011 in western New York (Schmale *et al*. 2011, Spolti *et al*. 2014a). These results suggest that other genetic loci may contribute to this population shift. Therefore, the conflicting results of these studies highlight the need to subdivide Fg strains in the U.S. with high-resolution genome-wide analysis.

The objectives of this study are: 1) to construct and sequence GBS libraries from various Fg strains; 2) to extract high quality polymorphism markers from GBS reads for population analysis; 3) to detect patterns of subdivision of Fg strains in the U.S.; and 4) to measure the decay of linkage disequilibrium (LD) in selected subpopulations and compare the distribution of trait values between subpopulations. The hypothesis tested here was that neither trichothecene chemotype has a fitness advantage. This study provides more information on the Fg population in the U.S. and helps to clarify how different forms of trichothecene may affect fitness traits.

Materials and Methods

Fungal strains and phenotype data

Three hundred and sixty-five Fg strains were analyzed, including 42 from Uruguay, South America (SA), 131 from the Upper Midwest (UMW, 65 from Minnesota, 22 from Montana, and 44 from North Dakota) in the U.S., and 192 from New York (NY) (Table 4.1). All strains were single spored in Leslie lab at the Kansas State University following the method described in *The Fusarium Laboratory Manual* (Leslie and Summerell 2006). An analysis of

clonality was performed (Toomajian, unpublished). Fourteen previously measured saprophytic and pathogenic fitness traits from 38 of the strains from NY were retrieved (Spolti *et al*. 2014a), as were the sensitivities to fungicides tebuconazole and metconazole from the same set of strains (Spolti *et al*. 2014b).

Trichothecene genotype

Each strain was given a trichothecene genotype designation (NIV, 15ADON, 3ADON, and NX-2) based on three types of information: information provided by collaborators (Zeller *et al*. 2004, Ward *et al*. 2008, Puri and Zhong 2010, Liang *et al*. 2014), PCR assays of the *Tri12* and *Tri3* genes that result in characteristic band sizes for 15ADON, 3ADON, and NIV genotypes (Ward *et al*., 2002) (though the NX-2 genotypes gives the same band sizes as the 3ADON genotype in this assay) and a PCR assay of the *Tri1* gene designed in our lab that can distinguish NX-2 from the other genotypes, performed in our lab, and based on diagnostic single nucleotide polymorphisms (SNPs) recovered from our GBS data from the regions within and flanking the core trichothecene gene cluster and the *Tri1* gene (Toomajian, unpublished).

GBS library preparation

DNA of Fg strains was extracted and four GBS libraries were constructed using the method described in Chapter 2. All GBS libraries were sequenced using 100 bp single-end reads on an Illumina HiSeq2000.

Sequence reads alignment and SNP calling

The raw data were initially processed using the Geneious package and the custom Galaxy pipeline described in Chapter 2. As a final step in this pipeline, the trimmed reads were mapped to the *F. graminearum* PH-1 genome sequence (Broad Institute) using Bowtie2 (Langmead and Salzberg 2012). The BAM file generated from each Fg strain was sorted and indexed using

Picard-tools-1.113 (https://broadinstitute.github.io/picard/). SNPs were called using UnifiedGenotyper in the Genome Analysis Tool Kit (GATK)-3.3 (McKenna *et al*. 2010) with optimized parameters (stand_call_conf: 30, stand_emit_conf: 30, ploidy: 1, hets: 0.01, output_mode: EMIT_VARIANTS_ONLY).

Population structure analysis

The principal component analysis (PCA) was performed with filtered SNPs (filtering described in the Results section) using Smartpca with default settings in the EIGENSOFT package (Price *et al*. 2006). The result was visualized in R (R Core Team. 2017, https://www.R-project.org/). Population structure analysis was performed with the StrAuto (Chhatre 2012) program with optimized parameters (10,000 burn-in iterations, 20,000 Markov Chain Monte Carlo iterations, 10 proposed populations, 10 runs for each cluster *K,* and ploidy as 1), which automated several iterations (10 in our case) of structure analysis by the STRUCTURE (Pritchard *et al*. 2000) software. The compressed results file was uploaded to STRUCTURE HARVESTER (Earl and vonHoldt 2012) for estimation of the optimal number of inferred genetic clusters *K* using the Evanno method (Evanno *et al*. 2005). The result was plotted with CLUMPP (Jakobsson and Rosenberg 2007) and DISTRUCT (Rosenberg 2004).

Linkage disequilibrium analysis

Linkage disequilibrium (LD) was calculated for every SNP pair separated by less than 2 Mb and reported as r^2 values using PLINK v1.07 (Purcell *et al*. 2007) with options (allow-no-sex, no-fid, no-parents, no-sex, no-pheno, ld-window 99999, ld-window-kb 2000, missing-genotype -9, ld-window-r2 0).

Fitness trait analysis

To investigate systematic differences in fitness traits for different subsets of samples from New York (Spolti *et al*. 2014a), 38 isolates for which we had retrieved these published data (each isolate's average over replicates for each trait) were grouped based on PCR-determined trichothecene genotype or genome-wide population structure analysis determined in our study. For each group, the means and standard deviations were calculated. Statistical significance at the level of 0.05 was evaluated by a permutation test in R (version 3.3.3) with 10,000 randomized permutations.

Results

Processing GBS data

Four high quality GBS libraries of the collected Fg strains were constructed and sequenced on Illumina HiSeq, resulting in over 900 million raw reads. Following the same procedure in Chapter 2, filtered reads were mapped to the Fg PH1 genome assembly to create BAM files for each isolate. Over 23,000 SNPs were extracted from these BAM files by using the GATK pipeline. The 365 strains used in this study had only miniscule amounts of heterozygosity and were confirmed to be Fg by performing pairwise comparisons of genotype differences. The SNPs were filtered, first discarding all that were not biallelic, those where the minor allele was found in only a single isolate, and those where genotypes were called for less than 51% of total isolates. Of the remaining 14,247 SNPs, more than 6,000 were removed due to high LD with other SNPs on the same chromosome (PLINK software pruned SNPs on each chromosome until no pairs on the same chromosome had $r^2 \geq 0.8$). After this filtering, 8,245 high quality SNPs were used to perform a principal component analysis (PCA).

Population structure analysis

Forty-two strains from South America (SA) were identified as 15ADON (abbreviated 15 SA). Strains from the Upper Midwest (UMW) were divided into three groups, 96 15ADON (15 UMW, including 45 from Minnesota, 22 from Montana, and 29 from North Dakota), 32 3ADON (3 UMW, including 17 from Minnesota and 15 from North Dakota), and 3 NX-2 (all from Minnesota). Strains from New York (NY) were also divided into three groups, 124 15ADON (15 NY), 56 3ADON (3 NY), and 12 NX-2.

For the Principal Component Analysis (PCA), the three components that explained the greatest amount of genetic variation explained 29%, 11% and 7% respectively (Figure 4.1). At least 3 major groups were identified, one represented by 15 SA (42 strains) and two clusters in the U.S. (323 strains). The two U.S. subpopulations corresponded fairly well to the 15ADON and 3ADON chemotype designations but with some exceptions, which is consistent with previous studies (Schmale *et al*. 2011, Toomajian, unpublished). Within the 3ADON group, strains from UMW (3 UMW) and NY (3 NY) were clustered together but also included a few 15ADON strains (15 UMW and 15 NY, Figure 4.1). In contrast, the 15ADON group contained strains mainly from the UMW (15 UMW), NY (15 NY) and NX-2 chemotype strains, with the exception of a few 3ADON strains (3NY, Figure 4.1).

A complementary population structure analysis with STRUCTURE clearly suggested two different subpopulations (Figure 4.2), including subpopulation I (green, 15 SA, 3 UMW, and 3 NY) and subpopulation II (red, 15 UMW, 15 NY, and NX-2, Figure 4.3). Again some exceptions were apparent: some 3 UMW strains belong to subpopulation II (red lines among green lines in 3 UMW group, Figure 4.3), and some 15 UMW and 15 NY strains belong to subpopulation I (green lines among red lines, Figure 4.3).

Statistical analysis of fitness traits

Spolti *et al*. (2014a,b) detected no significant differences between 3ADON and 15ADON genotypes from NY in 14 saprophytic and pathogenic traits and sensitivities to 2 fungicides by comparing 25 Fg isolates from each group. Pairwise comparisons of these 50 strains with our genome wide GBS SNPs suggested some strains were clonal. After removing clonal strains, we were left with 38 out of 50 Fg strains to reanalyze for significant trait differences between grouped strains. One set of comparisons was performed on strains grouped by the trichothecene genotype (as originally done by Spolti *et al.*2014a,b) while the other was performed on strains grouped based on the inferred subpopulation from our PCA and STRUCTURE results that used genome-wide SNP genotypes. For the trichothecene genotype grouping, 21 strains were 15ADON strains and 17 were 3ADON strains (Spolti *et al.* 2014a). Only one trait (amount of 15ADON) showed a significant difference ($p = 0.0001$) between these two genotype groups (Table 4.2). Similarly, we had 20 subpopulation II strains and 18 subpopulation I strains based on our PCA and STRUCTURE results (Figure 4.4). However, in this case six traits showed significant differences ($p < 0.05$) between these two categories (Table 4.2), including mycelial growth rate, temperature sensitivity, amount of DON, amount of 3ADON, total amount of trichothecene (DON + 15ADON + 3ADON), and sensitivity to fungicide tebuconazole. Together, the significant difference in these six traits suggests that subpopulation I strains have a fitness advantage over subpopulation II strains, but the near-lack of significant differences between the trichothecene chemotype groups suggests that the fitness advantage of the subpopulation I strains is not due to the presence of the 3ADON *TRI* haplotype.

Analyses of linkage disequilibrium

The average extent of LD along chromosomes in Fg populations is important for their evolution as well as for future work on association mapping in these populations. We observed clear population structure in the U.S. Fg samples. Because population structure patterns are expected to create or increase the amount of LD observed in such an admixed population, we chose to analyze LD patterns within subpopulations with little evidence of structure. In particular, we defined as one subpopulation a set of 26 3-UMW strains that tightly clustered together in the PCA (Figure 4.1) and as a second subpopulation a set of tightly clustered 76 15-UMW strains (Figure 4.1). The observed pattern of LD decay with physical distance between SNPs was somewhat different (Figure 4.5) between the 3ADON and 15ADON subpopulations. In the 15-UMW subpopulation, LD was lower overall (average $r^2 \sim 0.2$ within 10 kb), and decayed to half of its maximal value by 20 kb and decayed to background levels by 200 kb (Figure 4.5). In contrast, the 3-UMW subpopulation had somewhat higher LD (average $r^2 > 0.4$ within 10 kb) and LD decayed more slowly to background levels by 400 kb, but like the 15-UMW subpopulation LD decayed to half of its maximal value by 20 kb.

Discussion

As an important plant disease, Fg reduces crop yield and causes contaminated food and feed all over the world. We genotyped over 300 Fg strains from the U.S. and a few strains from South America using our Genotyping-by-Sequencing (GBS) protocol adapted for *Fusarium*. Like genotyping studies in other fungi (Milgroom *et al*. 2014, Leboldus *et al*. 2015, Talas and McDonald 2015), our GBS protocol provided a feasible and cost effective procedure for population studies in Fg.

Most previous population studies of Fg in America included only a relatively small set of markers, such as AFLP markers, RFLP markers, SSR markers, etc. (Ward *et al*. 2008, Schmale *et al*. 2011, Burlakoti *et al*. 2011, Liang *et al*. 2014, Kelly *et al*. 2015). These kinds of markers may be distributed unevenly or at insufficient density across the Fg genome, which may result in biased conclusions (Talas and McDonald 2015). We conducted the first population study of Fg in America with over 8,000 SNP markers across the Fg genome and over 300 Fg strains. Our results confirmed that 3ADON and 15ADON are the two predominant chemotypes in the UMW and NY, with a small number of NX-2 isolates found in each region (Ward *et al*. 2008, Schmale *et al*. 2011, Liang *et al*. 2014, Varga *et al*. 2015, Kelly *et al*. 2015). In addition, few NIV strains were detected in these areas, but they were not included in this study.

Principal component analysis (PCA) and STRUCURE analysis are two important tools to analyze population structure (Reich *et al*. 2008, Porras-Hurtado *et al*. 2013). Our PCA result indicated the existence of two major subpopulations in the U.S. (Figure 4.1), which was confirmed by the STRUCURE analysis (Figure 4.3). Although the PCA results suggested strains from 15 SA grouped into a distinct cluster, which was close to the 3ADON group (15 SA clustered with 3ADON according to PC1, but PC3 separated these two), the STRUCURE result indicated that they all belong to one single subpopulation I (Figure 4.1, 4.3). This finding suggests that the 15 SA population is genetically more similar to the US 3ADON population (*i.e*., 3 UMW and 3 NY), and they might be derived from the same or from closely related populations. The NX-2 toxin is a novel type A trichothecene, which is structurally more like 3ADON (Varga *et al*. 2015). However, PCA and STRUCURE analysis revealed that NX-2 isolates are genetically more like US 15ADON isolates (*i.e*., 15 UMW and 15 NY). The close

genetic relationship and the recent identification suggest that isolates with the NX-2 chemotype might be derived from 15ADON chemotype ancestors in the U.S.

The results of several studies indicated that 3ADON strains were more pathogenic than 15ADON strains (Ward *et al*. 2008, Puri and Zhong 2010, von der Ohe *et al*. 2010, Gilbert *et al*. 2010, Foroud *et al*. 2012, Walkowiak *et al*. 2015). In contrast, a recent study with strains from NY indicated no significant difference between 3ADON strains and 15ADON strains (Spolti *et al*. 2014a,b). However, our analysis that grouped these same strains based on the subpopulations determined by PCA and STRUCTURE analyses that used genome-wide GBS SNPs indicated that the significant fitness advantage appears to be due to the subpopulation I background and does not result directly from the 3ADON/15ADON differences found at the main trichothecene gene cluster (Table 4.2).

Linkage disequilibrium (LD) is the non-random co-inheritance of alleles at different loci in a given population (Nordborg and Tavare 2002). The genome wide pattern of LD provides information on the history of a population (Slatkin 2008). By plotting r^2 (the correlation coefficient, a common LD measure) against physical distance, we can determine the rate of LD decay with physical distance (Gaut and Long 2003). The LD decay with distance is an important factor in the design of genome-wide association studies (GWAS) in human (Easton *et al*. 2007), Drosophila (Burke et al. 2014), plant (Thornsberry *et al*. 2001), and fungal populations (Palma-Guerrero *et al*. 2013). Both our subpopulations showed a rapid decay of LD to half of its maximum value within 20 kb. However, the decay in LD to background levels ($r^2 \approx 0$) occurred at around 400 kb and 200 kb in a 3ADON subpopulation (n=26) and a 15ADON subpopulation (n=76) respectively, in the UMW (Figure 4.5). Obviously, the 3ADON subpopulation had a higher r^2 value, *i.e*., higher LD. Two possible reasons might be responsible for it. One is that the

3ADON subpopulation has a smaller effective population size than the 15ADON subpopulation. This explanation is consistent with the history of 3ADON as a rare genotype in the US that has only recently increased in frequency. Secondly, strains in the 15ADON subpopulation appeared to be more tightly clustered together compared with those in the 3ADON subpopulation (Figure 4.1). This suggests that some population subdivision may remain in the 3ADON subpopulation, resulting in higher LD levels. A rapid decay over a short physical distance (LD decay to $r^2 < 0.2$ within 10 kb) was detected in a single meta-population (13 field populations that were separated by space but consisted of the same species) of 213 Fg strains with a high degree of recombination in Germany (Talas and McDonald 2015). Our subpopulations show a similar rate of decay to $r^2 < 0.2$, though we could still detect low levels of LD out to several hundred kb that might be due to a smaller effective population size and/or population structure in our subpopulations.

Fg, as an important plant pathogen, causes huge economic loss in the cereal industry in the U.S. No population studies of Fg isolates in the U.S. make use of genome-wide high-density markers. To address this gap, we demonstrated a feasible high throughput genotyping method for Fusaria population studies. Through genotyping over 300 Fg strains using our GBS protocol, we generated over 8000 high quality SNP markers for population genomic analysis. As a result, we detected two major subpopulations of Fg in the U.S. and determined the physical extent of LD using selected subpopulations. With a small set of strains collected from NY, we may have resolved the apparent conflict from previous studies regarding a potential fitness advantage of 3ADON genotypes. This confirmed our hypothesis that neither trichothecene chemotype has a clear fitness advantage. However, isolates originating from one subpopulation may on average have a fitness advantage over isolates from the other subpopulation. Our genome-wide SNP

markers, LD analysis, and population analysis provide valuable resources for further GWAS analysis in Fg to identify genes underlying important disease traits.

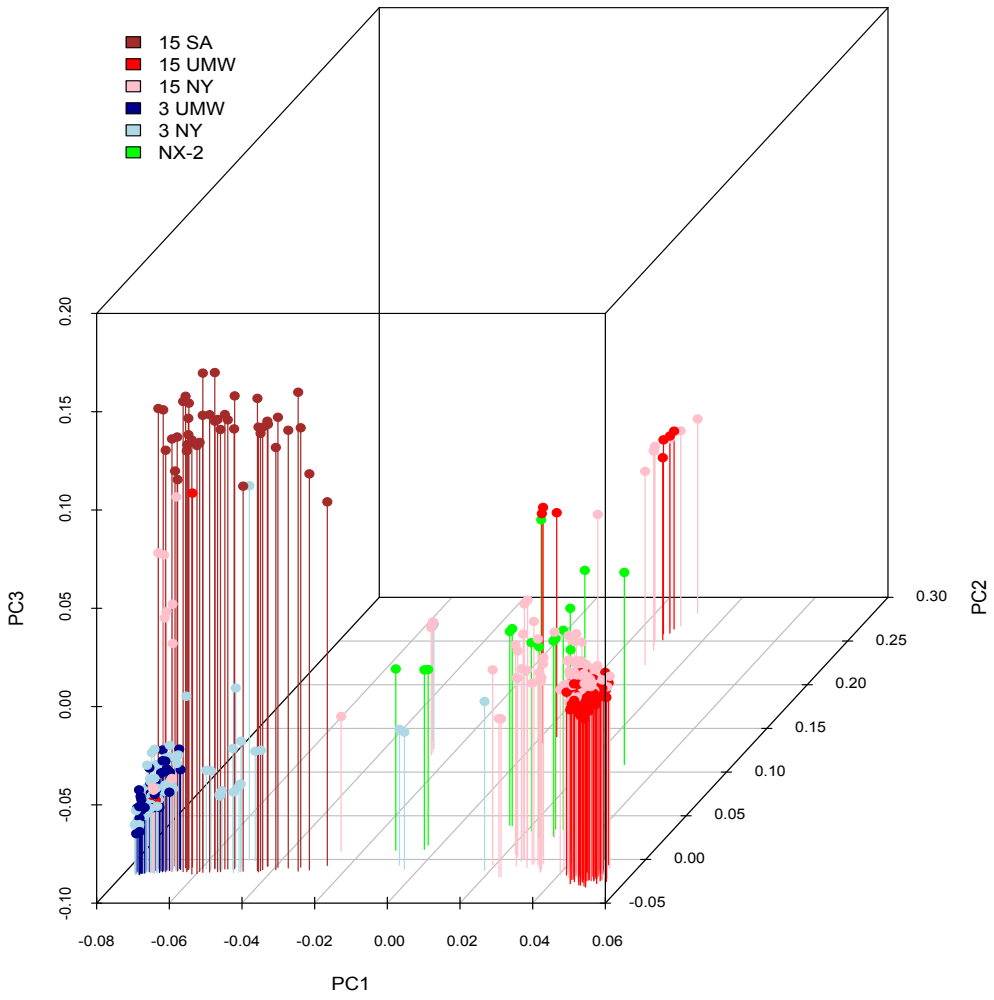

Figure 4.1 Principal component analysis.

The three dimensions shown explained 29%, 11% and 7% of the genetic variation detected among 365 strains in the analysis. 15 SA: 15ADON from South America, 15 UMW: 15ADON from Upper Midwest, 15 NY: 15ADON from New York, 3 UMW: 3ADON from Upper Midwest, 3 NY: 3ADON from New York, NX-2: new chemotype from Upper Midwest and New York.

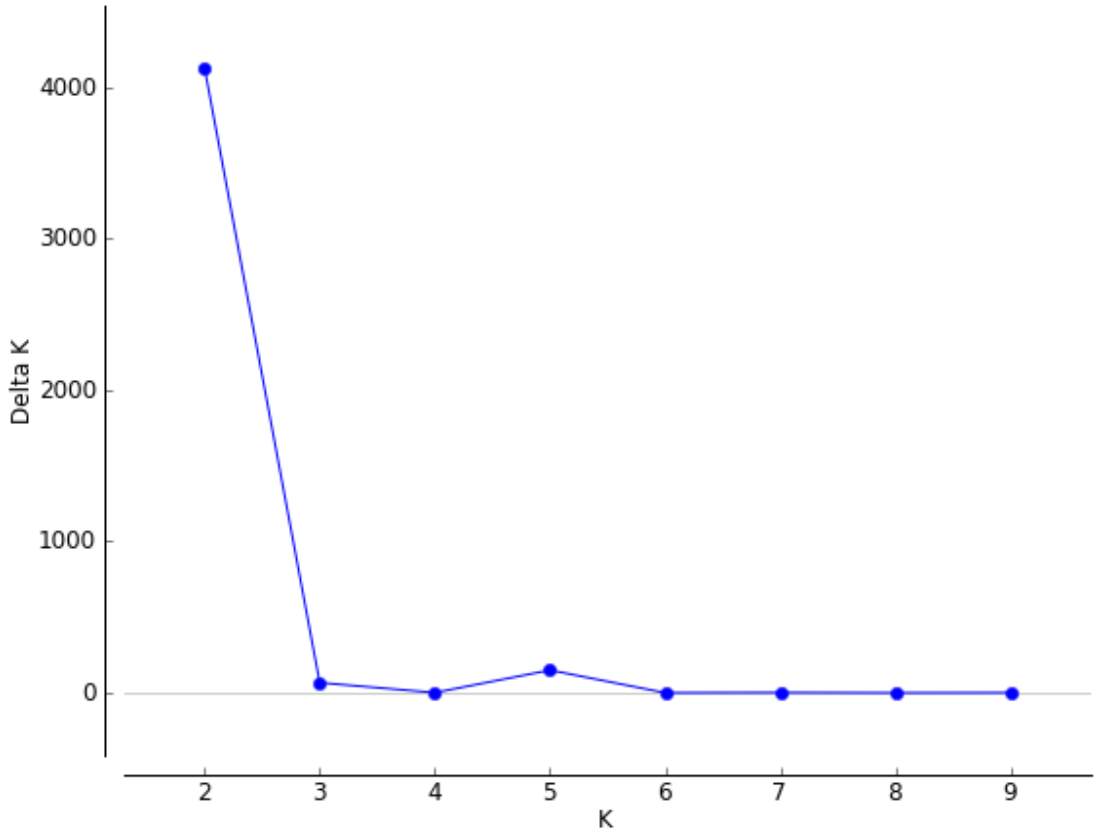

Figure 4.2 Delta K value distribution of each proposed number of populations.

K was the proposed number of populations. Delta K was calculated based on the rate of change in the log probability of data between successive K values (Evanno *et al*. 2005) in Structure Harvester (Earl and vonHoldt 2012).

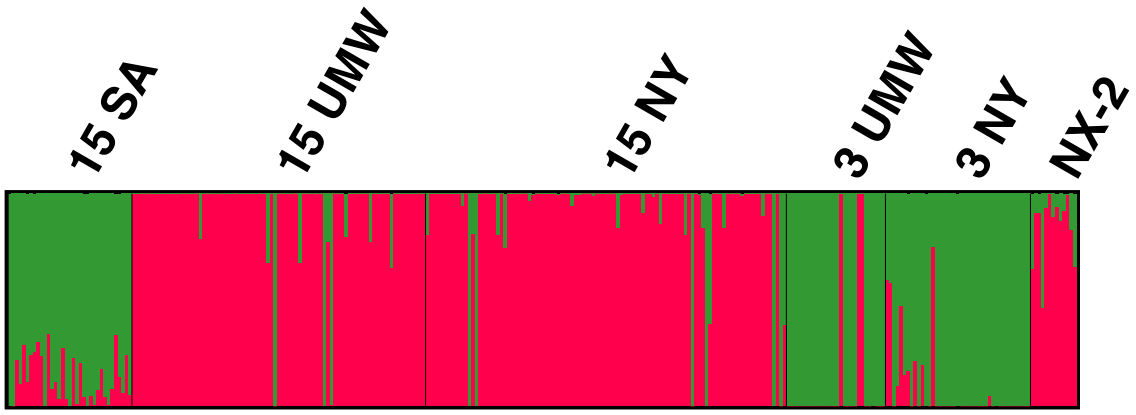

Figure 4.3 Structure analysis.

The plot suggested two different subpopulations (I-green and II-red) among 365 strains in the analysis. 15 SA: 15ADON from South America, 15 UMW: 15ADON from Upper Midwest, 15 NY: 15ADON from New York, 3 UMW: 3ADON from Upper Midwest, 3 NY: 3ADON from New York, NX-2: new chemotype from Upper Midwest and New York.

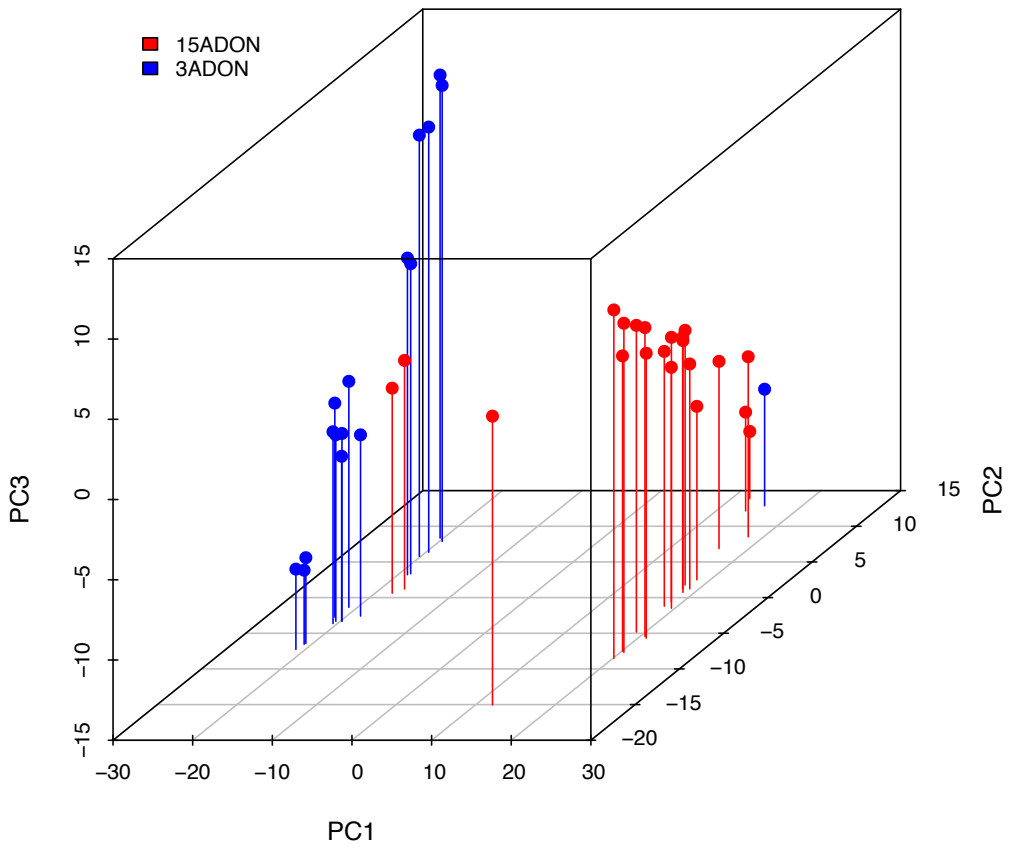

Figure 4.4 Principal component analysis of 38 *F. graminearum* strains from New York.

Two distinct subpopulations correlated with trichothecene chemotype (15ADON and 3ADON) were detected in these small dataset.

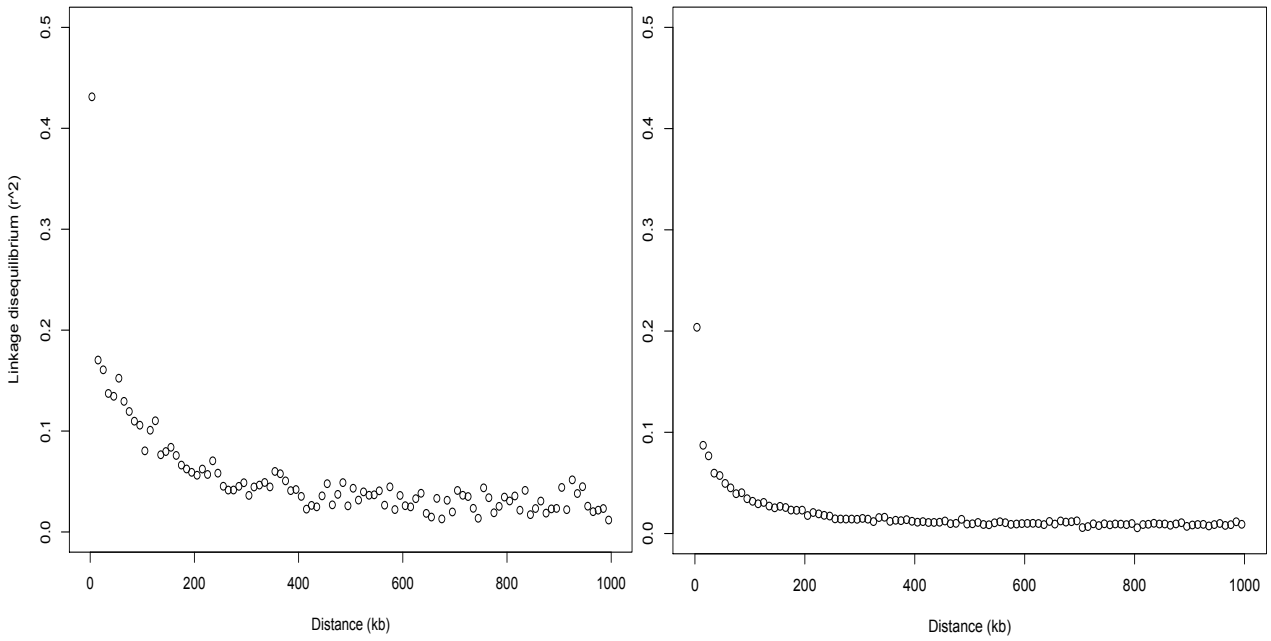

Figure 4.5 Linkage disequilibrium (LD).

Pairwise LD as measured by average r^2 for all pairs of SNPs in the genome that are separated by a given physical distance (*e.g.*, between 0 and 10 kb apart, or between 10 and 20 kb apart) in the subpopulation of 26 3UMW Fg strains (left panel) and the subpopulation of 76 15UMW Fg strains (right panel). The *x*-axis distances are grouped into 10 kb intervals.

Table 4.1 *F. graminearum* isolates used in this study.

TRI: trichothecene, SA: South America, NY: New York, MN: Minnesota, MT: Montana, ND: North Dakota.

KSU ID	location	TRI genotype	Reference	KSU ID	location	TRI genotype	Reference
17205	SA	15ADON	Toomajian unpublished	19089	MT	15ADON	Toomajian unpublished
17209	SA	15ADON	Toomajian unpublished	19090	MT	15ADON	Toomajian unpublished
17211	SA	15ADON	Toomajian unpublished	19091	MT	15ADON	Toomajian unpublished
17216	SA	15ADON	Toomajian unpublished	19092	MT	15ADON	Toomajian unpublished
17221	SA	15ADON	Toomajian unpublished	19093	MT	15ADON	Toomajian unpublished
17224	SA	15ADON	Toomajian unpublished	19094	MT	15ADON	Toomajian unpublished
17228	SA	15ADON	Toomajian unpublished	19095	MT	15ADON	Toomajian unpublished
17229	SA	15ADON	Toomajian unpublished	19096	MT	15ADON	Toomajian unpublished
17231	SA	15ADON	Toomajian unpublished	19099	MT	15ADON	Toomajian unpublished
17237	SA	15ADON	Toomajian unpublished	19104	MT	15ADON	Toomajian unpublished
17241	SA	15ADON	Toomajian unpublished	19105	MT	15ADON	Toomajian unpublished
17243	SA	15ADON	Toomajian unpublished	19106	MT	15ADON	Toomajian unpublished
17244	SA	15ADON	Toomajian unpublished	19107	MT	15ADON	Toomajian unpublished
17250	SA	15ADON	Toomajian unpublished	19109	MT	15ADON	Toomajian unpublished
17252	SA	15ADON	Toomajian unpublished	19110	MT	15ADON	Toomajian unpublished
17254	SA	15ADON	Toomajian unpublished	19111	MT	15ADON	Toomajian unpublished
17259	SA	15ADON	Toomajian unpublished	19113	MT	15ADON	Toomajian unpublished
17260	SA	15ADON	Toomajian unpublished	19114	MT	15ADON	Toomajian unpublished
17261	SA	15ADON	Toomajian unpublished	19115	MT	15ADON	Toomajian unpublished
17262	SA	15ADON	Toomajian unpublished	19116	MT	15ADON	Toomajian unpublished
17266	SA	15ADON	Toomajian unpublished	19117	MT	15ADON	Toomajian unpublished
17269	SA	15ADON	Toomajian unpublished	19118	MT	15ADON	Toomajian unpublished
17344	SA	15ADON	Toomajian unpublished	19127	NY	15ADON	Kuhnem et al. 2015
17348	SA	15ADON	Toomajian unpublished	19128	NY	15ADON	Kuhnem et al. 2015
17359	SA	15ADON	Toomajian unpublished	19129	NY	15ADON	Kuhnem et al. 2015
17365	SA	15ADON	Toomajian unpublished	19130	NY	15ADON	Kuhnem et al. 2015
17366	SA	15ADON	Toomajian unpublished	19131	NY	15ADON	Kuhnem et al. 2015
17372	SA	15ADON	Toomajian unpublished	19132	NY	15ADON	Kuhnem et al. 2015
17375	SA	15ADON	Toomajian unpublished	19133	NY	15ADON	Kuhnem et al. 2015
17376	SA	15ADON	Toomajian unpublished	19135	NY	15ADON	Kuhnem et al. 2015
17377	SA	15ADON	Toomajian unpublished	19136	NY	15ADON	Kuhnem et al. 2015
17380	SA	15ADON	Toomajian unpublished	19137	NY	15ADON	Kuhnem et al. 2015
17386	SA	15ADON	Toomajian unpublished	19139	NY	15ADON	Kuhnem et al. 2015

17388	SA	15ADON	Toomajian unpublished		19140	NY	15ADON	Kuhnem et al. 2015
17391	SA	15ADON	Toomajian unpublished		19141	NY	15ADON	Kuhnem et al. 2015
17392	SA	15ADON	Toomajian unpublished		19144	NY	15ADON	Kuhnem et al. 2015
17393	SA	15ADON	Toomajian unpublished		19145	NY	15ADON	Kuhnem et al. 2015
17401	SA	15ADON	Toomajian unpublished		19146	NY	15ADON	Kuhnem et al. 2015
17402	SA	15ADON	Toomajian unpublished		19150	NY	15ADON	Kuhnem et al. 2015
17404	SA	15ADON	Toomajian unpublished		19151	NY	15ADON	Kuhnem et al. 2015
17406	SA	15ADON	Toomajian unpublished		19153	NY	15ADON	Kuhnem et al. 2015
17410	SA	15ADON	Toomajian unpublished		19154	NY	15ADON	Kuhnem et al. 2015
19155	NY	15ADON	Kuhnem et al. 2015		23299	NY	15ADON	Kuhnem et al. 2015
19156	NY	3ADON	Kuhnem et al. 2015		23300	NY	15ADON	Kuhnem et al. 2015
23256	ND	3ADON	Puri and Zhong 2010		23301	NY	3ADON	Kuhnem et al. 2015
23257	ND	3ADON	Puri and Zhong 2010		23302	NY	15ADON	Kuhnem et al. 2015
23258	ND	3ADON	Puri and Zhong 2010		23303	NY	15ADON	Kuhnem et al. 2015
23259	ND	3ADON	Puri and Zhong 2010		23304	NY	15ADON	Kuhnem et al. 2015
23260	ND	3ADON	Puri and Zhong 2010		23305	NY	15ADON	Kuhnem et al. 2015
23261	ND	3ADON	Puri and Zhong 2010		23306	NY	15ADON	Kuhnem et al. 2015
23262	ND	15ADON	Puri and Zhong 2010		23307	NY	15ADON	Kuhnem et al. 2015
23263	ND	3ADON	Puri and Zhong 2010		23308	NY	15ADON	Kuhnem et al. 2015
23264	ND	3ADON	Puri and Zhong 2010		23309	NY	15ADON	Kuhnem et al. 2015
23265	ND	3ADON	Puri and Zhong 2010		23310	NY	15ADON	Kuhnem et al. 2015
23266	ND	3ADON	Puri and Zhong 2010		23311	NY	15ADON	Kuhnem et al. 2015
23267	ND	3ADON	Puri and Zhong 2010		23312	NY	15ADON	Kuhnem et al. 2015
23268	ND	15ADON	Puri and Zhong 2010		23313	NY	15ADON	Kuhnem et al. 2015
23269	ND	15ADON	Puri and Zhong 2010		23314	NY	15ADON	Kuhnem et al. 2015
23270	ND	3ADON	Puri and Zhong 2010		23315	NY	15ADON	Kuhnem et al. 2015
23271	ND	3ADON	Puri and Zhong 2010		23316	NY	15ADON	Kuhnem et al. 2015
23272	ND	3ADON	Puri and Zhong 2010		23317	NY	15ADON	Kuhnem et al. 2015
23276	NY	3ADON	Kuhnem et al. 2015		23318	NY	15ADON	Kuhnem et al. 2015
23277	NY	3ADON	Kuhnem et al. 2015		23319	NY	3ADON	Kuhnem et al. 2015
23278	NY	3ADON	Kuhnem et al. 2015		23320	NY	15ADON	Kuhnem et al. 2015
23279	NY	3ADON	Kuhnem et al. 2015		23321	NY	15ADON	Kuhnem et al. 2015
23280	NY	3ADON	Kuhnem et al. 2015		23322	NY	15ADON	Kuhnem et al. 2015
23281	NY	15ADON	Kuhnem et al. 2015		23323	NY	15ADON	Kuhnem et al. 2015
23282	NY	15ADON	Kuhnem et al. 2015		23324	NY	15ADON	Kuhnem et al. 2015
23283	NY	15ADON	Kuhnem et al. 2015		23325	NY	3ADON	Kuhnem et al. 2015
23284	NY	15ADON	Kuhnem et al. 2015		23326	NY	3ADON	Kuhnem et al. 2015
23285	NY	15ADON	Kuhnem et al. 2015		23327	NY	15ADON	Kuhnem et al. 2015
23286	NY	15ADON	Kuhnem et al. 2015		23328	NY	15ADON	Kuhnem et al. 2015

23287	NY	15ADON	Kuhnem et al. 2015		23329	NY	15ADON	Kuhnem et al. 2015
23288	NY	3ADON	Kuhnem et al. 2015		23330	NY	3ADON	Kuhnem et al. 2015
23289	NY	15ADON	Kuhnem et al. 2015		23331	NY	15ADON	Kuhnem et al. 2015
23290	NY	15ADON	Kuhnem et al. 2015		23332	NY	15ADON	Kuhnem et al. 2015
23291	NY	15ADON	Kuhnem et al. 2015		23333	NY	15ADON	Kuhnem et al. 2015
23292	NY	15ADON	Kuhnem et al. 2015		23334	NY	15ADON	Kuhnem et al. 2015
23293	NY	15ADON	Kuhnem et al. 2015		23335	NY	15ADON	Kuhnem et al. 2015
23294	NY	15ADON	Kuhnem et al. 2015		23336	NY	3ADON	Kuhnem et al. 2015
23295	NY	15ADON	Kuhnem et al. 2015		23337	NY	3ADON	Kuhnem et al. 2015
23296	NY	15ADON	Kuhnem et al. 2015		23338	NY	3ADON	Kuhnem et al. 2015
23297	NY	3ADON	Kuhnem et al. 2015		23339	NY	3ADON	Kuhnem et al. 2015
23298	NY	15ADON	Kuhnem et al. 2015		23340	NY	3ADON	Kuhnem et al. 2015
23341	NY	NX2	Kuhnem et al. 2015		23383	NY	15ADON	Kuhnem et al. 2015
23342	NY	3ADON	Kuhnem et al. 2015		23384	NY	3ADON	Kuhnem et al. 2015
23343	NY	NX2	Kuhnem et al. 2015		23385	NY	15ADON	Kuhnem et al. 2015
23344	NY	3ADON	Kuhnem et al. 2015		23386	NY	3ADON	Kuhnem et al. 2015
23345	NY	3ADON	Kuhnem et al. 2015		23387	NY	NX2	Kuhnem et al. 2015
23346	NY	NX2	Kuhnem et al. 2015		23388	NY	NX2	Kuhnem et al. 2015
23347	NY	3ADON	Kuhnem et al. 2015		23389	NY	NX2	Kuhnem et al. 2015
23348	NY	NX2	Kuhnem et al. 2015		23390	NY	NX2	Kuhnem et al. 2015
23349	NY	3ADON	Kuhnem et al. 2015		23391	NY	15ADON	Kuhnem et al. 2015
23350	NY	3ADON	Kuhnem et al. 2015		23392	NY	3ADON	Kuhnem et al. 2015
23351	NY	15ADON	Kuhnem et al. 2015		23393	NY	15ADON	Kuhnem et al. 2015
23352	NY	3ADON	Kuhnem et al. 2015		23394	NY	15ADON	Kuhnem et al. 2015
23353	NY	3ADON	Kuhnem et al. 2015		23395	NY	15ADON	Kuhnem et al. 2015
23354	NY	15ADON	Kuhnem et al. 2015		23396	NY	15ADON	Kuhnem et al. 2015
23355	NY	15ADON	Kuhnem et al. 2015		23397	NY	15ADON	Kuhnem et al. 2015
23356	NY	15ADON	Kuhnem et al. 2015		23398	NY	15ADON	Kuhnem et al. 2015
23357	NY	3ADON	Kuhnem et al. 2015		23399	NY	15ADON	Kuhnem et al. 2015
23358	NY	NX2	Kuhnem et al. 2015		23400	NY	15ADON	Kuhnem et al. 2015
23359	NY	3ADON	Kuhnem et al. 2015		23402	NY	15ADON	Kuhnem et al. 2015
23360	NY	15ADON	Kuhnem et al. 2015		23403	NY	15ADON	Kuhnem et al. 2015
23361	NY	15ADON	Kuhnem et al. 2015		23404	NY	15ADON	Kuhnem et al. 2015
23362	NY	15ADON	Kuhnem et al. 2015		23405	NY	15ADON	Kuhnem et al. 2015
23363	NY	15ADON	Kuhnem et al. 2015		23406	NY	15ADON	Kuhnem et al. 2015
23364	NY	15ADON	Kuhnem et al. 2015		23407	NY	15ADON	Kuhnem et al. 2015
23365	NY	15ADON	Kuhnem et al. 2015		23408	NY	15ADON	Kuhnem et al. 2015
23366	NY	NX2	Kuhnem et al. 2015		23409	NY	15ADON	Kuhnem et al. 2015
23367	NY	15ADON	Kuhnem et al. 2015		23410	NY	15ADON	Kuhnem et al. 2015

23368	NY	NX2	Kuhnem et al. 2015		23411	NY	15ADON	Kuhnem et al. 2015
23369	NY	15ADON	Kuhnem et al. 2015		23412	NY	15ADON	Kuhnem et al. 2015
23370	NY	15ADON	Kuhnem et al. 2015		23413	NY	15ADON	Kuhnem et al. 2015
23371	NY	3ADON	Kuhnem et al. 2015		23414	NY	15ADON	Kuhnem et al. 2015
23372	NY	15ADON	Kuhnem et al. 2015		23415	NY	15ADON	Kuhnem et al. 2015
23373	NY	15ADON	Kuhnem et al. 2015		23416	NY	15ADON	Kuhnem et al. 2015
23374	NY	15ADON	Kuhnem et al. 2015		23417	NY	15ADON	Kuhnem et al. 2015
23375	NY	3ADON	Kuhnem et al. 2015		23418	NY	15ADON	Kuhnem et al. 2015
23376	NY	15ADON	Kuhnem et al. 2015		23419	NY	15ADON	Kuhnem et al. 2015
23377	NY	15ADON	Kuhnem et al. 2015		23420	NY	15ADON	Kuhnem et al. 2015
23378	NY	15ADON	Kuhnem et al. 2015		23421	NY	15ADON	Kuhnem et al. 2015
23379	NY	3ADON	Kuhnem et al. 2015		23422	NY	3ADON	Kuhnem et al. 2015
23380	NY	15ADON	Kuhnem et al. 2015		23423	NY	3ADON	Kuhnem et al. 2015
23381	NY	15ADON	Kuhnem et al. 2015		23424	NY	3ADON	Kuhnem et al. 2015
23382	NY	15ADON	Kuhnem et al. 2015		23425	NY	3ADON	Kuhnem et al. 2015
23426	NY	3ADON	Kuhnem et al. 2015		23470	MN	3ADON	Liang et al. 2014
23427	NY	3ADON	Kuhnem et al. 2015		23471	MN	15ADON	Liang et al. 2014
23428	NY	3ADON	Kuhnem et al. 2015		23472	MN	3ADON	Liang et al. 2014
23429	NY	3ADON	Kuhnem et al. 2015		23473	MN	15ADON	Liang et al. 2014
23430	NY	3ADON	Kuhnem et al. 2015		23474	MN	15ADON	Liang et al. 2014
23431	NY	3ADON	Kuhnem et al. 2015		23475	MN	3ADON	Liang et al. 2014
23432	NY	3ADON	Kuhnem et al. 2015		23476	MN	15ADON	Liang et al. 2014
23433	NY	3ADON	Kuhnem et al. 2015		23477	MN	15ADON	Liang et al. 2014
23434	NY	3ADON	Kuhnem et al. 2015		23478	MN	15ADON	Liang et al. 2014
23435	NY	3ADON	Kuhnem et al. 2015		23479	MN	3ADON	Liang et al. 2014
23436	NY	3ADON	Kuhnem et al. 2015		23480	MN	15ADON	Liang et al. 2014
23437	NY	3ADON	Kuhnem et al. 2015		23481	MN	15ADON	Liang et al. 2014
23438	NY	3ADON	Kuhnem et al. 2015		23482	MN	15ADON	Liang et al. 2014
23439	NY	3ADON	Kuhnem et al. 2015		23483	MN	15ADON	Liang et al. 2014
23440	NY	3ADON	Kuhnem et al. 2015		23484	MN	15ADON	Liang et al. 2014
23441	NY	15ADON	Kuhnem et al. 2015		23485	MN	15ADON	Liang et al. 2014
23442	NY	3ADON	Kuhnem et al. 2015		23486	MN	15ADON	Liang et al. 2014
23443	NY	3ADON	Kuhnem et al. 2015		23487	MN	3ADON	Liang et al. 2014
23444	NY	3ADON	Kuhnem et al. 2015		23489	MN	NX2	Liang et al. 2014
23445	NY	NX2	Kuhnem et al. 2015		23490	MN	15ADON	Liang et al. 2014
23446	NY	15ADON	Kuhnem et al. 2015		23491	MN	15ADON	Liang et al. 2014
23447	MN	15ADON	Liang et al. 2014		23492	MN	NX2	Liang et al. 2014
23448	MN	15ADON	Liang et al. 2014		23493	MN	15ADON	Liang et al. 2014
23449	ND	15ADON	Liang et al. 2014		23494	MN	3ADON	Liang et al. 2014

23450	ND	15ADON	Liang et al. 2014		23495	MN	NX2	Liang et al. 2014
23451	ND	15ADON	Liang et al. 2014		23537	ND	15ADON	Liang et al. 2014
23452	ND	15ADON	Liang et al. 2014		23496	MN	15ADON	Liang et al. 2014
23453	ND	15ADON	Liang et al. 2014		23538	ND	15ADON	Liang et al. 2014
23454	ND	15ADON	Liang et al. 2014		23539	ND	15ADON	Liang et al. 2014
23455	ND	15ADON	Liang et al. 2014		23497	MN	15ADON	Liang et al. 2014
23456	ND	15ADON	Liang et al. 2014		23498	MN	15ADON	Liang et al. 2014
23458	ND	15ADON	Liang et al. 2014		23540	ND	15ADON	Liang et al. 2014
23459	ND	15ADON	Liang et al. 2014		23541	ND	15ADON	Liang et al. 2014
23460	ND	15ADON	Liang et al. 2014		23542	ND	15ADON	Liang et al. 2014
23461	ND	15ADON	Liang et al. 2014		23543	ND	15ADON	Liang et al. 2014
23462	ND	15ADON	Liang et al. 2014		23499	MN	3ADON	Liang et al. 2014
23463	ND	15ADON	Liang et al. 2014		23500	MN	3ADON	Liang et al. 2014
23464	ND	15ADON	Liang et al. 2014		23501	MN	15ADON	Liang et al. 2014
23466	ND	15ADON	Liang et al. 2014		23502	MN	15ADON	Liang et al. 2014
23467	ND	15ADON	Liang et al. 2014		23503	MN	15ADON	Liang et al. 2014
23468	ND	15ADON	Liang et al. 2014		23504	MN	15ADON	Liang et al. 2014
23469	MN	3ADON	Liang et al. 2014		23505	MN	15ADON	Liang et al. 2014
23514	MN	3ADON	Liang et al. 2014		23506	MN	3ADON	Liang et al. 2014
23515	MN	15ADON	Liang et al. 2014		23507	MN	15ADON	Liang et al. 2014
23516	MN	15ADON	Liang et al. 2014		23509	MN	15ADON	Liang et al. 2014
23517	MN	15ADON	Liang et al. 2014		23510	MN	3ADON	Liang et al. 2014
23518	MN	15ADON	Liang et al. 2014		23511	MN	15ADON	Liang et al. 2014
23519	MN	15ADON	Liang et al. 2014		23512	MN	15ADON	Liang et al. 2014
23520	MN	15ADON	Liang et al. 2014		23513	MN	15ADON	Liang et al. 2014
23521	MN	15ADON	Liang et al. 2014		23536	ND	15ADON	Liang et al. 2014
23522	MN	3ADON	Liang et al. 2014		23530	MN	15ADON	Liang et al. 2014
23524	MN	3ADON	Liang et al. 2014		23531	MN	3ADON	Liang et al. 2014
23525	MN	15ADON	Liang et al. 2014		23532	MN	15ADON	Liang et al. 2014
23526	MN	15ADON	Liang et al. 2014		23533	MN	15ADON	Liang et al. 2014
23527	MN	15ADON	Liang et al. 2014		23534	MN	15ADON	Liang et al. 2014
23528	MN	3ADON	Liang et al. 2014		23535	ND	3ADON	Liang et al. 2014
23529	MN	3ADON	Liang et al. 2014					

Table 4.2 Mean fitness traits of 38 *F. graminearum* strains from New York.

Strains were grouped based on PCR-based trichothecene genotype or genome wide population structure result. Fitness traits data are shown as means ± standard deviation (Spolti *et al*., 2014a,b). Significant trait differences were detected between 15ADON groups and 3ADON groups at level of 0.05.

Variable	Population structure		p-value	Trichothecene genotype		p-value
	"15ADON" (n=20)	"3ADON" (n=18)		15ADON (n=21)	3ADON (n=17)	
Growth rate	7.64±1.74	8.71±1.57	0.0278	7.88±1.75	8.48±1.69	0.1511
Ascosp_CA	213.79±230.19	216.24±167.12	0.4846	175.52±133.34	263.66±256.54	0.0830
Ascosp_CS	31.28±35.05	37.66±30.09	0.2852	33.23±35.25	35.62±29.8	0.4111
Temp_sens	3.21±1.04	3.78±0.74	0.0350	3.37±0.98	3.62±0.9	0.2093
Severity(%)	39.91±21.81	39.31±23.5	0.4728	41.99±21.85	36.71±23.2	0.2358
Below(%)	43.83±26.24	41.68±28.19	0.4045	46.08±26.37	38.78±27.65	0.2022
Above(%)	26.72±22	27.45±23.04	0.4678	28.93±21.91	24.77±23.01	0.2883
Test weight(g)	14.01±4.43	13.58±3.62	0.3821	13.48±4.56	14.2±3.34	0.3047
FDK(%)	50.63±35.43	57.47±33.31	0.2700	55.55±35.53	51.79±33.33	0.3600
Damage(%)	47.34±16.67	48.97±13.62	0.3777	49.33±17.14	46.61±12.54	0.2911
DON(ug/g)	42.1±36.2	78.82±68.67	0.0185	56.26±55.24	63.49±59.32	0.3531
3ADON(ug/g)	0.59±0.73	1.13±0.93	0.0261	0.71±0.78	1.01±0.95	0.1391
15ADON(ug/g)	0.62±0.71	0.31±0.88	0.1232	0.84±0.93	0.02±0.07	0.0001
Total(ug/g)	43.31±37.33	80.25±69.99	0.0230	57.81±56.67	64.51±60.22	0.3666
EC50TEB(mg/l)	0.91±0.28	1.59±1.7	0.0032	1.23±1.59	1.23±0.51	0.5549
EC50MET(mg/l)	0.33±0.22	0.33±0.15	0.4786	0.29±0.19	0.37±0.19	0.1022

p-value < 0.05, 10,000 permutation test in R

Growth rate: mycelial growth rate at 25℃ (millimeters/day), Ascosp_CA: ascospore carrot agar (ascospores/mm^2), Ascosp_CS: ascospore corn stalk (CFU/plate), Temp_sens: temperature sensitivity, Below: percentage of diseased spikelets below the inoculation, Above: percentage of diseased spikelets above the inoculation, Test weight: tested thousand-kernel weight, FDK: Fusarium-damaged kernels, Damage: relative weight damage, Total: (DON+3-ADON+15-ADON), EC50TEB and EC50MET: sensitivity to fungicide tebuconazole (EC50TEB) and metconazole (EC50MET)

References

Alexander, N.J., McCormick, S.P., Waalwijk, C., Van Der Lee, T., and Proctor, R.H. 2011. The genetic basis for 3-ADON and 15-ADON trichothecene chemotypes in *Fusarium*. Fungal Genetics and Biology, 48(5): 485-495.

Altschul, S., Gish, W., Miller, W., Myers, E., and Lipman, D. 1990. Basic Local Alignment Search Tool. Journal of Molecular Biology, 215(3):403-410.

Aoki T., O'Donnell K., and Geiser D. M. 2014. Systematics of key phytopathogenic *Fusarium* species: current status and future challenges. Journal of General Plant Pathology, 80:189–201.

Arias S. L., Theumer M. G., Mary V. S., and Rubinstein H. R. 2012. Fumonisins: probable role as effectors in the complex interaction of susceptible and resistant maize hybrids and *Fusarium verticillioides*. Journal of Agricutrual Food Chemestry, 60(22):5667-5675.

Arango Isaza R. E., Diaz-Trujillo C., Dhillon B., Aerts A., Carlier J., Crane C. F., de Jong T. V., de Vries I., Dietrich R., Farmer A. D., Fereira C. F., Garcia S., Guzman M., Hamelin R. C., Lindquist E. A., Mehrabi R., Quiros O., Schmutz J., Shapiro H., Reynolds E., Scalliet G., Souza M., Jr., Stergiopoulos I., Van der Lee T. A. J., De Wit P. J. G. M., Zapater M., Zwiers L., Grigoriev I. V., Goodwin S. B., and Kema G. H. J. 2016. Combating a global threat to a clonal crop: banana black sigatoka pathogen *pseudocercospora fijiensis* (synonym *Mycosphaerella fijiensis*) genomes reveal clues for disease control. PLoS Genetetics, 12(8):e1005876.

Bai, G.H., Desjardins, A.E., Plattner, R.D. 2002. Deoxynivalenol-nonproducing *Fusarium graminearum* causes initial infection, but does not cause disease spread in wheat spikes. Mycopathologia, 153(2), 91-98.

Bayram O., and Braus G. H. 2011. Coordination of secondary metabolism and development in fungi: the velvet family of regulatory proteins. FEMS Microbiology Review, 36(1):1-24.

Boehm, E.W.A., Ploetz, R.C., and Kistler, H.C. 1994. Statistical analysis of electrophoretic karyotype variation among vegetative compatibility groups of *Fusarium oxysporum* f. sp. *cubense*. Molecular Plant Microbe Interactions, 7(2): 196-207.

Bömke, C., Rojas, M.C., Gong, F., Hedden, P., and Tudzynski, B. 2008. Isolation and characterization of the gibberellin biosynthetic gene cluster in *Sphaceloma manihoticola*. Applied and Environmental Microbiology, 74(17): 5325-5339.

Bömke, C., and Tudzynski, B. 2009. Diversity, regulation, and evolution of the gibberellin biosynthetic pathway in fungi compared to plants and bacteria. Phytochemistry, 70(15): 1876-1893.

Boutet, G., Carvalho, S.A., Falque, M., Peterlongo, P., Lhuillier, E., Bouchez, O., Lavaud, C., Pilet-Nayel, M., Rivière, N., and Baranger, A. 2016. SNP discovery and genetic mapping using genotyping by sequencing of whole genome genomic DNA from a pea RIL population. BMC Genomics, 17(1): 121.

Broman K. W. and Sen S. 2009. A guide to QTL mapping with R/qtl. Springer, New York.

Broman KW, Wu H, Sen Ś, Churchill GA 2003 R/qtl: QTL mapping in experimental crosses. Bioinformatics, 19:889-890.

Burke, M.K., King, E.G., Shahrestani, P., Rose, M.R., and Long, A. D. 2014. Genome-wide association study of extreme longevity in *Drosophila melanogaster*. Genome Biology and Evolution, 6(1), 1-11.

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(6), 687-695.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. 2008. BLAST+: architecture and applications. BMC Bioinformatics ,10:421.

Carle, G.F., and Olson, M.V. 1985. An electrophoretic karyotype for yeast. Proceedings of the National Academy of Sciences USA, 82:3756-3760.

Catchen J. M., Amores A., Hohenlohe P., Cresko W., and Postlethwait J. H. 2011. Stacks: building and genotyping loci de novo from short-read sequences. G3: Genes|Genomes|Genetics, 1 3, 171–182.

Cerdá-Olmedo, E., Fernández-Martín, R. and Ávalos, J. 1994. Genetics and gibberellin production in *Gibberella fujikuroi*. Antonie van Leeuwenhoek 65(3): 217-225.

Chhatre, V.E. 2012. StrAuto ver0.3.1: A Python utility to automate Structure analysis. Available at http://www.crypticlineage.net/pages/software.html (January 2016)

Chiara, M., Fanelli, F., Mulè, G., Logrieco, A. F., Pesole, G., Leslie, J. F., Horner, D.S., and Toomajian, C. 2015. Genome sequencing of multiple isolates highlights subtelomeric genomic diversity within *Fusarium fujikuroi*. Genome Biology and Evolution, 7(11): 3062-3069.

Choi J., Kim K., Jeon J., and Lee Y. 2013. Fungal plant cell wall-degrading enzyme database: a platform for comparative and evolutionary genomics in fungi and Oomycetes. BMC Genomics, 14:S7.

Coleman J. J., Rounsley S. D., Rodriguez-Carres M., Kuo A., Wasmann C. C., Grimwood J., Schmutz J., Taga M., White G. J., Zhou S., Schwartz D. C., Freitag M., Ma L., Danchin E. G. J., Henrissat B., Coutinho P. M., Nelson D. R., Straney D., Napoli C. A., Barker B.

M., Gribskov M., Rep M., Kroken S., Molnar I., Rensing C., Kennell J. C., Zamora J., Farman M. L., Selker E. U., Salamov A., Shapiro H., Pangilinan J., Lindquist E., Lamers C., Grigoriev I. V., Geiser D. M., Covert S. F., Temporini E., and VanEtten H. D. 2009. The genome of *Nectria haematococca*: contribution of supernumerary chromosomes to gene expansion. PLoS Genetics, 5(8): e1000618.

Correll J. C., Klittich C. J. R., and Leslie J. F. 1987. Nitrate non-utilizing mutants of *Fusarium oxysporum* and their use in vegetative compatibility tests. Phytopathology 77: 1640-1646.

Cumagun, C.J.R., Bowden, R.L., Jurgenson, J.E., Leslie, J.F., and Miedaner, T. 2004. Genetic mapping of pathogenicity and aggressiveness of *Gibberella zeae* (*Fusarium graminearum*) toward wheat. Phytopathology, 94(5): 520-526.

Cuomo C. A., Gueldener U., Xu J., Trail F., Turgeon B. G., Di Pietro A., Walton J. D., Ma L., Baker S. E., Rep M., Adam G., Antoniw J., Baldwin T., Calvo S., Chang Y., 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., Mitterbauer R., Muehlbauer G., Muensterkoetter M., Nelson D., O'Donnell K., Ouellet T., Qi W., Quesneville H., Roncero M. I. G., Seong K., Tetko I. V., Urban M., Waalwijk C., Ward T. J., Yao J., Birren B. W., and Kistler H. C. 2007. The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. Science, 317(5843), 1400-1402.

Curtis P., and Cross B. 1954. Gibberellic acid - a new metabolite from the culture filtrates of *Gibberella fujikuroi*. Chemical Industries. 35:1066-1066.

Desjardins, A. E. 2006. *Fusarium* mycotoxins: chemistry, genetics, and biology. APS Press. St. Paul, Minnesota.

Desjardins, A.E., Plattner, R.D., and Gordon, T.R. 2000. *Gibberella fujikuroi* mating population A and *Fusarium subglutinans* from teosinte species and maize from Mexico and Central America. Mycological Research, 104(7): 865-872.

Desjardins, A., Plattner R., and Nelson P. 1997. Production of fumonisin B-1 and moniliformin by *Gibberella fujikuroi* from rice from various geographic areas. Applied Environmental Microbiology, 63 (5):1838-1842.

De Vos, L., Myburg, A.A., Wingfield, M.J., Desjardins, A.E., Gordon, T.R., and Wingfield, B.D. 2007. Complete genetic linkage maps from an interspecific cross between *Fusarium circinatum* and *Fusarium subglutinans*. Fungal Genetics and Biology, 44(8): 701-714.

De Vos, L., Van der Nest, M.A., Van der Merwe, N.A., Myburg, A.A., Wingfield, M.J., and Wingfield, B.D. 2011. Genetic analysis of growth, morphology and pathogenicity in the F 1 progeny of an interspecific cross between *Fusarium circinatum* and *Fusarium subglutinans*. Fungal Biology, 115(9): 902-908.

de Wit, P. J. G. M., Mehrabi R., van den Burg H. A., and Stergiopoulos I. 2009. Fungal effector proteins: past, present and future. Molecular Plant Pathology 10(6):735-747.

du Toit L., Inglis D., and Pelter G. 2003. *Fusarium proliferatum* pathogenic on onion bulbs in Washington. Plant Disease, 87(6):750-750.

Earl, D.A. and vonHoldt, B.M. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources, 4(2): 359-361.

Easton D. F., Pooley K. A., Dunning A. M., Pharoah P. D. P., Thompson D., Ballinger D. G., Struewing J. P., Morrison J., Field H., Luben R., Wareham N., Ahmed S., Healey C. S., Bowman R., Meyer K. B., Haiman C. A., Kolonel L. K., Henderson B. E., Le Marchand

L., Brennan P., Sangrajrang S., Gaborieau V., Odefrey F., Shen C., Wu P., Wang H., Eccles D., Evans D. G., Peto J., Fletcher O., Johnson N., Seal S., Stratton M. R., Rahman N., Chenevix-Trench G., Bojesen S. E., Nordestgaard B. G., Axelsson C. K., Garcia-Closas M., Brinton L., Chanock S., Lissowska J., Peplonska B., Nevanlinna H., Fagerholm R., Eerola H., Kang D., Yoo K., Noh D., Ahn S., Hunter D. J., Hankinson S. E., Cox D. G., Hall P., Wedren S., Liu J., Low Y., Bogdanova N., Schuermann P., Doerk T., Tollenaar R. A. E. M., Jacobi C. E., Devilee P., Klijn J. G. M., Sigurdson A. J., Doody M. M., Alexander B. H., Zhang J., Cox A., Brock I. W., MacPherson G., Reed M. W. R., Couch F. J., Goode E. L., Olson J. E., Meijers-Heijboer H., van den Ouweland A., Uitterlinden A., Rivadeneira F., Milne R. L., Ribas G., Gonzalez-Neira A., Benitez J., Hopper J. L., McCredie M., Southey M., Giles G. G., Schroen C., Justenhoven C., Brauch H., Hamann U., Ko Y., Spurdle A. B., Beesley J., Chen X., Mannermaa A., Kosma V., Kataja V., Hartikainen J., Day N. E., Cox D. R., Ponder B. A. J., SEARCH Collaborators, kConFab, and AOCS Management Grp. 2007. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature, 447(7148), 1087-1093.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., and Mitchell, S.E. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One, 6(5): e19379.

Evanno, G., Regnaut, S., and Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology, 14(8): 2611-2620.

Foroud, N.A., McCormick, S.P., MacMillan, T., Badea, A., Kendra, D.F., Ellis, B.E., and Eudes, F. 2012. Greenhouse studies reveal increased aggressiveness of emergent Canadian *Fusarium graminearum* chemotypes in wheat. Plant Disease, 96(9): 1271-1279.

Gale L., Bryant J., Calvo S., Giese H., Katan T., O'Donnell K., Suga H., Taga M., Usgaard T., Ward T., and Kistler H. 2005. Chromosome complement of the fungal plant pathogen *Fusarium graminearum* based on genetic and physical mapping and cytological observations. Genetics, 171(3): 985-1001.

Gale, L. R., Ward, T. J., Balmas, V., and Kistler, H. C. 2007. Population subdivision of *Fusarium graminearum* sensu stricto in the upper Midwestern United States. Phytopathology, 97(11), 1434-1439.

Gaut, B. S., and Long, A. D. 2003. The lowdown on linkage disequilibrium. The Plant Cell, 15(7), 1502-1506.

Geiser D. M., Aoki T., Bacon C. W., Baker S. E., Bhattacharyya M. K., Brandt M. E., Brown D. W., Burgess L. W., Chulze S., Coleman J. J., Correll J. C., Covert S. F., Crous P. W., Cuomo C. A., De Hoog G. S., Di Pietro A., Elmer W. H., Epstein L., Frandsen R. J. N., Freeman S., Gagkaeva T., Glenn A. E., Gordon T. R., Gregory N. F., Hammond-Kosack K. E., Hanson L. E., Jimenez-Gasco M. d. M., Kang S., Kistler H. C., Kuldau G. A., Leslie J. F., Logrieco A., Lu G., Lysoe E., Ma L., McCormick S. P., Migheli Q., Moretti A., Munaut F., O'Donnell K., Pfenning L., Ploetz R. C., Proctor R. H., Rehner S. A., Robert V. A. R. G., Rooney A. P., bin Salleh B., Mercedes Scandiani M., Scauflaire J., Short D. P. G., Steenkamp E., Suga H., Summerell B. A., Sutton D. A., Thrane U., Trail F., Van Diepeningen A., VanEtten H. D., Viljoen A., Waalwijk C., Ward T. J., Wingfield M. J., Xu J., Yang X., Yli-Mattila T., and Zhang N. 2013. One fungus, one name:

defining the genus *Fusarium* in a scientifically robust way that preserves longstanding use. Phytopathology, 103(5): 400-408.

Gilbert, J., Brûlé-Babel, A., Guerrieri, A.T., Clear, R.M., Patrick, S., Slusarenko, K., and Wolfe, C. 2014. Ratio of 3-ADON and 15-ADON isolates of Fusarium graminearum recovered from wheat kernels in Manitoba from 2008 to 2012. Canadian Journal of Plant Pathology, 36(1): 54-63.

Gilbert, J., Clear, R. M., Ward, T. J., Gaba, D., Tekauz, A., Turkington, T. K., Woods, S.M. and O'Donnell, K. 2010. Relative aggressiveness and production of 3-or 15-acetyl deoxynivalenol and deoxynivalenol by *Fusarium graminearum* in spring wheat. Canadian Journal of Plant Pathology, 32(2), 146-152.

Goecks J., Nekrutenko A., and Taylor J. 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biology 11(8), R86.

Harrison, L.R., Colvin, B.M., Green, J.T., Newman, L.E. and Cole, J.R. 1990. Pulmonary edema and hydrothorax in swine produced by fumonisin B1, a toxic metabolite of *Fusarium moniliforme*. Journal of Veterinary Diagnostic Investigation 2: 217-221.

Hoffmeister, D., and Keller, N.P. 2007. Natural products of filamentous fungi: enzymes, genes, and their regulation. Natural Product Reports, 24(2): 393-416.

Howson, W. T., R. C. McGinnis, and W. L. Gordon. 1963. Cytological studies on the perfect stages of some species of Fusarium. Canadian Journal of Genetics Cytology, 5:60–64.

Jakobsson, M., and 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(14): 1801-1806.

Jansen, C., Von Wettstein, D., Schäfer, W., Kogel, K. H., Felk, A., & Maier, F. J. 2005. Infection patterns in barley and wheat spikes inoculated with wild-type and trichodiene synthase gene disrupted *Fusarium graminearum*. Proceedings of the National Academy of Sciences USA, 102(46), 16892-16897.

Jeong, H., Lee, S., Choi, G.J., Lee, T., & Yun, S.-H. 2013. Draft genome sequence of *Fusarium fujikuroi* B14, the causal agent of the bakanae disease of rice. Genome Announcements, 1(1): e00035–13.

Jurgenson J., Bowden R., Zeller K., Leslie J., Alexander N., and Plattner R. 2002a. A genetic map of *Gibberella zeae* (*Fusarium graminearum*). Genetics 160 (4):1451-1460.

Jurgenson J., Zeller K., and Leslie J. 2002b. Expanded genetic map of *Gibberella moniliformis* (*Fusarium verticillioides*). Applied Environmental Microbiology, 68 (4):1972-1979.

Jurka J., Kapitonov V., Pavlicek A., Klonowski P., Kohany O., and Walichiewicz J. 2005. Repbase update, a database of eukaryotic repetitive elements. Cytogenetics Genome Research, 110(1-4):462-467.

Kanehisa M., Araki M., Goto S., Hattori M., Hirakawa M., Itoh M., Katayama T., Kawashima S., Okuda S., Tokimatsu T., and Yamanishi Y. 2008. KEGG for linking genomes to life and the environment. Nucleic Acids Research, 36:D480-D484.

Kawaide, H. 2006. Biochemical and molecular analyses of gibberellin biosynthesis in fungi. Bioscience, Biotechnology, and Biochemistry, 70(3): 583-590.

Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Mentjies P., and Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12), 1647-1649.

Kellerman, T.S., Marasas, W.F.O., Thiel, P.G., Gelderblom, W.C.A., Cawood, M. and Coetzer, J.A.W. 1990. Leukoencephalomalacia in two horses induced by oral dosing of fumonisin B1. Onderstepoort Journal of Veterinary Research 57: 269-275.

Kelly A. C., Clear R. M., O'Donnell K., McCormick S., Turkington T. K., Tekauz A., Gilbert J., Kistler H. C., Busman M., and Ward T. J. 2015. Diversity of Fusarium head blight populations and trichothecene toxin types reveals regional differences in pathogen composition and temporal dynamics. Fungal Genetics and Biology, 82: 22-31.

Kelly A., Proctor R. H., Belzile F., Chulze S. N., Clear R. M., Cowger C., Elmer W., Lee T., Obanor F., Waalwijk C., and Ward T. J. 2016. The geographic distribution and complex evolutionary history of the NX-2 trichothecene chemotype from *Fusarium graminearum*. Fungal Genetics and Biology, 95, 39-48.

Kirkpatrick M. 2010. How and why chromosome inversions evolve. PLoS. Biology, 8(9):e1000501.

Kistler, H.C., Rep, M., and Ma, L.J. 2013. Structural dynamics of *Fusarium* genomes. *Fusarium*: Genomics, Molecular and Cellular Biology. Caister Academic Press. Poole, U.K.

Koonin E., Fedorova N., Jackson J., Jacobs A., Krylov D., Makarova K., Mazumder R., Mekhedov S., Nikolskaya A., Rao B., Rogozin I., Smirnov S., Sorokin A., Sverdlov A., Vasudevan S., Wolf Y., Yin J., and Natale D. 2004. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. Genome Biology, 5(2):R7.

Kubicek C. P., Starr T. L., and Glass N. L. 2014. Plant cell wall-degrading enzymes and their secretion in plant-pathogenic fungi. Annual Review Phytopathology, 52:427-451.

Kuhnem, P.R., Spolti, P., Del Ponte, E.M., Cummings, J.A., and 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(5), 695-699.

Kurtz, S., Phillippy, A., Delcher, A., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S. 2004. Versatile and open software for comparing large genomes. Genome Biology, 5(2):R12.

Kvas, M., Marasas, W.F.O., Wingfield, B.D., Wingfield, M.J. and Steenkamp, E.T. 2009. Diversity and evolution of *Fusarium* species in the *Gibberella fujikuroi* complex. Fungal Diversity, 34: 1-21.

Langmead, B., and Salzberg, S.L. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods, 9(4): 357-359.

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

Lee J., Jurgenson J. E., Leslie J. F., and Bowden R. L. 2008. Alignment of genetic and physical maps of *Gibberella zeae*. Applied and Environmental Microbiology, 74(8):2349-2359.

Leslie, J.F., Anderson, L.L., Bowden, R.L. and Lee, Y.W. 2007. Inter-and intra-specific genetic variation in *Fusarium*. International Journal of Food Microbiology 119: 25-32.

Leslie J. F. and Summerell B. A. 2006. The *Fusarium* laboratory manual. Blackwell Publishing, Ames, Iowa.

Leslie J. F., Zeller K. A., Wohler M., and Summerell B. A. 2004. Interfertility of two mating populations in the *Gibberella fujikuroi* species complex. European Journal of Plant Pathology, 110: 610-618.

Li L., Stoeckert C., and Roos D. 2003. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Research. 13 9:2178-2189.

Liang, J.M., Xayamongkhon, H., Broz, K., Dong, Y., McCormick, S.P., Abramova, S., Ward, T.J., Ma, Z.H. and Kistler, H. C. 2014. Temporal dynamics and population genetic structure of *Fusarium graminearum* in the upper Midwestern United States. Fungal Genetics and Biology, 73: 83-92.

Link H.F. 1809. Observationes in ordines plantarum naturals, Dissetatio I. Magazin der Gesellschaft Naturforschenden Freunde, Berlin 3:3–42.

Lowe T., and Eddy S. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Research, 25(5):955-964.

Lyons E., Pedersen B., Kane J., Freeling M. 2008. The value of nonmodel genomes and an example using SynMap within CoGe to dissect the hexaploidy that predates rosids. Tropical Plant Biology, 1(3):181-190.

Ma L., Geiser D. M., Proctor R. H., Rooney A. P., O'Donnell K., Trail F., Gardiner D. M., Manners J. M., and Kazan K. 2013. *Fusarium* Pathogenomics. Annual Review of Microbiology, 67:399-416.

Ma L., van der Does H. C., Borkovich K. A., Coleman J. J., Daboussi M., Di Pietro A., Dufresne M., Freitag M., Grabherr M., Henrissat B., Houterman P. M., Kang S., Shim W., Woloshuk C., Xie X., Xu J., 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., Li L., Manners J. M., Miranda-Saavedra D., Mukherjee M., Park G., Park J., Park S., Proctor R. H., Regev A., Carmen Ruiz-Roldan M., 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., and Rep M. 2010. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. Nature, 464(7287), 367-373.

Malonek, S., Bömke, C., Bornberg-Bauer, E., Rojas, M.C., Hedden, P., Hopkins, P. and Tudzynski, B. 2005c. Distribution of gibberellin biosynthetic genes and gibberellin production in the *Gibberella fujikuroi* species complex. Phytochemistry 66: 1296-1311.

Malonek, S., Rojas, M.C., Hedden, P., Gaskin, P., Hopkins, P., and Tudzynski, B. 2005a. Functional characterization of two cytochrome P450 monooxygenase genes, P450-1 and P450-4, of the gibberellic acid gene cluster in *Fusarium proliferatum* (*Gibberella fujikuroi* MP-D). Applied and Environmental Microbiology, 71(3): 1462-1472.

Malonek, S., Rojas, M.C., Hedden, P., Hopkins, P., and Tudzynski, B. 2005b. Restoration of gibberellin production in *Fusarium proliferatum* by functional complementation of enzymatic blocks. Applied and Environmental Microbiology, 71(10): 6014-6025.

Marasas, W.F.O. 2001. Discovery and occurrence of the fumonisins: A historical perspective. Environmental Health Perspectives, 109: 239-243.

Marchler-Bauer, A., and Bryant, S.H. (2004). CD-Search: protein domain annotations on the fly. Nucleic Acids Research, 32(suppl 2): W327-W331.

Marchler-Bauer A., Lu S., Anderson J. B., Chitsaz F., Derbyshire M. K., DeWeese-Scott C., Fong J. H., Geer L. Y., Geer R. C., Gonzales N. R., Gwadz M., Hurwitz D. I., Jackson J.

D., Ke Z., Lanczycki C. J., Lu F., Marchler G. H., Mullokandov M., Omelchenko M. V., Robertson C. L., Song J. S., Thanki N., Yamashita R. A., Zhang D., Zhang N., Zheng C., and Bryant S. H. 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Research, 39(suppl 1): D225-D229.

Mascher M., Wu S., Amand P.S., Stein N., Poland J. 2013. Application of Genotyping-by-Sequencing on semiconductor sequencing platforms: a comparison of genetic and reference-based marker ordering in barley. PLoS One 8(10): e76925.

McCann M., Roberts K. 1991. Architecture of the primary cell wall. The cytoskeletal basis of plant growth and form. Academic Press. Ann Arbor, Michigan.

McCormick, S.P., Stanley, A.M., Stover, N.A., and Alexander, N.J. 2011. Trichothecenes: from simple to complex mycotoxins. Toxins, 3(7): 802-814.

McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K., Altshuler D., Gabriel S., Daly M., and DePristo M. A. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 20(9): 1297-1303.

Melen K., Krogh A., and von Heijne G. 2003. Reliability measures for membrane protein topology prediction algorithms. Journal of Molecular Biology, 327(3):735-744.

Michielse, C.B., Pfannmüller, A., Macios, M., Rengers, P., Dzikowska, A., and Tudzynski, B. 2014. The interplay between the GATA transcription factors AreA, the global nitrogen regulator and AreB in *Fusarium fujikuroi*. Molecular Microbiology, 91(3): 472-493.

Mihlan M., Homann V., Liu T., and Tudzynski B. 2003. AREA directly mediates nitrogen regulation of gibberellin biosynthesis in *Gibberella fujikuroi*, but its activity is not affected by NMR. Molecular Microbiology, 47(4):975-991.

Milgroom, M.G., del Mar Jiménez-Gasco, M., García, C.O., Drott, M.T., and Jiménez-Díaz, R.M. 2014. Recombination between clonal lineages of the asexual fungus *Verticillium dahliae* detected by genotyping by sequencing. PLoS One, 9(9): e106740.

Mohamed Nor, N.M.I. 2014. Genetics of Southeast Asian populations and interspecific hybrids of *Fusarium spp*. Ph.D. Dissertation. Department of Plant Pathology, Kansas State University, Manhattan, Kansas.

Morris G. P., Ramu P., Deshpande S. P., Hash C. T., Shah T., Upadhyaya H. D., Riera-Lizarazu O., Brown P. J., Acharya C. B., Mitchell S. E., Harriman J., Glaubitz J. C., Buckler E. S., and Kresovich S. 2013. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proceedings of the National Academy of Sciences USA, 110(2): 453-458.

Nahlik K., Dumkow M., Bayram O., Helmstaedt K., Busch S., Valerius O., Gerke J., Hoppert M., Schwier E., Opitz L., Westermann M., Grond S., Feussner K., Goebel C., Kaever A., Meinicke P., Feussner I., and Braus G. H. 2010. The COP9 signalosome mediates transcriptional and metabolic response to hormones, oxidative stress protection and cell wall rearrangement during fungal development. Molecular Microbiology, 78(4):964-979.

Ng M., Vergara I. A., Frech C., Chen Q., Zeng X., Pei J., and Chen N. 2009. OrthoClusterDB: an online platform for synteny blocks. BMC Bioinformatics 10:192.

Niehaus E., Janevska S., von Bargen K. W., Sieber C. M. K., Harrer H., Humpf H., and Tudzynski B. 2014a. Apicidin F: Characterization and genetic manipulation of a new secondary metabolite gene cluster in the rice pathogen *Fusarium fujikuroi*. PLoS One 9(7):e103336.

Niehaus, E.M., von Bargen, K.W., Espino, J.J., Pfannmüller, A., Humpf, H.U., and Tudzynski, B. 2014b. Characterization of the fusaric acid gene cluster in Fusarium fujikuroi. Applied Microbiology and Biotechnology, 98(4): 1749-1762.

Niehaus E., Muensterkoetter M., Proctor R. H., Brown D. W., Sharon A., Idan Y., Oren-Young L., Sieber C. M., Novak O., Pencik A., Tarkowska D., Hromadova K., Freeman S., Maymon M., Elazar M., Youssef S. A., El-Shabrawy E. S. M., Shalaby A. B. A., Houterman P., Brock N. L., Burkhardt I., Tsavkelova E. A., Dickschat J. S., Galuszka P., Gueldener U., and Tudzynski B. 2016. Comparative "Omics" of the *Fusarium fujikuroi* Species Complex Highlights Differences in Genetic Potential and Metabolite Synthesis. Genome Biology and Evolution, 8(11):3574-3599.

Nordborg, M., and Tavare, S. 2002 Linkage disequilibrium: What history has to tell us. Trends in Genetics 18: 83-90.

O'Donnell, K., Cigelnik, E., Nirenberg, H.I. 1998. Molecular systematics and phylogeography of the Gibberella fujikuroi species complex. Mycologia 90:465–493.

O'Donnell, K., Ward, T.J., Geiser, D.M., Kistler, H.C., and Aoki, T. (2004). Genealogical concordance between the mating type locus and seven other nuclear genes supports formal recognition of nine phylogenetically distinct species within the *Fusarium graminearum* clade. Fungal Genetics and Biology, 41(6): 600-623.

Ohm R. A., Feau N., Henrissat B., Schoch C. L., Horwitz B. A., Barry K. W., Condon B. J., Copeland A. C., Dhillon B., Glaser F., Hesse C. N., Kosti I., LaButti K., Lindquist E. A., Lucas S., Salamov A. A., Bradshaw R. E., Ciuffetti L., Hamelin R. C., Kema G. H. J., Lawrence C., Scott J. A., Spatafora J. W., Turgeon B. G., de Wit P. J. G. M., Zhong S., Goodwin S. B., and Grigoriev I. V. 2012. Diverse lifestyles and strategies of plant

pathogenesis encoded in the genomes of eighteen dothideomycetes fungi. PLoS Pathogens, 8 12:e1003037.

Ott, J. 1991. Analysis of human genetic linkage. Johns Hopkins, Baltimore, MD.

Palma-Guerrero, J., Hall, C. R., Kowbel, D., Welch, J., Taylor, J.W., Brem, R.B., and Glass, N.L. 2013. Genome wide association identifies novel loci involved in fungal communication. PLoS Genetics, 9(8), e1003669.

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

Phinney, B.O. 1983. The history of gibberellins. In: Crozier A (ed) The biochemistry and physiology of gibberellins. Praeger, New York.

Poland J.A., Brown P.J., Sorrells M.E., and Jannink J-L. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme Genotyping-by-Sequencing Approach. PLoS One, 7(2): e32253.

Porras-Hurtado, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, Á., and Lareu, M. 2013. An overview of STRUCTURE: applications, parameter settings, and supporting software. Frontiers in Genetics, 4, 98.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics, 38(8): 904-909.

Pritchard, J.K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics, 155(2): 945-959.

Puhalla, J.E., and Spieth, P.T. 1985. A comparison of heterokaryosis and vegetative incompatibility among varieties of *Gibberella fujikuroi* (*Fusarium moniliforme*). Experimental Mycology, 9(1): 39-47.

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., and Sham P. C. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics, 81(3): 559-575.

Puri, K.D., and 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(10): 1007-1014.

Rademacher, W. 1997. Gibberellins. In: Anke T, Ed. Fungal Biotechnology. London: Chapman & Hall.

R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reich, D., Price, A.L., and Patterson, N. 2008. Principal component analysis of genetic data. Nature Genetics, 40(5), 491-492.

Rep, M., and Kistler, H.C. 2010. The genomic organization of plant pathogenicity in *Fusarium* species. Current Opinion in Plant Biology, 13(4): 420-426.

Rosenberg, N.A. 2004. DISTRUCT: a program for the graphical display of population structure. Molecular Ecology Notes, 4(1): 137-138.

Schmale, D.G., Wood‐Jones, A.K., Cowger, C., Bergstrom, G.C., and Arellano, C. 2011. Trichothecene genotypes of *Gibberella zeae* from winter wheat fields in the eastern USA. Plant Pathology, 60(5): 909-917.

Schwartz, D.C., and Cantor, C.R. 1984. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel-electrophoresis. Cell. 37, 67-75.

Schwechheimer C., Serino G., Callis J., Crosby W., Lyapina S., Deshaies R., Gray W., Estelle M., and Deng X. 2001. Interactions of the COP9 signalosome with the E3 ubiquitin ligase SCFTIR1 in mediating auxin response. Science, 292(5520), 1379-1382.

Selker, E. U., Cambareri, E. B., Jensen, B. C., and Haack, K. R. 1987. Rearrangement of duplicated DNA in specialized cells of *Neurospora*. Cell, 51(5), 741-752.

Simao, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19):3210-3212.

Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J.M., and Birol, I. 2009. ABySS: A parallel assembler for short read sequence data. Genome Res. 19(6):1117-1123.

Slatkin, M. 2008. Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. Nature reviews. Genetics, 9(6), 477-485.

Smit, A.F., Hubley, R., and Green, P. RepeatMasker Open-3.0. 1996-2010. (http://www.repeatmasker.org)

Sonah H., Bastien M., Iquira E., Tardivel A., Legare G., Boyle B., Normandeau E., Laroche J., Larose S., Jean M., and Belzile F. 2013. An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. PLoS One, 8(1): e54603.

Spolti, P., Del Ponte, E. M., Cummings, J. A., Dong, Y., and 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.

Spolti, P., Del Ponte, E. M., Dong, Y., Cummings, J. A., and 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 Disease. 98:607-613.

Sperschneider J., Gardiner D. M., Thatcher L. F., Lyons R., Singh K. B., Manners J. M., and Taylor J. M. 2015. Genome-wide analysis in three *Fusarium* pathogens identifies rapidly evolving chromosomes and genes associated with pathogenicity. Genome Biology and Evolution, 7(6):1613-1627.

Stankovic S., Levic J., Petrovic T., Logrieco A., and Moretti A. 2007. Pathogenicity and mycotoxin production by *Fusarium proliferatum* isolated from onion and garlic in Serbia. European Jorunal of Plant Pathology, 118(2):165-172.

Steffenson, B. J. 2003. Fusarium head blight of barley: Impact, epidemics, management, and strategies for identifying and utilizing genetic resistance. Fusarium Head Blight of Wheat and Barley. The American Phytopathological Society, St. Paul, MN.

Studt L., Troncoso C., Gong F., Hedden P., Toomajian C., Leslie J. F., Humpf H-U., Rojas M. C., and Tudzynski B. 2012. Segregation of secondary metabolite biosynthesis in hybrids of *Fusarium fujikuroi* and *Fusarium proliferatum*. Fungal Genetics Biology, 49: 567-577.

Suga, H., Ikeda, S., Taga, M., Kageyama, K., Hyakumachi, M. 2002. Electrophoretic karyotyping and gene mapping of seven formae speciales in *Fusarium solani*. Current Genetics, 41(4): 254-260.

Taga, M., Waalwijk, C., Flier, W. G., and Kema, G. H. J. 2003. Cytological karyotyping of somatic chromosomes from *Phytophthora infestans*, *Mycosphaerella graminicola*, and *Fusarium* spp. Fungal Genetics Newsletter, 50, 468.

Takahashi N., Kitamura H., Kawarada A., Seta Y., Takai M., Tamura S., and Sumiki Y. 1955. Biochemical studies on bakanae fungus .34. Isolation of gibberellins and their properties. Bulletin of the Agricultural Chemical Society of Japan, 19(4):267-277.

Talas, F., and McDonald, B.A. 2015. Genome-wide analysis of *Fusarium graminearum* field populations reveals hotspots of recombination. BMC Genomics, 16(1), 996.

Talbot, N. J. 2003. On the trail of a cereal killer: exploring the biology of *Magnaporthe grisea*. Annual Reviews in Microbiology, 57(1), 177-202.

Thornsberry, J.M., Goodman, M.M., Doebley, J., Kresovich, S., Nielsen, D., and Buckler, E.S. 2001. Dwarf8 polymorphisms associate with variation in flowering time. Nature Genetics, 28(3), 286-289.

Troncoso, C., González, X., Bömke, C., Tudzynski, B., Gong, F., Hedden, P., and Rojas, M. C. 2010. Gibberellin biosynthesis and gibberellin oxidase activities in *Fusarium sacchari*, *Fusarium konzum* and *Fusarium subglutinans* strains. Phytochemistry, 71(11), 1322-1331.

Tsavkelova, E. A., Boemke, C., Netrusov, A. I., Weiner, J., and Tudzynski, B. 2008. Production of gibberellic acids by an orchid-associated *Fusarium proliferatum* strain. Fungal Genetics Biology, 45: 1393-1403.

Tsavkelova, E., Oeser, B., Oren-Young, L., Israeli, M., Sasson, Y., Tudzynski, B., and Sharon, A. 2012. Identification and functional characterization of indole-3-acetamide-mediated IAA biosynthesis in plant-associated *Fusarium* species. Fungal Genetics and Biology, 49(1): 48-57.

Tudzynski B. 1999. Biosynthesis of gibberellins in *Gibberella fujikuroi*: biomolecular aspects. Applied Microbiology Biotechnology, 52: 298-310.

Tudzynski B. 2005. Gibberellin biosynthesis in fungi: genes, enzymes, evolution, and impact on biotechnology. Applied Microbiology Biotechnology, 66: 597–611.

Tudzynski, B. 2014. Nitrogen regulation of fungal secondary metabolism in fungi. Frontiers in Microbiology, 5: 656.

Tudzynski B., and Holter K. 1998. Gibberellin biosynthetic pathway in *Gibberella fujikuroi*: Evidence for a gene cluster. Fungal Genetics and Biology, 25(3):157-170.

Tudzynski, B., Homann, V., Feng, B., and Marzluf, G.A. 1999. Isolation, characterization and disruption of the *areA* nitrogen regulatory gene of *Gibberella fujikuroi*. Molecular and General Genetics, 261(1): 106-114.

Ueno, Y. 1980. Trichothecene mycotoxins mycology, chemistry, and toxicology. In Advances in Nutritional Research. Springer US.

Varga E., Wiesenberger G., Hametner C., Ward T. J., Dong Y., Schoefbeck D., McCormick S., Broz K., Stueckler R., Schuhmacher R., Krska R., Kistler H. C., Berthiller F., and Adam G. 2015. New tricks of an old enemy: isolates of *Fusarium graminearum* produce a type A trichothecene mycotoxin. Environmental Microbiology, 17(8): 2588-2600.

Vollrath, D., and Davis, R.W. 1987. Resolution of DNA molecules greater than 5 megabases by contour-clamped homogeneous electric fields. Nucleic Acids Research. 15(19):7865–7876.

von der Ohe C., Gauthier V., Tamburic-Ilincic L., Brule-Babel A., Fernando W. G. D., Clear R., Ward T. J., and 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. European Journal of Plant Pathology, 127(3), 407-417.

Walkowiak, S., Bonner, C. T., Wang, L., Blackwell, B., Rowland, O., and Subramaniam, R. 2015. Intraspecies interaction of Fusarium graminearum contributes to reduced toxin production and virulence. Molecular plant-microbe interactions, 28(11), 1256-1267.

Ward, T. J., Bielawski, J. P., Kistler, H. C., Sullivan, E., and O'Donnell, K. 2002. Ancestral polymorphism and adaptive evolution in the trichothecene mycotoxin gene cluster of phytopathogenic *Fusarium*. Proceedings of the National Academy of Sciences USA, 99(14), 9278-9283.

Ward T. J., Clear R. M., Rooney A. P., O'Donnell K., Gaba D., Patrick S., Starkey D. E., Gilbert J., Geiser D. M., and 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 Genetics and Biology, 45(4): 473-484.

Wiemann P., Brown D. W., Kleigrewe K., Bok J. W., Keller N. P., Humpf H., and Tudzynski B. 2010. FfVel1 and FfLae1, components of a *velvet*-like complex in *Fusarium fujikuroi*, affect differentiation, secondary metabolism and virulence. Molecular Microbiology, 77 (4):972-994.

Wiemann P., Sieber C. M. K., Von Bargen K. W., Studt L., Niehaus E., Espino J. J., Huss K., Michielse C.B., Albermann S., Wagner D., Bergner S. V., Connolly L. R., Fischer A., Reuter G., Kleigrewe K., Bald T., Wingfield B. D., Ophir R., Freeman S., Hippler M., Smith K. M., Brown D. W., Proctor R. H., Munsterkotter M., Freitag M., Humpf H-U., Guldener U., and Tudzynski B. 2013. Deciphering the cryptic genome: Genome-wide analyses of the rice pathogen *Fusarium fujikuroi* reveal complex regulation of secondary metabolism and novel metabolites. PLoS Pathogens, 9: 1-35.

Windels, C.E. 2000. Economic and social impacts of Fusarium head blight: changing farms and

rural communities in the Northern Great Plains. Phytopathology, 90:17–21.

Wingfield B. D., Steenkamp E. T., Santana Q. C., Coetzee M. P. A., Bam S., Barnes I., Beukes C. W., Chan W. Y., de Vos L., Fourie G., Friend M., Gordon T. R., Herron D. A., Holt C., Korf I., Kvas M., Martin S. H., Mlonyeni X. O., Naidoo K., Phasha M. M., Postma A., Reva O., Roos H., Simpson M., Slinski S., Slippers B., Sutherland R., van der Merwe N. A., van der Nest M. A., Venter S. N., Wilken P. M., Yandell M., Zipfel R., and Wingfield M. J. 2012. First fungal genome sequence from Africa: a preliminary analysis. South African Journal of Science, 108(1-2): 1-9.

Xu J., and Leslie J. 1996. A genetic map of *Gibberella fujikuroi* mating population A (*Fusarium moniliforme*). Genetics 143(1):175-189.

Xu J. R., Yan K.Y., Dickman M.B. and Leslie J.F. 1995. Electrophoretic karyotypes distinguish the biological species of *Gibberella-fujikuroi (Fusarium* section *Liseola*). Molecular Plant-Microbe Interaction, 8: 74-84.

Xu Z., and Wang H. 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Research, 35:W265-W268.

Zdobnov, E.M., and Apweiler, R. 2001. InterProScan–an integration platform for the signature-recognition methods in InterPro. Bioinformatics, 17(9): 847-848.

Zeller, K. A., Bowden, R. L., and Leslie, J. F. 2003. Diversity of epidemic populations of *Gibberella zeae* from small quadrats in Kansas and North Dakota. Phytopathology, 93(7), 874-880.

Zeller, K. A., Bowden, R. L., and Leslie, J. F. 2004. Population differentiation and recombination in wheat scab populations of *Gibberella zeae* from the United States. Molecular Ecology, 13(3), 563-571.

Appendix A - Galaxy workflow

Step 1: FASTQ Groomer

File to groom
select at runtime
Input FASTQ quality scores type
Sanger
Advanced Options
Hide Advanced Options

Step 2: FASTQ Quality Trimmer

FASTQ File
Output dataset 'output_file' from step 1
Keep reads with zero length
False
Trim ends
3' only
Window size
20
Step Size
1
Maximum number of bases to exclude from the window during aggregation
1
Aggregate action for window
min score
Trim until aggregate score is
>=
Quality Score
20.0

Step 3: Clip

Library to clip
Output dataset 'output_file' from step 2
Minimum sequence length (after clipping, sequences shorter than this length will be discarded)
15
Source
Enter custom sequence
Enter custom clipping sequence
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT
enter non-zero value to keep the adapter sequence and x bases that follow it
0
Discard sequences with unknown (N) bases
No
Output options
Output both clipped and non-clipped sequences

Step 4: Filter by quality

Library to filter
Output dataset 'output' from step 3
Quality cut-off value
25
Percent of bases in sequence that must have quality equal to / higher than cut-off value
50

Step 5: Filter FASTQ

FASTQ File
Output dataset 'output' from step 4
Minimum Size
20
Maximum Size
0
Minimum Quality
10.0
Maximum Quality
0.0
Maximum number of bases allowed outside of quality range
2
This is paired end data
False
Quality Filter on a Range of Bases

Step 6: Filter FASTQ

FASTQ File
Output dataset 'output_file' from step 5
Minimum Size
20
Maximum Size
0
Minimum Quality
20.0
Maximum Quality
0.0
Maximum number of bases allowed outside of quality range
4
This is paired end data
False
Quality Filter on a Range of Bases

Step 7: Bowtie2

Is this library mate-paired?
Single-end
FASTQ file
Output dataset 'output_file' from step 6
Write all reads that could not be aligned to a file (uses --un for single-end and --un-conc for paired-ends)
True

Will you select a reference genome from your history or use a built-in index?
Use one from the history
Select the reference genome
select at runtime
Parameter Settings
Use Defaults
Specify the read group for this file?
No