Comparative genomics and prediction of genome features in *Fusarium oxysporum* f. sp. *vasinfectum* isolates to aid in the identification of virulence factors specific for cotton

by

Seungyeon Seo

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 7, 2021

Keywords: Fusarium wilt on cotton, comparative genomics, virulence factors

Approved by

Dr. Jeffrey Coleman, Associate Professor of Entomology and Plant Pathology
Dr. Leonardo De La Fuente, Professor of Entomology and Plant Pathology
Dr. Kira L. Bowen, Professor of Entomology and Plant Pathology
Dr. Neha Potnis, Assistant Professor of Entomology and Plant Pathology
Dr. Leslie R. Goertzen, Associate Professor of Biological Sciences

Abstract

Fusarium oxysporum f. sp *vasinfectum* (*Fov*), the causal agent of Fusarium wilt of cotton (*Gossypiym* species), is causing a devastating decline in worldwide cotton production. Although many factors associated with pathogenicity of *F. oxysporum* species have been revealed, individual members in the *F. oxysporum* species complex (FOSC) display host specificity. This indicates that characterized virulence factors may not function similarly in other isolates virulent on different plant hosts. Thus, individual *Fusarium* isolates may need to be evaluated independently on each host plant. The genomes of five *Fov* isolates, including a representative of the highly virulent race 4 genotype, were sequenced at a high coverage to obtain reference quality assemblies. Various bioinformatic approaches were employed to identify genomic features of *Fov*.

In Chapter 2, the core (CC) and accessory (AC) components of the genome of the five *Fov* isolates were identified. Scaffolds indicated as part of the AC were searched for syntenic regions between the five *Fov* genomes. Except for the *Fov* race 4 (89-1A) genome, AC regions of the remaining four *Fov* genomes shared syntenic loci. Homologs in *Verticillium* species that share the same cotton host with *Fov* that were confirmed to be involved in virulence were used to narrow down a key region in the ACs for pathogenicity on cotton. As a result, about 360 kb of the genomes of TF1, LA3B, and 14-004 was inferred to be important for conferring pathogenicity on cotton. In addition, the most virulent strain, *Fov* 89-1A had unique genomic features; a larger genome size and proportion of repetitive elements than the other four *Fov* genotypes included in the study.

In Chapter 3, comparative genomics was applied to identify potential secreted protein sets from six *Fov* genomes (cotton) and ten publicly available *F. oxysporum* genomes to uncover putative secreted proteins which could be contributing to virulence (fungal effectors) on cotton. Among the *Fov* isolates, the highly virulent race 4 is of particular interest since no highly resistant

cotton cultivars are commercially available and it is able to infect plants without the presence of nematodes. To narrow down unique putative effector gene specific for *Fov, in silico* analysis was universally applied to identify proteins containing a signal peptide and exclude those with transmembrane domains. After filtering all the secretomes to only proteins less than 300 amino acids in length, the refined secretome of each individual FOSC was analyzed using BLASTP against of the SSP set of 687 proteins from *Fov* race 4 (89-1A) to cluster SSPs. A total of 78 SSPs were shared between at least three isolates from cotton but were absent in all the other FOSC genomes not pathogenic on cotton. Four mutants of SSP genes, *F02527*, *F02627*, *F07669/07679* and *F11624* were chosen for gene deletion to evaluate their contribution to fungal pathogenicity on cotton. Three pathogenicity assays, stem, hydroponic, and pot, were conducted with the Pima cultivar DP744 to assess whether these SSPs had any influence on the virulence to cotton. SSP *F11624* was confirmed to serve as a virulence factor on cotton and the gene expression was up-regulated during infection.

In Chapter 4, *in silico* analysis was performed to identify secondary metabolite (SM) clusters to elucidate their biological significance within the FOSC. Twenty-one genomes of members of the FOSC including six *Fov* genomes were queried to identify putative secondary metabolite biosynthetic gene clusters. Core and accessory regions of all the FOSC genomes were defined using the well-established *Fol* 4287 genome as a reference for chromosomal structure. Altogether, comparative analysis of the SM cluster types revealed 28 clusters in the core regions are shared in all 21 FOSC genomes. Furthermore, to understand whether specific clusters exist in the *Fov* isolates including race 7 (NRRL 25433), the type of cluster, uniqueness, and location in sub-telomeric regions were manually curated. The putative siderophore cluster in the middle of chromosome 11 (CHR11) harboring the core gene encoded by *NRPS6* was observed in all *Fov* and

Fol 4287 isolates and has high conservation. Additional putative siderophore clusters encoding a homolog of *NRPS1* on CHR12 were found in both *Fov* 14-004 and LA3B isolates. The *Fov* 14-004 genome contains many unique SM clusters not observed in the other *Fov* genomes, and four *Fov* isolates, TF1, 89-1A, 25433, and LA3B had one to three SM clusters encoded in the accessory region.

In conclusion, the *Fov* genomes provide a necessary resource for comparative genomic analyses to identify genes and genomic features that are involved in pathogenicity on cotton and may ultimately be used to develop improved management strategies. Elucidating the function of many putative SSPs could be valuable in identifying novel and isolate-specific effector genes in *Fov* and more importantly could facilitate identification of corresponding resistance gene(s) in cotton for use in breeding programs. Such fundamental studies are necessary for developing novel methods to control the threat caused by *Fusarium* and SMs for agricultural biosecurity.

Acknowledgments

I am pleased to share 5 years of my hard work and both personal and professional growth here in the department of Plant Pathology at Auburn university. I would like to express my gratitude to all the amazing people who are my personal heroes and have made it possible to accomplish more than I dreamed of. There were many challenging times, but I was able to rely not only on myself but also on the wonderful people who were there. I am lucky to receive a full scholarship to get my Ph.D. under the supervision of Dr. Jeffrey Coleman. I was so passionate about comparative analysis especially using several genomic data before coming to Auburn. He set up the environments where I could get in quickly. I have enjoyed all functional annotation using bioinformatics and wet-lab experiments such as CRISPR-Cas9 mediated transformation. I have encountered many failures, but all the experiences have come alive for my future assets. Thanks for your guide and patience. I thank my graduate committee members, Dr. Kira Bowen, Dr. Leonardo De La Fuente, Dr. Leslie Goertzen, and Dr. Neha Potnis, and the outside reader for my dissertation, Dr. Alex Harkess for their contribution of time to this project. In particular, the comments and computational script for the ortholog groups in FOSC from Dr. Neha Potnis helped a lot. Most computational skills were achieved from the auburn bioinformatic bootcamp 2017 and biology courses taught by Dr. Tonia Schwartz and Dr Laurie Stevison. Dr. David Young of Alabama supercomputer and other administrators of Auburn hopper system set up analyzing environments in their servers and helped for trouble shooting many times. Thank to Jane Farr, a director of Research/Extension Center for my cotton seeds. I appreciate Laura Wendell and Dr. Yohannes Mehari who helped set up my first bench work for even minor tasks like PCR and gel running. Dr. Sahar Hasim opened my chance to begin the mutant generation experiments for my second chapter. Without her, I could not start my experiments at all. I am thankful to my lab mates,

Ambika Pokhrel, Frances Loyo, Kellie Bourguignon, Miranda Otero, and Dr. Qiang Wang. Ambika and I started our degrees, master, and Ph.D. together. We have experienced so many challenges, but enjoyed our time, and made good memories. I have shared our tiny office with Frances and Kellie, and I enjoyed our discussion time very much. I got a lot of fresh ideas and learned English from them. Hongcai, an exchanging student from China, taught to purify the Cas9 proteins. Thanks to her, I got the enough proteins after nine times. My friends and professors who are back in Korea and met at the conferences, gave me wonderful advice and encouraged me when I couldn't get the mutants in more than 20 trials. To all my Auburn friends, I appreciate you all for sharing your precious time with me. I learned so many things and made happy memories I will keep forever. Lastly and most importantly, a big thank goes out to all my joyful family, particularly my parents Beomseok Seo and Younghee Choi, my siblings Eunjin and Changwoo Seo for their support and love during my degree. Thank you all.

Table of Contents

Abstract ii

Acknowledgments v

List of Tables xii

List of Figures xiv

List of Abbreviations xvii

Chapter 1 Review Literature 1

1.1 *Fusarium* species 1

1.2 Disease cycle and symptoms 2

1.3 *Fusarium oxysporum* f. sp. *vasinfectum* 2

1.4 Comparative genomic analysis 4

1.4.1 Genomics 5

1.4.2 Fungal genomics 5

1.4.3 Comparative genomics or pathogenomics 6

1.5 Accessory chromosomes 8

1.6 Fungal virulence factors 10

Reference 13

Chapter 2 Comparative genomics and prediction of the accessory genome regions in cotton-infecting *Fusarium oxysporum formae speciales* 20

Abstract 20

2.1 Introduction 21

2.2 Materials and methods 22

2.2.1 Selection of *Fov* isolates for this study 22

2.2.2 Genome sequencing, assembly, and basic genome statistics 23
2.2.3 Functional annotation of genomes and protein-coding genes 24
2.2.3.1 Repetitive elements 24
2.2.3.2 Prediction of secreted proteins 24
2.2.3.3 Prediction of carbohydrate utilization enzymes 24
2.2.3.4 Other functional analysis 25
2.2.4 Investigation of the structure of the *Fov* genome 25
2.2.4.1 Defining core chromosomes 25
2.2.4.2 Mitochondrial genome 26
2.2.4.3 Identification of accessory genomic regions 26
2.2.5 Pan-genome analysis of *F. oxysporum* strains (tree building and ortholog analysis) 27
2.2.6 Identification of homologous genes of *Verticillium dahliae Vd*991 in the five *Fov* isolates 28
2.3 Results 28
2.3.1 High quality genome assemblies of five *F. oxysporum* f. sp. *vasinfectum* (*Fov*) 28
2.3.2 The phylogenetic tree using orthologues shared within FOSC 29
2.3.3 The compartmentalization of five *Fov* genomes 30
2.3.3.1 Mitochondrial genome 30
2.3.3.2 Genomic features sharing common characteristics in the core chromosomes of five *Fov* isolates 30
2.3.3.3 Identification of the accessory genome from the five *Fov* isolates 31
2.3.4 Genome and gene features of *Fov* genomes 33
2.3.5 Homologs between the cotton pathogen *Verticillium* and the five *Fov* genomes 35

2.4 Discussion 36

Reference 41

Chapter 3 Identification and partial characterization of putative small secreted proteins in five *Fusarium oxysporum* f. sp. *vasinfectum* (*Fov*) isolates uncovers several involved in virulence on cotton 65

Abstract 65

3.1 Introduction 67

3.2 Materials and methods 69

3.2.1 Data collection of FOSC proteomes 69

3.2.2 Secretome prediction and comparison 69

3.2.3 Annotation of candidate SSPs in *Fov* race 4 69

3.2.4 Search of *SIX* gene homologs in *Fov* strains 70

3.2.5 Mutant generation 70

3.2.5.1 Cas9 protein purification 70

3.2.5.2 sgRNA synthesis 71

3.2.5.3 *In vitro* cleavage 72

3.2.5.4 Donor DNA (50bp) 72

3.2.5.5 Protoplasting 73

3.2.5.6 Fungal transformation 73

3.2.6 Pathogenicity experiments in cotton 74

3.2.6.1 Stem assay 74

3.2.6.2 Hydroponic assay 75

3.2.6.3 Pot assay 76

3.2.7 Gene expression 77
3.3 Results 78
3.3.1 *in silico* analysis to identify SSPs in the genomes of 16 *F. oxysproum* isolates 78
3.3.2 Identification of *Fov* specific SSP 78
3.3.3 Identification of homologs of *SIX* genes in *Fov* 80
3.3.4 SSP F11624 is a virulence factor on cotton and the gene expression is up-regulated during infection 80
3.4 Discussion 82
Reference 88
Chapter 4 Profiling of secondary metabolites of *Fusarium oxysporum* genomes based on chromosome regions and unique clusters in isolates from cotton 114
Abstract 114
4.1 Introduction 116
4.2 Materials and methods 118
4.2.1 Genome sequence and bioinformatics analysis 118
4.2.2 Mapping SM gene clusters to genomes 118
4.2.3 *Fov* specific bioinformatics analysis 119
4.3. Results 120
4.3.1 Inventory of putative secondary metabolite cluster types in 21 FOSC isolates 120
4.3.2 Common clusters in core regions 121
4.3.3 SM clusters in accessory regions 122
4.3.4 Identification of *Fov* specific clusters on core regions 123
4.3.5 Additional studies of the SM clusters among *Fov* isolates 125

4.3.5 Additional studies of the SM clusters among *Fov* isolates 126

4.4 Discussion 127

Reference 135

List of Tables

Chapter 2 Comparative genomics and prediction of the accessory genome regions in cotton-infecting *Fusarium oxysporum formae speciales*

Table 1 Summary of genome features of five *Fusarium oxysporum* f. sp. *vasinfectum* isolates 45

Table 2 Information of FOSC genomes analyzed in this study. 46

Table 3 Mitochondrial genome assemblies of each *Fov* isolate 47

Table 4 Statistical analysis of the genomes and protein-coding genes among the five *F. oxysporum* f. sp. *vasinfectum* genomes 48

Table 5 Number of homologs in *Fov* genomes to genes of interest in *Verticillium* that is virulent on cotton 49

Table 6 Detailed information of homologs of Vd991 in the five *Fov* genomes 50

Chapter 3 Identification and partial characterization of putative small secreted proteins in five *Fusarium oxysporum* f. sp. *vasinfectum* (*Fov*) isolates uncovers several involved in virulence on cotton

Table 1 FOSC genome information used in this study 92

Table 2 The number of predicted proteins at each step of the pipeline for the 16 FOSC proteomes 93

Table 3 Details of the 79 shared *Fov* SSP candidates in *Fov* race 4 potentially important for pathogenicity 94

Table 4 Presence of *F. oxysporum SIX* homologs in five *F. oxysporum* f. sp. *vasinfectum* isolates 97

Table 5 Primer sequences used in this study for generating knock-out mutants 98

Chapter 4 Profiling of secondary metabolites of *Fusarium oxysporum* genomes based on chromosome regions and unique clusters in isolates from cotton

Table 1 Details of the FOSC genomes used in this study 139

Table 2 The number of secondary metabolite cluster types from antiSMASH in 21 FOSC genomes. 140

Table 3 Profile of secondary metabolite clusters in each core chromosomes 141

Table 4 Information about the known clusters from antiSMASH 142

Table 5 Secondary metabolite cluster in potential accessory scaffolds in FOSC 143

Table 6 Analyses of predicted secondary metabolites in the accessory regions of the *Fov* genomes 141

Table 7 The core genes responsible for siderophore biosynthesis in *Fol* 4287 and *Fov* genomes 146

List of Figures

Chapter 2 Comparative genomics and prediction of the accessory genome regions in cotton-infecting *Fusarium oxysporum formae speciales*

Figure 1 Bioinformatic pipeline for mining the structure and protein-coding genes in five *F. oxysporum* f. sp. *vasinfectum* (*Fov*) genomes. 52

Figure 2 Distribution of orthologues from 21 FOSC isolates 53

Figure 3 Phylogenetic relationships of the five *Fov* genomes and other *F. oxsyporum* species 55

Figure 4 Genome comparison of each *Fov* and *Fol* 4287 56

Figure 5 Synteny relationships in the accessory (lineage-specific) regions among the five *Fov* strains 58

Figure 6 Visualization of genomic features and protein-coding gene annotation of five *Fov* isolates 60

Figure 7 Repeat content in the five *Fov* genomes 61

Figure 8 Relative distribution of CAZymes in the *Fov* genomes 62

Figure 9 Phylogenetic relationships between proteins encoding homologous genes from *Fov* genomes and lineage-specific (LS) genes from *V. dahlia* Vd991 63

Chapter 3 Identification and partial characterization of putative small secreted proteins in five *Fusarium oxysporum* f. sp. *vasinfectum* (*Fov*) isolates uncovers several involved in virulence on cotton

Figure 1 Bioinformatic pipeline for mining key small secreted proteins (SSPs) of *F. oxysporum* f. sp. *vasinfectum* (*Fov*) 100

Figure 2 Profiles of small secreted proteins (SSPs) in 16 FOSC genomes 101

Figure 3 Distribution of orthologues of predicted SSPs relative to *Fov* race 4 within 16 FOSC genomes 102
Figure 4 Graphical presentation of annotated 78 small secreted protein candidates 103
Figure 5 Amino acid sequence alignments of *SIX* genes with homologs in *Fov* 104
Figure 6 The schematic representation of Cas9 protein purification method using a HisPur Ni-NTA column 105
Figure 7 Schematic diagrams of sgRNA and donor DNA construction and *F. oxysporum* transformation by PEG-mediated delivery of Cas9 RNPs 106
Figure 8 Gene expression of selected four Fov race 4 SSPs in infected root 109
Figure 9 PCR amplification of the target gene for the template (WT) for *in vitro* cleavage and for the validation of the transformants 110
Figure 10 Assessment of SSP mutants in a stem lesion virulence assay on Pima cultivar DP744 111
Figure 11 Virulence assay of SSPs mutants on the cotton cultivar, DP744 using hydroponic system 112
Figure 12 Assessment of virulence of putative SSP mutants using a pot assay in Pima cultivar, DP744 113

Chapter 4 Profiling of secondary metabolites of *Fusarium oxysporum* genomes based on chromosome regions and unique clusters in isolates from cotton

Figure 1 Physical distribution of common secondary metabolite clusters on core regions of *Fol* 4287 and five *Fov* genomes 147
Figure 2 Distribution of gene clusters common, unique, and located in sub-telomeric regions on the *Fov* core chromosomes 148

Figure 3 Sequence comparison of a putative siderophore cluster of *Fol* 4287 and six *Fov* genomes including Fov25433 149

List of Abbreviations

Fo	*Fusarium oxysporum*
FOSC	*Fusarium oxysporum* Species Complex
f. sp.	*formae speciales*
Fol	*Fusarium oxysporum* f. sp. *lycopersici*
Fov	*Fusarium oxysporum* f. sp. *vasinfectum*
spp. or sp.	Species
SSPs	Small secreted proteins
SMs	Secondary metabolites
CCs	Core chromosomes
ACCs	Accessory chromosomes
LS	Lineage-specific
CAZymes	Carbohydrate-active enzymes
mtDNA	Mitochondrial DNA
CHR	Chromosome
SIX	Secreted-in xylem
RNPs	Ribonucleoproteins
RTop	Regeneration top
H_2O_2	Hydrogen peroxide
cDNA	Complementary DNA
qRT-PCR	Quantitative reverse transcription PCR
VCGs	Vegetative compatibility groups
mimp	Miniature Impala

CWDEs	Cell wall degrading enzymes
NRPSs	Non-ribosomal peptide synthetases
PKSs	Polyketide synthases
smCOGs	Secondary metabolism Clusters of Orthologous Groups
adh	Alcohol dehydrogenase

Chapter 1. Literature Review

1.1 *Fusarium* species

The genus *Fusarium* is taxonomically classified within the Nectriaceae family, Hypocreales Order, Sordariomycetes Class and Ascomycota phylum of fungi (Snyder & Hansen, 1940). Since the initial description of the *Fusarium* genus by Link (Link, 1809), studies have indicated over 1,000 species, varieties, and *formae speciales* belong to the genus, based on the canoe or banana shaped conidia (Edel-Hermann & Lecomte, 2019). Wollenweber and Reinking in 1935, reduced the number to 65 species and 22 forms and varieties with the modern taxonomic concept, which was further narrowed to only nine species within the genus (Snyder & Hansen, 1940). However, emerging molecular phylogenetic studies have indicated that a monophyletic *Fusarium* genus consists of at least 23 clades referred to as species complexes (Geiser et al., 2013).

Fusarium species are one of the largest groups of filamentous fungi and one of the most common and widespread in the world containing non-pathogenic, human/animal, and plant pathogens (Alabouvette et al., 2009; Anaissie & Nucci, 2002; T. R. Gordon & Martyn, 1997). Among the plant pathogens within *Fusarium*, the *F. sambucinum* (containing *F. graminearum*) and the *F. oxysporum* species complexes (FOSC) are considered some of the most agronomically important plant pathogens (Edel-Hermann & Lecomte, 2019). *F. graminearum* is a major problem in cereal crops and a few non-cereal species causing Fusarium head blight (FHB) (Goswami & Kistler, 2004). On the other hand, *F. oxysporum* has a wide host range, impacting over one hundred agriculturally important plant species including banana, cotton, soybean, melon, onion, tomato, as well as ornamental crops (cyclamen and orchids). Therefore, the term *formae speciales* (f. sp.) is used to denote host specificity, where pathogenic isolates on the same host are grouped together within the same *formae speciales*. For example, *F. oxysporum* f. sp. *lycopersici* (*Fol*) is pathogenic

to tomato (*Solanum lycoprsicum*), and *F. oxysporum* f. sp. *vasinfectum* (*Fov*), the main focus of this dissertation, is a group of isolates pathogenic on cotton (*Gossypium* sp.). Altogether, *F. oxysporum* can be divided into over 140 *formae speciales* depending on the host (Edel-Hermann & Lecomte, 2019).

1.2 Disease cycle and symptoms

In general, wilt diseases by members of the FOSC initiate at the roots with soil borne inoculum. Although mycelial fragments may be viable in soil, chlamydospores are more persistent and are considered the primary inoculum (T. R. Gordon & Martyn, 1997). The spores germinate and the germ-tube elongates towards exudates released from plant roots such as sugars and amino acids (Schroth & Hildebrand, 1964). The germ-tube grows until it contacts the root tip and the zone of root elongation. Fungal hyphal branching follows and a compact mycelial mat is formed on the root surface. The hyphal tip eventually penetrates through a small opening generated by degradative enzymes and enters the vascular system (Michielse & Rep, 2009). *F. oxysporum* is considered a hemibiotroph because when it enters the plant roots at the beginning stage of infection, the surrounding plant cells are alive. However, once the hyphae reach the plant vascular system, the fungus colonizes the vascular tissue with hyphae and produce spores (Nelson, 1981). Consequently, it blocks the transport of water and nutrients. This triggers chlorosis of plant leaves in the early stage of the disease cycle, followed by massive wilting, necrosis, and ultimately death of plants (L. J. Ma et al., 2013). As a result of infection, a necrotrophic phase follows.

1.3 *Fusarium oxysporum* f. sp. *vasinfectum*

The FOSC includes soilborne fungi, and usually produce spores such as microconidia, macroconidia, and chlamydospores which are generated from hyphae (Thomas R. Gordon, 2017). No sexual stage has been reported for these fungi (Leslie & Summerell, 2006). *F. oxysporum*

Schlechtend. f. sp. *vasinfectum* (Atk.) Snyd. & Hans (*Fov*) is the causal agent of Fusarium wilt of cotton (*Gossypium* species) which is found throughout the southeastern United States (Hand et al., 2021; Kochman, 1995). In the past decade, Fusarium wilt has been prevalent in all cotton-producing regions around the world, even emerging in Australia in 1993 (Kochman, 1995). *Fov* is capable of infecting all four domesticated cotton species *Gossypium hirsutum*, *G. barbadense*, *G. arboretum*, and *G. herbaceum* causing serious economic losses worldwide.

Fov has been divided into six races, 1, 2, 3, 4, 6, and 8 (Cianchetta et al., 2015; Edel-Hermann & Lecomte, 2019). In the past, *Fov* isolates were originally separated into eight races based roughly on their geographical distribution. The races 1 and 2 were reported in the United States, and the other races were initially identified in other major cotton growing regions throughout the world including race 3 in Egypt, race 4 in India, race 5 in Sudan, race 6 in Brazil, and races 7 and 8 in China (G. M. Armstrong & Armstrong, 1960, 1978; J. K. Armstrong & Armstrong, 1958; Chen et al., 1985; Ibrahim, 1966). A host differential test was eventually established to differentiate between the races of *Fov* by assessing individual race virulence to a range of cotton lines (*Gossypium*) and up to five non-cotton hosts including soybean, tobacco, and yellow lupin (Davis et al., 1996). The separation of race 1 and 2 was primarily by the ability to cause wilt on soybean and tobacco (G. M. Armstrong & Armstrong, 1960, 1978). Whereas the race 7 and 8 isolates from China were grouped separately through examining the virulence on cotton, tobacco, alfalfa, okra, and soybean (Chen et al., 1985). The Australian *Fov* isolates were first identified from diseased plants in 1993 and had pathogenicity similar to *Fov* race 6 from Brazil (Kochman, 1995). However, through the use of aesculin hydrolysis tests the Australian isolates were able to clear the medium after an initial color change allowing differentiation between the Australian isolates and those in races 1, 2, and 6 (Davis et al., 1996). While genetic markers have

been used for the identification of *Fov* races, several genomic loci have been developed for *Fov* race classification including translation elongation factor (TEF-1α), β-tubulin (BT), nitrate reductase (NIR), mitochondrial small subunit (mtSSU) rDNA, and phosphate permease (PHO) (Cianchetta et al., 2015; Holmes et al., 2009; Moricca et al., 1998; Ortiz et al., 2017; Skovgaard et al., 2001). These genetic markers uncovered and resolved the redundancy that existed between the geographically defined races in particular between races 3 and 5 which were identical and race 4 from India and race 7 from China were indistinguishable (Davis et al., 2006; Nirenberg et al., 1994; Skovgaard et al., 2001). Since specific DNA sequences have been used for the identification of *Fov* isolates, the race terminology has begun to be replaced with the more accurate term "genotype" and is mostly only used for historical reference (R. M. Davis et al., 2006).

Five of the "classical" defined races 1, 2, 3, 4, and 8 have been recognized in the United States (Holmes et al., 2009; Kim et al., 2005). Race 4 is one of the most concerning genotype in the U.S. since its initial report in California in 2001 (Kim et al., 2005). The race 4 genotype is highly virulent on most of the commercial Pima cultivars (*G. barbadense* L.) and unlike other *Fov* genotypes, race 4 is capable of causing disease in the absence of co-infection with nematodes (Kim et al., 2005). While *Fov* race 4 is comparatively less virulent on Upland cotton cultivars, there are no commercial Upland cultivars (*G. hirsutum*) that are highly resistant to race 4 (Hutmacher et al., 2013; Kim et al., 2005). More recently, several *Fov* genotypes have been found in the Southeastern U.S., including LA108, LA110, LA112, and LA127/140 which were identified based on their unique partial EF-1α sequences (Holmes et al., 2009). Additionally, *Fov* isolates have been obtained from diseased cotton that have identical *EF-1α* sequences as the highly virulent race 4 genotype; however other genetic loci separate them into the closely related genotype MDS-12.

1.4 Comparative genomic analysis

1.4.1 Genomics

In 2001, the first draft sequence of the human genome was completed by whole-genome shotgun sequencing, where DNA fragments were cloned and sequenced and later reassembled into a "continuous" genome (Abdellah et al., 2004). This project provided the stimulus to dramatically improve the number and speed of sequencing of many model organisms, and the development of sequencing technologies and computational programs to further explore the genomes (Bansal & Boucher, 2019; Pareek et al., 2011). Next-generation or second-generation sequencing (NGS) methods have been commercially available from Roche/454 (Pareek et al., 2011) and Illumina/Solexa (Bentley et al., 2008), and while they produce much shorter reads, these technologies are capable of generating a massive volume of data at a fraction of the cost (Metzker, 2010). However, genomes generated using short-read length technologies have been poorly suited for understanding the biological relevance in genomic regions rich in repeats and/or duplications (Metzker, 2010). The third-generation sequencing (TGS) technologies by Pacific Biosciences (PacBio; Rhoads & Au, 2015) and Oxford Nanopore Technologies (ONT; Branton et al., 2008), produces long-read length sequences to overcome these limitations; however, these TGS technologies have shortcomings such as high error rates and increased cost per base (Schadt et al., 2010). Many scientists have taken advantage of the second and third sequencing technologies to improve the genome quality; in particular the combination of Illumina and PacBio sequence data is considered one of the best options to get an accurate and high-quality genome assembly (Utturkar et al., 2014).

1.4.2 Fungal genomics

High throughput technologies have stimulated large scale genome projects to increase the diversity of species with complete genomes available, including insects and plants (Leebens-Mack

et al., 2019; Robinson et al., 2011). Since the first fungal genome, *Saccharomyces cerevisiae* in 1996 (Goffeau et al., 1996), multiple fungal genome sequencing projects have launched including the Fungal Genome Initiative in 2002 and 2003, the Fungal Genomics Program, and especially, the 1000 Fungal Genomes (1KFG) Project in 2011 and have been contributing to fungal genome diversity (Aylward et al., 2017). The 1KFG is one of the representative projects for fungal genome sequencing, which aims to sequence and annotate two species from each of the known 500 fungal families (Martin et al., 2011). Still, sequencing of fungal genomes is on-going by various scientific groups having different interests and needs. Most of the genomes have been accumulated in online genomes repositories such as MycoCosm (Grigoriev et al., 2012, 2014), NCBI Genome (www.ncbi.nlm.nih.gov/genome), the universal cataloguing database, Genomes OnLine Database (GOLD; Reddy et al., 2015), etc.

Among publicly available fungal genomes, a high proportion belongs to pathogenic and medically important fungi; especially, plant pathogenic species affecting food security are predominant, exceeding the number of human pathogens (Aylward et al., 2017). For example, *Magnaporthe oryzae*, causing a serious disease on cultivated rice, ranks second only after the model yeast *S. cerevisiae* in the number of completed whole-genome sequencing list of fungal species in the GOLD. The species *Fusarium oxysporum* causing wilt disease of most commercially important crops is also well represented on the list (Aylward et al., 2017). As of January 2021, the genomes of 37 species in the genus of *Fusarium* have been completed by whole genome sequencing, and 22 strains in the *F. oxysporum* species complex (FOSC) were collected in GOLD database. However, in the NCBI database, there are 69 *Fusarium* genome reports.

1.4.3 Comparative genomics or pathogenomics

Beyond the analysis of a single genome, the plethora of genomes has inspired comparative genomic analyses between fungal isolates and species. Among the publicly available 1090 fungal genomes, the size of the most of fungal genomes ranges from 30 to 40 Mb. Frequently, plant pathogens possess an above average genome size when compared to the average of fungi (Aylward et al., 2017). Especially, plant pathogenic basidiomycetes usually have a larger average genome size (57 Mb) than plant pathogenic ascomycetes (39 Mb) and the rest of fungal genomes (35 Mb). However, the number of predicted open reading frames (ORFs) encoded within these genomes only results in slight differences between plant pathogens and other fungi. The fact that the number of protein coding genes does not correlate with the genome size is most likely a result of repetitive sequences and/or transposable elements (Kidwell, 2002). In fungal plant pathogens, the larger genome due to these repetitive sequences indicates that fungi have more high frequency recombination loci, and the fungi are under evolutionary pressure. This pressure drives fungi to adapt to new host plants, facilitating the rapid development of novel genes or chromosome rearrangement (Wostemeyer & Kreibich, 2002). Comparative genomic analysis is a useful application to further understand the lifestyle, evolutionary relationships, as well as fungal pathogenicity between or within a fungal species.

Being one of the most destructive plant pathogens for many economically important crops, members of the FOSC have been well studied on a molecular level. Over 200 different *Fusarium* genomes are publicly available in the fungal genome database (JGI; https://genome.jgi.doe.gov/portal/) and their genome sizes range from 44 to 65 Mb (Grigoriev et al., 2012). Searching for species-specific gene sets and finding orthologs in other isolates is a common way to compare the genomes of two or more different fungi. *Fusarium* genomes can be used for comparative genomic analysis for in-depth understanding of 1) stimulus-based shifts in

lifestyles, 2) infection strategies, 3) pathogenicity-related factors and their origin and acquisition, 4) genome organization and evolution, 5) genetic diversity within and between species complexes and in specific regions such as the accessory genome, 6) risk prediction for the emergence of new strains and future disease outbreaks, and 7) the potential of developing fungicide resistance (Cuomo et al., 2007; L. J. Ma et al., 2010; Sperschneider et al., 2015; Taylor et al., 2016; van Dam et al., 2017, 2018).

1.5 Accessory chromosomes

Accessory chromosomes (AC), also called supernumerary, B, lineage-specific, dispensable, or mini-chromosome, are widely reported among plant, animal, and fungal species, conversely to the core chromosomes (CCs) which are essential for primary fungal growth and encode genes related to cellular metabolism and reproduction (Croll & McDonald, 2012; Dong et al., 2015; Raffaele & Kamoun, 2012). Loss of one of the accessory chromosomes rarely affects the normal growth of the fungus but could result in loss of some functions such as pathogenicity on a specific host depending on the genes encoded on the chromosome (Coleman et al., 2009; L. J. Ma et al., 2010). The ACs harbor genetic features to allow them to colonize specific niches such as in planta (Dong et al., 2015). The ACs usually have a different structural composition compared to CCs such as 1) small size (usually less than 3 Mb), 2) %GC content, 3) alternative codon usage bias, 4) a lack of synteny, and 5) a large number of transposable elements (TEs) (Croll & McDonald, 2012; Wostemeyer & Kreibich, 2002). In an evolutionary aspect, the TE-enriched regions are hotspots under strong evolutionary pressure to adapt to a new condition. Given this feature, the two-speed genome concept was developed and describes the different evolutionary rates of between CCs and AC.

Fungal ACs are able to be visualized using pulse-field gel electrophoresis for the karyotype profile (Zolan, 1995). The development of whole-genome sequencing technologies has facilitated the discovery of ACs in genomes by eliminating the CCs based on sequence similarity to reference genome(s). An ideal representative reference genome to classify core and accessory chromosomes in the FOSC is *F. oxysporum* f. sp. *lycopersici* 4287 (*Fol*) which has an optical map and a high-quality genome that has been well annotated (L. J. Ma et al., 2010). The genome of this *Fol* isolate was composed of 15 chromosomes; where eleven chromosomes are syntenic to the chromosomes of *F. verticillioides*, and *F. graminearum*, except for the telomere-proximal regions; and four chromosomes, 3, 6, 14, and 15 were unique sequences specific for the *Fol* 4287 genome (L. J. Ma et al., 2010).

The relationship between accessory chromosomes and pathogenicity has been well established in the genus *Fusarium*. Early studies focused on individual genes that resided on ACs. For example, the gene encoding the cytochrome P450 pisatin demethylase (PDA) which detoxifies the phytoalexin pisatin produced by pea (*Pisum sativum*) is located on a small 1.6 Mb chromosome in *F. vanettenii* (formerly *Nectria haematococca* MPVI or *F. solani* f. sp. *pisi*) and in *F. oxysporum* f. sp. *pisi* (Coleman et al., 2011). When an isolate that is nonpathogenic on pea and does not have a functional copy of *PDA*, *F. oxysporum* f. sp. *lini*, was transformed with *FoPDA1* the isolate had increased virulence on pea. Studies within the past decade have demonstrated entire ACs are associated with host-specific pathogenicity in multiple *forma speciales* of *F. oxysporum* (L. J. Ma et al., 2010; van Dam et al., 2017). In *Fol* 4287, a chromosome (CHR14) is required for pathogenicity on tomato, encoding *in planta*-secreted proteins and virulence factors. When this chromosome was horizontally transferred to the non-pathogenic *F. oxysporum* biocontrol isolate Fo47, the strain acquired the ability to infect tomato (L. J. Ma et al., 2010). Another example is *F.*

oxysporum f. sp. *radicis-cucumerinum* (*Forc*), the causal agent of root and stem rot in cucurbits, where a chromosome when transferred to the nonpathogenic Fo47 isolate increased virulence on cucurbits (van Dam et al., 2017).

1.6 Fungal virulence factors

Pathogen-plant interactions have been studied from the aspect of both the host immune response and pathogen virulence and disease development, and both perspectives are complicated due to the various factors involved. From the pathogen perspective, studies on the metabolism or infection strategies utilized by the microbe are frequently a main topic. To cause disease on plants, phytopathogens use sophisticated strategies throughout the process of penetration, infection, and colonization to reduce plant defense mechanisms (van der Does & Rep, 2007). Virulence factors (virulence determinant or pathogenicity factors) are necessary for disease development, and the disruption of those factors is correlated with a reduction of virulence or complete loss of disease symptoms (van de Wouw & Howlett, 2011). A plethora of virulence factors have been identified in phytopathogenic fungi, but common ones include secreted proteins (effectors), genes that are involved in infection, and secondary metabolites.

Effectors are small secreted proteins (SSPs) from the pathogen that enter host plant cells (Kamoun, 2009). They have various roles in disease development but usually enable the pathogen to overcome the host defense system, e.g., pattern-triggered immunity, and to promote effector-triggered susceptibility to spread disease symptoms for the benefit of the pathogen. In turn, plants have evolved resistance (*R*) genes to perceive the corresponding effectors, and upon effector recognition, induce the defense system, effector-triggered immunity. Pathogens continuously develop effector molecules that interfere with recognition/defense during their interaction with the host plant (Jones & Dangl, 2006). This has been explained by the zig-zag model of Jones and

Dangl, comprising the gene-for-gene model (Flor, 1971). Due to the various lifestyles of fungi, including bio-, hemibio- necro-trophic, endophytic, and mycorrhizal, the definition of effectors is expanding as secreted microbial products that facilitate the establishment of various plant-microbe associations ranging from beneficial to detrimental (Kloppholz et al., 2011). SSPs are known to be responsible for disease development as virulence factors and are small sized proteins (usually less than 200 amino acids) and may have a large even number of cysteine residues (Rep, 2005). Using these features, secreted proteins have been identified from fungal genomes *in silico* with other features, such as the presence of a signal peptide, lack of a transmembrane domain, and in some instances the presence of a conserved motif such as the RXLR motif of oomycete effectors (Feldman et al., 2020; Lu & Edwards, 2016). However, the lack of conserved domains or motifs has hampered the identification of the secreted proteins using bioinformatic comparative analysis (lo Presti et al., 2015). Therefore, all resources from both genomic data and in planta expression data should be used for the comprehensive understanding of fungal effectors.

Since plant pathogens secrete various small proteins or chemicals into the host cells during infection, the secretion to the plant xylem has been thoroughly studied. The xylem sap of *Fol* 4278 infected tomato plants was collected and proteins were identified, and the corresponding genes referred as the *Secreted In Xylem (SIX)* genes (Rep, 2005; Rep et al., 2004; Schmidt et al., 2013). Among them, five genes, *SIX1, 3, 4, 5,* and *6*, have been demonstrated to encode effector proteins in planta (Gawehns et al., 2014; Houterman et al., 2009; L. Ma et al., 2015; Rep et al., 2004; Takken & Rep, 2010; Thatcher et al., 2012). All fourteen *SIX* genes are located on the AC 14 of *Fol* (Vlaardingerbroek et al., 2016). The chromosome of *Forc* encodes four *SIX* gene homologs and *SIX6* contributes to virulence in *Forc* (van Dam & Rep, 2017). Further analysis of the upstream region of *SIX* genes have shown an association with Miniature Inverted-repeat Transposable

Elements (MITEs), which are non-autonomous transposable elements (Schmidt et al., 2013). These characteristics can provide clues to the identification of the same lineage fungal species, e.g., *F. oxysporum* f. sp. *cubense* (van Dam et al., 2017).

Reference

Abdellah, Z., Ahmadi, A., Ahmed, S., Aimable, M., Ainscough, R., Almeida, J., Almond, C., Ambler, A., Ambrose, K., Ambrose, K., Andrew, R., Andrews, D., Andrews, N., Andrews, D., Apweiler, E., Arbery, H., Archer, B., Ash, G., Ashcroft, K., … Kamholz, S. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, *431*(7011), 931–945. https://doi.org/10.1038/nature03001

Alabouvette, C., Olivain, C., Migheli, Q., & Steinberg, C. (2009). Microbiological control of soil-borne phytopathogenic fungi with special emphasis on wilt-inducing *Fusarium oxysporum*. *New Phytologist*, *184*(3), 529–544. https://doi.org/10.1111/j.1469-8137.2009.03014.x

Anaissie, E., & Nucci, M. (2002). Cutaneous infection by Fusarium species in healthy and immunocompromised hosts: Implications for diagnosis and management. *Clinical Infectious Diseases*, *35*(8), 909–920. https://doi.org/10.1086/342328

Armstrong, G. M., & Armstrong, J. K. (1960). American, Egyptian, and Indian cotton-wilt fusaria: their pathogenicity and relationship to other wilt fusaria. In *United States Department of Agriculture*.

Armstrong, G. M., & Armstrong, J. K. (1978). A new race (race 6) of the cotton-wilt *Fusarium* from Brazil. In *Plant Disease Reporter* (pp. 421–423).

Armstrong, J. K., & Armstrong, G. M. (1958). A race of the cotton-wilt *Fusarium* causing wilt on Yelredo soybean and flue-cured tobacco. In *The Plant Disease Reporter* (Vol. 42, pp. 147–151).

Aylward, J., Steenkamp, E. T., Dreyer, L. L., Roets, F., Wingfield, B. D., & Wingfield, M. J. (2017). A plant pathology perspective of fungal genome sequencing. *IMA Fungus*, *8*(1), 1–15. https://doi.org/10.5598/imafungus.2017.08.01.01

Bansal, V., & Boucher, C. (2019). Sequencing Technologies and Analyses: Where Have We Been and Where Are We Going? *IScience*, *18*, 37–41. https://doi.org/10.1016/j.isci.2019.06.035

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Keira Cheetham, R., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray, S. J., … Smith, A. J. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, *456*(7218), 53–59. https://doi.org/10.1038/nature07517

Branton, D., Deamer, D. W., Marziali, A., Bayley, H., Benner, S. A., Butler, T., di Ventra, M., Garaj, S., Hibbs, A., Huang, X., Jovanovich, S. B., Krstic, P. S., Lindsay, S., Ling, X. S., Mastrangelo, C. H., Meller, A., Oliver, J. S., Pershin, Y. v., Ramsey, J. M., … Schloss, J. A. (2008). The potential and challenges of nanopore sequencing. *Nature Biotechnology*, *26*(10), 1146–1153. https://doi.org/10.1038/nbt.1495

Chen, Q., Ji, X., & Sun, W. (1985). Identification of races of cotton wilt *Fusarium* in China. In *Agric. Sci. China* (Vol. 6).

Cianchetta, A. N., Allen, T. W., Hutmacher, R. B., Kemerait, R. C., Kirkpatrick, T. L., Lawrence, G. W., Lawrence, K. S., Mueller, J. D., Nichols, R. L., Olsen, M. W., Overstreet, C., Woodward, J. E., & Davis, R. M. (2015). Plant pathology and nematology: Survey of *Fusarium oxysporum* f. sp. *vasinfectum* in the United States. *The Journal of Cotton Science*, *19*, 328–336.

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., … VanEtten, H. D. (2009). The Genome of *Nectria haematococca*: Contribution of Supernumerary Chromosomes to Gene Expansion. *PLoS Genetics*, *5*(8), e1000618. https://doi.org/10.1371/journal.pgen.1000618

Coleman, J. J., Wasmann, C. C., Usami, T., White, G. J., Temporini, E. D., McCluskey, K., & van Etten, H. D. (2011). Characterization of the gene encoding pisatin demethylase (*FoPDA1*) in *Fusarium oxysporum*. *Molecular Plant-Microbe Interactions*, *24*(12), 1482–1491. https://doi.org/10.1094/MPMI-05-11-0119

Croll, D., & McDonald, B. A. (2012). The accessory genome as a cradle for adaptive evolution in pathogens. *PLoS Pathogens*, *8*(4). https://doi.org/10.1371/journal.ppat.1002608

Cuomo, C. A., Güldener, U., Xu, J. R., Trail, F., Turgeon, B. G., di Pietro, A., Walton, J. D., Ma, L. J., Baker, S. E., Rep, M., Adam, G., Antoniw, J., Baldwin, T., Calvo, S., Chang, Y. L., DeCaprio, D., Gale, L. R., Gnerre, S., Goswami, R. S., … Kistler, H. C. (2007). The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science*, *317*(5843), 1400–1402. https://doi.org/10.1126/science.1143708

Davis, R. D., Moore, N. Y., & Kochman, J. K. (1996). Characterisation of a population of *Fusarium oxysporum* f. sp. *vasinfectum* causing wilt of cotton in Australia. In *Aust. J. Agric. Res* (Vol. 47, pp. 1143–1156).

Davis, R. M., Colyer, P. D., Rothrock, C. S., & Kochman, J. K. (2006). Fusarium wilt of cotton: Population diversity and implications for management. *Plant Disease*, *90*(6), 692–703. https://doi.org/10.1094/PD-90-0692

Dong, S., Raffaele, S., & Kamoun, S. (2015). The two-speed genomes of filamentous pathogens: Waltz with plants. In *Current Opinion in Genetics and Development* (Vol. 35, pp. 57–65). Elsevier Ltd. https://doi.org/10.1016/j.gde.2015.09.001

Edel-Hermann, V., & Lecomte, C. (2019). Current status of *Fusarium oxysporum formae speciales* and races. *Phytopathology*, *109*(4), 512–530. https://doi.org/10.1094/PHYTO-08-18-0320-RVW

Feldman, D., Yarden, O., & Hadar, Y. (2020). Seeking the Roles for Fungal Small-Secreted Proteins in Affecting Saprophytic Lifestyles. *Frontiers in Microbiology*, *11*, 455. https://doi.org/10.3389/fmicb.2020.00455

Flor, H. H. (1971). Current status of the gene-for-gene concept. *Annual Review of Phytopathology*, *9*, 275–296. www.annualreviews.org

Gawehns, F., Houterman, P. M., Ichou, F. A., Michielse, C. B., Hijdra, M., Cornelissen, B. J. C., Rep, M., & Takken, F. L. W. (2014). The *Fusarium oxysporum* effector *Six6* contributes to virulence and suppresses *I-2*-mediated cell death. *Molecular Plant-Microbe Interactions*, *27*(4), 336–348. https://doi.org/10.1094/MPMI-11-13-0330-R

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., … 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. https://doi.org/10.1094/PHYTO-07-12-0150-LE

Goffeau, A., Barrell, G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., & Oliver, S. G. (1996). Life with 6000 genes. *Science*, *274*(5287), 546–567. https://doi.org/10.1126/science.274.5287.546

Gordon, T. R., & Martyn, R. D. (1997). The evolutionary biology of *Fusarium oxysporum*. *Annual Review of Phytopathology*, *35*(1), 111–128. https://doi.org/10.1146/annurev.phyto.35.1.111

Gordon, Thomas R. (2017). *Fusarium oxysporum* and the *Fusarium* Wilt Syndrome. *Annual Review of Phytopathology*, *55*(1), 23–39. https://doi.org/10.1146/annurev-phyto-080615-095919

Goswami, R. S., & Kistler, H. C. (2004). Heading for disaster: *Fusarium graminearum* on cereal crops. *Molecular Plant Pathology*, *5*(6), 515–525. https://doi.org/10.1111/j.1364-3703.2004.00252.x

Grigoriev, I. v., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., Smirnova, T., Nordberg, H., Dubchak, I., & Shabalov, I. (2014). MycoCosm portal: Gearing up for 1000 fungal genomes. *Nucleic Acids Research*, *42*(D1), D699–D704. https://doi.org/10.1093/nar/gkt1183

Grigoriev, I. v., Nordberg, H., Shabalov, I., Aerts, A., Cantor, M., Goodstein, D., Kuo, A., Minovitsky, S., Nikitin, R., Ohm, R. A., Otillar, R., Poliakov, A., Ratnere, I., Riley, R., Smirnova, T., Rokhsar, D., & Dubchak, I. (2012). The Genome Portal of the Department of Energy Joint Genome Institute. *Nucleic Acids Research*, *40*(D1), D26–D32. https://doi.org/10.1093/nar/gkr947

Hand, C., Culpepper, S., Harris, G., Kemerait, B., Liu, Y., Perry, C., Porter, W., Roberts, P., Smith, A., Virk, S., & Bag, S. (2021). *Georgia Cotton Production Guide* (P. Roberts, Ed.). The University of Georgia Cooperative Extention. http://www.ugacotton.com/

Holmes, E. A., Spurgeon, D. W., Bennett, R. S., Colyer, P. D., & Davis, R. M. (2009). New genotypes of *Fusarium oxysporum* f. sp. *vasinfectum* from the Southeastern United States. *Plant Disease*, *93*(12), 1298–1304. https://doi.org/10.1094/PDIS-93-12-1298

Houterman, P. M., Ma, L., van Ooijen, G., de Vroomen, M. J., Cornelissen, B. J. C., Takken, F. L. W., & Rep, M. (2009). The effector protein Avr2 of the xylem-colonizing fungus *Fusarium oxysporum* activates the tomato resistance protein I-2 intracellularly. *The Plant Journal*, *58*(6), 970–978. https://doi.org/10.1111/j.1365-313X.2009.03838.x

Hutmacher, R. B., Ulloa, M., Wright, S. D., Campbell, B. T., Percy, R., Wallace, T., Myers, G., Bourland, F., Weaver, D., Chee, P., Thaxton, P., Zhang, J., Smith, W., Dever, J., Kuraparthy, V., Bowman, D., Jones, D., & Burke, J. (2013). Elite Upland Cotton Germplasm-Pool Assessment of Fusarium Wilt Resistance in California. *Agronomy Journal*, *105*(6), 1635–1644. https://doi.org/10.2134/agronj2013.0264

Ibrahim, F. M. (1966). A new race of Cotton-wilt *Fusarium* in the Sudan Gezira. *Empire Cotton Growing Review*, *43*(4), 296–299.

Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. *Nature*, *444*(7117), 323–329. https://doi.org/10.1038/nature05286

Kamoun, S. (2009). The Secretome of Plant-Associated Fungi and Oomycetes. In *The Mycota* (pp. 173–180). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-87407-2_9

Kidwell, M. G. (2002). Transposable elements and the evolution of genome size in eukaryotes. *Genetica*, *115*(1), 49–63. https://doi.org/10.1023/A:1016072014259

Kim, Y., Hutmacher, R. B., & Davis, R. M. (2005). Characterization of California isolates of *Fusarium oxysporum* f. sp. *vasinfectum*. *Plant Disease*, *89*(4), 366–372. https://doi.org/10.1094/PD-89-0366

Kloppholz, S., Kuhn, H., & Requena, N. (2011). A secreted fungal effector of glomus intraradices promotes symbiotic biotrophy. *Current Biology*, *21*(14), 1204–1209. https://doi.org/10.1016/j.cub.2011.06.044

Kochman, J. K. (1995). Fusarimn wilt in cotton - a new record in Australia. *Australasian Plant Pathology*, *24*(1), 74. https://doi.org/10.1071/APP9950074

Leebens-Mack, J. H., Barker, M. S., Carpenter, E. J., Deyholos, M. K., Gitzendanner, M. A., Graham, S. W., Grosse, I., Li, Z., Melkonian, M., Mirarab, S., Porsch, M., Quint, M., Rensing, S. A., Soltis, D. E., Soltis, P. S., Stevenson, D. W., Ullrich, K. K., Wickett, N. J., DeGironimo, L., … Wong, G. K. S. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature*, *574*(7780), 679–685. https://doi.org/10.1038/s41586-019-1693-2

Leslie, J. F., & Summerell, B. A. (2006). The Fusarium Laboratory Manual. In J. F. Leslie & B. A. Summerell (Eds.), *The Fusarium Laboratory Manual*. Blackwell Publishing. https://doi.org/10.1002/9780470278376

Link, H. F. (1809). *Observationes in ordines plantarum naturals, Dissetatio I. Magazin der Gesellschaft Naturforschenden Freunde Berlin*.

lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., Zuccaro, A., Reissmann, S., & Kahmann, R. (2015). Fungal effectors and plant susceptibility. *Annual Review of Plant Biology*, *66*, 513–545. https://doi.org/10.1146/annurev-arplant-043014-114623

Lu, S., & Edwards, M. C. (2016). Genome-Wide Analysis of Small Secreted Cysteine-Rich Proteins Identifies Candidate Effector Proteins Potentially Involved in *Fusarium graminearum* −Wheat Interactions. *Phytopathology®*, *106*(2), 166–176. https://doi.org/10.1094/PHYTO-09-15-0215-R

Ma, L., Houterman, P. M., Gawehns, F., Cao, L., Sillo, F., Richter, H., Clavijo-Ortiz, M. J., Schmidt, S. M., Boeren, S., Vervoort, J., Cornelissen, B. J. C., Rep, M., & Takken, F. L. W. (2015). The *AVR2–SIX5* gene pair is required to activate *I-2*-mediated immunity in tomato. *New Phytologist*, *208*(2), 507–518. https://doi.org/10.1111/nph.13455

Ma, L. J., Geiser, D. M., Proctor, R. H., Rooney, A. P., O'donnell, K., Trail, F., Gardiner, D. M., Manners, J. M., & Kazan, K. (2013). *Fusarium* Pathogenomics. *Annual Review of Microbiology*, *67*, 399–416. https://doi.org/10.1146/annurev-micro-092412-155650

Ma, L. J., van der Does, H. C., Borkovich, K. A., Coleman, J. J., Daboussi, M. J., di Pietro, A., Dufresne, M., Freitag, M., Grabherr, M., Henrissat, B., Houterman, P. M., Kang, S., Shim, W. B., Woloshuk, C., Xie, X., Xu, J. R., Antoniw, J., Baker, S. E., Bluhm, B. H., … Rep, M. (2010). Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature*, *464*(7287), 367–373. https://doi.org/10.1038/nature08850

Martin, F., Cullen, D., Hibbett, D., Pisabarro, A., Spatafora, J. W., Baker, S. E., & Grigoriev, I. v. (2011). Sequencing the fungal tree of life. *New Phytologist*, *190*(4), 818–821. https://doi.org/10.1111/j.1469-8137.2011.03688.x

Metzker, M. L. (2010). Sequencing technologies the next generation. *Nature Reviews Genetics*, *11*(1), 31–46. https://doi.org/10.1038/nrg2626

Michielse, C. B., & Rep, M. (2009). Pathogen profile update: *Fusarium oxysporum*. *Molecular Plant Pathology*, *10*(3), 311–324. https://doi.org/10.1111/j.1364-3703.2009.00538.x

Moricca, S., Ragazzi, A., Kasuga, T., & Mitchelson, K. R. (1998). Detection of *Fusarium oxysporum* f. sp. *vasinfectum* in cotton tissue by polymerase chain reaction. *Plant Pathology*, *47*(4), 486–494. https://doi.org/10.1046/j.1365-3059.1998.00262.x

Nelson, P. E. (1981). Life Cycle and Epidemiology of *Fusarium oxysporum*. In Marshall E. Mace, A. A. Bell, & C. H. Beckman (Eds.), *Fungal Wilt Diseases of Plants* (pp. 51–80). Academic Press. https://doi.org/10.1016/b978-0-12-464450-2.50008-5

Nirenberg, H. I., Ibrahim, G., & Michail, S. H. (1994). Race identity of three isolates of *Fusarium oxysporum* Schlecht. f. sp. *vasinfectum* (Atk.) Snyd. & Hans, from Egypt and the Sudan. *Journal of Plant Diseases and Protection*, *101*(6), 594–597.

Ortiz, C. S., Bell, A. A., Magill, C. W., & Liu, J. (2017). Specific PCR Detection of *Fusarium oxysporum* f. sp. *vasinfectum* California Race 4 Based on a Unique *Tfo1* Insertion Event in the *PHO* Gene. *Plant Disease*, *101*(1), 34–44. https://doi.org/10.1094/PDIS-03-16-0332-RE

Pareek, C. S., Smoczynski, R., & Tretyn, A. (2011). Sequencing technologies and genome sequencing. *Journal of Applied Genetics*, *52*(4), 413–435. https://doi.org/10.1007/s13353-011-0057-x

Raffaele, S., & Kamoun, S. (2012). Genome evolution in filamentous plant pathogens: Why bigger can be better. *Nature Reviews Microbiology*, *10*(6), 417–430. https://doi.org/10.1038/nrmicro2790

Reddy, T. B. K., Thomas, A. D., Stamatis, D., Bertsch, J., Isbandi, M., Jansson, J., Mallajosyula, J., Pagani, I., Lobos, E. A., & Kyrpides, N. C. (2015). The Genomes OnLine Database (GOLD) v.5: A metadata management system based on a four level (meta)genome project classification. *Nucleic Acids Research*, *43*(D1), D1099–D1106. https://doi.org/10.1093/nar/gku950

Rep, M. (2005). Small proteins of plant-pathogenic fungi secreted during host colonization. *FEMS Microbiology Letters*, *253*(1), 19–27. https://doi.org/10.1016/j.femsle.2005.09.014

Rep, M., van der Does, H. C., Meijer, M., van Wijk, R., Houterman, P. M., Dekker, H. L., de Koster, C. G., & Cornelissen, B. J. C. (2004). A small, cysteine-rich protein secreted by *Fusarium oxysporum* during colonization of xylem vessels is required for *I-3*-mediated resistance in tomato. *Molecular Microbiology*, *53*(5), 1373–1383. https://doi.org/10.1111/j.1365-2958.2004.04177.x

Rhoads, A., & Au, K. F. (2015). PacBio Sequencing and Its Applications. *Genomics, Proteomics and Bioinformatics*, *13*(5), 278–289. https://doi.org/10.1016/j.gpb.2015.08.002

Robinson, G. E., Hackett, K. J., Purcell-Miramontes, M., Brown, S. J., Evans, J. D., Goldsmith, M. R., Lawson, D., Okamuro, J., Robertson, H. M., & Schneider, D. J. (2011). Creating a buzz about insect genomes. *Science*, *331*(6023), 1386. https://doi.org/10.1126/science.331.6023.1386

Schadt, E. E., Turner, S., & Kasarskis, A. (2010). A window into third-generation sequencing. *Human Molecular Genetics*, *19*(R2), R227–R240. https://doi.org/10.1093/hmg/ddq416

Schmidt, S. M., Houterman, P. M., Schreiver, I., Ma, L., Amyotte, S., Chellappan, B., Boeren, S., Takken, F. L. W., & Rep, M. (2013). MITEs in the promoters of effector genes allow prediction of novel virulence genes in *Fusarium oxysporum*. *BMC Genomics*, *14*(1), 119. https://doi.org/10.1186/1471-2164-14-119

Schroth, M. N., & Hildebrand, D. C. (1964). Influence of Plant Exudates on Root-Infecting Fungi. *Annual Review of Phytopathology*, *2*(1), 101–132. https://doi.org/10.1146/annurev.py.02.090164.000533

Skovgaard, K., Nirenberg, H. I., O'Donnell, K., & Rosendahl, S. (2001). Evolution of *Fusarium oxysporum* f. sp. *vasinfectum* races inferred from multigene genealogies. *Phytopathology*, *91*(12), 1231–1237. https://doi.org/10.1094/PHYTO.2001.91.12.1231

Snyder, W. C., & Hansen, H. N. (1940). The species concept in *Fusarium*. *American Journal of Botany*, *27*(2), 64–67. https://doi.org/10.1002/j.1537-2197.1940.tb14217.x

Sperschneider, J., Gardiner, D. M., Thatcher, L. F., Lyons, R., Singh, K. B., Manners, J. M., & 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. https://doi.org/10.1093/gbe/evv092

Takken, F., & Rep, M. (2010). The arms race between tomato and *Fusarium oxysporum*. *Molecular Plant Pathology*, *11*(2), 309–314. https://doi.org/10.1111/j.1364-3703.2009.00605.x

Taylor, A., Vágány, V., Jackson, A. C., Harrison, R. J., Rainoni, A., & Clarkson, J. P. (2016). Identification of pathogenicity-related genes in *Fusarium oxysporum* f. sp. *cepae*. *Molecular Plant Pathology*, *17*(7), 1032–1047. https://doi.org/10.1111/mpp.12346

Thatcher, L. F., Gardiner, D. M., Kazan, K., & Manners, J. M. (2012). A highly conserved effector in *Fusarium oxysporum* is required for full virulence on *Arabidopsis*. *Molecular Plant-Microbe Interactions*, *25*(2), 180–190. https://doi.org/10.1094/MPMI-08-11-0212

Utturkar, S. M., Klingeman, D. M., Land, M. L., Schadt, C. W., Doktycz, M. J., Pelletier, D. A., & Brown, S. D. (2014). Evaluation and validation of de novo and hybrid assembly techniques to derive high-quality genome sequences. *Bioinformatics*, *30*(19), 2709–2716. https://doi.org/10.1093/bioinformatics/btu391

van Dam, P., de Sain, M., ter Horst, A., van der Gragt, M., & Rep, M. (2018). Use of comparative genomics-based markers for discrimination of host specificity in *Fusarium oxysporum*. *Applied and Environmental Microbiology*, *84*(1), 1868–1885. https://doi.org/10.1128/AEM.01868-17

van Dam, P., Fokkens, L., Ayukawa, Y., van der Gragt, M., ter Horst, A., Brankovics, B., Houterman, P. M., Arie, T., & Rep, M. (2017). A mobile pathogenicity chromosome in *Fusarium oxysporum* for infection of multiple cucurbit species. *Scientific Reports*, *7*(1), 1–15. https://doi.org/10.1038/s41598-017-07995-y

van Dam, P., & Rep, M. (2017). The Distribution of Miniature Impala Elements and *SIX* Genes in the *Fusarium* Genus is Suggestive of Horizontal Gene Transfer. *Journal of Molecular Evolution*, *85*(1–2), 14–25. https://doi.org/10.1007/s00239-017-9801-0

van de Wouw, A. P., & Howlett, B. J. (2011). Fungal pathogenicity genes in the age of 'omics.' *Molecular Plant Pathology*, *12*(5), 507–514. https://doi.org/10.1111/j.1364-3703.2010.00680.x

van der Does, H. C., & Rep, M. (2007). Virulence genes and the evolution of host specificity in plant-pathogenic fungi. *Molecular Plant-Microbe Interactions*, *20*(10), 1175–1182. https://doi.org/10.1094/MPMI-20-10-1175

Vlaardingerbroek, I., Beerens, B., Schmidt, S. M., Cornelissen, B. J. C., & Rep, M. (2016). Dispensable chromosomes in *Fusarium oxysporum* f. sp. *lycopersici*. *Molecular Plant Pathology*, *17*(9), 1455–1466. https://doi.org/10.1111/mpp.12440

Wollenweber, H. W., & Reinking, O. (1935). The Fusaria. *The Fusaria*.

Wostemeyer, J., & Kreibich, A. (2002). Repetitive DNA elements in fungi (Mycota): Impact on genomic architecture and evolution. *Current Genetics*, *41*(4), 189–198. https://doi.org/10.1007/s00294-002-0306-y

Zolan, M. E. (1995). Chromosome-length polymorphism in fungi. *Microbiology and Molecular Biology Reviews*, *59*(4).

Chapter 2.

Comparative genomics and prediction of the accessory genome regions in cotton-infecting *Fusarium oxysporum formae speciales*

Abstract

Fusarium oxysporum f. sp. *vasinfectum* (*Fov*) is an important plant pathogen responsible for vascular wilt disease on cotton. The genomes of five *Fov* isolates, including a representative of the highly virulent race 4 genotype, were sequenced at a high coverage to obtain reference quality assemblies. Following the development of sequencing technologies, various bioinformatic approaches were employed to identify chromosomal features of *Fov*. The chromosomes of *Fov* were divided into two different categories, the core (CC) and accessory (AC). The CCs usually contain orthologous genes and other components which are essential to the normal life of the fungus. On the other hand, ACs are primarily composed of non-essential genes that confer specific environmental or niche colonizing traits, and therefore if an AC is lost, the fungus can still survive. Scaffolds identified as part of AC were searched to identify syntenic regions between the five *Fov* genomes. Except for *Fov* race 4 (89-1A), AC regions of the remaining four *Fov* genomes shared syntenic loci. Homologs of *Verticillium* species that share the same cotton host with *Fov* and confirmed to be involved in virulence were used to narrow down a key locus in ACs for pathogenicity on cotton. As a result, the genomes of TF1, LA3B, and 14-004 shared about 360 kb that could be important for conferring pathogenicity on cotton. In addition, the most virulent strain *Fov* 89-1A has unique genomics features; a larger genome size and proportion of repetitive elements than the other four *Fov* strains. These genomes provide a necessary resource for comparative genomic analyses to identify genes and genome features that are involved in pathogenicity on cotton and may ultimately be used to identify improved management strategies.

2.1 Introduction

The ascomycete fungi within the *Fusarium oxysporum* species complex (FOSC) are asexual soilborne phytopathogens (Geiser et al., 2013). Collectively, members of the FOSC have a broad host range including high-value agricultural crops such as banana, cotton, melon, onion, and tomato, as well as opportunistic pathogens of animals and humans (Ma et al., 2013; Summerell, 2019). In this regard, the term *forma specialis* (f. sp.) is used to explain host specificity, where a specific *F. oxysporum* isolate is a pathogen on only one or a few host plant species (Edel-Hermann & Lecomte, 2019). The main focus of this study, *F. oxysporum* f. sp. *vasinfectum* (*Fov*) is pathogenic on cotton (*Gossypium* species) and is responsible for a devastating vascular wilt disease, Fusarium wilt (Atkinson, 1892).

To date, many genomes of members of the FOSC have been completed, and comparative genomic analyses have revealed that *F. oxysporum* are phylogenetically diverse and host specificity is polyphyletic (O'Donnell et al., 2009; Zhang et al., 2020). *F. oxysporum* is known to have accessory (also referred to as supernumerary, lineage-specific, B-, or conditionally dispensable) chromosomes that harbor various genomic variations and key pathogenicity factors (Coleman et al., 2009; Ma et al., 2010; Miao et al., 1991). In addition, these accessory chromosomes can be experimentally horizontally transferred to other isolates within the FOSC. The best example of this phenomenon is chromosome 14 in *F. oxysporum* f. sp. *lycopersici* (*Fol* 4287) which is pathogenic on tomato (*Solanum lycopersicum*) (Ma et al., 2010). This accessory chromosome carries important genes encoding effector proteins (*Secreted-in xylem*) and when this chromosome is transferred to an isolate non-pathogenic on tomato, it is able to confer pathogenicity on tomato to the recipient isolate (van Dam et al., 2017; Vlaardingerbroek et al., 2016).

Among *Fov*, a race 7 isolate (NRRL 25433) from China was the first representative from this *forma specialis* sequenced as part of a FOSC genome sequencing project by the Broad Institute. This *Fov* genome was utilized for comparative genomic analysis and phylogenetic relationships between isolates within the FOSC, and subsequent molecular studies were conducted including the identification and characterization of genes of interest (Wang et al., 2018). Despite many advantages, the genome was generated by Illumina sequencing technologies resulting in short length reads and consisted of over 900 scaffolds. The genome fragmentation has hampered in-depth studies in the genomic regions around repeat-rich regions, such as sub-telomeric regions (AT-rich) or genomic hot spots for gene duplication or chromosomal rearrangement (repeats and transposable elements).

In this study, we sequenced five genomes of *F. oxysporum* f. sp. *vasinfectum* isolates collected from cotton plants (*Gossypium* spp.) displaying symptoms indicative of Fusarium wilt using PacBio RS II sequencing technology. The objectives were to: 1) complete five *Fov* genome assemblies and validate the genome quality; 2) identify chromosome structures dividing core, accessory (or lineage-specific), and mitochondrial regions; 3) annotate and identify protein-coding genes that could be contributing to pathogenicity and serve as virulence factors; and 4) perform phylogenetic analysis of 21 FOSC isolates that have different host specificity using orthologous family sets. This study was not limited to search genomic features in each *Fov* genome but sought to define specific genomic regions and virulence factors for the host through comparative analysis between *Fov* genomes.

2.2 Materials and methods

2.2.1 Selection of *Fov* isolates for this study

Fusarium oxysporum f. sp. *vasinfectum* (*Fov*) isolates from five different genotypes were selected based on their significant impact on cotton yield (Table 1). These five *Fov* isolates were collected from infected cotton presenting wilt symptoms, and after isolation, the fungi were grown and maintained on Potato Dextrose Agar (PDA) medium.

2.2.2 Genome sequencing, assembly, and basic genome statistics

High-quality genomic DNA of the five *Fov* genomes was extracted and sequenced using PacBio RS II (Pacific BioSciences, CA, USA). Single molecule real-time (SMRT) bell DNA template libraries of ~8- to 10-kb average insert size for fungal samples were prepared according to the manufacturer's specification. The SMRT sequencing was conducted on the PacBio RS II according to standard protocols (Chin et al., 2013). The SMRT filtered reads trimming adapters and filtering the subreads from the company were sent as final products. *De novo* assemblies of raw reads were performed using Canu v.1.4 (Koren et al., 2017) with the default settings except the expectedSize that was set for the race 4 genome at 65 Mb, the MDS-12 genome at 55 Mb, and the remaining *Fov* genomes at 50 Mb. The quality of the assemblies was validated via QUAST v.4.6.3 and BUSCO v.3.0.1 using the dikarya ortholog gene lineage dataset and the -sp parameter, fusarium_graminearum (Mikheenko et al., 2016; Simão et al., 2015). The genome data has been deposited in the National Center for Biotechnology Information GenBank database under the accession numbers and BioSamples (Tables 1 and 2).

The five genome assemblies were annotated using Funannotate v.1.5.0 pipeline (Palmer, 2017) with raw transcriptomic data sets generated by *Fov* race 4 (89-1A) samples. Prior to Funannotate, the repeated regions in the assembled genomes were softmasked by RepeatMasker v.3.2.7. The extraction, assembly, and initial annotation of these genomes have been previously published (Seo et al., 2020).

2.2.3 Functional annotation of genomes and protein-coding genes

2.2.3.1 Repetitive elements

Repetitive sequences in the five *Fov* genomes were identified with RepeatModeler v.1.0.11 (http://www.repeatmasker.org) using the default parameters. The resulting repeat families served as the library for RepeatMasker v.4.0.6 (http://www.repeatmasker.org) program to calculate the frequency and identify the location of each repeat family in the assembly. Using the repetitive element information, telomeric repeats (CCCTAA) at the end of each scaffold were searched manually (van Dam et al., 2017)

2.2.3.2 Prediction of secreted proteins

To predict small secreted proteins (SSPs), two programs, SignalP v.4.1 and TMHMM v.2 were applied to define proteins having a signal peptide but lacking transmembrane domains (Moller et al., 2001; Petersen et al., 2011). Within the defined secretomes of the five *Fov* genomes, proteins having a sequence length of ≤300 amino acids were chosen. Another effector prediction program, EffectorP v.2.0.1 (Sperschneider, Dodds, Gardiner, et al., 2018) was used followed by ApoplastP (Sperschneider, Dodds, Singh, et al., 2018) using the five *Fov* proteomes as an input.

2.2.3.3 Prediction of carbohydrate utilization enzymes

To search putative carbohydrate-active enzymes (CAZymes), a hidden Markov model (HMM)-based routine was applied. The HMM profiles (release 8.0) were downloaded from dbCAN database (Yin et al., 2012). The Hmmscan in HMMER v.3.0 package (Eddy, 2011) was used to search each *Fov* secretome as queries. Next, the hmmscan-parser script trimmed the primary results. To increase the accuracy of the CAZyme annotation for fungal species specifically,

significant hits when compared with the database were recognized in the set of putative CAZymes using BLAST (e-value < 1e-15 and similarity ≥35%).

2.2.3.4 Other functional analysis

To investigate gene clusters potentially involved in the synthesis of secondary metabolites, the five *Fov* genomes were submitted through the web-based antiSMASH v.3.0 program (Blin et al., 2017). To search homologs of known pathogen-host interaction (PHI) genes and membrane transport proteins, the reference data were downloaded from PHI database v.4.9 (Urban et al., 2020) and TCDB recently modified at Oct 2020 (Saier et al., 2016). Using BlastP (e-value of 0.001), the final homologs were acquired. Overall scheme of the methods can be seen in Figure 1.

2.2.4 Investigation of the structure of the *Fov* genomes

2.2.4.1 Defining core chromosomes

The previously characterized *F. oxysporum* f. sp. *lycopersici* 4287 assembly (Ma et al., 2010) was used as reference to identify core and accessory (or lineage-specific) regions in *Fov* genomes using MUMmer v.3.22 (PROmer –mum, delta-filter -g) (Kurtz et al., 2004). The assembly of *Fol* 4287 was obtained from NCBI (GCA_000149955.2) and repeats were masked by RepeatMasker v.3.2.7 as with the *Fov* assemblies. After the alignment between *Fol* and each *Fov* genome, the percentage of covering bps in each scaffold was calculated and scaffolds having a cutoff over 30% identity were designated as core chromosomes (Williams et al., 2016). The scaffold names from *Fov* that were similar to the *Fol* genome were changed following the chromosome information of *Fol* 4287.

The near chromosome-level assembly quality of the *Fov* genomes allowed the core regions to be separated for an in-depth study of sub-telomeric regions. The resulting tab-separated files

from MUMmer previously used for characterizing genome compartments, had all the coordinates which were the start and stop sites of both the reference and the query. After extracting the coordinates of one or two scaffolds of each *Fov* genome assigned to each of the 11 core chromosome of *Fol* 4287 manually, 100 kb windows were constructed for each core chromosome merging interspecific nucleotide alignments via BedTools v.2.29.2 (merge –d 100000) (Quinlan & Hall, 2010). Based on the resultant MUMmer plots, the direction of scaffolds showing blue lines, was adjusted. Core regions of the five *Fov* genomes were categorized into matching regions with *Fol* 4287 and the gaps without any matches, and visualized as linear bar plots by the R package, chromoMAP (Anand & Rodriguez Lopez, 2019).

2.2.4.2 Mitochondrial genome

The entire mitochondrial sequences, previously analyzed for members of the *Fusarium* genus (Al-Reedy et al., 2012), were obtained from NCBI; *F. oxysporum* F11 (AY945289) and VPR1 (AY874423), *F. verticillioides* 7600 (JN041210), *F. vanettenii* 77-13-4 (JN041209), and *F. graminearum* PH-1 (DQ364632). All the scaffolds of five *Fov* genomes were searched against the mitochondrial genomes by BlastN with the default setting.

2.2.4.3 Identification of accessory genomic regions

After identifying the core scaffolds and the mitochondrial genome, the remaining scaffolds with less than 30% identity were considered to comprise the accessory genome, and the scaffolds named as ACC and ordered according to size. The ACC scaffolds were searched for syntenic regions between the five *Fov* genomes, and alignments were generated via the NUCmer algorithm with default settings in the MUMmer program v.3.22 (Kurtz et al., 2004). The results were further filtered by the delta-filter utility with cutoffs, >10 kb in length and >80% identity. To see the

overall alignment of patterns between syntenic regions, continuous alignments were generated ignoring gaps less than 50 kb. The linear syntenic regions were visualized in R using the package karyoploteR (Gel & Serra, 2017). Information concerning the GC content, repetitive regions, locations of predicted genes, gene clusters of secondary metabolites, and other genomic information generated during the previous annotation, was divided into core and accessory chromosomes separately to identify potential distinguishable patterns.

2.2.5 Pan-genome analysis of *F. oxysporum* strains (tree building and ortholog analysis)

To determine the phylogenetic placement of the sequenced *Fov* isolates within the FOSC, 21 protein sets representing the five *Fov* and 16 other isolates from the FOSC were chosen based on the quality of the genome and proteome data (Table 2) (McCarthy & Fitzpatrick, 2019). OrthoMCL v.2.0.9 (Li et al., 2003) was utilized to group the proteomes of the 21 FOSC isolates, and 458,206 proteins filtering 51 poor quality proteins were blasted at an e-value cutoff of 10^{-5}. Among the 25,559 orthogroups, only groups having a single copy gene for all 21 genomes were chosen to visualize the distribution of syntenic orthologues by the R package UpSetR (Conway et al., 2017). Of all the orthogroups, 5,479 groups (21.4%) were observed as common single copy genes in all 21 genomes and were extracted for phylogenetic analysis. Protein sequences in individual ortholog groups were aligned by MUCSCLE v.3.8 (Edgar, 2004) with the default parameters and trimmed by BMGE v.1.12 (Criscuolo & Gribaldo, 2010) with a more relaxed index (-m BLOSUM30). The alignments were concatenated into a single supermatrix composed of 2,258,063 amino acids. To generate the phylogenetic tree, a PhyML algorithm (Guindon et al., 2010) was chosen to handle the large sequence and increase the speed using JTT (Jones-Taylor-Thornton) model of amino acid evolution with a bootstrap of 500 replicates. The final consensus tree was drawn using FigTree v.1.4.4 (Rambaut, 2012).

2.2.6 Identification of homologous genes of *Verticillium dahliae Vd*991 in the five *Fov* isolates

Seven genes (*VEDA_05193-05199*) involved in virulence on cotton and resided in the lineage-specific regions of *V. dahliae* (*Vd*991) and had similarities with *F. oxysporum* f. sp. *vasinfectum* race 7 (NRRL 25433) were obtained from GenBank (Chen et al., 2018). Putative homologs in the five *Fov* proteomes were identified using BlastX with a cutoff e-value of 0.001 and >75% identity. Annotated information of the identified homologs including the chromosome location, and both start and stop sites identified. To see the phylogenetic distance of each gene between *Fov* and *Verticillium* species if valid matching homologs were found, maximum likelihood trees between the genes from *Vd*991 and homologs from the six *Fov* isolates were constructed using PhyML with 1000 bootstraps and the default settings after aligning protein sequences via ClusterW (Thompson et al., 2003). For most of the trees, a homolog from *Magnaporthe oryzae* was used as the outgroup, except for one gene, VEDA_05193, where a homolog from *F. verticillioides* was used as the outgroup following the previous study in *V. dahlia* (Chen et al., 2018). The consensus trees were drawn using FigTree v.1.4 (Rambaut, 2012).

2.3 Results

2.3.1 High quality genome assemblies of five *F. oxysporum* f. sp. *vasinfectum* (*Fov*)

The five genomes of *F. oxysporum* f. sp. *vasinfectum* (*Fov*) had an average coverage over 188×, ranging from 163× in *Fov* TF1 to 238× in *Fov* LA3B (Table 1). The five *Fov* genomes were assembled using CANU (Koren et al., 2017), and the draft genomes of isolates TF1, LA3B, and 14-004 were about 50 Mb in genome size and the LA1E isolate had a slightly larger genome size of 53 Mb. The race 4 isolate 89-1A had distinctive genomic characteristics when compared to the other *Fov* isolates, in particular a larger genome size, 63 Mb, that was distributed in 94 scaffolds

(Table 1). The genomes of three isolates TF1, LA3B, and 14-004 were encompassed in less than 20 scaffolds. However, *Fov* LA1E (MDS-12) was composed of 26 scaffolds while the *Fov* 89-1A (race 4) genome had a total of 94 scaffolds. Despite this, the %GC content was almost the same at ~47% for all the genomes. BUSCO analysis indicated that the genomes had over 98% completeness which exceeded other *Fusarium* genomes including the well referenced genomes of *F. graminearum* PH1 (94.7%) and *F. oxysporum* (94.2%) (Niehaus et al., 2017).

By combining transcriptomic data generated from the *Fov* race 4 (89-1A) isolate after a treatment of gossypol (Pokhrel, 2018), a phytoalexin of cotton, 17,975-20,222 protein encoding genes were predicted by Funannotate (Palmer, 2017) in the *Fov* genomes. Overall, the *Fov* race 4 genome encoded a larger number of genes, as it had the largest genome size. During the beginning of the annotation by Funannotate, RepeatMasker (default) predicted repetitive regions and masked the entire genome to reduce the misannotation, and the repeat-masked genomes contained 6.25-8.61% repetitive regions except for *Fov* race 4. The genome of 89-1A consisted of 18.44% repetitive regions, which was 2-3 times larger than the other *Fov* isolates. These results have been previously published (Seo et al., 2020) and are summarized in Table 1.

2.3.2 The phylogenetic tree using orthologues shared within FOSC

To better understand the evolutionary relationship between the *Fov* isolates and other members within the FOSC, orthologue families were grouped from 21 FOSC proteomes (Table 2). Out of 25,559 orthogroup families, 5,479 single-copy proteins (composed of a total of 2,258,063 amino acids) were common among all 21 FOSC proteomes and were concatenated to generate a maximum-likelihood phylogeny with 500 bootstrap replicates (Figure 2). The phylogenetic analysis indicated that the ability to cause disease on cotton (f. sp. *vasinfectum*) is polyphyletic in the FOSC (Figure 3). The FOSC was divided into two large clades except for two isolates of *F.*

oxysporum f. sp. *cubense* tropical race 4 which were located outside these clades. Among the six *Fov* isolates, five *Fov* isolates were clustered in the same clade and *Fov* LA127/140 (14-004) was in the different subclade. The *Fov* race 4 and 7 and the MDS-12 isolates were closely related in the same subclade but separate from the *Fov* race 1 and LA108 isolates.

2.3.3 The compartmentalization of five *Fov* genomes

2.3.3.1 Mitochondrial genome

Four complete mitochondrial genomes from other Fusarium species and incomplete mitochondrial DNA (mtDNA) from *F. oxysporum* VPR1 were used to identify the mitochondrial DNA (mtDNA) from each *Fov* isolate (Table 3). A single scaffold was identified in each *Fov* genome corresponding to the mtDNA. Despite that the lengths of the two *F. oxysporum* mtDNA were shorter (~34 kb) than the mtDNA of the *Fov* genomes (~74 kb), the entire DNA sequence of the two other *F. oxysporum* mtDNA were more than 98% similar to the mtDNA specific scaffold for each *Fov* (Table 3). In addition, the GC content (%) was distinguishable between the nuclear DNA and the mtDNA. The GC content of all five *Fov* mtDNA was predicted to be 31-32%, while the average GC content of the nuclear genome was 46-47% (Table 4). Due to the different genomic characteristics of the mtDNA, the mtDNA and protein-coding genes were excluded from further analysis.

2.3.3.2 Genomic features sharing common characteristics in the core chromosomes of five *Fov* isolates

Whole genome comparison of the *Fov* genomes to the *Fol* genome by the MUMmer program indicated 11 to 14 scaffolds of the *Fov* genomes corresponded to the eleven core chromosomes of *Fol* 4287, ranging in size from 45 to 48 Mb (Figure 4a and Table 4). To better

examine the sub-telomeric regions of the core chromosomes in the *Fov* isolates, continuous alignments were generated from the coordinates between each *Fov* isolate and *Fol* 4287 by Bedtools (Quinlan & Hall, 2010). The regions colored blue indicate conservation with the corresponding region of the core chromosomes of *Fol* 4287, while regions that lack similarity/synteny are indicted in dark gray (Figure 4b). The sub-telomeric regions at the ends of the core chromosomes were found to have the highest variability within the core chromosomes and were usually found to be more variable on the smaller core chromosomes. Chromosomes (CHR) 02, 04, and 05 in all five *Fov* genomes were syntenic and had high similarity with the corresponding chromosomes in *Fol* almost to the telomere. However, the remaining chromosomes had regions that were less conserved. In addition, CHR02 in *Fov* race 4 and MDS12 contained regions that lacked conservation in the middle of the chromosome (Figure 4), while CHR04 in *Fov* TF1 and CHR05 in *Fov* 14-004 had a gap of approximately 200 kb. In most instances, a single scaffold of the *Fov* assembly was assigned to one chromosome of *Fol* 4287, however two separate *Fov* scaffolds consisted of one *Fol* core chromosome for chromosome 4 in *Fov* TF1, chromosomes 5, 7, and 12 in 89-1A, and chromosomes 1, 5, and 7 in MDS-12 and are indicated with an asterisk (*) at the end of the chromosomes (Figure 4b).

2.3.3.3. Identification of the accessory genome from the five *Fov* isolates

Previously, four accessory (AC or lineage-specific) chromosomes were identified in *F. oxysporum* f. sp. *lycopersici* (*Fol* 4287) representing approximately 29% of the entire genome: chromosomes 3, 6, 14, and 15 (Ma et al., 2010). Lineage-specific regions in *Fol* 4278 are enriched for sequences that lack synteny to various other FOSC genomes (Armitage et al., 2018; Ma et al., 2010; Zhang et al., 2020). About 23% percent (over 14 Mb) of the *Fov* race 4 genome is composed of AC regions, more than twice as large as the AC component of the other four *Fov* genomes

(Table 4). The *Fov* MDS-12 genome had the second largest AC component, 11% (over 6 Mb) while the AC sizes of the *Fov* race 1, LA108, and LA127/140 genomes were under 5 Mb and made up 8- 9% of each genome.

To reveal the potential shared genome features in the accessory regions, a genome-wide synteny comparison was performed by MUMmer (Langner et al., 2021). Compared to the core chromosomes, the accessory regions were highly fragmented, primarily due to the high percentage of repetitive sequences present (Table 4). There were several syntenic regions identified with an identity cutoff over 80% (Figure 5a). However, the *Fov* race 4 isolate, 89-1A, contained no syntenic regions when compared to the AC genome of three isolates, *Fov* TF1, LA3B, and 14-004. When compared to the more closely related genome of *Fov* MDS-12, there were several syntenic areas observed at the end of scaffolds, but the size was negligible indicating the *Fov* race 4 has no significant syntenic regions with the other *Fov* isolates.

Outside of regions less than 100 kb in size, the syntenic blocks found between the four *Fov* genomes ranged from 250 to 600 kb, and three regions were significant (Figure 5a). The first syntenic region was shared between TF1, LA3B, and 14-004, where the locus underwent an inversion event in the LA3B genome. An additional syntenic region was located on the same accessory scaffold between TF1 and 14-004. The second significant syntenic region accounted for almost entire scaffolds of TF1 (ACC02) and 14-004 (ACC03) which had syntenic regions with two scaffolds of MDS-12 (ACC03 and ACC05). The shared syntenic region between TF1 and 14-004 had sequence identity over 97%. Although the chromosomal inversion events in ACC03 and the almost entire ACC05 scaffold in MDS12 were observed, these two scaffolds (ACC03 and ACC05) had sequence identity about 94 and 69% with other two isolates, TF1 and 14-004. Lastly, two scaffolds of 14-004 (ACC02 and 04) and one scaffold, ACC01 of MDS12 had a syntenic

region with two inversion events. These results indicate that chromosomal rearrangements occurred within the accessory genome of four *Fov* isolates.

2.3.4 Genome and gene features of *Fov* genomes

The assembles of the five *Fov* genomes had different presence/absence patterns of telomeric repeats at the end of each scaffold. Three *Fov* genomes, isolates TF1, LA3B, and MDS-12, probably had complete chromosome-level scaffolds for 10, 7, and 6 chromosomes, respectively, as each end had at least over 100 bp of telomeric repeats (CCCTAA). Most of the scaffolds having telomeric sequences on both ends for the three isolates were designated as core chromosomes: CHR05, 07-13 (8) in TF1, CHR01, 07-08, 10-13 (7) in LA3B, and CHR02, 04, 08, 10, and 12 (5) in MDS-12. In the case of the *Fov* MDS-12 genome, two scaffolds had high sequence similarity to the core chromosome 01 of *Fol* 4287, each had telomeric repeats at one end and were able to be joined to construct a complete telomere-to-telomere CHR01 (data not shown). Intriguingly, both *Fov* TF1 and MDS-12 had accessory scaffolds containing telomeres on both ends, ACC01-02 in TF1, and ACC02 in MDS-12. The remaining scaffolds in the three isolates had telomeric repeats on at least one end (Table 1). Only two scaffolds had telomeric repeats on both ends in *Fov* 14-004, where most of the scaffolds (12) had at least one telomeric end. Interestingly, telomeric repeats were absent from the ends of all 94 scaffolds in the *Fov* race 4 (89-1A) genome (Table 1).

With the exception of the *Fov* race 4 genome, the four other *Fov* genomes had a similar size of repetitive elements, ranging from 3.16 to 4.49 Mb, where approximately 2-3 Mb resided in the core genomic regions (Table 4). However, the repetitive sequences of the race 4 genome were significantly different, and the size of the repetitive sequences was approximately three times larger than that of the other *Fov* genomes. In addition, the total repetitive sequences in the accessory regions of the race 4 genome were twice as large than the repeats found in the core

regions, and approximately four to seven times larger than the repeats in the accessory regions of other the four *Fov* genomes (Figure 6c). The repeat content for these genomes was mainly composed of DNA and long terminal repeat (LRT) elements, and contained simple repeats and long interspaced nuclear elements (LINE) with small sections, which contributed to the repeat variations in the four genomes. Furthermore, although the repeat size was large in *Fov* 89-1A, the repeat content mainly consisted of DNA elements and LRT elements like the other *Fov* genomes and had a smaller number of simple repeats compared to the other four *Fov* genomes (Figure 7a-b).

Functional annotation of the protein-coding genes in the five *Fov* genomes identified several potential features contributing to pathogenicity and could be serving as virulence factors including genes encoding secreted proteins, carbohydrate-active enzymes (CAZymes), secondary metabolites (SMs), pathogen-host interaction (PHI) proteins, and transporters, which may play important functions in pathogenicity (Table 4). As with the genome size, the *Fov* race 4 (89-1A) isolate had the highest number in all of these categories with the exception of putative CAZymes and secondary metabolite gene clusters (Table 4). Three different approaches were taken to identify putative fungal effectors including a universal pipeline (SignalP and TMHMM), EffectorP2, and ApoplastP (Figure 6d-f), where EffectopP2 predicted the highest number of putative effectors. Several putative effectors were located in the accessory regions in all five *Fov* genomes. The estimated number of CAZymes and SM clusters were similar for all five of the *Fov* genomes and most of the enzymes and SMs were located on scaffolds of the core genome (Figure 6g-h). Among the classes of CAZymes, proteins containing carbohydrate-binding modules (CBM) and glycosyltransferase (GT) enzymes were rarely present, while the glycoside hydrolase (GH)

family had the highest number (over 180) of enzymes followed by enzymes with auxiliary activities (AA), the carbohydrate esterases (CE), and the polysaccharide lyases (PL) (Figure 8).

2.3.5 Homologs between the cotton pathogen *Verticillium* and the five *Fov* genomes

Genome analysis of the causative agent of Verticillium wilt of cotton, *V. dahliae* (*Vd*991) revealed lineage-specific regions which encoded 32 homologs that were also present in *Fusarium* spp., including *SIX* (Secreted-in xylem) genes (Chen et al., 2018). Further analysis showed seven clustered LS genes (*VEDA_05193-05199*) were homologues to genes in the genome of the *Fov* race 7 (NRRL 25433) isolate. Homologues for all seven of the *Verticullium* genes were searched for in the five *Fov* proteomes (e-value $<1e^{-3}$, and identities >75%) (Table 5). One or two homologs were identified in most of the *Fov* genomes with the exception of *VEDA_05194* which was absent in all the *Fov* genomes, while the genome of the *Fov* MDS-12 isolate lacked homologs to many of the genes (Table 5). Interestingly, three *Fov* isolates, TF1 (ACC01), LA3B (ACC03), and 14-004 (ACC01) had a single-copy homolog of four of the *Vd*991 genes (*VEDA_05193, VEDA_05196-8*) and two copies of *VEDA_05199* located in a 12 kb cluster on an accessory scaffold (Table 6). Furthermore, these clusters were found in one of the syntenic blocks among the accessory regions between the three isolates (Figure 5b). In the case of the *Fov* race 4 (89-1A) isolate, four of the *Vd*991 LS genes (*VEDA_05196-9*) had homologs in two different accessory scaffolds, ACC02 and 60.

Phylogenetic analyses of the five *Vd*991 genes and the homologues from the five sequenced *Fov* genomes, and the previously identified corresponding homologs from the race 7 (25433) genome were conducted (Figure 9). Phylogenetic trees of four genes (*VEDA_05193, 05196-8*) indicated that the *Vd*991 genes were more closely related to the homologues from the *Fov* race 4 and 7 genomes than the other *Fov* genomes, and that homologs of *Fov* race TF1, LA3B, and 14-

004 for four genes grouped together in the same sub-clade (Figure 9a-d). In the case of *VEDA_05199*, two homologs were present in *Fov* TF1, 89-1A, LA3B, and 14-004 (Table 5 and Figure 9e). One homolog of *VEDA_05199* from the *Fov* race 7 genome and two copies from the *Fov* race 4 genomes grouped together (Figure 9e).

2.4 Discussion

The diversity of the sequenced *Fov* race 1, 4, LA108, LA127/140, and MDS-12 isolates provides insight into the molecular basis and evolutionary relationship within the *F. oxysporum* species complex, particularly the virulence factors contributing to host specificity for cotton. In this study, the long-read sequences enabled the assembled genomes to have near chromosome-level quality, as the high coverage with long reads increased the probability to have the entire sequence of large repetitive regions or at the end of chromosomes e.g., centromeres and telomeres (van Dam et al., 2017). Thus, the generation of the complete, single-scaffold (or contig) genomes improved N50 lengths and scaffold numbers.

The *Fov* genomes were compartmentalized with structural conservation similar to the *F. oxysporum* f. sp. *lycopersici* 4287 genome. Eleven to fourteen scaffolds of the five *Fov* genomes were designated to compose the core genome, a single scaffold represented the mitochondrial genome, and the rest of scaffolds were classified as potential accessory (or lineage-specific) genomic regions. Most of the core regions have high sequence similarity with the *F. oxysporum* f. sp. *lycopersici* 4287 genome. In minor cases, 120-200 kb regions in the middle of CHR02 (*Fov* 89-1A and MDS-12), CHR04 (TF1) and CHR05 (14-004) lacked significant conservation with the corresponding locus in the *Fol* 4287 genome. Furthermore, the classical telomeric repeats (CCCTAA) were found in most of the scaffolds of the core genome existed on at least one end of the scaffold, with the exception of scaffolds of *Fov* 89-1A and 14-004. While an optical map can

be used to validate the assembled genomes of the *Fov* isolates, the scaffolds constructed by long sequence reads with high coverage can be defined as core chromosomes after comparison with the *Fol* 4287 assembly and the presence of telomeric repeats (Ma et al., 2010; Zhang et al., 2020).

The phylogeny generated using 55 conserved single copy orthologs within the *Fusarium* genus (Zhang et al., 2020) was similar to the single-copy 5,479 orthologous genes phylogeny in this study, clearly indicating a polyphyletic nature for virulence on cotton. The orthologs for the phylogenetic trees were located in the core regions rather than accessory. In the phylogeny, *F. oxysporum* f. sp. *cubense* TR4, responsible for Panama wilt of bananas, was distinctly separate from the other FOSC isolates. The phylogenetic tree revealed two major clades, where the first clade included FOSC isolates pathogenic to tomato (f. sp. *lycopersici* and *radices-lycopersici*) and melon (f. sp. *melonis*), and Fo47 used for biological control while the second clade had two subclades, containing most of the cotton isolates (f. sp. *vasinfectum*) as well as FOSC isolates that have various plant hosts such as *Pisum* sp. (*pisi*), *Raphanus* sp. (*raphani*), *Brassica* sp. (*conglutinans*), and *Arabidopsis* sp. (*arabidopsis*). Focusing on the six *Fov* isolates, two divergent lineages were responsible for the evolution of the *Fov* isolates. The *Fov* 14-004 isolate is uniquely placed in the clade with other FOSC isolates having a host range other than cotton with 100% bootstrap support. The race 4, 7, and MDS-12 genotypes, and the *Fov* race 1 and LA108 genotypes were divided in two separate lineages.

The question of how these asexual pathogens have evolved can be partially addressed using the emerging genomic data and analyzing the well-assembled chromosome structure. Comparison of the *Fov* sequence similarities with *Fol* 4287 allowed the potential core chromosomes/genome to be identified and the remaining scaffolds designated as potential accessory chromosomes. Potential accessory chromosomes were found in all five of the *Fov* genomes using comparative

genomic analysis, indicating all major lineages of FOSC carry accessory chromosomes. In addition, the comparison with core regions revealed that the accessory scaffolds of each *Fov* isolate were likely not originated from core chromosomes through rearrangements or segmental duplications. Although the accessory regions were more variable than the core regions, they share common genomic features with each other such as lower gene density and higher amount of repetitive sequence. For example, in the *Fov* LA3B genome, the first assigned accessory scaffold (ACC01) by size has a larger size than the smallest core chromosome (CHR13), 3.2 and 2.8 Mb, respectively. However, the number of predicted genes encoding proteins (790 and 1089), putative effectors (53 and 113), and CAZymes (9 and 24), were distinguishably smaller in ACC01 when compared to core CHR13. The amount of repetitive sequences of the largest accessory scaffold, (ACC01, 837 kb) were two or three times more than on the smallest core chromosome (CHR13, 304 kb). A high proportion of repeats in the accessory scaffolds of the *Fov* genomes was common. The smaller size of scaffolds in the accessory regions is most likely due to the high repeat content which can contribute to sequencing errors or misassembles. However, among the five sequenced *Fov* genomes, three accessory scaffolds had classical telomeric repeats on both ends, and repeats were present at either end of many scaffolds identified as part of the accessory genome of several of the *Fov* isolates except race 4, where none of the putative accessory scaffolds had telomeric repeats.

Analysis of all the accessory regions of the five *Fov* isolates revealed several syntenic regions shared between the genomes that were absent in *Fov* race 4 (89-1A). In *Fov* 14-004, almost all of the accessory scaffolds had synteny with loci in other *Fov* isolates. *Fov* LA3B has a single syntenic locus with *Fov* TF1 and 14-004, but the rest of the accessory regions were unique to its genome. Importantly, the shared syntenic block between three isolates, TF1, LA3B, and 14-004 contains clustered homologs with five genes in *V. dahlia Vd*991, which is responsible for wilt

disease on agriculturally important crops including cotton (Chen et al., 2018). The comparative analyses of the genome of the highly virulenct cotton *V. dahlia* isolate *Vd*991 identified 64 genes in LS regions that were specific to *Vd*991 when compared to isolates, JR2 (tomato) and *Vd*Ls.17 (lettuce). Interestingly, 32 of these genes have homologs in 15 *F. oxysporum* isolates and in particular, seven of these *Vd*991 genes had high similarity with genes in the *Fov* race 7 genome. Gene deletion mutants of this LS region containing the gene clusters in *Vd*991 had a reduction of disease symptoms on cotton (Chen et al., 2018). The genomic study assumed that some LS genes or regions in *Vd*991 were probably acquired from other fungi, possibly *F. oxysporum* f. sp. *vasinfectum*, which share ecological niches such as the soil and the plant vascular system during coinfection on cotton (van Dam et al., 2017).

The phylogenetic placement of *Fov* 14-004 suggests virulence on cotton emerged independently when compared to the other *Fov* isolates that have a close phylogenetic relatedness. As the accessory genome of this isolate has synteny in most of the accessory regions, *Fov* 14-004 may have acquired the lineage-specific chromosomes during co-existence with other *Fov* isolates. It is plausible that several mechanisms, e.g., a horizontal chromosome transfer, has conferred a non-pathogenic *F. oxysporum* isolate to become pathogenic on cotton. Likewise, the genome features in the core genome including the size, GC content, the number of genes, effectors, CAZymes and the other annotation metrics are consistent with the core genome of other *Fov* isolates, suggesting genome stabilization in the core regions. However, *Fov* 89-1A has a high amount of variability when compared to the other *Fov* isolates, in particular a surprisingly larger amount of repetitive sequences and a lack of synteny in the accessory regions. It is possible that the *Fov* race 4 genomic regions have various recombination hot spots which increases the chances for the genome to adapt to the environment and increase pathogenicity on hosts at telomeric or

sub-telomeric regions by chromosomal rearrangements and segmental duplications (Fokkens et al., 2018). The completion of three genomes which have many telomeres appear to be near complete and could serve as a reference for comparison with other FOSC isolates.

In summary, almost complete genomes of *Fov* isolates provide insight to host specificity on cotton and increase the genome diversity in the FOSC. Most genetic features are still unknown, limiting our complete understanding of the roles in pathogenesis and host adaption. Our comparative genomics analysis among *Fov* genomes provide insights into the genetic variation between the genotypes of *Fov* and lays the basis for studying the genomic structures of *Fov* and can broaden the known pathogenicity factors.

Reference

Al-Reedy, R. M., Malireddy, R., Dillman, C. B., & Kennell, J. C. (2012). Comparative analysis of *Fusarium* mitochondrial genomes reveals a highly variable region that encodes an exceptionally large open reading frame. *Fungal Genetics and Biology*, *49*(1), 2–14. https://doi.org/10.1016/j.fgb.2011.11.008

Anand, L., & Rodriguez Lopez, C. M. (2019). ChromoMap: An R package for Interactive Visualization and Annotation of Chromosomes. In *bioRxiv* (p. 605600). *bioRxiv*. https://doi.org/10.1101/605600

Armitage, A. D., Taylor, A., Sobczyk, M. K., Baxter, L., Greenfield, B. P. J., Bates, H. J., Wilson, F., Jackson, A. C., Ott, S., Harrison, R. J., & Clarkson, J. P. (2018). Characterisation of pathogen-specific regions and novel effector candidates in *Fusarium oxysporum* f. sp. *cepae*. *8*(1), 13530. https://doi.org/10.1038/s41598-018-30335-7

Atkinson, G. F. (1892). *Some Diseases of Cotton* (Vol. 41). Agricultural Experiment Station.

Blin, K., Wolf, T., Chevrette, M. G., Lu, X., Schwalen, C. J., Kautsar, S. A., Suarez Duran, H. G., de Los Santos, E. L. C., Kim, H. U., Nave, M., Dickschat, J. S., Mitchell, D. A., Shelest, E., Breitling, R., Takano, E., Lee, S. Y., Weber, T., & Medema, M. H. (2017). AntiSMASH 4.0 - improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Research*, *45*(W1), W36–W41. https://doi.org/10.1093/nar/gkx319

Chen, J. Y., Liu, C., Gui, Y. J., Si, K. W., Zhang, D. D., Wang, J., Short, D. P. G., Huang, J. Q., Li, N. Y., Liang, Y., Zhang, W. Q., Yang, L., Ma, X. F., Li, T. G., Zhou, L., Wang, B. L., Bao, Y. M., Subbarao, K. v., Zhang, G. Y., & Dai, X. F. (2018). Comparative genomics reveals cotton-specific virulence factors in flexible genomic regions in *Verticillium dahliae* and evidence of horizontal gene transfer from *Fusarium*. *New Phytologist*, *217*(2), 756–770. https://doi.org/10.1111/nph.14861

Chin, C. S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E. E., Turner, S. W., & Korlach, J. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods*, *10*(6), 563–569. https://doi.org/10.1038/nmeth.2474

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., … VanEtten, H. D. (2009). The Genome of *Nectria haematococca*: Contribution of Supernumerary Chromosomes to Gene Expansion. *PLoS Genetics*, *5*(8), e1000618. https://doi.org/10.1371/journal.pgen.1000618

Conway, J. R., Lex, A., & Gehlenborg, N. (2017). UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics*, *33*(18), 2938–2940. https://doi.org/10.1093/bioinformatics/btx364

Criscuolo, A., & Gribaldo, S. (2010). BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments.

Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Computational Biology*, *7*(10), 1002195. https://doi.org/10.1371/journal.pcbi.1002195

Edel-Hermann, V., & Lecomte, C. (2019). Current status of *Fusarium oxysporum formae speciales* and races. In *Phytopathology* (Vol. 109, Issue 4, pp. 512–530). American Phytopathological Society. https://doi.org/10.1094/PHYTO-08-18-0320-RVW

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. https://doi.org/10.1186/1471-2105-5-113

Fokkens, L., Shahi, S., Connolly, L. R., Stam, R., Schmidt, S. M., Smith, K. M., Freitag, M., & Rep, M. (2018). The multi-speed genome of *Fusarium oxysporum* reveals association of histone modifications with sequence divergence and footprints of past horizontal chromosome transfer events. In *bioRxiv* (p. 465070). bioRxiv. https://doi.org/10.1101/465070

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., … 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. https://doi.org/10.1094/PHYTO-07-12-0150-LE

Gel, B., & Serra, E. (2017). karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics*, *33*(19), 3088–3090. https://doi.org/10.1093/bioinformatics/btx346

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307–321. https://doi.org/10.1093/sysbio/syq010

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. *Genome Research*, *27*(5), 722–736. https://doi.org/10.1101/gr.215087.116

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., & Salzberg, S. L. (2004). Open Access Versatile and open software for comparing large genomes (Vol. 5, Issue 2). http://www.tigr.org/software/mummer.

Langner, T., Harant, A., Gomez-Luciano, L. B., Shrestha, R. K., Malmgren, A., Latorre, S. M., Burbano, H. A., Win, J., & Kamoun, S. (2021). Genomic rearrangements generate hypervariable mini-chromosomes in host-specific isolates of the blast fungus. *PLoS Genetics*, *17*(2), e1009386. https://doi.org/10.1371/JOURNAL.PGEN.1009386

Li, L., Stoeckert, C. J., & Roos, D. S. (2003). OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Research*, *13*(9), 2178–2189. https://doi.org/10.1101/gr.1224503

Ma, L. J., Geiser, D. M., Proctor, R. H., Rooney, A. P., O'donnell, K., Trail, F., Gardiner, D. M., Manners, J. M., & Kazan, K. (2013). *Fusarium* Pathogenomics. https://doi.org/10.1146/annurev-micro-092412-155650

Ma, L. J., van der Does, H. C., Borkovich, K. A., Coleman, J. J., Daboussi, M. J., di Pietro, A., Dufresne, M., Freitag, M., Grabherr, M., Henrissat, B., Houterman, P. M., Kang, S., Shim, W. B., Woloshuk, C., Xie, X., Xu, J. R., Antoniw, J., Baker, S. E., Bluhm, B. H., … Rep, M. (2010). Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature*, *464*(7287), 367–373. https://doi.org/10.1038/nature08850

McCarthy, C. G. P., & Fitzpatrick, D. A. (2019). Pan-genome analyses of model fungal species. *Microbial Genomics*, *5*(2), e000243. https://doi.org/10.1099/mgen.0.000243

Miao, V. P., Covert, S. F., & Vanetten, H. D. (1991). A fungal gene for antibiotic resistance on a dispensable ("B") chromosome. *Science*, *254*(5039), 1773–1776. https://doi.org/10.1126/science.1763326

Mikheenko, A., Valin, G., Prjibelski, A., Saveliev, V., & Gurevich, A. (2016). Icarus: visualizer for *de novo* assembly evaluation. *Bioinformatics*, *32*(21), 3321–3323. https://doi.org/10.1093/bioinformatics/btw379

Moller, S. M. ¨, Croning, M. D. R., & Apweiler, R. (2001). Evaluation of methods for the prediction of membrane spanning regions. In *BIOINFORMATICS* (Vol. 17, Issue 7).

Niehaus, E. M., Kim, H. K., Münsterkötter, M., Janevska, S., Arndt, B., Kalinina, S. A., Houterman, P. M., Ahn, I. P., Alberti, I., Tonti, S., Kim, D. W., Sieber, C. M. K., Humpf, H. U., Yun, S. H., Güldener, U., & Tudzynski, B. (2017). Comparative genomics of geographically distant *Fusarium fujikuroi* isolates revealed two distinct pathotypes correlating with secondary metabolite profiles. *PLoS Pathogens*, *13*(10), e1006670. https://doi.org/10.1371/journal.ppat.1006670

O'Donnell, K., Gueidan, C., Sink, S., Johnston, P. R., Crous, P. W., Glenn, A., Riley, R., Zitomer, N. C., Colyer, P., Waalwijk, C., Lee, T. van der, Moretti, A., Kang, S., Kim, H. S., Geiser, D. M., Juba, J. H., Baayen, R. P., Cromey, M. G., Bithell, S., … Sarver, B. A. J. (2009). A two-locus DNA sequence database for typing plant and human pathogens within the *Fusarium oxysporum* species complex. *Fungal Genetics and Biology*, *46*(12), 936–948. https://doi.org/10.1016/j.fgb.2009.08.006

Palmer, J. (2017). *Funannotate: eukaryotic genome annotation pipeline* (1.7.0). https://funannotate.readthedocs.io/en/latest/

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: Discriminating signal peptides from transmembrane regions. In *Nature Methods* (Vol. 8, Issue 10, pp. 785–786). https://doi.org/10.1038/nmeth.1701

Pokhrel, A. (2018). Transcriptomic analysis to identify candidate genes conferring gossypol tolerance in *Fusarium oxysporum* f. sp. *vasinfectum* Title.

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Rambaut, A. (2012). FigTree *v1.4*. http://ci.nii.ac.jp/naid/10030433668/en/

Saier, M. H., Reddy, V. S., Tsu, B. v., Ahmed, M. S., Li, C., & Moreno-Hagelsieb, G. (2016). The Transporter Classification Database (TCDB): Recent advances. *Nucleic Acids Research*, *44*(D1), D372–D379. https://doi.org/10.1093/nar/gkv1103

Seo, S., Pokhrel, A., & Coleman, J. J. (2020). The genome sequence of five genotypes of *Fusarium oxysporum* f. sp. *vasinfectum*: A resource for studies on Fusarium wilt of cotton. *Molecular Plant-Microbe Interactions*, *33*(2), 138–140. https://doi.org/10.1094/MPMI-07-19-0197-A

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. v., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Sperschneider, J., Dodds, P. N., Gardiner, D. M., Singh, K. B., & Taylor, J. M. (2018). Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. *Molecular Plant Pathology*, *19*(9), 2094–2110. https://doi.org/10.1111/mpp.12682

Sperschneider, J., Dodds, P. N., Singh, K. B., & Taylor, J. M. (2018). ApoplastP: prediction of effectors and plant proteins in the apoplast using machine learning. *New Phytologist*, *217*(4), 1764–1778. https://doi.org/10.1111/nph.14946

Summerell, B. A. (2019). Annual Review of Phytopathology Resolving *Fusarium*: Current Status of the Genus. https://doi.org/10.1146/annurev-phyto-082718

Thompson, J. D., Gibson, Toby. J., & Higgins, D. G. (2003). Multiple Sequence Alignment Using ClustalW and ClustalX. *Current Protocols in Bioinformatics*, *00*(1), 2.3.1-2.3.22. https://doi.org/10.1002/0471250953.bi0203s00

Urban, M., Cuzick, A., Seager, J., Wood, V., Rutherford, K., Venkatesh, S. Y., de Silva, N., Martinez, M. C., Pedro, H., Yates, A. D., Hassani-Pak, K., & Hammond-Kosack, K. E. (2020). PHI-base: The pathogen-host interactions database. *Nucleic Acids Research*, *48*(D1), D613–D620. https://doi.org/10.1093/nar/gkz904

van Dam, P., Fokkens, L., Ayukawa, Y., van der Gragt, M., ter Horst, A., Brankovics, B., Houterman, P. M., Arie, T., & Rep, M. (2017). A mobile pathogenicity chromosome in *Fusarium oxysporum* for infection of multiple cucurbit species. *Scientific Reports*, *7*(1), 1–15. https://doi.org/10.1038/s41598-017-07995-y

Vlaardingerbroek, I., Beerens, B., Rose, L., Fokkens, L., Cornelissen, B. J. C., & Rep, M. (2016). Exchange of core chromosomes and horizontal transfer of lineage-specific chromosomes in *Fusarium oxysporum*. *Environmental Microbiology*, *18*(11), 3702–3713. https://doi.org/10.1111/1462-2920.13281

Wang, Q., Cobine, P. A., & Coleman, J. J. (2018). Efficient genome editing in *Fusarium oxysporum* based on CRISPR/Cas9 ribonucleoprotein complexes. *Fungal Genetics and Biology*, *117*, 21–29. https://doi.org/10.1016/j.fgb.2018.05.003

Williams, A. H., Sharma, M., Thatcher, L. F., Azam, S., Hane, J. K., Sperschneider, J., Kidd, B. N., Anderson, J. P., Ghosh, R., Garg, G., Lichtenzveig, J., Corby Kistler, H., Shea, T., Young, S., Buck, S.-A. G., Kamphuis, L. G., Saxena, R., Pande, S., Ma, L.-J., … Singh, K. B. (2016). Comparative genomics and prediction of conditionally dispensable sequences in legume-infecting *Fusarium oxysporum formae speciales* facilitates identification of candidate effectors. https://doi.org/10.1186/s12864-016-2486-8

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., & Xu, Y. (2012). dbCAN: A web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research*, *40*(W1), W445–W451. https://doi.org/10.1093/nar/gks479

Zhang, Y., Yang, H., Turra, D., Zhou, S., Ayhan, D. H., DeIulio, G. A., Guo, L., Broz, K., Wiederhold, N., Coleman, J. J., Donnell, K. O., Youngster, I., McAdam, A. J., Savinov, S., Shea, T., Young, S., Zeng, Q., Rep, M., Pearlman, E., … Ma, L. J. (2020). The genome of opportunistic fungal pathogen *Fusarium oxysporum* carries a unique set of lineage-specific chromosomes. *Communications Biology*, *3*(1), 1–12. https://doi.org/10.1038/s42003-020-0770-2

Table 1. Summary of genome features of five *Fusarium oxysporum* f. sp. *vasinfectum* isolates.

Genotype	**Race 1**	**Race 4**	**LA108**	**LA127/140**	**MDS-12**
Isolate	**TF1**	**89-1A**	**LA3B**	**14-004**	**LA 1E**
Coverage	163×	178×	238×	179×	184×
# of scaffolds	17	94	17	18	26
Genome size (bp)	50,043,501	63,323,888	50,437,808	51,686,574	53,532,301
GC content	47.40%	47.65%	47.48%	47.35%	47.44%
N50 (bp)	4,223,171	3,569,865	4,733,764	4,700,795	3,332,013
L50	5	7	5	5	7
Longest contig (bp)	6,421,299	6,635,028	6,382,753	6,464,399	5,403,155
Contigs > 25 kb	17	86	17	17	25
BUSCO completeness	98.80%	98.70%	98.80%	98.50%	98.20%
# of predicted genes	17,975	20,222	18,132	18,195	19,054
Repetitive sequence (bp)	3,125,885	11,679,362	3,585,278	3,914,995	4,607,798
Repetitive sequence	6.25%	18.44%	7.11%	7.57%	8.61%
NCBI accession	VINL00000000	VINP00000000	VINM00000000	VINO00000000	VINN00000000
BioSample	SAMN12175309	SAMN12175305	SAMN12175308	SAMN12175306	SAMN12175307
# of scaffolds containing telomeric repeats[a]					
on both ends	10	0	7	2	6
on one end	5	0	8	12	11
not observed	2	94	2	4	9

[a]: The number of scaffolds having telomeric repeats (CCCTAA) are counted.

Table 2. Information of FOSC genomes analyzed in this study.

f. sp.	Strain or NRRL #	Race	Host	Size (Mb)	GC%	Scaffolds	Accession	Name
Fo47	Fo47	-	Soil	49.6646	47.7	124	GCA_000271705.2	Fo47
Fo5176	Fo5176	-	*Brassica* sp.	54.7676	47.8	7858	GCA_000222805.1	Fo5176
cepae	Fus2	-	*Allium* sp.	53.4258	47.72	34	GCA_003615085.1	FoceFus2
conglutinans	54008	2	*Brassica* sp.	53.5754	47.7	2552	GCA_000260215.2	Foco54008
cubense	N2	1	*Musa* sp.	47.6574	48	1341	GCA_000350345.1	Focu1
	54006	TR4		46.5538	47.5	418	GCA_000260195.2	FocuTR4
	-	TR4		48.5884	47.5	15	GCA_007994515.1	FocuTF4UK
lycopersici	4287	-	*Solanum* sp.	61.3869	48.43	114	GCA_000149955.2	Fol4287
	MN25	2		48.6374	47.7	388	GCA_000259975.2	FolMN25
melonis	26406	-	*Cucumis* sp.	54.0343	47.5	1146	GCA_000260495.2	Fom26406
NRRL 32931	NRRL 32931	-	Clinical isolate	47.9063	47.6	168	GCA_000271745.2	FoNRRL32931
pisi	HDV247	-	*Pisum* sp.	55.1882	47.6	472	GCA_000260075.2	FopHDV247
raphani	54005	3	*Raphanus* sp.	53.4994	47.8	1218	GCA_000260235.2	For54005
radicis-cucumerinum	Forc016	-	Cucurbit sp.	52.9083	47.68	33	GCA_001702695.2	Forc016
radicis-lycopersici	26381	-	*Solanum* sp.	49.3593	47.6	418	GCA_000260155.3	Forl26381
vasinfectum	TF1	1	*Gossypium* sp.	50.0435	47.4	17	GCA_009602505.1	FovTF1
	89-1A	4		63.3239	47.65	94	GCA_009602635.1	Fov891A
	25433	7		52.9144	47.7	985	GCA_000260175.2	Fov25433
	LA3B	LA108		50.4378	47.48	17	GCA_009602545.1	FovLA3B
	10-004	LA127/140		51.6366	47.35	18	GCA_009602625.1	Fov10004
	LA 1E	MDS12		53.5323	47.44	26	GCA_009602615.1	FovMDS12

Table 3. Mitochondrial genome assemblies of each *Fov* isolate.

Fusarium species	Strain name	Race	Scaffold name	Size (bp)	Gene number	ORFs	GC (%)
F. graminearum	PH-1		-	95,676	15	-	31.8
F. oxysporum	F11		-	34,476	15	-	31.0
	VPR1[a]		-	33,396	-	-	-
F. solani	MPVI		-	62,978	15	-	28.9
F. verticillioides	7600		-	53,753	15	-	32.6
F. oxysporum f. sp. *vasinfectum*[*]	TF1	race 1	16	74,123	21	170	32.17
	89-1A	race 4	48	71,655	20	177	31.54
	LA3B	LA108	16	72,249	19	172	31.99
	14-004	LA127/140	17	78,579	22	171	32.61
	MDS-12	LA 1E	22	65,853	23	172	32.59

[a]: incompletely sequenced mtGenome lacking detailed information than other *Fusarium* species.

[*]: predicted mitochondrial genomes in *Fov* isolates.

Table 4. Statistical analysis of the genomes and protein-coding genes among the five *F. oxysporum* f. sp. *vasinfectum* genomes.

Race	**1**				**4**				**LA108**				**LA127/140**				**MDS-12**			
Isolate	**TF1**				**89-1A**				**LA3B**				**14-004**				**LA 1E**			
		CCs	**ACCs**	**mt**		**CCs**	**ACCs**	**mt**		**CCs**	**ACCs**	**mt**		**CCs**	**ACCs**	**mt**		**CCs**	**ACCs**	**mt**
Scaffolds	17	12	4	1	94	14	79	1	17	11	5	1	18	11	6	1	26	14	11	1
Genome size (Mb)	50.04	45.86	4.11	74,124bp	63.32	48.73	14.52	71,656bp	50.44	45.51	4.85	72,250bp	51.69	46.67	4.94	78,580bp	53.53	47.37	6.09	65,854bp
Coding genes	17,975	16,978	976	21	20,202	17,500	2,702	20	18,132	16,891	1,222	19	18,195	17,137	1,036	22	19,054	17,544	1,487	23
GC content (Mb)	23.72	21.76	1.94	23,849bp	30.18	23.23	6.92	22,598bp	23.95	21.62	2.30	23,113bp	24.48	22.12	2.33	25,626bp	25.40	22.54	2.83	21,461bp
GC (%)	47.40	47.44	47.24	32.17	47.65	47.68	47.66	31.54	47.48	47.51	47.43	31.99	47.35	47.39	47.25	32.61	47.44	47.58	46.52	32.59
Repeats (Mb)	3.16	2.10	1.05	2028bp	12.25	4.57	7.64	1950bp	3.50	2.31	1.18	2977bp	3.85	2.62	1.22	2121bp	4.69	2.90	1.80	1421bp
Repeats (%)	6.31	4.59	25.62		19.29	9.37	52.63		6.93	5.08	24.33		7.44	5.62	24.73		8.77	6.12	29.48	
SSPs	1,421	1,367	54	-	1,539	1,361	178	-	1,416	1,333	83	-	1,409	1,342	67	-	1,430	1,347	83	-
SSPs_300	577	548	29	-	687	565	122	-	597	553	44	-	600	563	37	-	603	565	38	-
EffectorP2	2,305	2,151	154	-	2,716	2,257	459	-	2,329	2,138	191	-	2,352	2,200	152	-	2,526	2,311	215	-
ApoplastP	746	726	20	-	815	718	97	-	733	694	39	-	722	695	27	-	737	706	31	-
CAZymes	306	298	8	-	320	298	22	-	310	296	14	-	310	299	11	-	311	294	17	-
Secondary metabolites	41	39	2	-	40	38	2	-	39	38	1	-	46	46	-	-	40	40	-	-
PHI	7,347	7,032	313	2	7,898	7,139	758	1	7,363	6,969	393	1	7,512	7,173	338	1	7,646	7,161	485	-
Transporters	4,576	4,404	170	2	4,762	4,399	362	1	4,569	4,347	221	1	4,667	4,480	186	1	4,694	4,443	251	-

Table 5. Number of homologs in *Fov* genomes to genes of interest in *Verticillium* that is virulent on cotton.

Vd gene	Accession number	Protein length (aa)	FovTF1	Fov891A	FovLA3B	Fov14-004	FovMDS12
VEDA_05193	ATD50206.1	317	1	2	1	1	1
VEDA_05194	ATD50207.1	313	-	-	-	-	-
VEDA_05195	ATD50211.1	292	1	1	-	1	-
VEDA_05196	ATD50212.1	512	1	2	1	1	-
VEDA_05197	ATD50208.1	346	1	2	1	1	-
VEDA_05198	ATD50209.1	333	2	2	1	2	1
VEDA_05199	ATD50210.1	395	2	2	2	2	-

Table 6. Detailed information of homologs of Vd991 in the five *Fov* genomes.

Fov **(genotype)**	***Fov*** **(isolate)**	***Vd*** **gene**	**Gene**	**Location**	**Length (bp)**	**Start**	**Stop**	**% identity**	**E-value**
Race 1	TF1	*VEDA_05193*	*F11225*	ACC_01	2629525	2272165	2273280	91.105	0
		VEDA_05196	*F11244*			2276655	2278631	94.912	0
		VEDA_05197	*F11259*			2279539	2280576	93.333	0
		VEDA_05198	*F11269*			2281262	2282263	95.796	0
		VEDA_05199	*F11276*			2282988	2283729	89.041	9.37E-145
			F11281			2283751	2284257	96.396	3.19E-75
		VEDA_05195	*F00138*	ACC_02	1305674	43734	45867	79.672	2.60E-167
		VEDA_05198	*F00176*			52392	53414	83.183	0
Race 4	891A	*VEDA_05196*	*F01624*	ACC_02	1173354	161666	164893	99.217	0
		VEDA_05197	*F01610*			159550	161525	99.133	0
		VEDA_05198	*F01597*			158461	159549	98.799	0
		VEDA_05199	*F01580*			156488	157821	97.468	0
		VEDA_05196	*F00108*	ACC_60	30248	6441	9668	99.217	0
		VEDA_05197	*F00081*			4325	6300	99.133	0
		VEDA_05198	*F00056*			3236	4324	98.799	0
		VEDA_05199	*F00034*			1332	2596	96.216	0
		VEDA_05193	*F00372*	ACC_53	36136	27056	28348	99.73	0
			F00082	ACC_02	1173354	4930	6259	99.73	0
		VEDA_05195	*F00208*	ACC_02	1173354	14936	17053	88	0
LA108	LA3B	*VEDA_05193*	*F01798*	ACC_03	585529	378441	379556	91.105	0
		VEDA_05196	*F01765*			373090	375066	94.912	0
		VEDA_05197	*F01749*			371145	372182	93.333	0
		VEDA_05198	*F01740*			369747	370460	93.119	1.88E-151
		VEDA_05199	*F01728*			367995	368736	89.041	9.39E-145
			F01725			367467	367890	96.396	3.20E-75

LA127/140	14004	*VEDA_05193*	*F06269*	ACC_01	1529817	1276324	1277438	87.871	0
		VEDA_05196	*F06246*			1270973	1272949	94.912	0
		VEDA_05197	*F06234*			1269028	1270065	93.333	0
		VEDA_05198	*F06226*			1267341	1268342	95.796	0
		VEDA_05199	*F06222*			1265875	1266616	89.041	9.28E-145
			F06218			1265347	1265856	96.396	3.16E-75
		VEDA_05195	*F00133*	ACC_03	1306839	44025	46158	79.672	2.58E-167
		VEDA_05198	*F00176*			52685	53479	82.576	2.14E-160
MDS12	LA1E	*VEDA_05193*	*F00269*	ACC_03	929984	52883	55016	79.672	2.68E-167
		VEDA_05198	*F00314*			61543	62565	83.183	0

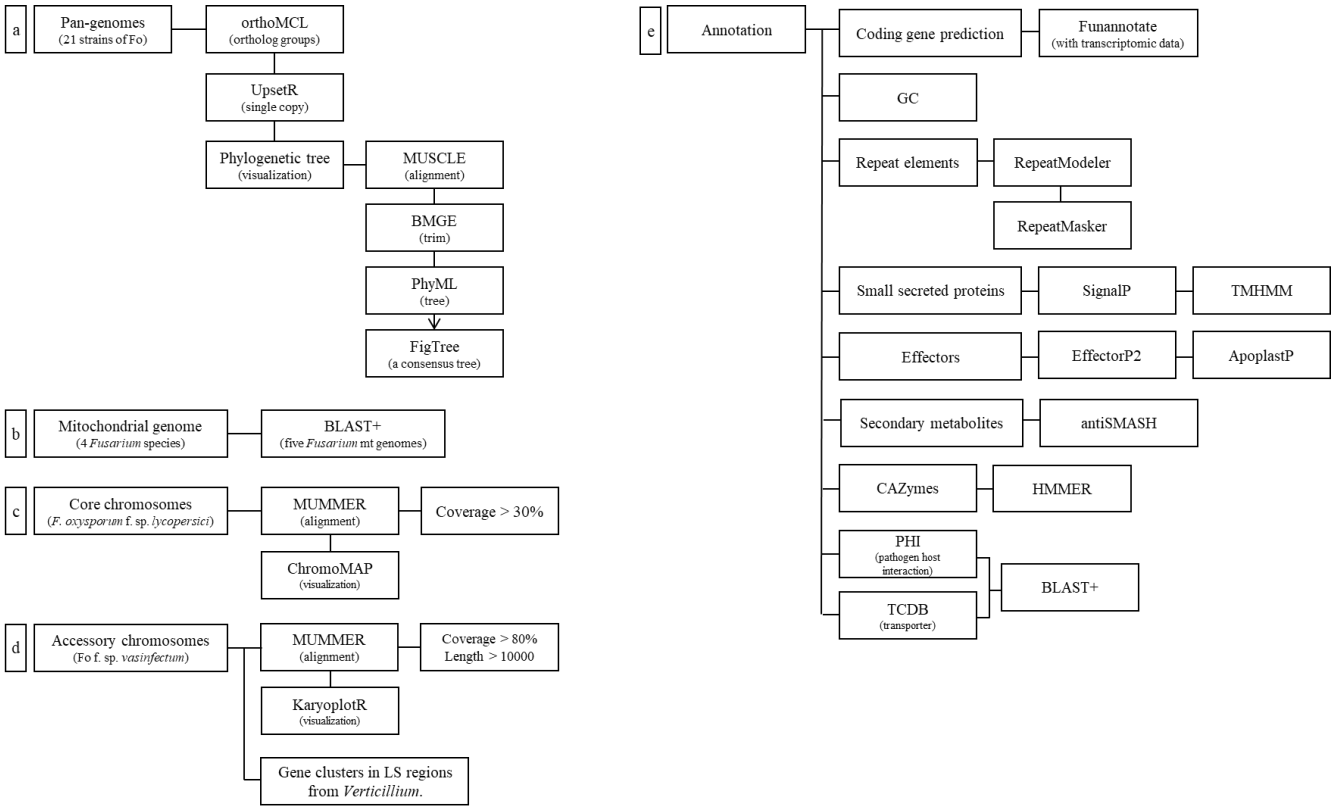


Figure 1. Bioinformatic pipeline for mining the structure and protein-coding genes in five *F. oxysporusm* f. sp. *vasinfectum* (*Fov*) genomes. To identify phylogenetic relationships between the five *Fov* and within the other members of the FOSC, pan-genome analysis was performed defining orthologs via orthoMCL and drawing the maximum likelihood tree using FigTree after aligning (Muscle) and trimming (BMGE) proteins sequences of the 21 FOSC genomes (a). The genome structures of five *Fov* isolates was assessed by assigning mitochondrial genomes with high similarities from other five known *Fusarium* species by BlastP (b). Core and accessory chromosomes/scaffolds were divided using the sequence similarity with *F. oxysporum* f. sp. *lycopersici* 4278 using MUMmer (c and d). All genome sequences and proteomes were annotated to identify basic statistics of genes and proteins, and to find putative virulence factors with various programs (e).

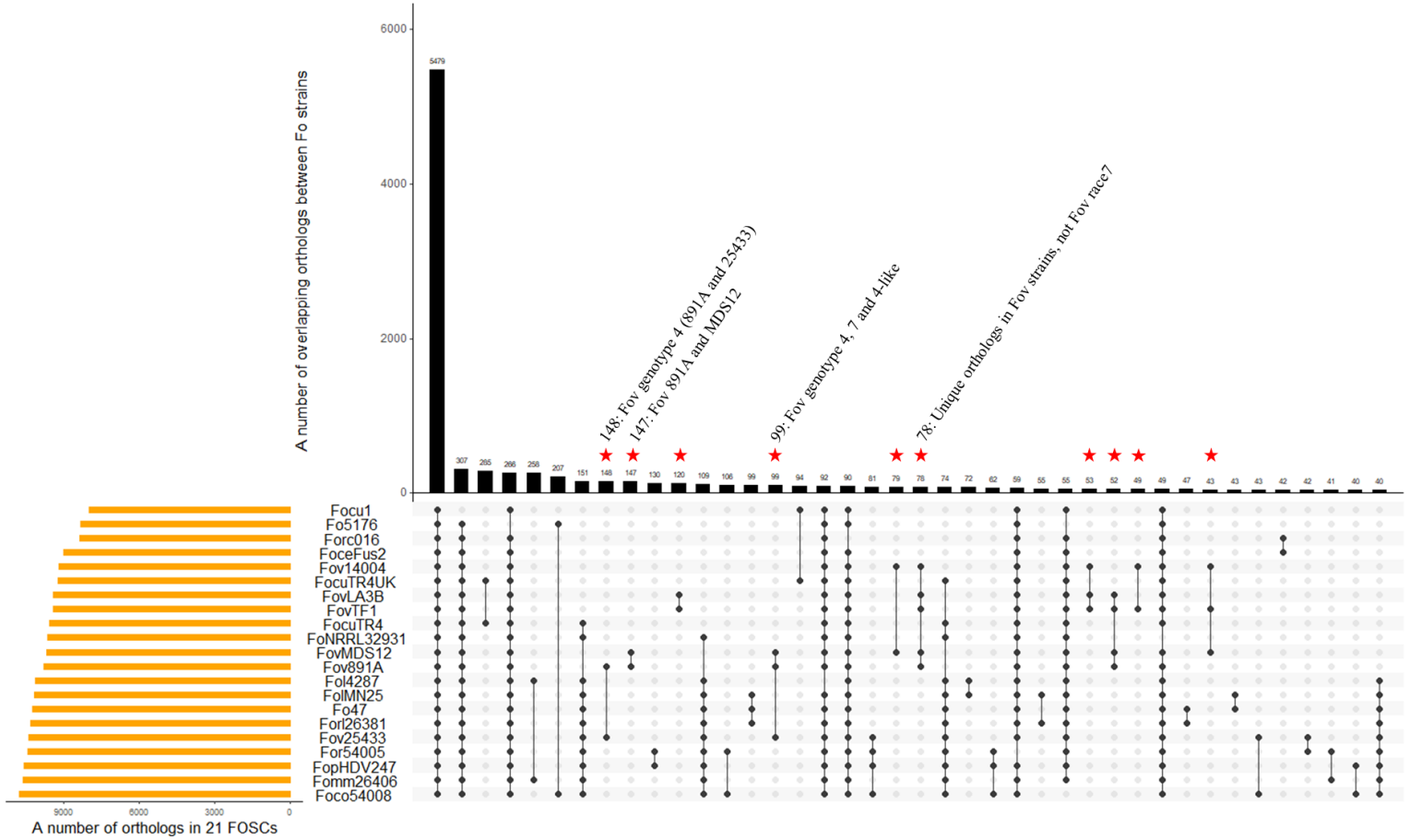


Figure 2. Distribution of orthologues from 21 FOSC isolates. The sets of single-copy orthologs were predicted and adjusted by OrthoMCL in all 21 FOSC genomes. A total of 17,203 orthologs were identified. Each bar and number above indicate how many orthologs are present within each set of FOSC genomes. Among the 17,203 orthologs, 5,479 ortholog groups were found in all 21 FOSC

genomes. The red stars indicate orthologs that are shared between cotton isolates. The graph was drawn with the UpSet R language package.

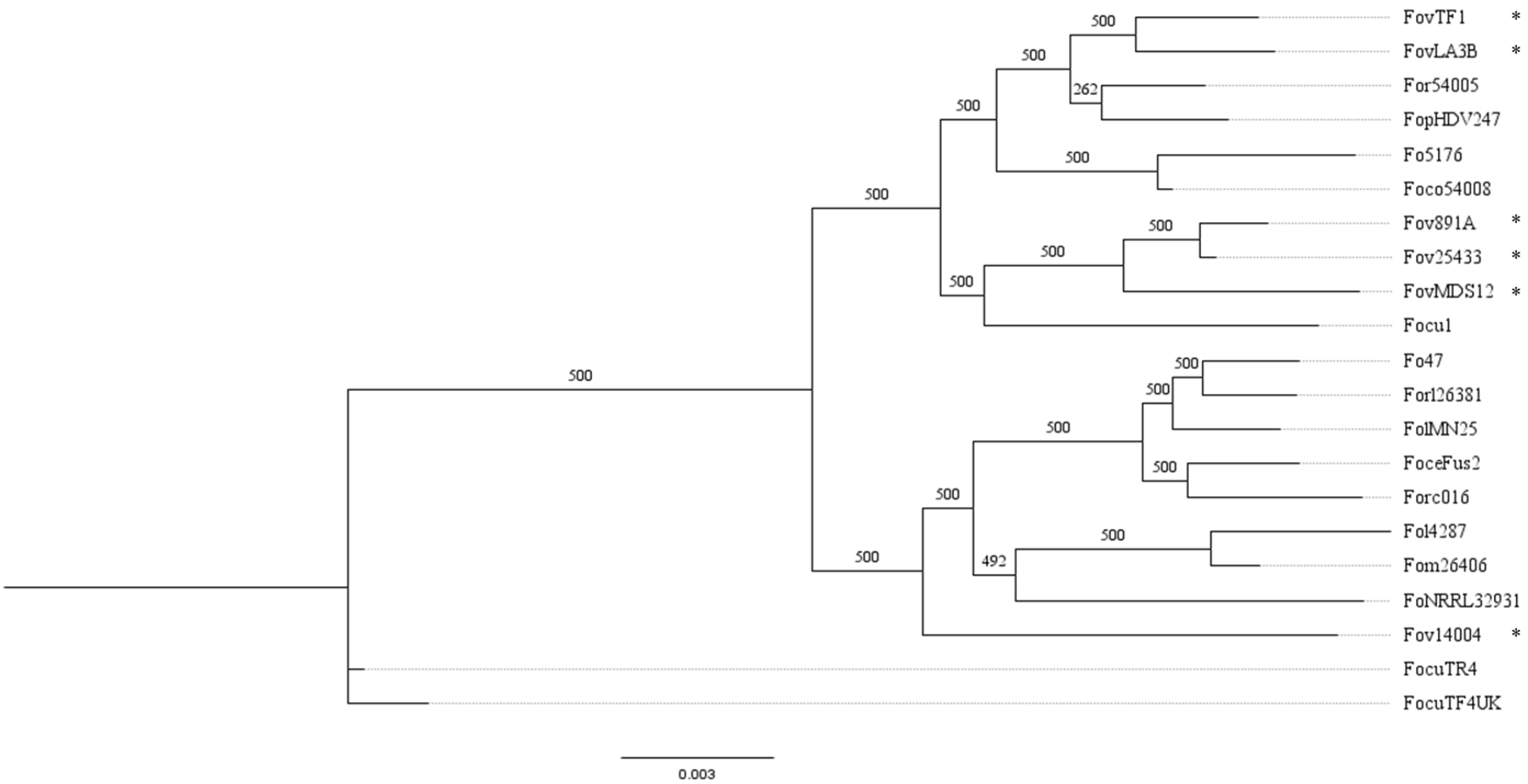


Figure 3. Phylogenetic relationships of the five *Fov* genomes and other *F. oxsyporum* species. Maximum-likelihood phylogeny analysis was conducted to understand the evolutionary relationships of 21 FOSC. The phylogenetic tree is based on 5,479 concatenated single-copy orthologues that are shared among all analyzed species. The robustness of the tree was assessed using 500 bootstrap replicates. *F. oxysporum* f. sp. *vasinfectum* strains were marked by asterisks.

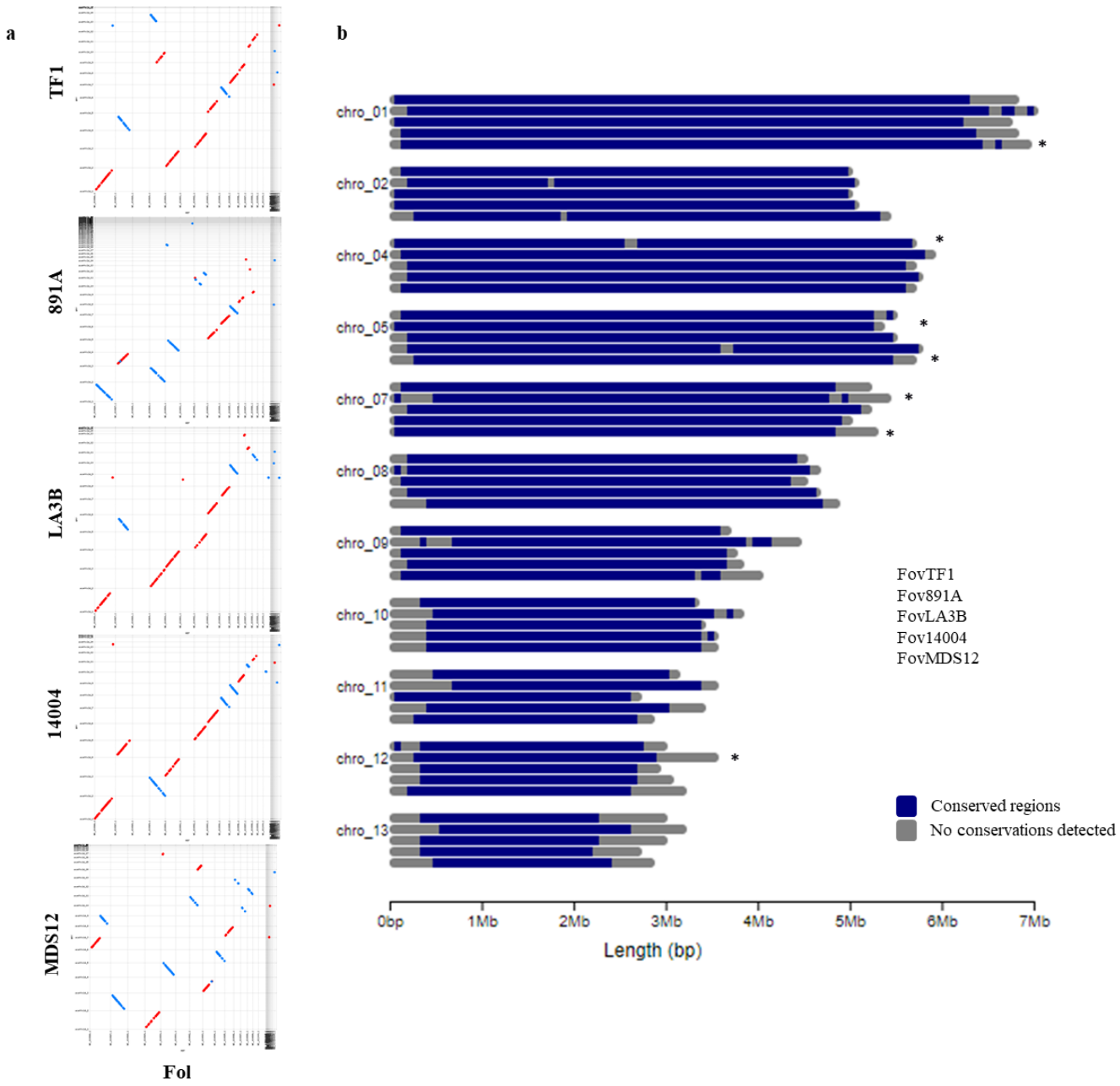


Figure 4. Genome comparison of each *Fov* and *Fol* 4287. a) Genome-level alignment of the five *Fov* genomes with *Fol* 4287 using MUMmer. Red lines indicate high sequence similarities between two scaffolds in the same direction, while blue lines show two sequences align together in the opposite direction. Thus, all blue lines are reversed to visualize conserved regions between each *Fov* and *Fol* 4287 chromosome. b) Syntenic regions on the core chromosomes of the five *Fov* isolates. The eleven core chromosomes of *Fol* 4287 match to scaffolds in each individual *Fov*

genome. Conserved regions are colored dark blue and the remaining unmatched regions are shown in gray. The graph was created by the R language package, chromoMAP. The symbol (*) indicates two *Fov* scaffolds cover one *Fol* chromosome.

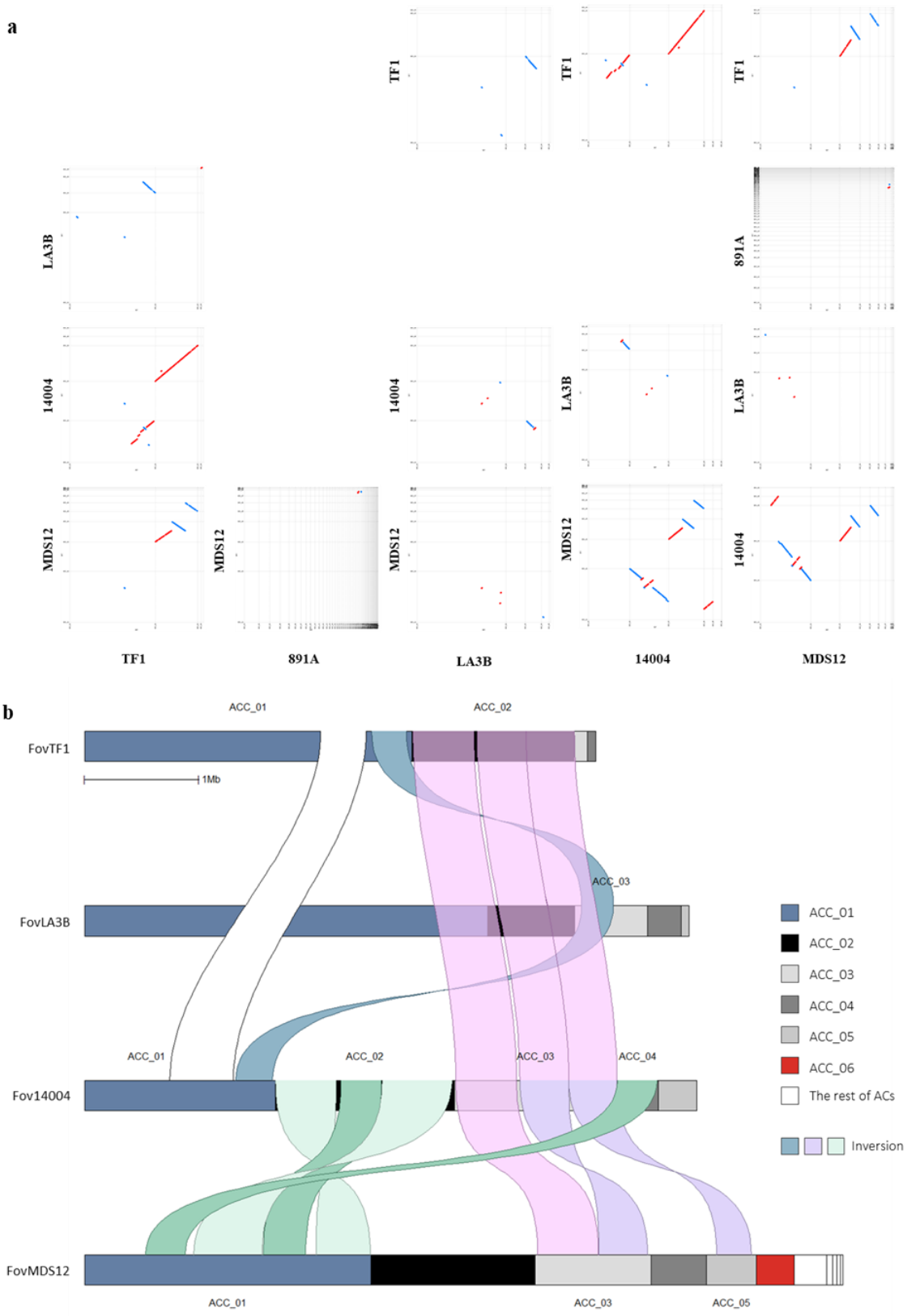


Figure 5. Synteny relationships in the accessory (lineage-specific) regions among the five *Fov*

strains. a) MUMmer plots aligning the accessory scaffolds between the five *Fov* isolates. Red and blue lines illustrate the same and opposite directions between matched scaffolds. b) Anchoring syntenic blocks between four *Fov* accessory genomes. Physical location of the syntenic blocks were visualized, and bar plots were created by the R language package, karyoplotR. The accessory chromosomes were similarly colored according to their size.

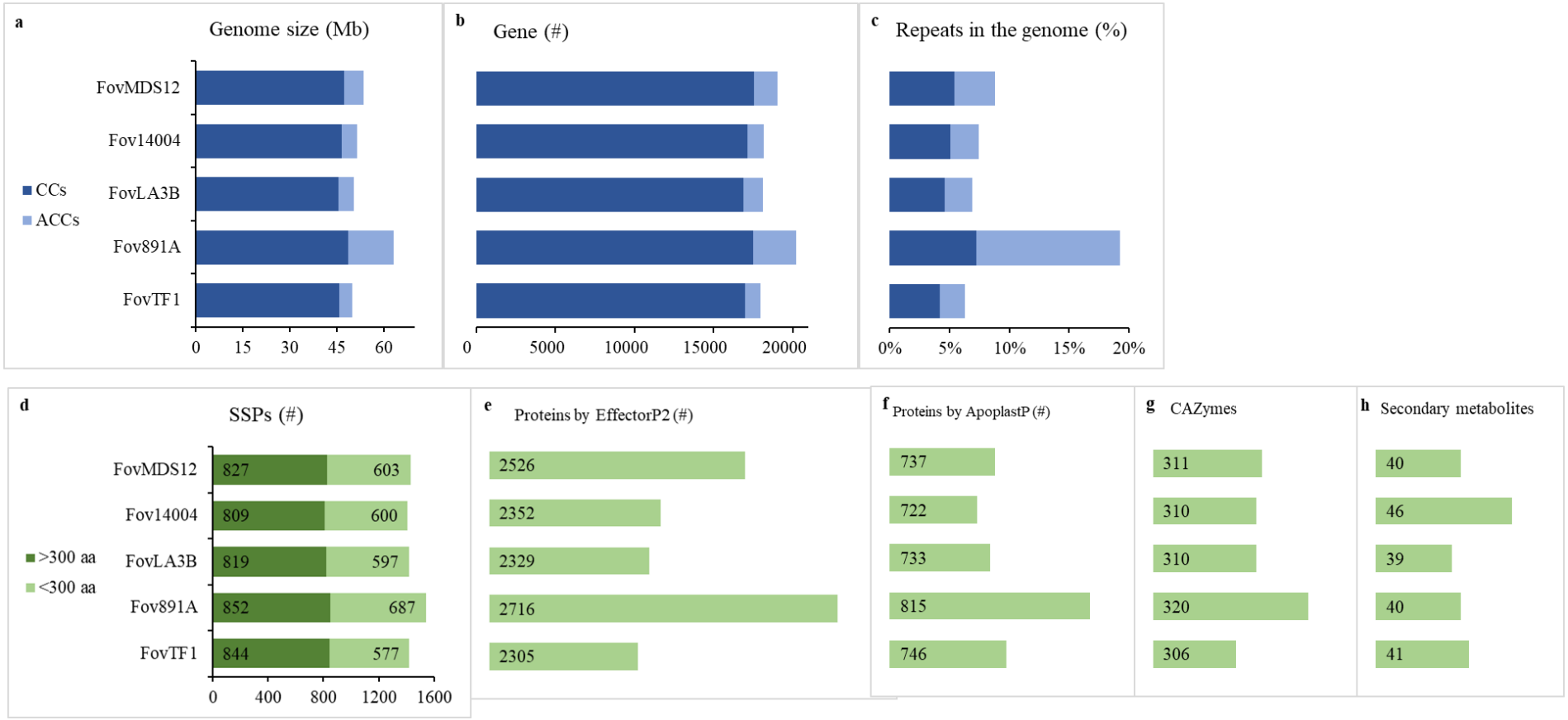


Figure 6. Visualization of genomic features and protein-coding gene annotation of five *Fov* isolates. a-c) The results were depicted based on location in the core and accessory genomes with different colors, dark and white blue. a) Genome size. b) Number of predicted protein-coding genes. c) Percentage of repeats. d) Number of predicted small secreted proteins, size over and under 300 amino acids. Number of predicted proteins by EffectorP2 (e) and ApoplastP (f). g) Number of predicted carbohydrate-active enzymes (CAZymes). h) Number of predicted secondary metabolites by antiSMASH. All bar plots were manually drawn.

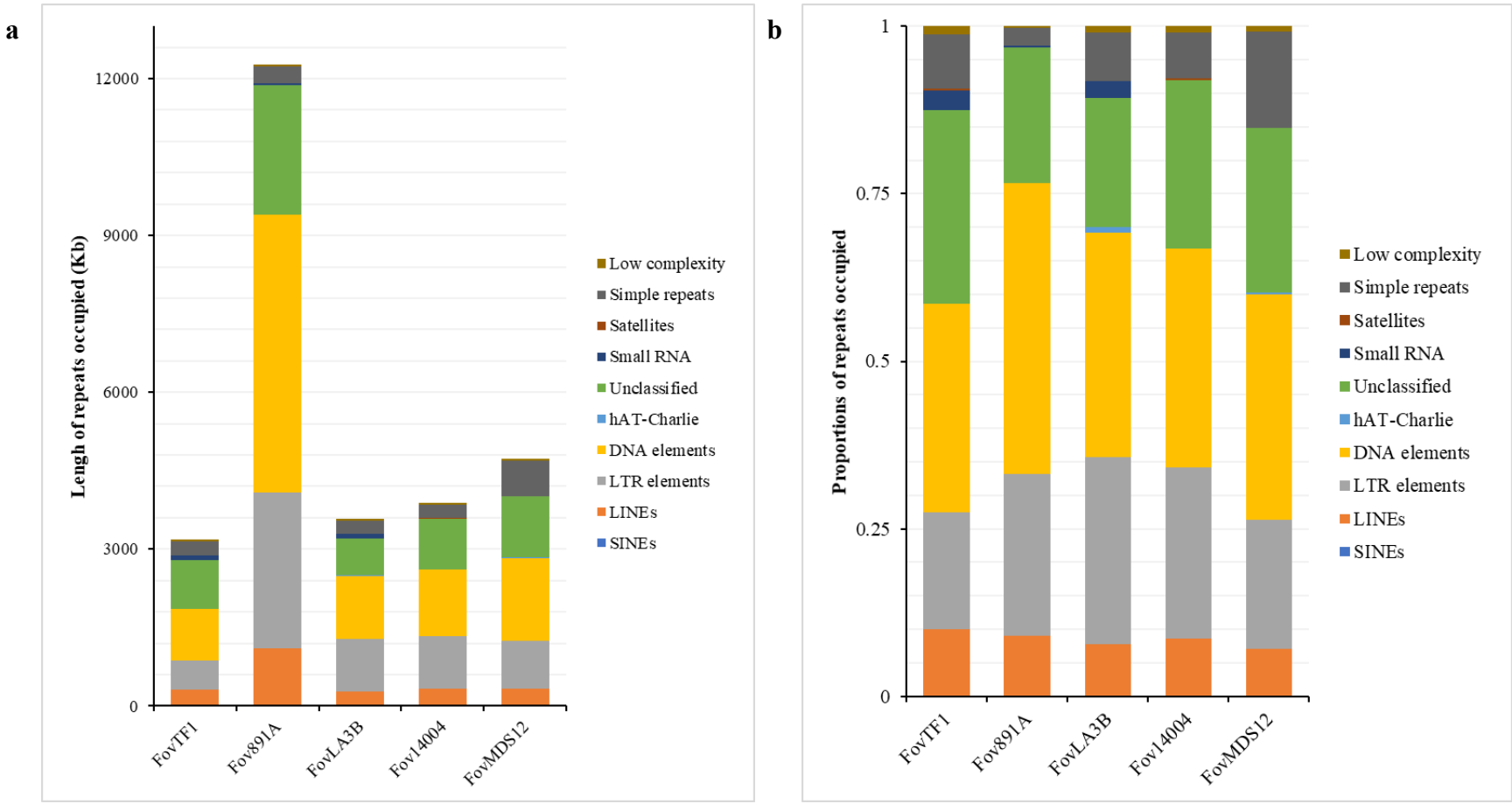
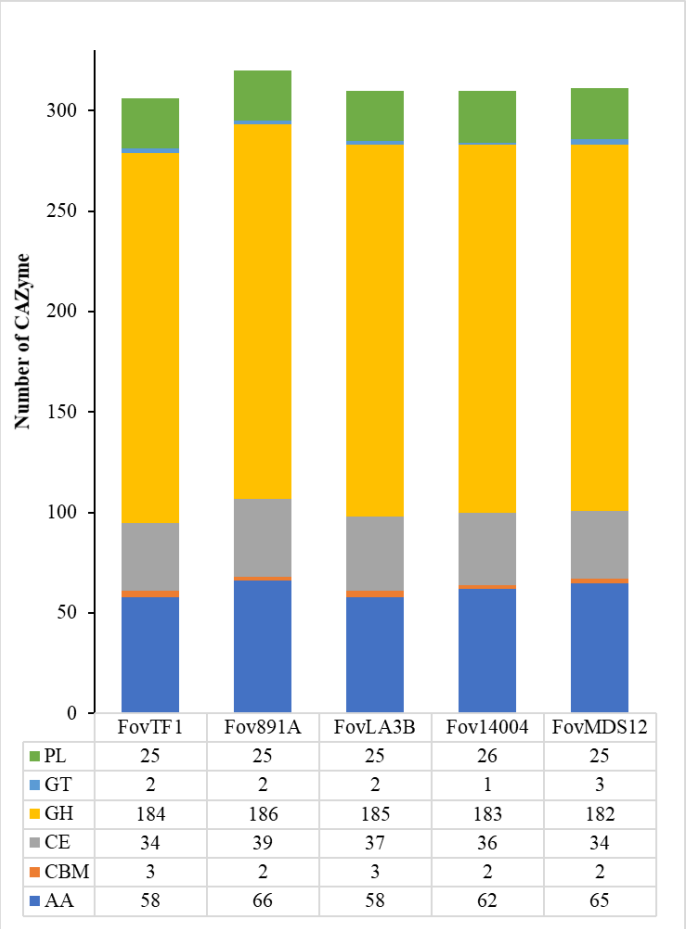


Figure 7. Repeat content in the five *Fov* genomes. a) The length of repeats in each genome. Individual colors indicate the types of repetitive elements in the five *Fov* genomes. b) The proportion of each type of repetitive elements occupied in the total repeats.



	FovTF1	Fov891A	FovLA3B	Fov14004	FovMDS12
PL	25	25	25	26	25
GT	2	2	2	1	3
GH	184	186	185	183	182
CE	34	39	37	36	34
CBM	3	2	3	2	2
AA	58	66	58	62	65

Figure 8. Relative distribution of CAZymes in the *Fov* genomes. Each color indicates the functional category of CAZymes in the five *Fov* genomes, as divided into representing auxiliary activities (AA), carbohydrate-binding modules (CBM), carbohydrate esterases (CE), glycoside hydrolases (GH), glycosyltransferases (GT), and polysaccharide lyases (PL).

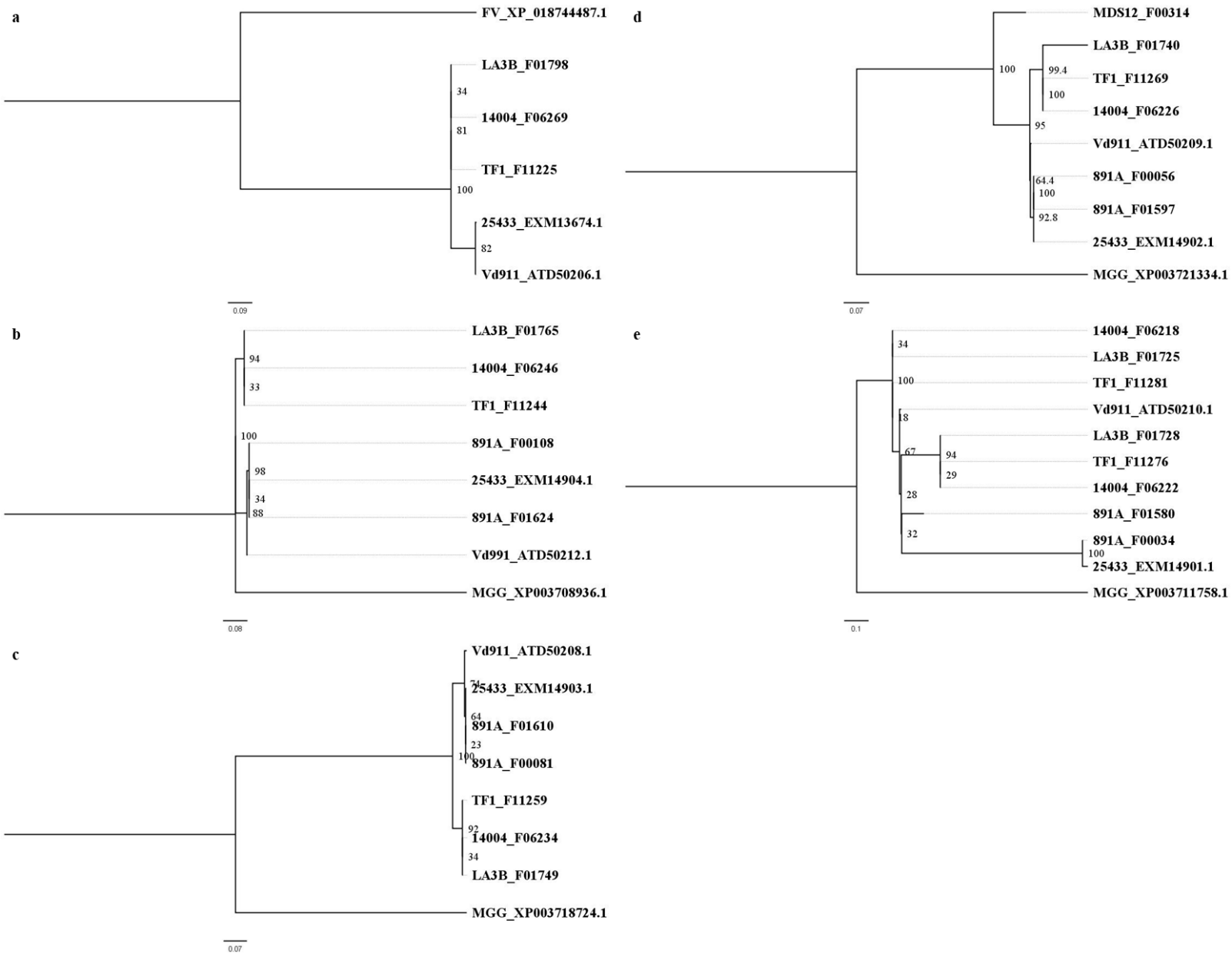

a
FV_XP_018744487.1
LA3B_F01798
34
14004_F06269
81
TF1_F11225
100
25433_EXM13674.1
82
Vd911_ATD50206.1
0.09
b
LA3B_F01765
94
14004_F06246
33
TF1_F11244
100
891A_F00108
98
25433_EXM14904.1
34
88
891A_F01624
Vd991_ATD50212.1
MGG_XP003708936.1
0.08
c
Vd911_ATD50208.1
25433_EXM14903.1
64
891A_F01610
23
100
891A_F00081
TF1_F11259
92
14004_F06234
34
LA3B_F01749
MGG_XP003718724.1
0.07
d
MDS12_F00314
LA3B_F01740
100
99.4
TF1_F11269
100
14004_F06226
95
Vd911_ATD50209.1
64.4
891A_F00056
100
891A_F01597
92.8
25433_EXM14902.1
MGG_XP003721334.1
0.07
e
14004_F06218
34
LA3B_F01725
100
TF1_F11281
Vd911_ATD50210.1
18
LA3B_F01728
67
94
TF1_F11276
29
14004_F06222
28
891A_F01580
32
891A_F00034
100
25433_EXM14901.1
MGG_XP003711758.1
0.1

Figure 9. Phylogenetic relationships between proteins encoding homologous genes from *Fov* genomes and lineage-specific (LS) genes from *V. dahlia* Vd991. Phylogenetic analysis of VEDA_05193 (a), VEDA_05196 (b), VEDA_05197 (c), VEDA_05198 (d), and VEDA_05199 (e). Phylogenetic analyses were performed on amino acid sequences of proteins from *Fov* homologs and lineage-specific genes in Vd991. Five trees were generated using the maximum likelihood analysis with 1000 bootstraps if there were meaningful hits by BlastP. One protein for the outgroup was from *Fusarium verticillioides* (a), and the others were from *Magnaporthe oryzae* (b-e) following the previous study (Chen et al., 2017). *V. dahliae* proteins are marked with an asterisk (*).

Chapter 3.

Identification and partial characterization of putative small secreted proteins in five *Fusarium oxysporum* f. sp. *vasinfectum* (*Fov*) isolates uncovers several involved in virulence on cotton

ABSTRACT

Fusarium wilt of cotton is a serious problem in cotton production fields. *Fusarium oxysporum* f. sp. *vasinfectum* race 4 (*Fov*) is the causal agent, affecting all genetically diverse domesticated cotton cultivars, including *Gossypium hirsutum* (Upland) and *G. barbadense* (Pima) worldwide. However, little is known about what confers pathogenicity to cotton in *Fov* genetically. To uncover putative secreted proteins (fungal effectors) which could be contributing to virulence on cotton plants in the United States, comparative genomics was applied to identify potential secreted protein sets from six *Fov* genomes (cotton) and ten *F. oxysporum* genomes publicly available in the *F. oxysporum* species complex (FOSC) group. Among the *Fov* isolates, the highly virulent race 4 is of particular concern since no highly resistant cotton cultivars are available and it is able to infect plants without the presence of nematodes. To narrow down unique putative effector genes specific for *Fov* especially race 4, an *in silico* analysis was universally applied to identify proteins containing a signal peptide and exclude those with transmembrane domains. After filtering all the secretomes to those under 300 amino acids in length, the refined secretome of each individual FOSC was analyzed using BLASTP against the set of 687 SSPs from *Fov* race 4 (89-1A) to cluster the SSPs. As a result, a total of 78 SSPs were shared between at least three isolates from cotton but were absent in the other FOSC genomes not pathogenic on cotton. Four SSP genes, *F02527*, *F02627*, *F07669/07679*, and *F11624* were chosen for gene deletion experiments to evaluate the contribution to fungal pathogenicity on cotton. Three pathogenicity assays, stem, hydroponic, and

pot, were conducted with the Pima cultivar DP744 to assess whether the four successfully generated mutants in *Fov* race 4 had any influence on the virulence to cotton. SSP *F11624* was confirmed to serve as a virulence factor on cotton and the gene expression was up-regulated during infection. Putative small secreted protein profiles were revealed in this study, and in future studies elucidating the function of many putative SSPs could be valuable in identifying novel and isolate-specific effector genes in *Fov* and more importantly could be used to identify any corresponding resistance gene(s) in cotton that could be of use in breeding programs.

3.1 Introduction

Plants have evolved defense mechanisms to impede infections by pathogens such as bacteria and fungi. Plants have two major defense systems, where the first process requires pattern recognition receptors (PRRs) on the plant cell surface to detect pathogen-associated molecular patterns (PAMPs) and mount an immune response called PAMP-triggered immunity (PTI). This is a broad-spectrum form of defense initiating a rapid series of defense reactions in plants. However, pathogens have evolved several mechanisms against the plant PTI by evading recognition, suppressing signaling pathways, and delivering effectors. As a second defense system host plants have products of resistance (*R*) genes that recognize pathogen effectors, resulting in a heightened defense referred as effector-triggered immunity (ETI). ETI usually results in the hypersensitive response (HR), which is the local cell death of plant cells around the infection to stop the spread of the pathogen (Jones & Dangl, 2006).

Many plant pathogens possess a set of effectors that they utilize during host infection (Kamoun, 2009). Identifying effectors and their function is critical in developing crop resistance. As genome sequencing technologies have advanced, the number of bioinformatic tools have continued to increase and become more specialized (de Jonge, 2012). Most effectors do not contain conserved domains or motifs which hampers the identification of effectors in genome sequence data (lo Presti et al., 2015). In phytopathogenic fungi, effectors are usually small secreted proteins (SSPs) with basic properties including having a low molecular weight, secreted from the fungal cell, and can be enriched with cysteine residues (Feldman et al., 2020; Lu & Edwards, 2016).

Wilt diseases are a serious problem in worldwide cotton production areas. The soil-borne plant pathogen, *Fusarium oxypsorum* f. sp. *vasinfectum* (*Fov*) is the causal agent of Fusarium wilt on commercial cotton cultivars, including Pima and Upland varieties in the United States (Davis

et al., 2006). *Fov* spores germinate in response to plant exudates and the germ-tube elongates until it comes onto contact with the plant root. The fungus colonizes the roots by branching hyphae and produces conidia inside the vascular tissue (Hall et al., 2013). Extensive growth throughout the xylem vessels occurs, blocking water and nutrient transport, ultimately leading to wilting and eventual death of the infected cotton plant (Nelson, 1981). Although little is known about the molecular mechanisms of *Fov* on cotton, studies on other isolates of *F. oxysporum* pathogenic on a variety of host plants have identified numerous genes, generally encoding virulence factors such as small secreted proteins or secondary metabolites (Pusztahelyi et al., 2015). Studies with the xylem sap of *F. oxysporum* f. sp. *lycopersici* (*Fol*)-infected tomato plants revealed a set of 14 genes encoding small, cysteine-rich proteins named the *Secreted In Xylem* (*SIX*) genes (Rep, 2005; Rep et al., 2004; Schmidt et al., 2013). Further studies showed that 6 *SIX* proteins are important in colonization, disease symptom progression, and virulence in tomatoes (Gawehns et al., 2014; Houterman et al., 2009; Ma et al., 2015; Rep, 2005). In *Fol*, the 14 *SIX* genes are located on the lineage-specific chromosome 14, and this type of chromosome can be horizontally transferred, conferring pathogenicity to another FOSC isolate (Li et al., 2020; van Dam & Rep, 2017). Further research has uncovered the genome evolves at two-speed, where one consists of the core part required for normal fungal growth and the second a highly dynamic part, named as the accessory or lineage-specific part.

In an effort to uncover putative secreted proteins which could be contributing to virulence on cotton in the United States, comparative genomics was applied to identify potential secreted protein sets in FOSC isolates and to narrow down unique putative effector gene(s) specific for *Fov*. The highly virulent *Fov* race 4 is of particular concern since no highly resistant cotton cultivars are available and it is able to infect plants without the presence of nematodes.

3.2 Materials and methods

3.2.1 Data collection of FOSC proteomes

The proteomes of five *Fov* isolates were obtained from recently sequenced and annotated with transcriptomic data (Seo et al., 2020). The proteomes of eleven *F. oxysporum* isolates pathogenic on different hosts were obtained from publicly available sources at Genbank and the Broad Institute of Harvard and MIT (http://www.broadinstitute.org; Table 1).

3.2.2 Secretome prediction and comparison

To predict small secreted proteins (SSPs) encoded in 16 FOSC genomes, two different methods were employed (Figure 1). One is a two-step pipeline using SignalP v.4.1 (Petersen et al., 2011) and TMHMM v.2 (Moller et al., 2001) which was universally applied to identify proteins containing a signal peptide and exclude those with transmembrane domains. This set of putative secreted proteins was further filtered by selecting only those whose sequence length was ≤300 amino acids. The other method utilized the effector prediction program EffectorP2 (Sperschneider, Dodds, Gardiner, et al., 2018). To cluster SSPs, the refined secretome of each individual FOSC was analyzed using BLASTP (cutoff of E-value of 0.001 and identity over 95%) against of the SSP set from *Fov* race 4 (89-1A). The distribution of groups showing with a high sequence similarity was visualized via the R package UpSetR (Conway et al., 2017) to depict common SSPs between the 16 FOSC genomes. Among 687 proteins in the refined secretome of *Fov* race 4, only the SSPs common in at least three *Fov* isolates were selected for further analysis.

3.2.3 Annotation of candidate SSPs in Fov race 4

The protein length and location on the chromosomes were analyzed using a custom script. To obtain additional information relevant to effectors, the EffectorP2 program was used to predict proteins that have similarities with known fungal effectors, conserved protein domains were

identified using InterProScan v.4 (Quevillon et al., 2005), and ApoplastP (Sperschneider, Dodds, Singh, et al., 2018) was used to determine if the proteins are localized to the plant apoplast. The sequence for miniature impala (mimp) inverted repeats (IRs), "TT[TA]TTGCNNCCCACTGNN" was searched for 2-3 kb upstream of each SSP candidate (Schmidt et al., 2013). Transcriptomic data generated from the *Fov* race 4 isolate treated with gossypol (80 μg/mL), used during the gene annotation, was search for transcripts corresponding to putative SSPs in 89-1A (Pokhrel, 2018). Other details such as copy number and if the putative SSPs resided in a cluster in the genome were identified (Table 3).

3.2.4 Search of homologs of the *SIX* genes in *Fov*

The coding sequence of the 14 *SIX* genes of *Fol* was obtained from GenBank (CAE55870.1, CAE55868.3, CAJ83999.1, CAJ84000.1, ACN87967.1, ACN69116-69118.1, and AGG54050-54055.1). Homologues of the putative *SIX* genes were identified in the proteome of all the *Fov* isolates using BLASTX with the default settings. The annotation data of the predicted proteins in the five *Fov* isolates was included such as protein information, EffectorP2, ApoplstP, the existence of mimp sequences, and the blast result.

3.2.5 Mutant generation

3.2.5.1 Cas9 protein purification

The plasmid pHis-parallel1-NLSH2BCas9 (Addgene Catalog #112065) was used to purify Cas9 protein for this study. The plasmid was transformed into the *E. coli* Rosetta™ (DE3) strain for high expression. A single colony was inoculated in 10 ml LB liquid medium with 100 μg/ml ampicillin for 14 h at 37℃. The culture was transferred to 1 L of fresh LB liquid medium. When the OD_{600} value of the *E. coli* cell culture reached 0.6, IPTG was added to a final concentration of 0.3 mM and incubated overnight at 25℃, 160 rpm on a rotary shaker. After the protein expression

was checked through SDS-PAGE gel, the culture was centrifuged at 3000 rpm for 15 min. The pellet was suspended in 30 ml lysis buffer (20 mM Tris-HCl, 150 mM NaCl, 10 mM imidazole, 1 mM PMSF, pH 8.0) and lysed in an ultrasonic bath (5 s + 5 s, 25 min). The mixture was centrifugated at 3500 rpm, 4℃ for 2 hr. The supernatants were filtered through a 0.22 μm filter syringe (to remove unnecessary particles). The crude protein purification was further processed using HisPurTM Ni-NTA Chromatography Cartridges (Thermo Fisher Scientific), modifying the manufacturer protocols. The filtered mixture was incubated with resin for 4 h on a rotary shaker at 4℃. The column was washed twice with the buffer (20 mM Tris-HCl, 500 mM NaCl, 25 mM imidazole, pH 8.0), and the protein were eluted with the buffer (20 mM Tris-HCl, 500 mM NaCl, 500 mM imidazole, pH 8.0), and concentrated with Amicon® Ultra Centrifugal Filters to a final concentration of 660 μg/ml. To check the activity of Cas9 protein, *in vitro* cleavage was performed, and the final purified protein was stored at -30℃ until use.

3.2.5.2 sgRNA synthesis

A target-specific oligo to synthesis sgRNA was designed using the CCTop online prediction tool (https://cctop.cos.uni-heidelberg.de/; Stemmer et al., 2015) and synthesized *in vitro* using the commercially available EnGen sgRNA synthesis kit, *S. pyogenes* (New England Biolab) (Table 5). The one or two kb upstream sequences of the interest gene were provided (Figure 7a). At least three potential sites for the sgRNA with a high score of efficiency were chosen and blasted into the genome to remove multiple cut sites unintentionally. All the primers designed in this study were listed in Table 5 and the primers for sgRNA synthesis were named with sgR after the target gene number.

The target-specific oligos were diluted to 1 μM and mixed with other provided components, following the manufacturer instructions. The total 20 μl volume mixture was thoroughly mixed

and incubated at 37℃ for 4 hours. For DNase treatment, the tube was kept in ice and mixed with 2 µl of DNase (provided) adding 30 µl of nuclease-free water (or DEPC water). It was mixed and incubated at 37℃ for 30 min. The synthesized sgRNA were purified using chloroform and precipitated with pure ethanol by centrifugation for 10 min at 13000 g and 4℃. The pellet was suspended in 50 µl of DEPC water, and the concentration was measured by Nanodrop (product).

3.2.5.3 *in vitro* cleavage

To confirm the viability both synthesized sgRNA and purified Cas9 proteins, *in vitro* cleavage was conducted by generating Cas9 protein/sgRNA ribonucleoproteins (RNPs). The template was prepared by amplifying 2-3 kb regions including the cutting site. the primers for amplifying the template were listed in the Table 5 and named with TF (forward) and TR (reverse) after the target gene number.

3.2.5.4 Donor DNA (50bp)

A hygromycin B phosphotransferase selection cassette was used as the selectable market for Cas9-mediated gene disruption for this study. An 1865-bp region of the hygromycin cassette including 1026 bp of *hph* and the *gpdA* promoter and terminator was amplified from the plasmid pCWHyg1 (Wasmann & VanEtten, 1996). To generate the repair templates, the primer set for amplifying the hygromycin cassette had a 50-bp overhang region following the target-specific sequences for sgRNA synthesis. The HYGB forward primer was combined with 50 bp upstream sequences before the cut site by RNPs. The reverse sequences were complimented after adding 50 bp downstream nucleotide after the PAM site. All the primers used in this study were listed in Table 5 and the primers for the franking regions were named with _50F (forward) and _50R (reverse) after the target gene number. The amplified PCR products were purified using E.Z.N.A.® Cycle Pure Kit kit (Omega) and eluted using nuclease-free water (Figure 7b).

3.2.5.5 Protoplasting

The wild-type *Fov* 89-1A isolate was used in this study. To harvest conidia, the strain was cultured in Potato Dextrose Broth (PDB) or minimal medium (M100) media for 5 days at room temperature (25℃) or on agar plates. Conidia cultivated from liquid media were filtered through Nitex and collected by centrifugation at 3000 rpm for 5 min. About 10^7 conidia were inoculated into 100 ml of fresh PDB liquid medium and germinated on a 160-rpm shaker overnight. Fungal germlings were collected on sterile nylon mesh, placed in a 50 ml centrifuge tube, and treated with 3 mg/ml Kitlase (Wako Chemicals) in 0.7 M NaCl for 3-4 h. Fungal protoplasts were filtered through Nitex mesh and collected by centrifugation at 800 rpm for 15 min. Protoplasts were washed twice with SuTC buffer (20% sucrose, 50 mM Tris-HCl, pH 7 and 50 mM $CaCl_2$) and diluted into a concentration of 2×10^7 protoplasts/ml with SuTC buffer.

3.2.5.6 Fungal transformation

During the protoplast generation, the Cas9 protein/sgRNA RNPs were prepared in a total reaction volume of 50 µl (10× Cas9 nuclease buffer, 40 µg Cas9, 8 µg sgRNA and DEPC water). The mixture was incubated at 37℃ for 20 min. Two hundred µl of protoplast suspension, the Cas9 RNP mixture, and 6 µg of the donor DNA were mixed in the same tube and incubated at room temperature for 20 min. One ml of PSTC was added, gently suspended, and incubated again for 20 min at room temperature. Three ml of TB3 (20 % sucrose, 1% glucose, 0.3% yeast extract and 0.3% casamino acids) was added and the protoplasts placed on a shaker at 160 rpm for 18 to 20 hours. The regenerated protoplasts were centrifuged at 3000 rpm for 10 min and resuspended in 100 µl of SuTC. The SuTC protoplast suspension was mixed with 10 ml regeneration top (RTop) medium (55℃) and overlaid on top of PDA plates with 120 µg/ml hygromycin B. These plates were kept in the dark at room temperature. After 2-3 days, hyphal growth on the agar surface was

observed indicating that hygromycin resistant transformants. The transformant colonies were transferred onto the fresh PDA-hygromycin plates and screened by PCR to confirm the transformant was the desired mutant.

3.2.6 Pathogenicity experiments in cotton

Three different virulence assays were conducted on cotton: a stem lesion assay, a whole plant hydroponic assay, and a whole plant pot assay. All three assays were repeated at least three times, and seven to nine cotton plants were inoculated per treatment.

3.2.6.1 Stem lesion assay

Cotton seeds were sterilized with a 30% solution of commercial bleach (NaOCl) for 3 min and rinsed with sterile water one time. The seeds were then washed with 70% (v/v) ethanol for 5 min and rinsed several times with sterile water. The sterilized seeds were soaked in water for one or two days to increase the germination rate on a room temperature shaker at 150 rpm. To keep the seeds in the dark, the flask was entirely covered in aluminum foil. Each seedling was placed in a borosilicate glass culture tube which had been filled with sterile vermiculite and 10 ml of Hoagland solution. The plants were grown in a chamber with controlled temperature (21℃) and proper humidity in the dark.

Fov race 4 (89-1A) and the gene deletion mutants were grown on M100 agar plates for five days at room temperature for inoculum for the stem assay. Two *F. oxysporum* f. sp. *pisi* (pea pathogens, F23 and F79) that are nonpathogenic on cotton were used as negative controls. About a week after the seeds had been planted, the seedlings emerged from the surface and the cotton hypocotyls were elongated enough for inoculation. The stem was marked 3-5 cm above the soil line and a small wound was made with a sterile 30 G needle. Fungal cultures on the five days-old plates were sampled with a sterile cork borer and half of a 5 mm agar plug was attached to the

marked, wounded stem. The inoculated plants were moved to a clean, closed aquarium to maintain moisture, and the cotton plants were returned to the plant growth chamber for a week at 21℃. Lesions as disease symptoms on the stem were measured. Lesion sizes between the wild type and each mutant for disease severity were evaluated by a t-test at the 5% probability level.

3.2.6.2 Hydroponic assay

Cotton seeds were sterilized by 3% hydrogen peroxide (H_2O_2) for 8 hours and rinsed with sterile water three times. The sterilized seeds were soaked in water on a shaker at 150 rpm for one or two days to increase the germination rate. Once cotton roots germinated and reached a length of 5-10 mm, cotton coats were individually removed using sterile forceps. Eight seedlings were placed into one magenta box containing 70-80 ml of half strength Murashige & Skoog (MS) media with 1.5% w/v sucrose, and 2.6 g/L phytagel and the pH adjusted to 6.2 with 1N NaOH. The boxes were sealed with parafilm and placed in the plant growth chamber with the same conditions as above. Four or five days were required until the number and length of lateral roots grew sufficiently. During this time, black plastic containers with 8 holes in the lid were cleaned twice with 10% solution of commercial bleach (NaOCl) and rinsed with sterile water multiple times. Eight cotton plants were transferred into each container containing half strength MS media. The containers were kept in a tray under a 7-inch dome to keep the plants moist and returned to the plant growth chamber for the remainder of the assay.

The SSP mutant strains and wild-type *Fov* race 4 (89-1A) isolate were grown in 50 ml PDB liquid media for 5 days at room temperature (25℃). Conidia were filtered through the sterile Nitex mesh and centrifuged at 3000 rpm for 5 min. The harvested conidia were transferred into 100 ml of fresh PDB media and placed on a 160 rpm shaker overnight. Fifty ml of the cultures in PDB media were centrifuged to obtain fungal germlings. The remaining spent liquid culture media was

carefully decanted, and 20 ml of sterile water was added and the germlings suspended. After centrifugation, the water was discarded, and 5-10 ml of water was added. The number of germlings was counted using a hemocytometer and microscope. One ml of conidia (~10^6) was mixed with 50 ml of half strength MS media and the mixture was added to the container. After two days, the cotton plants were transferred to fresh half strength MS media and kept in the growth chamber for 2-3 weeks. Every week, the media in the plant container was replaced with fresh media. To evaluate the disease symptoms, the leaves and vascular tissue were observed and the severity was measured according to the following four disease index categories: 0 - no wilt symptoms, 1 - light yellowing of the leaves or black roots present, 2 - moderate necrosis and defoliation of lower leaves, or root discoloration (6-25%), 3 - severe leaf necrosis and defoliation, stunted growth, or root discoloration (26-50%), 4 - severe defoliation accompanied by pronounced stunting, chlorosis, and necrosis of the remaining leaves, or dead plant.

3.2.6.3 Pot assay

Preparation of the cotton seeds, *Fov* 89-1A isolate, and inoculum of the mutants was the same as previously described in the hydroponic assay. After sterilizing cotton seeds, they were placed in the magenta box containing the half strength MS medium with sucrose and agar (2.6 g/L). Four or five days later, eight cotton plants were transferred into the sterile black plastic container. Fungal germlings (10^6) of the SSP mutant strains and wild-type *Fov* race 4 (89-1A) isolate were inoculated. Two days prior to inoculation, the soil and pots were prepared. Pro-Mix BX potting soil was autoclaved at least three times and sterilized water was prepared as well. The pots were cleaned twice with a 10% solution of commercial bleach and rinsed with water until all the bleach was removed. After the pots were dried entirely, they were filled with moist potting

mix. One or two days after the fungal inoculation, cotton plants were transferred to individual pots. Eight pots were kept in a tray and covered with a 7-inch dome.

Symptoms were observed every week and watering occurred at the same time. After 2-3 weeks, the plant height and leaf phenotype/disease development were recorded. The plant roots were gently removed from the soil and rinsed carefully with water. After eliminating soil particles and the water, the fresh weight was measured by drying the samples in an oven at 70℃ for 3 days, after which the dry weight was measured.

3.2.7 Gene expression

The preparation of cotton plants and *Fov* race 4 isolate was the same as described in the hydroponic assay. After *Fov* race 4 was inoculated, three cotton roots were collected at the following 10 different time-points, 0, 4, 8, 12, 24, 36, 48, 72, 96, and 120 h post inoculation. As soon as the cotton roots were collected, they were placed in a -80℃ freezer until homogenization. Under liquid nitrogen, the three root samples were ground to a fine powder, which was used for RNA extraction using the E.Z.N.A.® Total RNA Kit I (Omega Bio-Tek), following the manufacturer's instructions. The concentration and purity of total RNA were assessed by gel electrophoresis before complementary DNA (cDNA) synthesis. One μg of total RNA was reverse transcribed using the QuantiTect Reverse Transcription Kit (Qiagen) following the manufacturer's instructions. All synthesized cDNA samples were stored at -20℃. The transcript expression of targeted genes was determined by quantitative reverse transcription PCR (RT-qPCR) on a Bio-Rad CFX96 instrument for the ten different time points using the primer pairs of individual target gene (Table 5). Two genes, *EF-1α* and *GAPDH,* from *F. oxysporum* were used as the housekeeping gene controls for comparison. The relative expression levels were calculated using the *ΔΔCt* method (Livak & Schmittgen, 2001).

3.3 Results

3.3.1 *in silico* analysis to identify SSPs in the genomes of 16 *F. oxysporum* isolates

In order to generate well-defined effector profiles of cotton isolates, the FOSC isolates that had a high-quality genome and annotation were selected from within and between *formae speciale*s of *F. oxysporum*. The protein sequences of five *Fov* isolates previously sequenced (Seo et al., 2020) and 11 FOSC isolates, were obtained from NCBI and the Broad Institute (Table 1) and used to query small secreted proteins (SSPs). Given that the proteomes were generated by different sequencing technologies and assembly pipelines, the number of predicted genes in the five *Fov* strains was smaller than in other FOSC isolates. For instance, *Fov* race 4 (89-1A) had the most predicted proteins of the *Fov* genomes (20,248), yet it still had fewer genes than the other FOSC genomes where they had more than 22,000 genes (Table 2). However, the number of SSPs in the *Fov* isolates was not significantly different from the number predicted in the other FOSC isolates (Figure 2 and Table 2). Within all the FOSC genomes included in this study, the number of putative secreted proteins that lack transmembrane domains ranged between 1,409 to 1,638. However most known effectors are <300 amino acids in length, and when limited to this size range the number of putative secreted proteins was between 569 to 687 (Table 2 and Figure 2). Interestingly, the *Fov* race 4 had the largest number of putative SSPs under 300 amino acids among the 16 FOSC genomes.

3.3.2 Identification of *Fov* specific SSPs

Identifying SSPs that are specific to cotton isolates is crucial for defining those that could be contributing to host-specificity. Using the data set of 687 putative SSP from *Fov* race 4 as a reference, orthologous genes from the other 15 FOSC isolates were identified by sequence similarity. Among the 687 SSP proteins in the *Fov* race 4 assessed for similar proteins, a total of

192 putative SSPs were common in all FOSC isolates (Figure 3). Sixty-two SSPs were only found in the *Fov* race 4 proteome and thirty putative SSPs were shared between the *Fov* race 4 and 7 isolates, which are closely related based on molecular analysis (Skovgaard et al., 2001). Importantly, 16 orthologs were identified that were unique in five *Fov* isolates and a total of 78 that were shared between at least three isolates from cotton but absent in all other FOSC genomes not pathogenic on cotton (Figure 3 and Table 3).

Most of the *Fov* specific SSPs (70 of 78) were found to reside on scaffolds predicted to be part of the core genome, while eight putative SSPs were predicted to reside in accessory regions. In addition to the above pipeline, EffectorP2 and ApoplastP confirmed/identified 37 and 27 proteins among the 78 SSPs, and 14 proteins were predicted by both programs (Figure 4a-b). A cutoff of 1 kb or less between genes was used to define of the SSPs resided in a cluster and identified two clusters each comprised of two genes. One cluster composed of two SSPs on CHR04; *F11885* predicted by EffectorP2 too and had a known domain associated with the phytotoxin cerato-platanin and *F11890* was uncharacterized by any programs here. Additionally, the third gene *F11879* was closely located with the two clustered genes in the upstream within 600 bp. It was predicted by the effector pipelines and ApoplastP program, but the length over 300 aa (347 aa) was out of the basis for the SSP candidates. The other cluster of *F07669* and *F07679* was located on CHR12 (Figure 4c and Table 3). Among the 78 *Fov* specific SSP encoding genes, only two of them had mimp-associated sequences upstream, *F02480* and *F07807*.

Previously obtained expression data of *Fov* 89-1A after treatment with the cotton plant defense chemical gossypol was included (Pokhrel, 2018), where three genes were up-regulated (*F02627*, *F04524* and *F08927*) and one was down-regulated (*F14403*) (Figure 5e). All four of these genes were predicted by Effectorp2 or ApoplastP, and the up-regulated gene *F08927* had a

ferritin-like domain. Of the 78 *Fov* specific putative SSPs, 20 of them contained annotated domains including an FAD binding domain, transcription factor Tfb2, desaturase, hydrolase, subtilase, transferase, and others (Figure 4f). Six of these *Fov* specific putative SSPs appear to have multiple similar genes and were eliminated from further functional analysis, e.g., generation of mutants, expression studies, etc.

3.3.3 Identification of homologs of SIX genes in Fov

Homologs of the 14 characterized *SIX* genes in *Fol* were searched for in each of the *Fov* proteomes. Only homologs of three *SIX* genes, *SIX1*, *SIX5*, and *SIX9*, were identified in the *Fov* isolates (Table 4). Three *Fov* isolates, TF1, 14-004, and LA1E encoded a homolog of *SIX1*, and was predicted to reside in the accessory genome of each of these isolates. A homolog of *SIX5* was predicted in the core genome of chromosome 12 in three *Fov* isolates, 89-1A, LA3B, and LA1E. The *Fov* race 4 (89-1A) genome contained three homologues of *SIX9*, and all three homologues were in the accessory regions of ACC05 and 08, and all three *SIX9* homologs were predicted by EffectorP2 and ApoplastP and had a mimps associated sequence located upstream of the coding sequence. Amino acid alignment of the *SIX* genes of *Fol* 4287 and homologs in *Fov* indicated the *SIX1* gene had sequence similarity (76%) only in the N-terminus from amino acid positions 1-60 (Figure 5a). Among the three homologs of the *SIX9* gene in the *Fov* race 4, two homologs (*F02116* and *F02577*) had over 90% identity, while *F02116* was almost 99% identical at the C-terminus with a single amino acid difference (Figure 5b). F02577 had 14 more amino acids at the N-terminus than Six9 of *Fol*, but the remaining amino acid sequence were almost the same as both *SIX9* and *F02116*.

3.3.4 SSP *F11624* is a virulence factor on cotton and the gene expression is up-regulated during infection

Using quantitative real-time reverse transcription-polymerase chain reaction (qRT-PCR), only the putative SSP *F11624* was found to be significantly expressed *in planta* during infection increasing at 72 hpi. The other four putative SSPs *F02527*, *F02627*, *F07669*, and *F07679* did not change significantly within the first week after infection (Figure 8).

Four mutants of *Fov* race 4 were generated to evaluate the contribution of the putative SSPs to fungal pathogenicity. These mutants were a diverse selection of the 79 SSPs where mutants were generated from two individual genes (*F02527* and *F02627*) that resided on accessory chromosomes, two genes (*F07669* and *F07679*) within a 1 kb proximity were concurrently mutated, and one gene (*F11624*) on a core chromosome hypothesized to influence the immune response of cotton (Table 3). A CRISPR/Cas9 system was used to induce a homologous recombination repair mechanism at the target site(s) and insert a hygromycin resistance cassette generating a mutant (Figures 6 and 7).

Three pathogenicity assays were conducted with the Pima cultivar DP744 to assess whether mutation of any of the putative SSPs in *Fov* race 4 had any influence on the virulence to cotton. Stem, hydroponic, and pot assays indicated that several of the putative SSPs did play a key role in pathogenicity. In the stem lesion assay, two transformants, Δ*F07669/F07679* and Δ*F11624* produced smaller lesion sizes when compared to the wild type inoculated plants (Figure 10a), and these two mutants, Δ*F07669/F07679* and Δ*F11624* were consistently significantly reduced in virulence in all stem lesion virulence assays by t-test (p-value <0.05) (Figure 10b). The symptoms caused by the Δ*F02527* mutant were variable having no statistical difference in the first trial, but a significant difference from the wild type on the last two trials. On the other hand, the mutant, Δ*F02627* had no significant reduction of virulence throughout all the experiments, where the lesion sizes were similar to the wild type 89-1A isolate (Figure 10).

In the hydroponic virulence assay, the symptoms of disease development, including chlorosis, and necrosis of leaves, wilting as well as discoloration of the vascular system, were observed as the *Fov* infection progressed (Figure 11a). The wild type and four mutants were able to successfully colonize the vascular tissues, as discoloration was evident after cutting the root. The defoliation and necrosis of the leaves was severe with cotton plants inoculated with the wild type and two mutants, Δ*F02527* and Δ*F02527* (Figure 11a), resulting in severely wilted or dead plants. Cotton plants inoculated with the Δ*F07669/F07679* and Δ*F11624* strains appeared to have less disease symptoms when compared to the wild-type 89-1A isolate (Figure 11b).

In the pot virulence assay with soil, typical Fusarium wilt symptoms were able to be observed 2-3 weeks after inoculation. Cotton plants inoculated with the 89-1A *Fov* isolate had stunted growth and most of the leaves were defoliated. As with the stem and hydroponic assays, both the Δ*F02527* and Δ*F02627* mutants had no reduction in virulence and displayed similar disease symptoms as the plants inoculated with the wild type *Fov* (Figure 12a). Furthermore, about half of the inoculated plants were dead two weeks after inoculation (Figure 12c). There was a slight reduction in virulence of the Δ*F07669/F07679* mutant, where most of plants were categorized in the second and third scale of disease index (DI), (Figure 12c). No plants were dead during the assay after inoculation with Δ*F11624*, and many of the cotton plants only showed minor symptoms particularly chlorosis of the leaves (Figure 12b-c).

3.4 Discussion

The *SIX* genes were initially identified from tomato sap during the colonization of *Fol* 4287 (Rep et al., 2004). So far, fourteen *SIX* genes have been described with no homology to each other (Houterman et al., 2007; Rep et al., 2004; Schmidt et al., 2013; Takken & Rep, 2010). The function of *SIX1* (*Avr3*), *SIX3* (*Avr2*), *SIX4* (*Avr1*), and *SIX5* have been identified, uncovering the

corresponding resistance genes in tomato. Five Six proteins, Six1, Six3, Six4, Six5, and Six6, have been shown to directly contribute to pathogenicity in gene-deletion analysis (Gawehns et al., 2014; Houterman et al., 2009; Ma et al., 2015; Rep et al., 2004; Takken & Rep, 2010; Thatcher et al., 2012). Several of these *SIX* genes have homologs in other FOSC isolates virulent on other host plants including *vasinfectum* isolates, NRRL 25433 and Australian isolates (Chakrabarti et al., 2011; van Dam et al., 2016). In addition, isolates from the same *forma specialis* within the FOSC typically have a similar suite of effectors (van Dam et al., 2016). While the *SIX* genes were not conserved in all *Fov* isolates in this study, a few homologs of three genes, *SIX1*, *5*, and *9* were observed among five *Fov* isolates, although their role in pathogenicity in cotton is not known. Additionally, out of all the homologs found in the five *Fov* isolates, only a single homolog of *SIX9* in *Fov* 89-1A had high similarity (99%), with only a change of a single amino acid.

F. oxysporum f. sp. *cepae* (*Foce*) isolates that are highly pathogenic on onion can have seven *SIX* genes, *(3, 5, 7, 10, 12, 9,* and *14*). In contrast, all *SIX* genes were absent in non-pathogenic isolates to the same host with a few exception (Taylor et al., 2016). *SIX9* in *Foce* isolates has 90% nucleotide identity with the *SIX9a* allele identified in a *F. oxysporum* isolate infecting Arabidopsis and brassica (Thatcher et al., 2012). In another *formae speciales* group pathogenic to banana, *cubense* (*Focu*), *SIX9a* homologues were identified in all 23 vegetative compatibility groups (VCGs) (Czislowski et al., 2018). The phylogenetic relationship between the VCGs and presence/absence of several of the common *SIX* genes suggests horizontal transfer of these virulence factors may explain the polyphyletic evolution within the *Focu* (Czislowski et al., 2018). The authors proposed that the VCGs of race 1 or 2 served as a donor and the race 4 was the recipient by the horizontal transfer of a subset of *SIX* genes. While investigating the *SIX* genes in *Focu*, two homologues, *SIX9a* and *b*, were identified in a *Fov* race 7 (25433). *Fov* race 7 is

phylogenetically closely related to the *Fov* race 4 included in this study which had three homologues of *SIX9*, two which were >90% identical in amino acid sequence (Figure 5b). The *SIX9* homologues were encoded on two accessory scaffolds (ACC05 and ACC08) suggesting those may be transferred to/from an outside source. The presence of *SIX9* only in the *Fov* race 4/7 clade could be a reason why this clade is the most virulent among the *Fov* isolates. A single homolog of *SIX5* was identified in *Fov* race 4 and two other *Fov* genotypes, although the similarity is quite low between the homologue and *SIX5* (34%). *SIX5* resides in the core genome on chromosome 12 for all three *Fov* isolates; however, this chromosome is dispensable in *Fol* (Vlaardingerbroek et al., 2016) and therefore it could also be dispensable in *Fov*.

Putative small secreted proteins (SSPs) were identified in six *Fov* isolates as well as eleven FOSC genomes which have different host specificity than cotton. While bioinformatics programs are available, it can be difficult to define putative SSPs (or effectors) using known fungal effectors due to the lack of sequence similarity or common functions. Many effectors could be involved in the determination of host-range differences between *formae speciales* (Chakrabarti et al., 2011). The different degrees of pathogenicity of *Fov* genotypes on cotton cultivars may be reflected in the presence/absence of effectors carried by each genotype (Czislowski et al., 2018). A universal pipeline to predict the secretomes of *F. oxysporum* isolates allowed comparison of the secretomes between *Fov* and other FOSC isolates to narrow down *formae speciales*-specific SSPs. The phylogenetic relationship between the *Fov* race 4 and 7 genomes was evident in the similarity between the populations of SSPs in these isolates. Not limited to only the *Fov* race 4 genome, 79 orthologs in at least three *Fov* isolates were further evaluated. Fungal effectors in *Fol* 4287 revealed some effectors were associated with transposable elements such as a miniature inverted-repeat transposable element referred as a miniature Impala (mimp) in the 5′ promoter region

(Schmidt et al., 2013). These mimps sequences have been used as markers to discover putative effectors in other FOSC isolates (Armitage et al., 2018; van Dam et al., 2016).

Candidate SSP genes were chosen for mutation primarily based on chromosomal location. *F02527*, residing on ACC04, was also predicted by EffectorP2 and ApoplastP, while *F02627* was on ACC12 and was predicted to localize to the apoplast during infection and was induced by gossypol treatment. Interestingly, *F02627* was not significantly expressed *in planta* after inoculation with *Fov* 89-1A. The clustered genes *F07669* and *F07679* that are within 1 kb of each other on chromosome 12 were included in the 78 Fov specific putative SSPs. While *F07669* has no similarity to other characterized proteins, *F07679* was predicted to localize to the plant apoplast and has a domain related to the carboxylesterase (CE) family. CEs are one of the carbohydrate activity enzymes (CAZymes) known as cell wall degrading enzymes (CWDEs) that act in the decomposition of plant biomass (Ospina-Giraldo et al., 2010). CAZymes are produced by many fungal pathogens and responsible for penetration and successful infection of their hosts (Zhao et al., 2013). The final gene chosen for mutation was *F11624* which resides on the core chromosome 4 and has not been previously described as a virulence factor however may influence the immune response of cotton (L. Shan and P. He, personal communication). The virulence of the gene-deletion mutants was evaluated on the highly susceptible Pima cotton cultivar DP744 with three different pathogenicity tests: a stem lesion assay, a hydroponic assay, and a more standard pot assay. Although all methods and measurements differed between the three assays, the results were fairly consistent, with a significant reduction in virulence on cotton plants inoculated with Δ*F11624*, and to a lesser extent Δ*F07669/F07679* While the accessory genome has been known to harbor host specific virulence factors including SSPs and effectors, the two putative SSPs that resided in the accessory genome of the *Fov* race 4 isolate were not involved in virulence on cotton.

In many cases, deletion of a single effector has little effect on virulence, suggesting functional redundancy of effectors (Marton et al., 2018).

Chlorosis, wilting, necrosis of the leaves, and eventually plant death are typical symptoms of cotton plants infected with *Fov* (Davis et al., 2006). *Fov* isolates can infect at any stage of cotton development. In the hydroponic and pot virulence assays, most of the leaves of WT inoculated plants were either senesced/wilted or had dropped within three weeks following the inoculation of the roots (Wang et al., 2021, and this study). Given that *Fov* is a soil-borne pathogen colonizing plant roots, the best way to evaluate the disease severity of the *Fov* mutant strains on cotton cultivars is using either the hydroponic or the pot assay. Previously it has been shown that a soil bioassay is more advantageous for the accurate quantification of a disease such as Fusarium wilt (Batson et al., 2021).

CRISPR/Cas9 mediated genome editing, an efficient and powerful tool for gene deletion, has been reported in many *Fusarium* species (Ferrara et al., 2019; Wang et al., 2018). Depending on the target gene, a 50- (*F02527*, *F02627*, and *F07669/F07679*) and 300-bp flanking region (*F11624*) was used for homologous recombination to insert the selectable hygromycin resistance cassette in the target genes for disruption. Despite the difference in size of the flanking regions, the mutation frequency was not significantly different between the two methods (data not shown).

While the two putative SSPs that reside in the accessory genome of the *Fov* race 4 isolate did not show a reduction in virulence on cotton when mutated, the remaining putative SSPs have not been evaluated and many have an unknown function. Elucidating the function of many putative SSPs could be valuable in identifying novel and isolate-specific effector genes in *Fov*. For example, *Fov* race 4 (89-1A) has sixty-two SSPs that were unique for this isolate, and the *Fov* race 4 and 7 proteomes have twenty-seven SSPs in common. Like the common *SIX* genes in the same *formae*

speciales are being used as a genetic marker for the classification of races, the profiles among the *Fov* isolates can be promising markers for the classification of *Fov* isolates from cotton. In addition, different patterns of effector profiles can determine the uniqueness of host specificity and virulence within the *Fov*. Altogether, *Fov* isolates have evolved independently to acquire pathogenicity on cotton. Additional functional analyses are needed to reveal the exact roles of SSPs in pathogen virulence and more importantly to identify any corresponding resistance gene(s) in cotton if they exist and can be utilized in breeding programs.

Reference

Armitage, A. D., Taylor, A., Sobczyk, M. K., Baxter, L., Greenfield, B. P. J., Bates, H. J., Wilson, F., Jackson, A. C., Ott, S., Harrison, R. J., & Clarkson, J. P. (2018). Characterisation of pathogen-specific regions and novel effector candidates in *Fusarium oxysporum* f. sp. *cepae*. *Scientific Reports*, *8*(1), 13530. https://doi.org/10.1038/s41598-018-30335-7

Batson, A. M., Fokkens, L., Rep, M., & du Toit, L. J. (2021). Putative Effector Genes Distinguish Two Pathogenicity Groups of *Fusarium oxysporum* f. sp. *spinaciae*. *Molecular Plant-Microbe Interactions®*, *34*(2), 141–156. https://doi.org/10.1094/MPMI-06-20-0145-R

Chakrabarti, A., Rep, M., Wang, B., Ashton, A., Dodds, P., & Ellis, J. (2011). Variation in potential effector genes distinguishing Australian and non-Australian isolates of the cotton wilt pathogen *Fusarium oxysporum* f.sp. *vasinfectum*. *Plant Pathology*, *60*(2), 232–243. https://doi.org/10.1111/j.1365-3059.2010.02363.x

Conway, J. R., Lex, A., & Gehlenborg, N. (2017). UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics*, *33*(18), 2938–2940. https://doi.org/10.1093/bioinformatics/btx364

Czislowski, E., Fraser-Smith, S., Zander, M., O'Neill, W. T., Meldrum, R. A., Tran-Nguyen, L. T. T., Batley, J., & Aitken, E. A. B. (2018). Investigation of the diversity of effector genes in the banana pathogen, *Fusarium oxysporum* f. sp. *cubense* , reveals evidence of horizontal gene transfer. *Molecular Plant Pathology*, *19*(5), 1155–1171. https://doi.org/10.1111/mpp.12594

Davis, R. M., Colyer, P. D., Rothrock, C. S., & Kochman, J. K. (2006). Fusarium wilt of cotton: Population diversity and implications for management. In *Plant Disease* (Vol. 90, Issue 6, pp. 692–703). The American Phytopathological Society. https://doi.org/10.1094/PD-90-0692

de Jonge, R. (2012). In silico identification and characterization of effector catalogs. *Methods in Molecular Biology*, *835*, 415–425. https://doi.org/10.1007/978-1-61779-501-5_25

Feldman, D., Yarden, O., & Hadar, Y. (2020). Seeking the Roles for Fungal Small-Secreted Proteins in Affecting Saprophytic Lifestyles. *Frontiers in Microbiology*, *11*, 455. https://doi.org/10.3389/fmicb.2020.00455

Ferrara, M., Haidukowski, M., Logrieco, A. F., Leslie, J. F., & Mulè, G. (2019). A CRISPR-Cas9 System for Genome Editing of *Fusarium proliferatum*. *Scientific Reports*, *9*(1), 1–9. https://doi.org/10.1038/s41598-019-56270-9

Gawehns, F., Houterman, P. M., Ichou, F. A., Michielse, C. B., Hijdra, M., Cornelissen, B. J. C., Rep, M., & Takken, F. L. W. (2014). The *fusarium oxysporum* effector six6 contributes to virulence and suppresses i-2-mediated cell death. *Molecular Plant-Microbe Interactions*, *27*(4), 336–348. https://doi.org/10.1094/MPMI-11-13-0330-R

Hall, C., Heath, R., & Guest, D. (2013). The infection process of *Fusarium oxysporum* f.sp. *vasinfectum* in Australian cotton. *Australasian Plant Pathology*, *42*(1), 1–8. https://doi.org/10.1007/s13313-012-0169-8

Houterman, P. M., Ma, L., van Ooijen, G., de Vroomen, M. J., Cornelissen, B. J. C., Takken, F. L. W., & Rep, M. (2009). The effector protein Avr2 of the xylem-colonizing fungus *Fusarium oxysporum* activates the

tomato resistance protein I-2 intracellularly. *The Plant Journal*, *58*(6), 970–978. https://doi.org/10.1111/j.1365-313X.2009.03838.x

Houterman, P. M., Speijer, D., Dekker, H. L., de Koster, C. G., Cornelissen, B. J. C., & Rep, M. (2007). The mixed xylem sap proteome of *Fusarium oxysporum*-infected tomato plants. *Molecular Plant Pathology*, *8*(2), 215–221. https://doi.org/10.1111/j.1364-3703.2007.00384.x

Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. In *Nature* (Vol. 444, Issue 7117, pp. 323–329). Nature Publishing Group. https://doi.org/10.1038/nature05286

Kamoun, S. (2009). The Secretome of Plant-Associated Fungi and Oomycetes. In *The Mycota* (pp. 173–180). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-87407-2_9

Li, J., Fokkens, L., Conneely, L. J., & Rep, M. (2020). Partial pathogenicity chromosomes in *Fusarium oxysporum* are sufficient to cause disease and can be horizontally transferred. *Environmental Microbiology*, *22*(12), 4985–5004. https://doi.org/10.1111/1462-2920.15095

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods*, *25*(4), 402–408. https://doi.org/10.1006/meth.2001.1262

lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., Zuccaro, A., Reissmann, S., & Kahmann, R. (2015). Fungal effectors and plant susceptibility. *Annual Review of Plant Biology*, *66*, 513–545. https://doi.org/10.1146/annurev-arplant-043014-114623

Lu, S., & Edwards, M. C. (2016). Genome-Wide Analysis of Small Secreted Cysteine-Rich Proteins Identifies Candidate Effector Proteins Potentially Involved in *Fusarium graminearum* −Wheat Interactions. *Phytopathology®*, *106*(2), 166–176. https://doi.org/10.1094/PHYTO-09-15-0215-R

Ma, L., Houterman, P. M., Gawehns, F., Cao, L., Sillo, F., Richter, H., Clavijo-Ortiz, M. J., Schmidt, S. M., Boeren, S., Vervoort, J., Cornelissen, B. J. C., Rep, M., & Takken, F. L. W. (2015). The *AVR2–SIX 5* gene pair is required to activate *I-2* -mediated immunity in tomato. *New Phytologist*, *208*(2), 507–518. https://doi.org/10.1111/nph.13455

Marton, K., Flajsman, M., Radisek, S., Kosmelj, K., Jakse, J., Javornik, B., & Berne, S. (2018). Comprehensive analysis of *Verticillium nonalfalfae in silico* secretome uncovers putative effector proteins expressed during hop invasion. *PLoS ONE*, *13*(6), e0198971. https://doi.org/10.1371/journal.pone.0198971

Moller, S. M. ¨, Croning, M. D. R., & Apweiler, R. (2001). Evaluation of methods for the prediction of membrane spanning regions. In *BIOINFORMATICS* (Vol. 17, Issue 7).

Nelson, P. E. (1981). Life Cycle and Epidemiology of *Fusarium oxysporum*. In *Fungal Wilt Diseases of Plants* (pp. 51–80). Elsevier. https://doi.org/10.1016/b978-0-12-464450-2.50008-5

Ospina-Giraldo, M. D., Griffith, J. G., Laird, E. W., & Mingora, C. (2010). The CAZyome of *Phytophthora* spp.: A comprehensive analysis of the gene complement coding for carbohydrate-active enzymes in species of the genus *Phytophthora*. *BMC Genomics*, *11*(1), 525. https://doi.org/10.1186/1471-2164-11-525

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: Discriminating signal peptides from transmembrane regions. In *Nature Methods* (Vol. 8, Issue 10, pp. 785–786). https://doi.org/10.1038/nmeth.1701

Pokhrel, A. (2018). Transcriptomic analysis to identify candidate genes conferring gossypol tolerance in Fusarium oxysporum f. sp. vasinfectum Title.

Pusztahelyi, T., Holb, I. J., & Pócsi, I. (2015). Secondary metabolites in fungus-plant interactions. *Frontiers in Plant Science*, *6*(AUG), 573. https://doi.org/10.3389/fpls.2015.00573

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., & Lopez, R. (2005). InterProScan: Protein domains identifier. *Nucleic Acids Research*, *33*(SUPPL. 2), W116–W120. https://doi.org/10.1093/nar/gki442

Rep, M. (2005). Small proteins of plant-pathogenic fungi secreted during host colonization. *FEMS Microbiology Letters*, *253*(1), 19–27. https://doi.org/10.1016/j.femsle.2005.09.014

Rep, M., van der Does, H. C., Meijer, M., van Wijk, R., Houterman, P. M., Dekker, H. L., de Koster, C. G., & Cornelissen, B. J. C. (2004). A small, cysteine-rich protein secreted by *Fusarium oxysporum* during colonization of xylem vessels is required for I-3-mediated resistance in tomato. *Molecular Microbiology*, *53*(5), 1373–1383. https://doi.org/10.1111/j.1365-2958.2004.04177.x

Schmidt, S. M., Houterman, P. M., Schreiver, I., Ma, L., Amyotte, S., Chellappan, B., Boeren, S., Takken, F. L. W., & Rep, M. (2013). MITEs in the promoters of effector genes allow prediction of novel virulence genes in *Fusarium oxysporum*. *BMC Genomics*, *14*(1), 119. https://doi.org/10.1186/1471-2164-14-119

Seo, S., Pokhrel, A., & Coleman, J. J. (2020). The genome sequence of five genotypes of *Fusarium oxysporum* f. sp. *vasinfectum*: A resource for studies on Fusarium wilt of cotton. *Molecular Plant-Microbe Interactions*, *33*(2), 138–140. https://doi.org/10.1094/MPMI-07-19-0197-A

Skovgaard, K., Nirenberg, H. I., O'Donnell, K., & Rosendahl, S. (2001). Evolution of *Fusarium oxysporum* f. sp. **vasinfectum** races inferred from multigene genealogies. *Phytopathology*, *91*(12), 1231–1237. https://doi.org/10.1094/PHYTO.2001.91.12.1231

Sperschneider, J., Dodds, P. N., Gardiner, D. M., Singh, K. B., & Taylor, J. M. (2018). Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. *Molecular Plant Pathology*, *19*(9), 2094–2110. https://doi.org/10.1111/mpp.12682

Sperschneider, J., Dodds, P. N., Singh, K. B., & Taylor, J. M. (2018). ApoplastP: prediction of effectors and plant proteins in the apoplast using machine learning. *New Phytologist*, *217*(4), 1764–1778. https://doi.org/10.1111/nph.14946

Stemmer, M., Thumberger, T., del Sol Keyer, M., Wittbrodt, J., & Mateo, J. L. (2015). CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. *PLOS ONE*, *10*(4), e0124633. https://doi.org/10.1371/journal.pone.0124633

Takken, F., & Rep, M. (2010). The arms race between tomato and *Fusarium oxysporum*. *Molecular Plant Pathology*, *11*(2), 309–314. https://doi.org/10.1111/j.1364-3703.2009.00605.x

Taylor, A., Vágány, V., Jackson, A. C., Harrison, R. J., Rainoni, A., & Clarkson, J. P. (2016). Identification of pathogenicity-related genes in *Fusarium oxysporum* f. sp. *cepae*. *Molecular Plant Pathology*, *17*(7), 1032–1047. https://doi.org/10.1111/mpp.12346

Thatcher, L. F., Gardiner, D. M., Kazan, K., & Manners, J. M. (2012). A highly conserved effector in *Fusarium oxysporum* is required for full virulence on Arabidopsis. *Molecular Plant-Microbe Interactions*, *25*(2), 180–190. https://doi.org/10.1094/MPMI-08-11-0212

van Dam, P., Fokkens, L., Schmidt, S. M., Linmans, J. H. J., Kistler, H. C., Ma, L.-J., & Rep, M. (2016). Effector profiles distinguish *formae speciales* of *Fusarium oxysporum*. *Environmental Microbiology*, *18*(11), 4087–4102. https://doi.org/10.1111/1462-2920.13445

van Dam, P., & Rep, M. (2017). The Distribution of Miniature Impala Elements and *SIX* Genes in the *Fusarium* Genus is Suggestive of Horizontal Gene Transfer. *Journal of Molecular Evolution*, *85*(1–2), 14–25. https://doi.org/10.1007/s00239-017-9801-0

Vlaardingerbroek, I., Beerens, B., Schmidt, S. M., Cornelissen, B. J. C., & Rep, M. (2016). Dispensable chromosomes in *Fusarium oxysporum* f. sp. *lycopersici*. *Molecular Plant Pathology*, *17*(9), 1455–1466. https://doi.org/10.1111/mpp.12440

Wang, Q., Cobine, P. A., & Coleman, J. J. (2018). Efficient genome editing in *Fusarium oxysporum* based on CRISPR/Cas9 ribonucleoprotein complexes. *Fungal Genetics and Biology*, *117*, 21–29. https://doi.org/10.1016/j.fgb.2018.05.003

Wang, Q., Pokhrel, A., & Coleman, J. J. (2021). The Extracellular Superoxide Dismutase Sod5 From *Fusarium oxysporum* Is Localized in Response to External Stimuli and Contributes to Fungal Pathogenicity. *Frontiers in Plant Science*, *12*, 294. https://doi.org/10.3389/fpls.2021.608861

Wasmann, C. C., & VanEtten, H. D. (1996). Transformation-mediated chromosome loss and disruption of a gene for pisatin demethylase decrease the virulence of *Nectria haematococca* on pea. *Molecular Plant-Microbe Interactions*, *9*(9), 793–803. https://doi.org/10.1094/MPMI-9-0793

Zhao, Z., Liu, H., Wang, C., & Xu, J. R. (2013). Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics*, *14*(1), 274. https://doi.org/10.1186/1471-2164-14-274

Table 1. FOSC genome information used in this study.

	f. sp.	**Strain**	**Race**	**Host**	**Protein number**	**Accession**	**Name***
Cotton isolates	*vasinfectum*	TF1	1	*Gossypium* sp.	17,932	GCA_009602505.1	FovTF1
		89-1A	4		20,248	GCA_009602635.1	Fov891A
		25433	7		25,216	GCA_000260175.2	Fov25433
		LA3B	LA108		18,097	GCA_009602545.1	FovLA3B
		14-004	LA127/140		18,119	GCA_009602625.1	Fov14004
		LA 1E	MDS12		18,990	GCA_009602615.1	FovMDS12
Non-cotton isolates	Fo47	Fo47	-	Soil	24,818	GCA_000271705.2	Fo47
	Fo5176	Fo5176	-	*Brassica* sp.	17,817	GCA_000222805.1	Fo5176
	conglutinans	54008	2	*Brassica* sp.	26,246	GCA_000260215.2	Foco54008
	cubense	54006	TR4	*Musa* sp.	22,487	GCA_000260195.2	FocuTR4
	lycopersici	4287	2	*Solanum* sp.	27,347	GCA_000149955.2	Fol4287
	lycopersici	MN25	3		24,733	GCA_000259975.2	FolMN25
	melonis	26406	1	cucurbit	26,719	GCA_000260495.2	Fom26406
	pisi	HDV247	-	*Pisum* sp.	26,378	GCA_000260075.2	FopHDV247
	radicis-lycopersici	26381	-	*Solanum* sp.	24,739	GCA_000260155.3	Forl26381
	raphani	54005	-	*Raphanus* sp.	25,666	GCA_000260235.2	For54005

*: The names of FOSC isolates when referred to in this chapter.

Table 2. The number of predicted proteins at each step of the pipeline for the 16 FOSC proteomes.

	Isolates	Cotton						Non-cotton									
	Race/name	Fov891A	FovTF1	Fov25433	FovLA3B	Fov14004	FovMDS12	Fo47	Fo5176	Foco54008	FocuTR4	Fol4287	FolMN25	Fom26406	FopHDV247	Forl26381	For54005
	Total proteins	20,248	17,932	25,216	18,097	18,119	18,990	24,818	27,268	26,246	22,487	27,347	24,733	26,719	26,378	24,739	25,666
	SignalP4	1,831	1,714	1,914	1,710	1,694	1,726	1,933	1,896	1,929	1,805	2,053	1,842	2,036	2,000	1,877	1,944
	TMHMM	3,564	3,406	4,612	3,429	3,426	3,536	4,732	4,656	4,648	4,253	4,902	4,685	4,962	4,864	4,717	4,643
	SPs -TMs[a]	1,539	1,421	1,558	1,416	1,409	1,430	1,540	1,534	1,559	1,443	1,638	1,451	1,612	1,617	1,495	1,578
	SPs -TMs (300)[b]	687	577	608	597	600	603	593	636	607	569	683	573	655	613	584	615
	EffectorP2	2,716	2,305	2,967	2,329	2,352	2,526	2,825	3,530	3,304	2,474	3,451	2,821	3,215	3,052	2,795	3,066
	EffectorP2 (300)	2,620	2,221	2,850	2,244	2,277	2,450	2,719	3,431	3,204	2,378	3,360	2,729	3,115	2,947	2,691	2,954
BlastP[c]	**SPs -TMs (300)**	687	428	501	423	384	487	351	388	408	341	341	331	371	404	341	388
	EffectorP2	2716	1454	1699	1458	1276	1768	1126	1279	1292	1083	1100	1115	1166	1284	1122	1272
	EffectorP2 (300)	2,620	1,388	1,613	1,388	1,215	1,697	1,060	1,215	1,226	1,025	1,048	1,061	1,109	1,214	1,057	1,209

[a] SPs: proteins having signal peptides, TMs: proteins having transmembrane domain and filtering out.

[b] A sequence size of ≤300 amino acids was chosen to narrow down candidate SSPs for the further study.

[c]: Cutoffs of the BlastP results are e-value over 0.001 and identity over 95%.

Table 3. Details of the 79 shared *Fov* SSP candidates in *Fov* race 4 potentially important for pathogenicity.

Location		#	Gene name	Nucleotide (nt)	Protein (aa)	EffectorP2 probability	ApoplastP probability	mimps	Clustered (~1kb)	Multicopy	Expressed	Annotation (PFAM)
Core	CHR01	5	*F03957*	587	173	0.76						FAD binding domain
			*F07762**	547	59	0.68	0.69					
			F14403	1136	162	0.828					▼	
			F17950	707	91							
			F18639	159	52		0.69			x		
	CHR02	4	*F04253*	565	170	0.853						
			F06089	1345	299		0.8					GDSL-like Lipase/Acylhydrolase family
			F13462	964	80	0.658						
			*F15422**	312	103	0.712	0.75					
	CHR04	11	*F03482**	738	100	0.921	0.84					
			F11013	401	109					x		
			F11885	494	145	0.577			**o**			Cerato-platanin
			F11890	544	100				**o**			
			F15167	422	106							
			F16350	943	248	0.617						Subtilase family
			*F17306**	315	104							
			F18550	1392	276		0.8					
			F18840	285	94							
			F19250	489	103	0.812	0.89					
			F19308	762	95	0.945						
	CHR05	8	*F02931*	582	111	0.748	0.67					
			F03306	440	108	0.881	0.91					
			*F03347***	493	144							
			*F02349**	432	128							
			F11571	1114	266		0.75					
			*F14823**	315	86							

		*F17600**	1376	194	0.623						
CHR07	4	*F12824**	405	134	0.823	0.78					
		F04278	387	128	0.933	0.56					
		F04942	329	73	0.842						
		F07886	421	70							
CHR08	5	*F04524*	640	103	0.919					▲	
		F05697	611	112	0.79						
		F07652	825	274	0.571						Transcription factor Tfb2
		F15583	813	121	0.754						
		F18309	691	208							Glycosyl transferase family 64 domain
CHR09	8	*F03989**	606	201	0.755						Prion-inhibition and propagation
		*F08489***	312	103	0.943	0.65					
		*F08927**	1738	300		0.66				▲	Ferritin-like domain
		F09282	1804	148	0.591						
		*F10458**	477	137							Common central domain of tyrosinase
		F15987	300	99	0.765						
		F17035	1012	90	0.967	0.78					
		F17457	507	168	0.579						
CHR10	3	*F15455**	1059	300							Eukaryotic-type carbonic anhydrase
		F15918	462	153							
		F16793	947	298	0.557						Peptidase dimerisation domain
CHR11	9	*F02480*	357	118			**o**				
		*F05633**	945	295							
		F08505	918	261							
		*F08642**	336	86							
		*F09356**	729	242		0.59					
		F10430	4084	79		0.61					
		F11795	399	132							
		F14365	207	68							

		F15822	425	107	0.856						
	CHR12	10	*F04015*	600	199	0.843	0.55					
			F04040	527	154							
			F07169	774	257		0.73					Glycosyl hydrolase family 61
			F10201	391	111	0.754						
			F11192	941	294							Enoyl-CoA hydratase/isomerase
			F01567	327	88	0.867						
			F02770	542	164		0.85					Cerato-platanin
			F07449	1241	287	0.552						lycosyl hydrolase family 61
			F07669	873	290			**o**				
			F07679[*]	846	223		0.55	**o**				Carboxylesterase family
	CHR13	4	*F04778*	900	299		0.76			x		Glycosyl hydrolase family 61
			F07895	321	106							
			F10689	1298	285							
			F12154	825	225							
Accessory	ACC_03	8	*F07807*	610	114	0.929		o				
	ACC_04		*F02527*	228	75	0.835	0.63					
	ACC_05		*F03915*	621	206		0.74			x		Glycosyl hydrolase family 61
	ACC_09		*F03098*	261	86					x		
	ACC_12		*F01135*	512	153							Fatty acid desaturase
	ACC_12		*F02627*	1064	271		0.67				▲	
	ACC_18		*F01823*	345	114	0.821	0.71					
	ACC_29		*F00981*	300	99	0.584	0.74			x		Glycosyl hydrolase family 61

[*]: Sixteen common SSPs among five cotton isolates are marked.

[**]: Two common SSPs in all the six cotton isolates including the Fov race 7 (25433) are stated.

Table 4. Presence of *F. oxysporum SIX* homologs in five *F. oxysporum* f. sp. *vasinfectum* isolates.

Fol	Accession number	Fov (race)	Fov (strain)	Location		Gene	Nucleotide (nt)	Protein (aa)	Start	Stop	EffectorP2 probability	ApoplastP probability	mimps	% identity	E-value
SIX1	CAE55870.1	1	TF1	ACC	2	F01297	179	60	291452	291631				76.271	2.68E-18
		LA127/140	14-004	ACC	3	F01221	179	60	292181	292360				76.271	2.65E-18
		MDS12	LA 1E	ACC	3	F01724	179	60	300600	300779				76.271	2.76E-18
SIX5	ACN87967.1	4	89-1A	CORE	12	F02664	725	95	305169	305893	0.924	0.85		34.146	1.92E-04
		LA108	LA3B	CORE	12	F01549	477	81	333079	333556	0.944	0.89		36.585	1.12E-05
		MDS12	LA 1E	CORE	12	F14910	443	81	2763525	2763968	0.944	0.87		35.366	5.93E-05
SIX9	AGG54050.1	4	89-1A	ACC	5	F02577	690	128	290427	291116	0.801	0.69	o	91.228	2.01E-77
		4	89-1A	ACC	8	F01701	357	118	170065	170421	0.945	0.73	o	53.043	1.70E-41
		4	89-1A	ACC	8	F02116	868	114	224415	225282	0.886	0.64	o	99.123	2.91E-83

Table 5. Primer sequences used in this study for generating knock-out mutants.

Primer names	Oligomer sequences (5' to 3')	Usage
F2527_sgR	CAAGGTAGCGGCGAGGAAGA	
F2527_50F	AGCCGATTCTTACTTTCTCTTTCAAAAACTCAAGATGAA GATCAGCGCCTTATATTAATGGAAGGGTATATACCA	
F2527_50R	TTAATGCATTGGACCTCGCAACCCTTTCCTCGCCGCTACC TTGGCTAGCGTGATGGCCGCCTCCTCCCATGTCCA	
F2527_TF	CCGGGTAAGTTGTGAGGAAA	
F2527_TR	TTAGCCGAGAACCGCTAGAA	
F2527_GEF	TTCCTCGCCGCTACCTT	
F2527_GER	GTGTTGGAATCGAAGCAGCT	
F2627_sgR	TGGGCGCAGAGACATTGCAA	
F2627_50F	CAATTATTTGCCGTTGTTGCTGCTTTTATGAGCACACTTA CAGAGACCTTTATATTAATGGAAGGGTATATACCA	
F2627_50R	TTAATGCATTGGACCTCGCAACCCTGGTAGTTTTGTTTTT ACCTTGATCTTTAGAGTAACATGCAGGAAAAGCGA	
F2627_TF	GCATGGGACTTCTGGTCATT	
F2627_TR	GCTTTAATCCCGGTTGATGA	
F2627_GEF	CATCTGCTCTGGTGGTAAGTG	
F2627_GER	ATAACCACAATGCCAGGAAGT	
F7669_sgRI	AGAGCCGGGAGAAACGATGG	Two small guide RNA oligos were used to generate the mutant knocking out two genes together.
F7679_sgRII	AATCGAAGATCACCTATCGG	
F7669_50F	GAATAGTCGAGACTGATTCGGTAGACGCGATAGAGCCG GGAGAAACGATGTATATTAATGGAAGGGTATATACCA	
F7679_50R	TTAATGCATTGGACCTCGCAACCCTCGATAGGTGATCTT CGATTGTACGTGTCGCACTACCCGATTCGATCTGCG	
F7669_TF	CCATGGAGAACATGTTGCAG	
F7679_TR	TGTTGTTGCCAGCTGAAGTC	
F7669_GEF	CTCCGACATCACCAACAGATAC	
F7669_GER	GAACCACCGGCATCAAATTC	
F7679_GEF	GACTCTTACAGACGACCACAAA	
F7679_GER	TCTTTCTCGACGTTGTCACG	
F11624_sgR	ACATGCGACACTGATCTCCG	
F11624_UpF	ATTCGAGCTCGGTACCCGGGCAATTGGTGGGCAGATTTT ATTG	These primers, UpF/R, HYGF/R, and DownF/R were needed for the construction of the plasmid for homologous recombination.
F11624_UpR	CAACAGAGCTTCTTCTCATACTATATTAATGG	
F11624_HYGF	CTTCTCATACTATATTAATGGAAGGGTATATACCACGC	
F11624_HYGR	CATTGGACCTCGCAACCCTCTTATGTACA	
F11624_DownF	TCGCAACCCTCTTATGTACACTTGTGGTTTCTCATCTACC	

F11624_DownR	ATGGAGTCGTCGGCGCTGTGGCGTAATCATGGTCATAG	
F11624_TF	CAAGCAAGAAATTTGGCAGT	
F11624_TR	CTCATCGGCGAACTGTATCT	
F11624_GEF	GGTACACGAACCCAGAACTC	
F11624_GER	ACAATGATGCCTCTGCTACC	
EF-1α_GEF	TGAGTTCGAGGCTGGTATCT	These two primer pairs were designed among housekeeping genes in *F. oxysporum* isolates used as the internal reference for quantification of transcript expression.
EF-1α_GER	CACTTGGTGGTGTCCATCTT	
GAPDH_GEF	CCACCGTCCACTCCTACACT	
GAPDH_GER	GAGCTCAGGAATGACCTTGC	

sgR: oligomer sequences for synthesizing the targeted small guide RNA; 50F/R: primers adding 50bp regions for homologous recombination; TF/R: primer pairs for amplifying 1-2 kb including the target region by sgRNA for *in vitro* cleavage; and GEF/R: primers for checking gene expression.

The genes, *F02527, F02627*, and *F07669/F07679* were knocked out using the 50-bp overhanging method having forward and reverse primers simply. On the other hand, the gene, *F11624* had a construct of the DNA template which had 300-bp up- and down-stream regions for homologues recombination. In addition, the primer pairs were designed containing the selective marker (HYBG) inside.

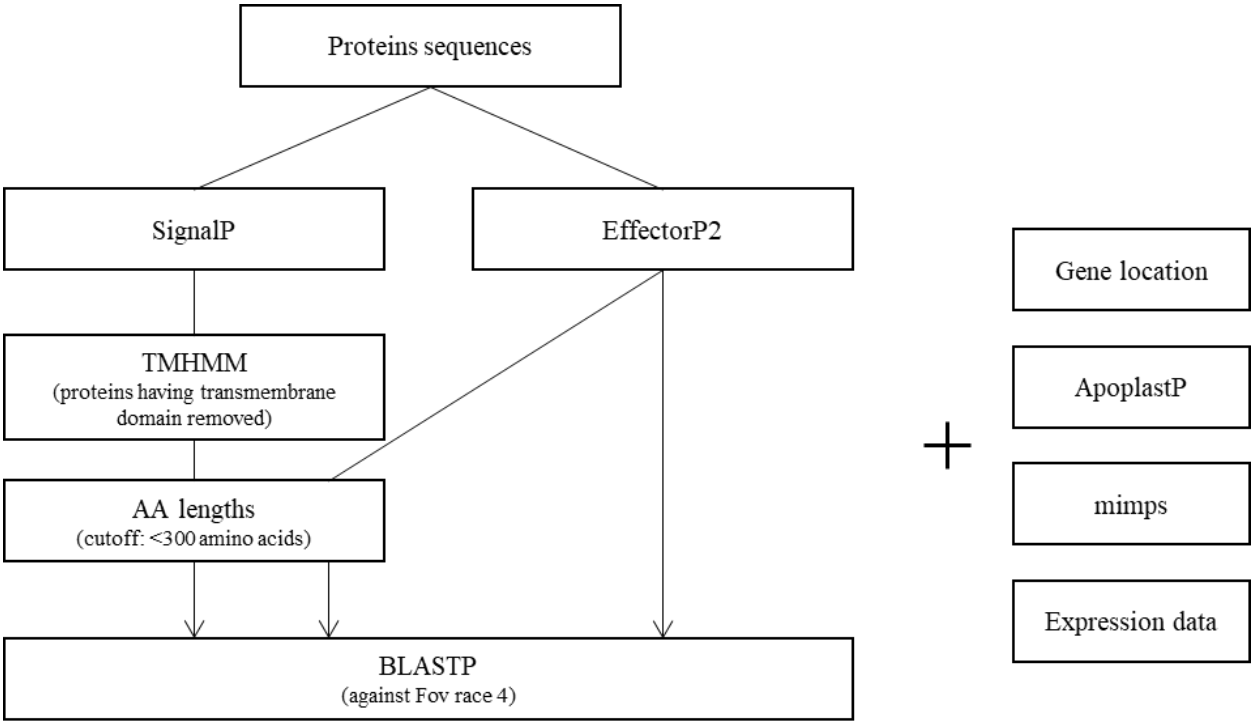


Figure 1. Bioinformatic pipeline for mining key small secreted proteins (SSPs) of *F. oxysporum* f. sp. *vasinfectum* (*Fov*). The commonly used pipelines including SignalP and TMHMM as well as EffectoP2 were used to build a comprehensive list of SSPs in FOSC isolates. Other annotation information by bioinformatics programs such as ApoplastP were included.

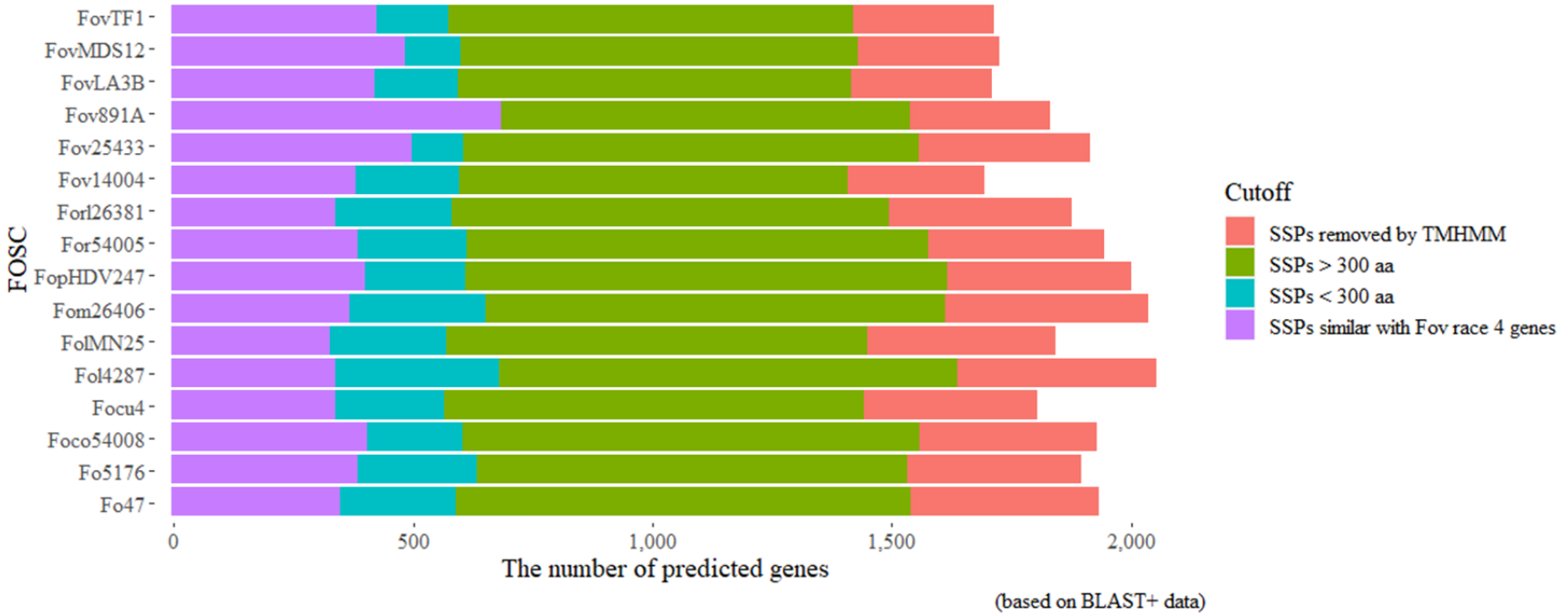


Figure 2. Profiles of small secreted proteins (SSPs) in 16 FOSC genomes. Each colored section corresponds to the number of proteins that are predicted by the criteria, the existence of transmembrane domains, protein sizes, and sequence similarity to *Fov* race 4.

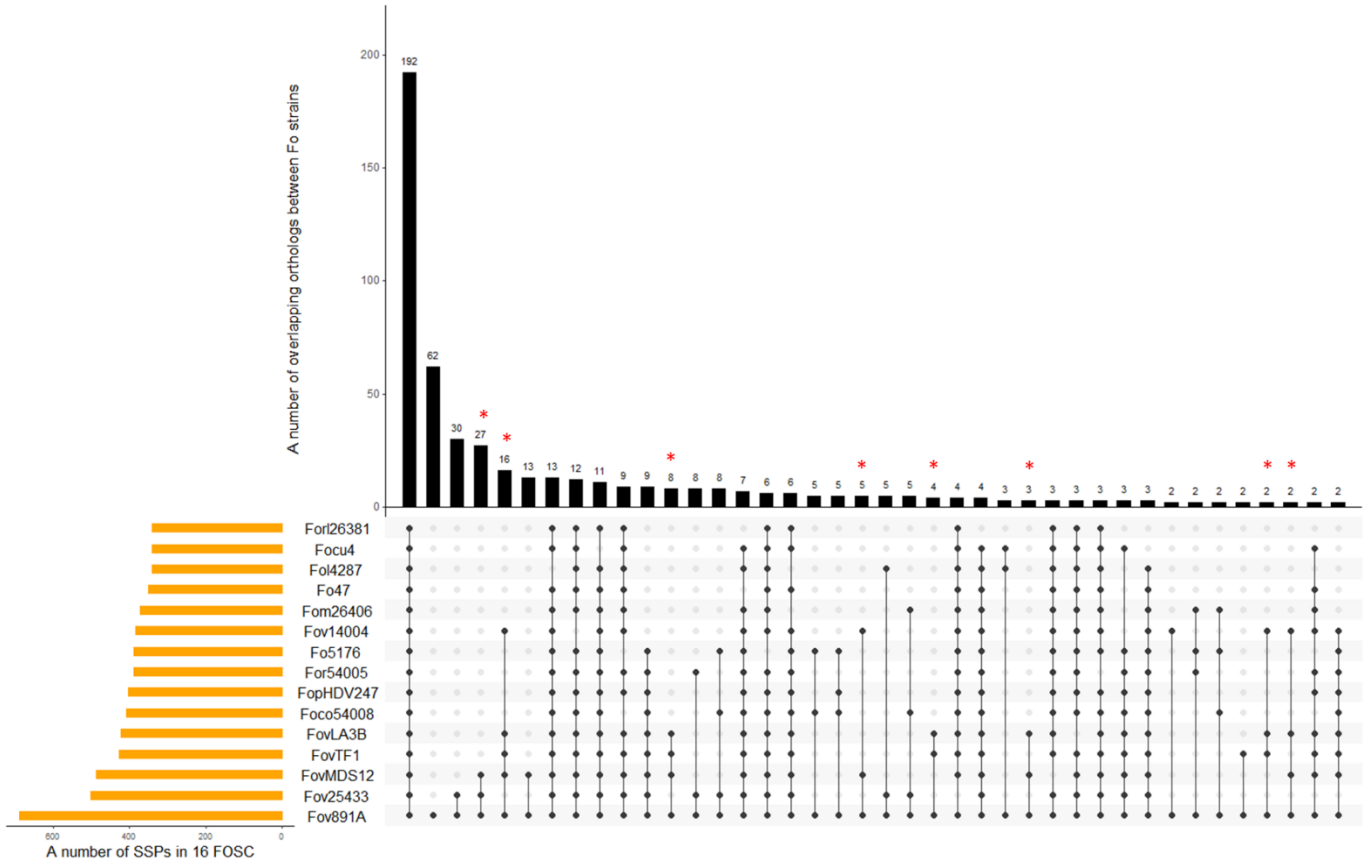


Figure 3. **Distribution of orthologues of predicted SSPs relative to *Fov* race 4 within 16 FOSC genomes**. The sets of SSPs in each FOSC isolate, were searched against 687 proteins of *Fov* race 4 to identify orthologs. The graph was drawn with the R language package, UpSetR. Each bar and above number indicate how many orthologs are intersected within Fo strains. The red stars are marked if the orthologs were shared between at least three cotton isolates. Due to the limited space, the remaining intersections with the lower numbers at the end are not included in the figure.

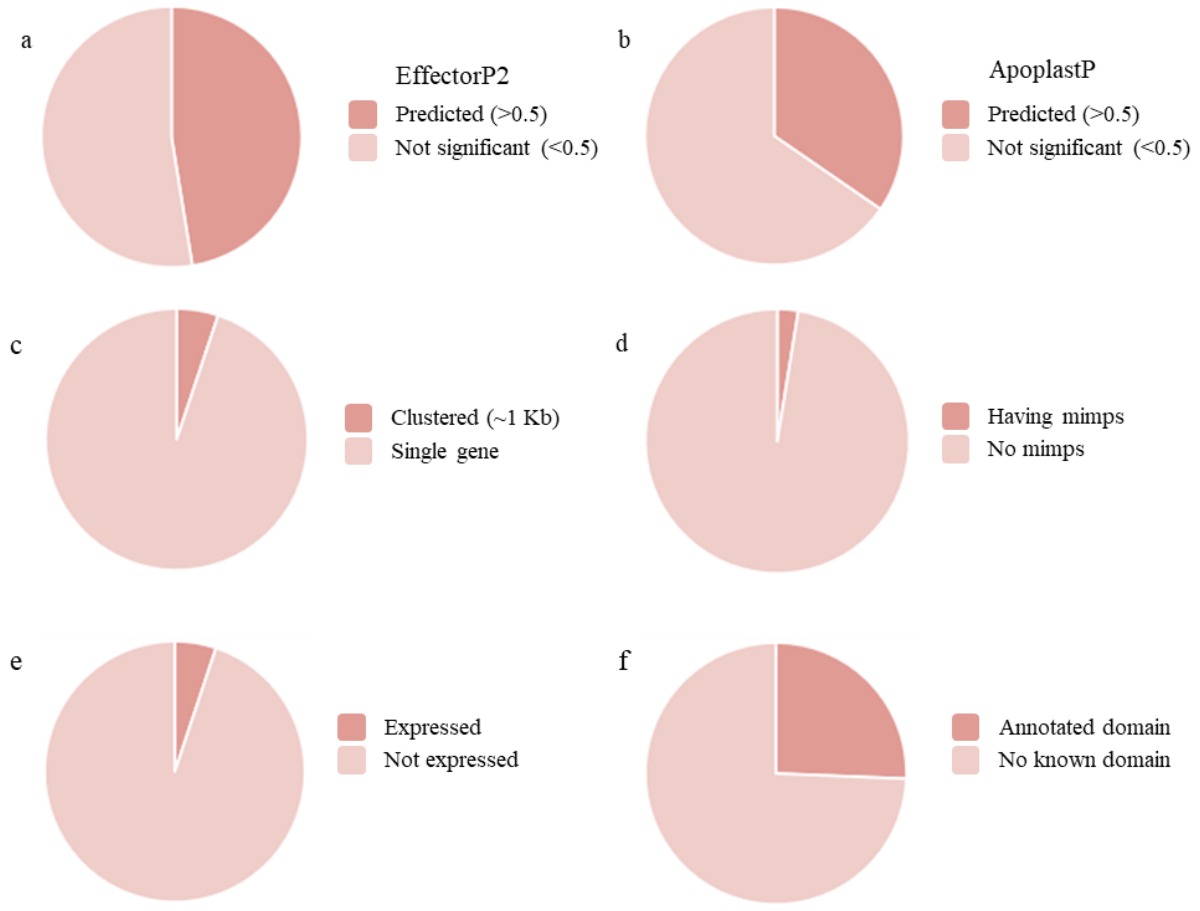


Figure 4. Graphical presentation of annotated 78 small secreted protein candidates. The pie chart shows the relative number of SSPs in each category, a-b) SSPs having significant probability predicted by EffectorP2 and ApoplastP programs, c) SSPs clustered where more than two genes reside within 1 kb, d) SSPs with mimps sequences near genes, e) SSPs expressed during gossypol treatment including up- and down-regulated genes, and f) SSPs having already known domains predicted by InterProScan.

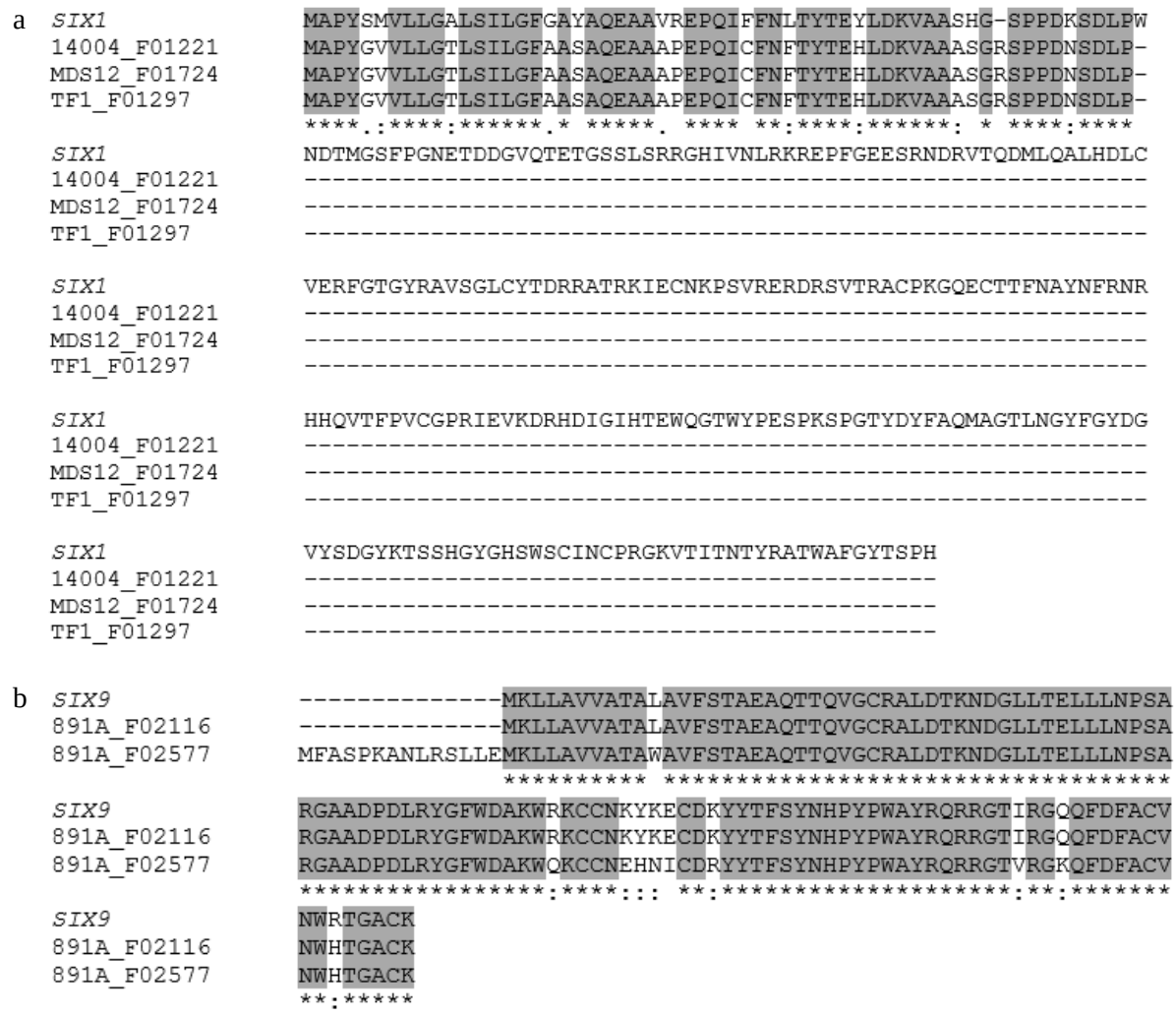


Figure 5. Amino acid sequence alignments of *SIX* genes with homologs in *Fov*. Only homologs over 75% identity are aligned with *Fol* a) *SIX1* and b) *SIX9*. The similar areas are shaded in light grey.

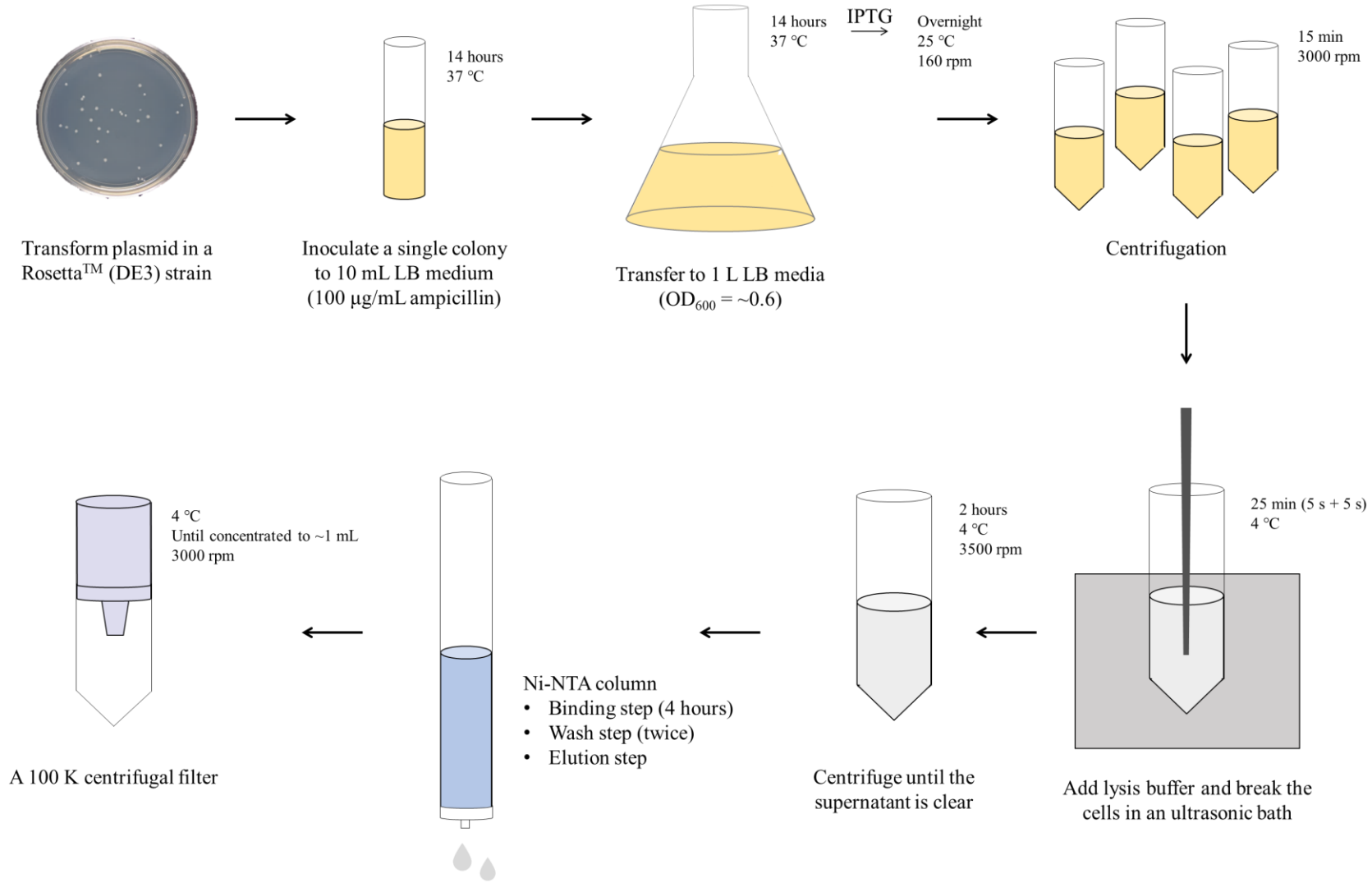
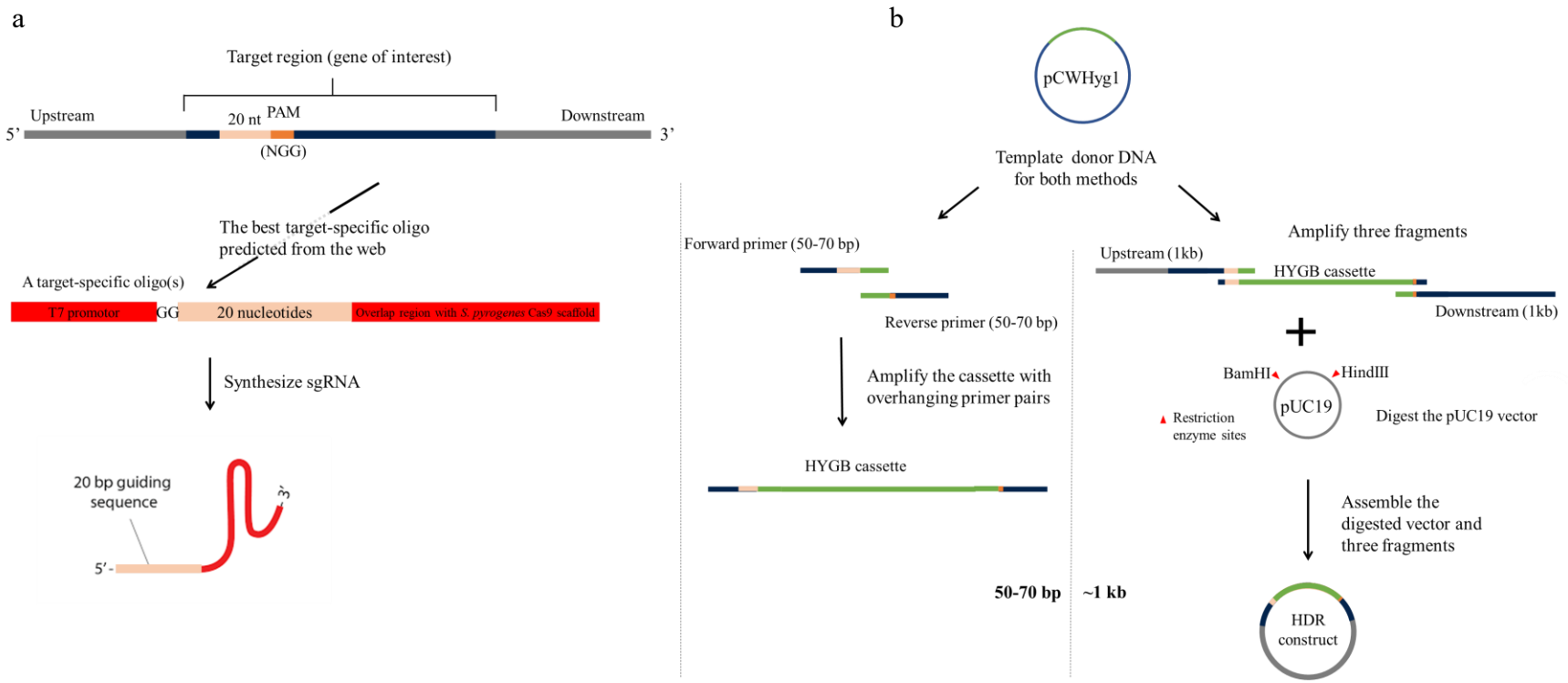


Figure 6. The schematic representation of Cas9 protein purification method using a HisPur Ni-NTA column.


a
Target region (gene of interest)
Upstream
20 nt
PAM
(NGG)
Downstream
5'
3'
The best target-specific oligo predicted from the web
A target-specific oligo(s)
T7 promotor
GG
20 nucleotides
Overlap region with S. pyrogenes Cas9 scaffold
Synthesize sgRNA
20 bp guiding sequence
5'-
3'
b
pCWHyg1
Template donor DNA for both methods
Forward primer (50-70 bp)
Reverse primer (50-70 bp)
Amplify the cassette with overhanging primer pairs
HYGB cassette
50-70 bp
Amplify three fragments
Upstream (1kb)
HYGB cassette
Downstream (1kb)
BamHI
HindIII
pUC19
Restriction enzyme sites
Digest the pUC19 vector
Assemble the digested vector and three fragments
~1 kb
HDR construct

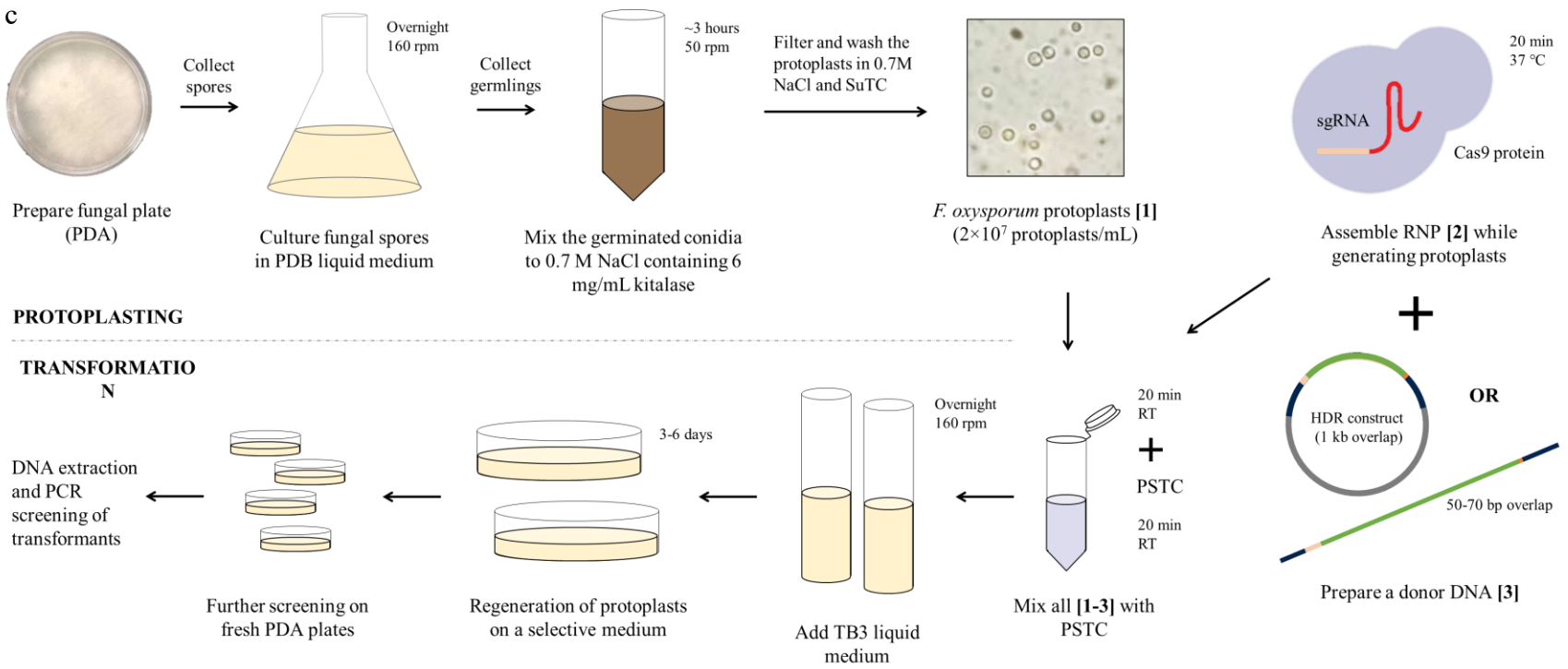


Figure 7. Schematic diagrams of sgRNA and donor DNA construction and *F. oxysporum* transformation by PEG-mediated delivery of Cas9 RNPs. a) One of the best sgRNA target site with PAM sequence is selected. To the 20-nucleotide target sequence (without NGG), T7 promoter to the 5′ end and an overlapping sequence with *S. pyrogenes* to the 3′end are added. The oligos of the target sgRNA are synthesized and the cleavage efficiency is checked after combining with Cas9 proteins. b) Donor DNA is constructed in two different ways using the pCWHyg1 plasmid. For the 50-70 bp strategy, the hygromycin cassette is amplified using primer pairs with the 50-70 bp overhanging regions. For the 1kb strategy, 1 kb upstream, 1 kb downstream, and the hygromycin cassette are amplified using different sets of overlapping primers (25bp). These three fragments are assembled into the digested pUC19 vector (digested with

restriction enzymes BamHI and HindIII). This construct is confirmed by PCR screening and sequencing with M13F and M13R primers. c) Protoplasts of *F. oxysporum* are generated using the chemical enzymes such as kitalase. They are counted to 2 x 10^7 protoplasts/ml and mix with assembled RNPs and donor DNA. The mixture regenerates in TB3 media overnight and plates with selective marker, hygromycin. The positive transformants are confirmed by amplifying the target gene sites and sequencing.

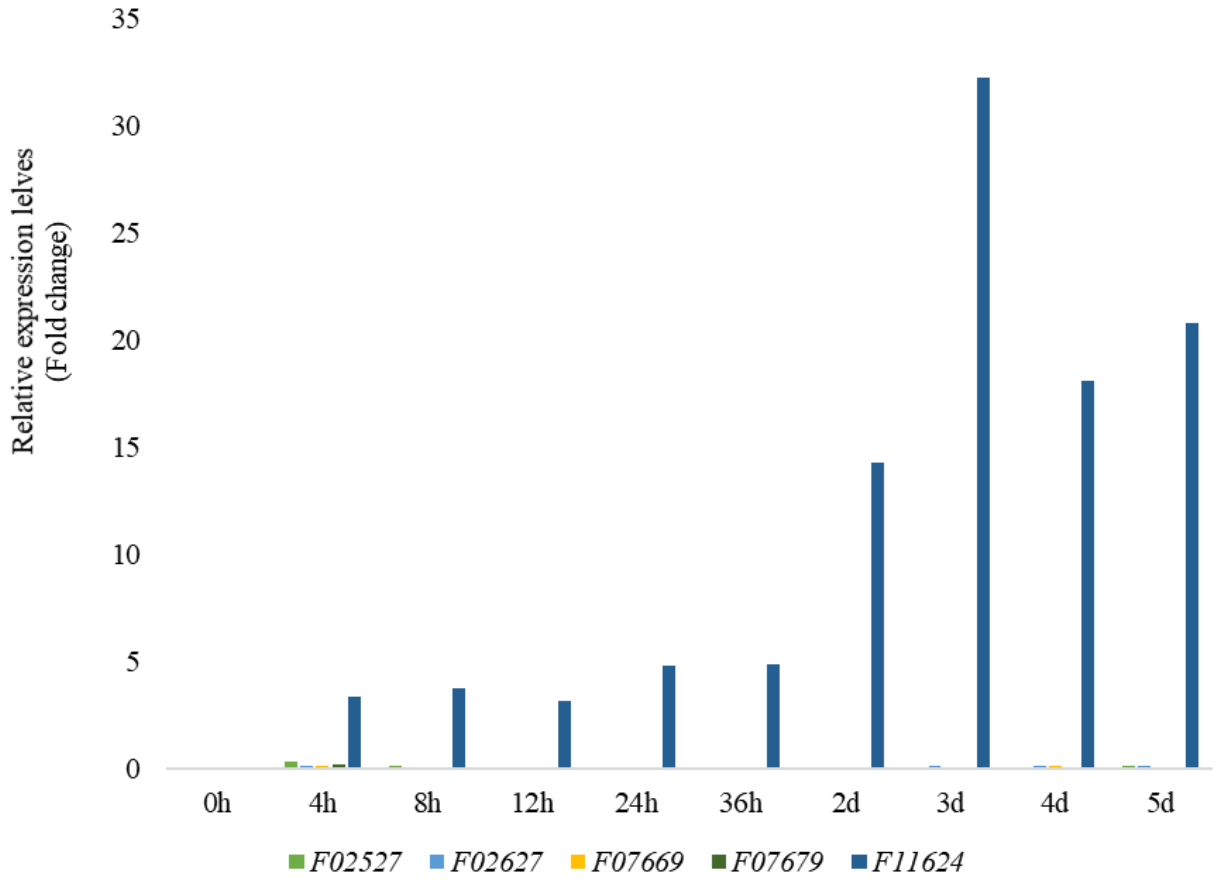


Figure 8. Gene expression of selected four Fov race 4 SSPs in infected root. FC FC, fold change in gene expression was measured by quantitative real-time PCR from Fov race 4 infected cotton (DP744) roots using two housekeeping genes, EF-1α and GAPDH from *F. oxysporum*.

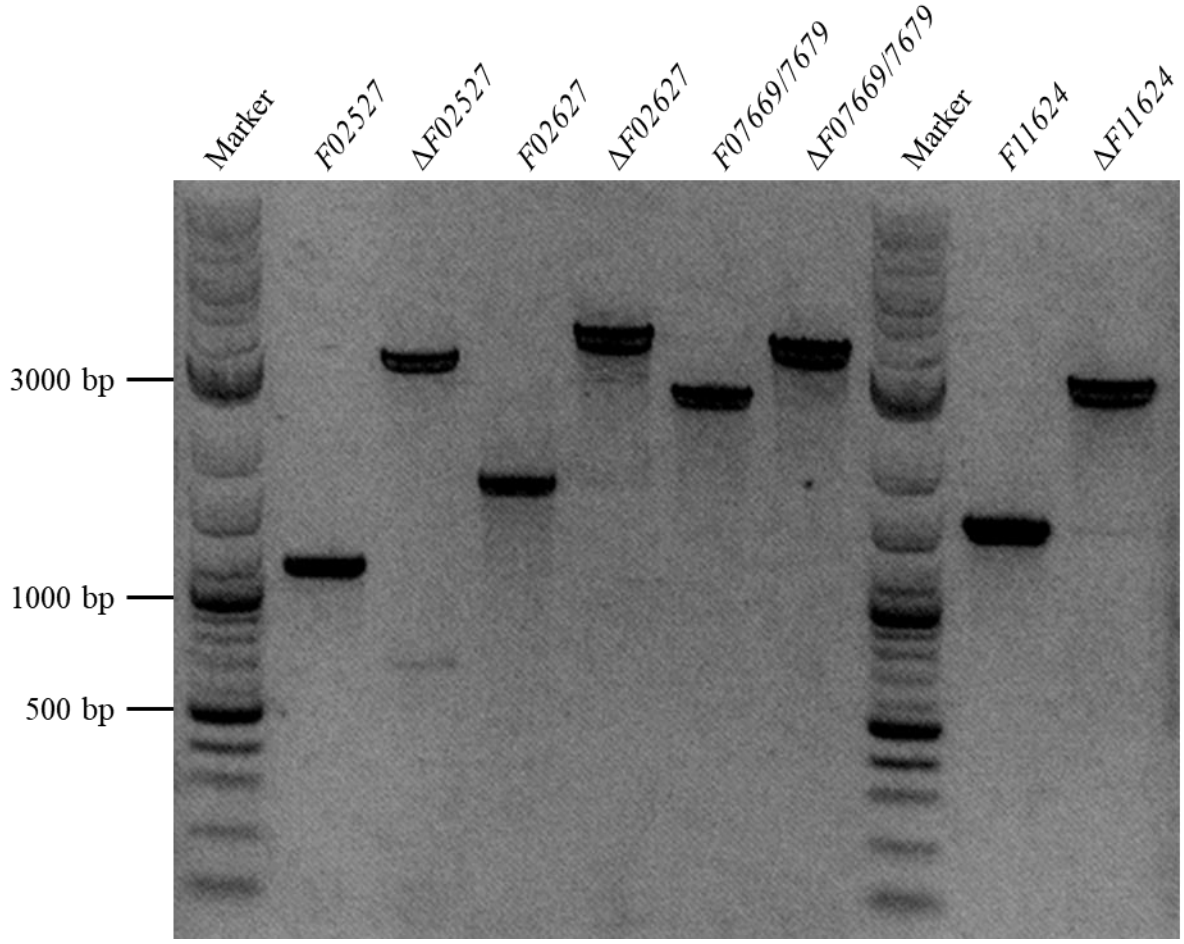


Figure 9. PCR amplification of the target gene for the template (WT) for *in vitro* cleavage and for the validation of the transformants. Expected amplicon size of the template was 1035 (*F02527*), 1560 (*F02627*), 2441 (*F7669/7679*), and 1376 bp (*F11624*). Expected amplicon size of the successfully generated gene-deletion mutants was 2898 (Δ*F02527*), 3424 (Δ*F02627*), 3259 (Δ*F7669/7679*), and 2844 (Δ*F11624*).

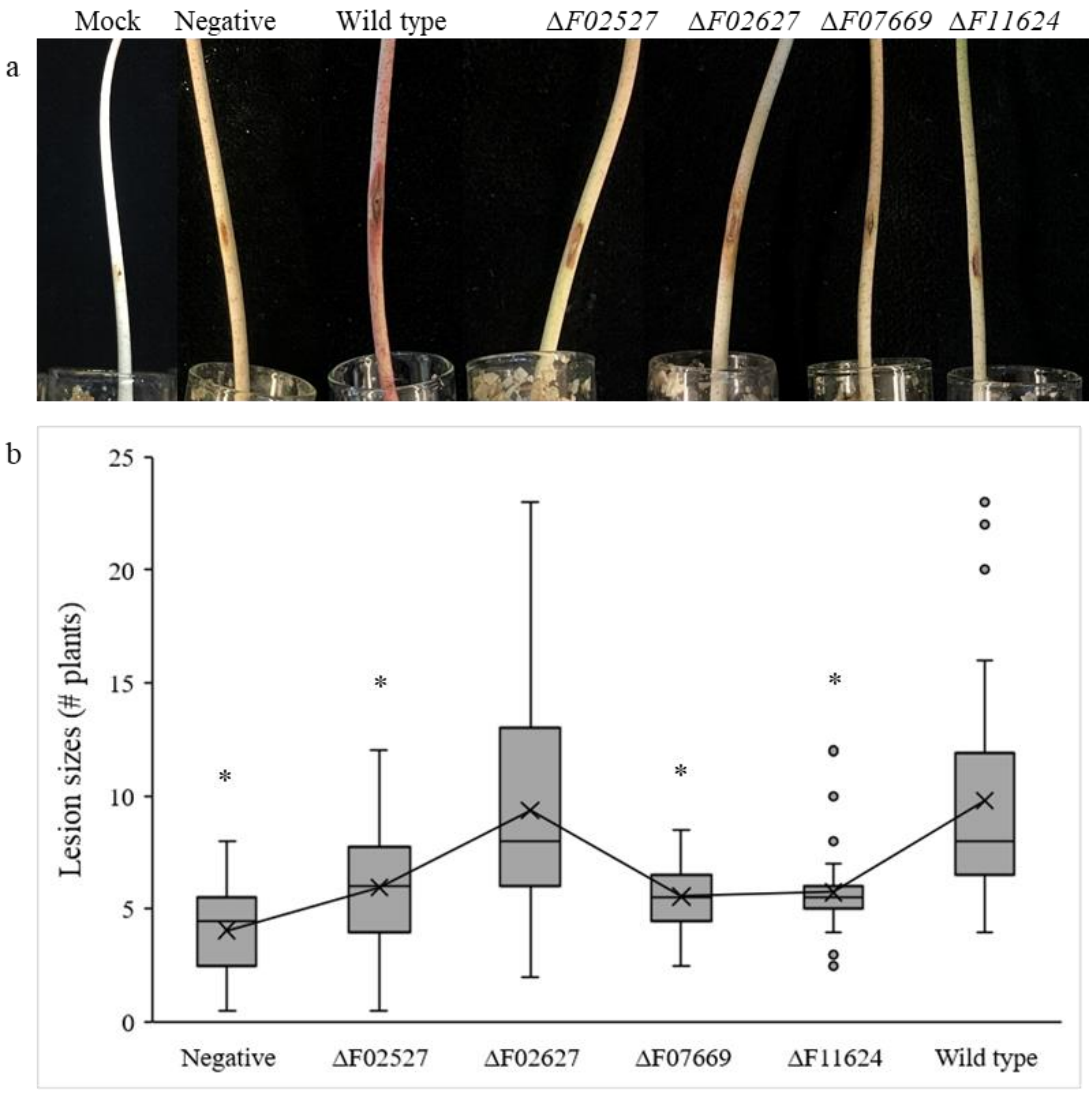


Figure 10. Assessment of SSP mutants in a stem lesion virulence assay on Pima cultivar DP744. a) Lesions produced 6-8 days after inoculation for (left to right) a mock, a FOSC isolate nonpathogenic on cotton (*F. oxysporum* f. sp. *pisi* F79), the 89-1A wild type, Δ*F02527,* Δ*F02627,* Δ*F07669/F07679*, and Δ*F11624* mutants. b) Barplots showing the distribution of lesion sizes (millimeters) based on 8-11 plants over three replicates. The symbol (*) indicates the lesion size was significantly different ($P \leq 0.05$) compared with wild type (*Fov* 89-1A).

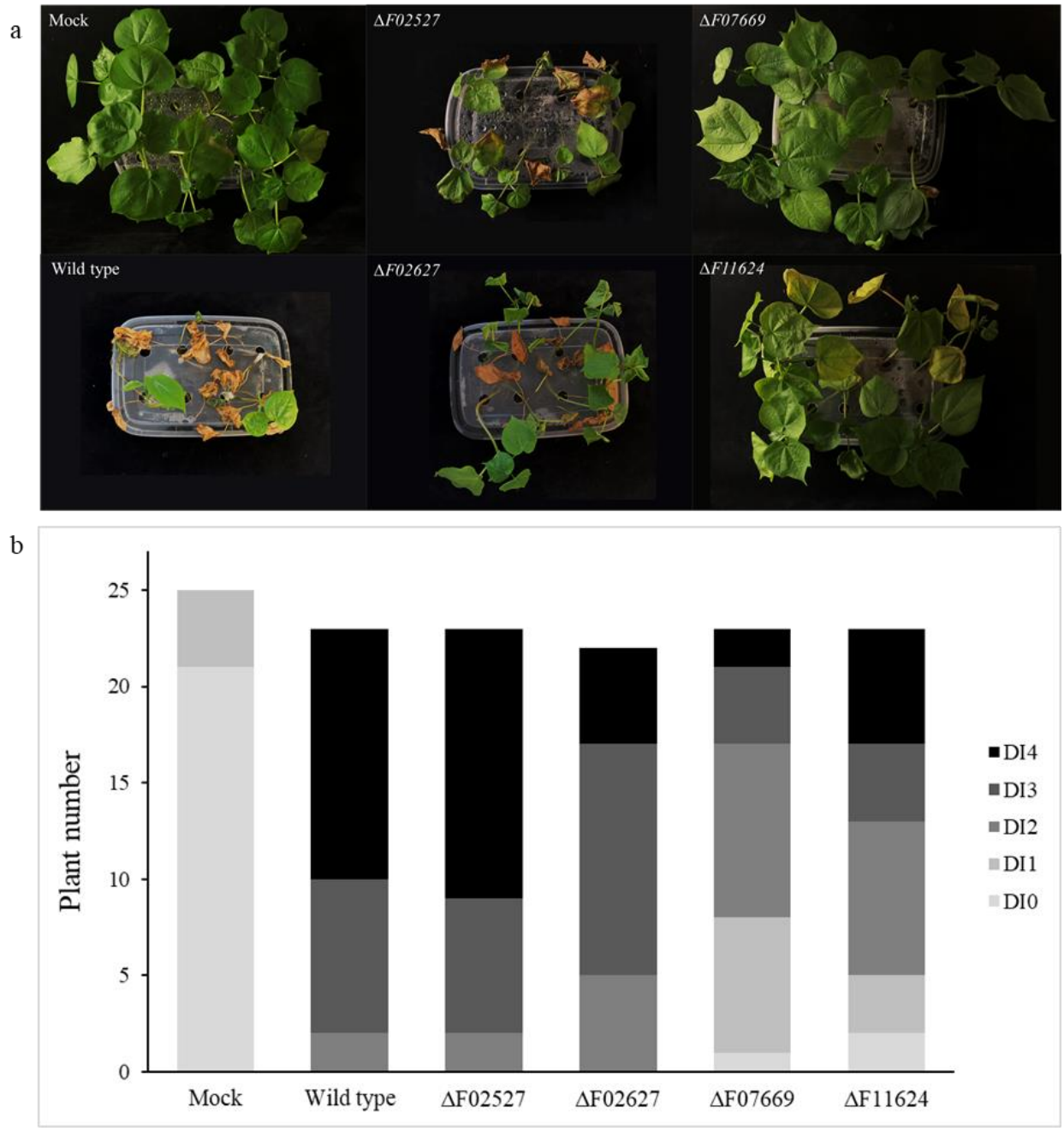


Figure 11. Virulence assay of SSPs mutants on the cotton cultivar, DP744 using hydroponic system. a) Symptoms on cotton leaves were evident two- to three-weeks post-inoculation, a mock, the wild type (89-1A), Δ*F02527,* Δ*F02627,* Δ*F07669/F07679*, and Δ*F11624* mutants. b) Disease index (DI) assigned to one of 0-4 categories. The assay was performed three times independently with 7-9 cotton plants.

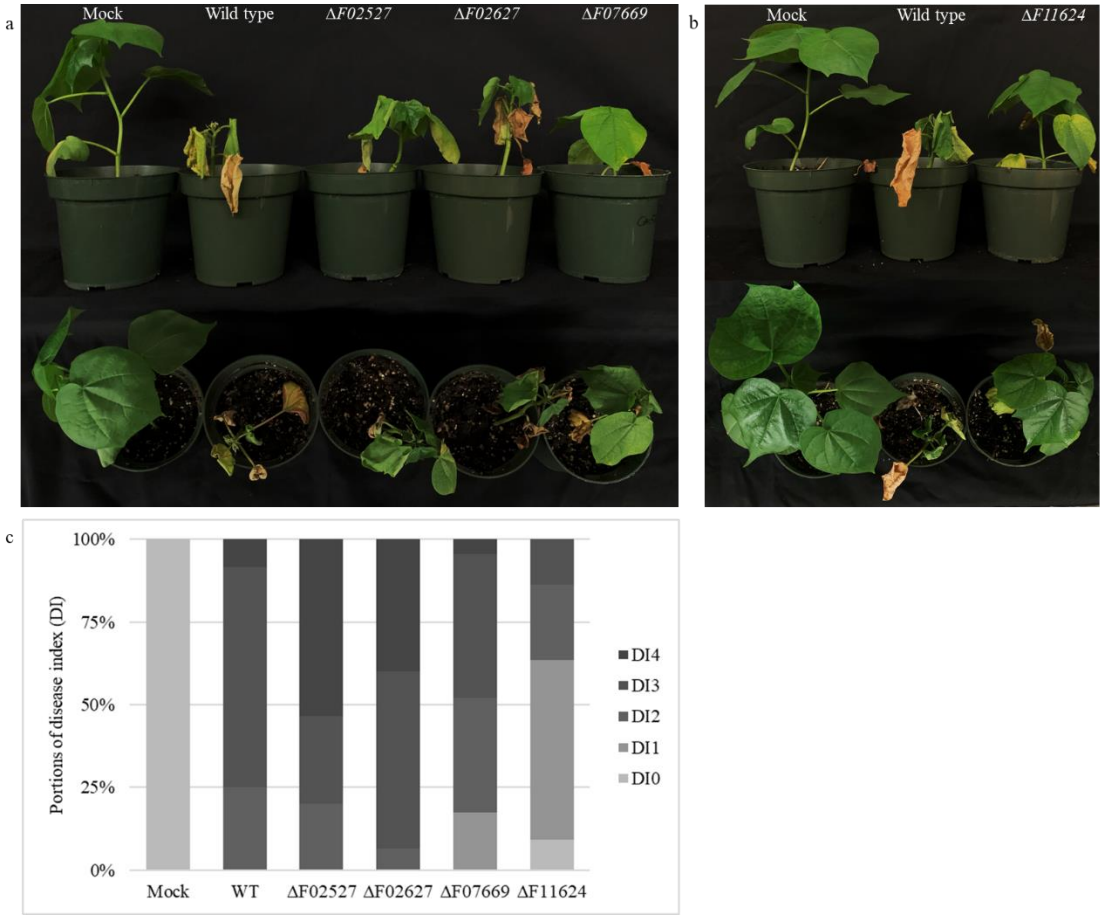


Figure 12. Assessment of virulence of putative SSP mutants using a pot assay in Pima cultivar, DP744. Two or three weeks after inoculation, the height of the cotton plants and symptoms of the leaves and roots were observed, for a mock, wild type (89-1A), Δ*F02527,* Δ*F02627,* and Δ*F07669/F07679* mutants a) and a mock, wild type, and the Δ*F11624* mutant b), and the disease index of each treatment was measured and visualized by barplots c). The pot assay was independently repeated twice with 7-9 cotton plants per treatment.

Chapter 4.

Profiling of secondary metabolites of *Fusarium oxysporum* genomes based on chromosome regions and unique clusters in isolates from cotton

Abstract

Fungal pathogens produce a broad spectrum of secondary metabolites (SMs). In fungi, SMs are not essential for growth, but provide an advantage to survive in unfavorable environmental conditions or to compete with other microorganisms for limited nutrients. In phytopathogenic fungi, SMs can serve as a virulence factor, contribuiting to infection and pathogenicity on plant hosts. However, in the *Fusarium oxysporum* species complex (FOSC), an agronomically important group of soil-borne fungi that cause vascular wilt disease on a variety of crops, in-depth studies of the secondary metabolites are rarely reported. Here, *in silico* analysis to identify secondary metabolite clusters was performed to further elucidate their biological significance within the FOSC. Twenty-one genomes of members of the FOSC including six from *F. oxysporum* f. sp. *vasinfectum* (*Fov*), a significant worldwide pathogen on cotton, were queried to identify putative SM biosynthetic gene clusters. Core and accessory regions of all the FOSC genomes were defined using the well-established *F. oxysporum* f. sp. *lycopersici* (*Fol*) 4287 genome as a reference for chromosomal structure. Altogether, comparative analysis of the SM cluster types revealed 28 clusters in the core regions are shared in all 21 FOSC genomes. Additionally, chromosome-level genomes of five *Fov* and *Fol* 4287 enabled visualizing the physical distribution of secondary metabolites in core regions. Furthermore, to understand whether specific clusters exist in the *Fov* isolates including race 7 (NRRL 25433), the type of cluster, uniqueness, and location in sub-telomeric regions were manually curated. The putative siderophore cluster in the middle of chromosome 11 (CHR12) harboring the core gene encoded by *NRPS6* was observed in all *Fov* and

Fol 4287 isolates and has high conservation. Additional putative siderophore clusters in CHR12 were found in both *Fov* 14-004 and LA3B isolates and the core gene is responsible is the *NRPS1*. The *Fov* 14-004 genome contains many unique SM clusters not observed in the other *Fov* genomes. Four isolates, TF1, 89-1A, 25433, and LA3B has one to three SM clusters encoded in the accessory region. Such fundamental studies are necessary for developing novel methods to control the threat caused by *Fusarium* and its mycotoxins and other SMs for agricultural biosecurity.

4.1 Introduction

Fungi are able to produce a broad spectrum of secondary metabolites (SMs), small organic molecules that are not essential for growth. SMs are useful for fungi to survive in unfavorable environmental conditions or to compete with other microorganisms for limited nutrients (Keller, 2019). In plant pathogenic fungi, SMs can serve as virulence factors, contributing to infection and pathogenicity on plant hosts (Pusztahelyi et al., 2015). For example, siderophores are small iron-chelating SMs that aid in iron uptake and binding of intracellular iron. They are needed for the careful management of iron due to the low solubility in the external environment and the high reactivity inside cells. Mutation of the key gene (*NRPS6*) in the biosynthetic gene cluster (BGC) producing an extracellular siderophore in *Fusarium graminearum* led to a reduction in virulence and an increase in H_2O_2 sensitivity (or oxidative stress), indicating siderophores play an important role in virulence (Oide et al., 2006). Among fungal secondary metabolites, mycotoxins (or phytotoxins) are one the most concerning due to adverse health effects to animals and humans. Deoxynivalenol, zearalenone, T-2 toxin, and fumonisin B1 are common examples among pathogenic *Fusarium* strains (Desjardins & Proctor, 2007). Fungi can be responsible for the contamination of plant hosts with one or several mycotoxins depending on the fungal isolate and environmental conditions.

Most fungi harbor the biosynthetic machinery capable of producing secondary metabolites, most importantly the key enzymes responsible for producing the "core" structure of the compounds, namely non-ribosomal peptide synthetases (NRPSs), polyketide synthases (PKSs), terpenes synthases, and prenyltransferases; while adjacent genes encode tailoring enzymes and transporters that provide methods to further diversify their molecular arsenal (Keller, 2019). During SM

biosynthesis, these key enzymes can be further modified by oxidation, reduction, methylation, or glycosylation (Graham-Taylor et al., 2020).

The *F. oxysporum* species complex (FOSC) comprises an agronomically important group of soil-borne filamentous fungi including many plant pathogens. The FOSC has a very wide range of plant hosts, but individual isolates have limited pathogenicity to one or only a few host species. Such isolates are grouped together into *formae speciales* (f. sp.) according to host specificity. For successful infection of their host plant, fungal pathogens secret a variety of virulence factors including effectors and secondary metabolites. Members of the FOSC are well-known for causing vascular wilt disease, but they are not considered major producers of mycotoxins that could harm humans and livestock (Munkvold, 2017). Nevertheless, these fungi have been reported to produce various secondary metabolites, including enniatins, fusaric acid, and moniliformin (Desjardins & Proctor, 2007). Following the drastic increase in available fungal genomes and the development of bioinformatic tools, comparative studies have revealed SM potentials as well as evolutionary relationships within *Fusarium* (Cairns & Meyer, 2017). Almost the entire repertoire of biosynthetic gene clusters, including key enzymes and adjunct genes, are capable of being predicted *in silico*. Such fundamental studies are necessary for developing novel methods to control the threat caused by *Fusarium* and its toxins for agricultural biosecurity.

In this study, an *in silico* prediction and a comparison analysis of SM biosynthetic gene clusters was conducted using the complete genomes of 21 *F. oxysporum* isolates from various host plants and habitats. The bioinformatic pipeline antiSMASH was used to identify the SM clusters in these fungi, and subsequently, chromosomes (or scaffolds) of the FOSC isolates were defined into core and accessory regions using the well characterized *F. oxysporum* f. sp. *lycopersici* 4287 (*Fol*) genome as a reference. All the clusters in the core chromosomes were classified and the

common clusters among the 21 *F. oxysporum* were identified. The remaining clusters which were not in the core regions were considered as part of the accessory genome. Our main interest was the SM clusters found in the genomes of *F. oxysporum* f. sp. *vasinfectum* (*Fov*). Therefore, the objectives of this study are: 1) identify putative SM biosynthetic gene clusters in the six *Fov* isolates including a previously sequenced race 7 isolate; 2) recognize which clusters are unique in the cotton isolates and absent in other FOSC isolates; and 3) investigate in detail the clusters that are specific for the *Fov* race 4 genome and could be contributing to high levels of virulence in cotton.

4.2 Materials and Methods

4.2.1 Genome sequence and bioinformatics analysis

The genomes of 16 *F. oxysporum* isolates were obtained from GenBank and were chosen based on the high genome quality and if the isolate has been thoroughly studied (Table 1). The five *Fov* genomes were previously sequenced and assembled and are publicly available (Seo et al., 2020). To investigate the secondary metabolites, the twenty-one FOSC genome sequences were analyzed by antiSMASH v.4 (Blin et al., 2017) with the built-in features including ClusterBlast, and KnownClusterBlast. The cutoff for a known cluster was over 40% similarity.

4.2.2 Mapping SM gene clusters to genomes

To determine the chromosomal location of each SM gene cluster, core and lineage-specific (LS or accessory) regions were identified in individual genomes using MUMmer v.3.22 (nucmer) (Kurtz et al., 2004) by aligning them to the *Fol* 4287 assembly. The resulting tab-separated files of the aligned regions to both the reference and the query were used to decide presence or absence regions in the genomes via Bedtools v.2.29.2 (genomecov) (Quinlan & Hall, 2010) where the

percentage covered by matching sequences of each scaffold (query) was calculated with a cutoff of 30% identity. Finally, the scaffolds (or contigs) over 30% were assigned to potential core chromosomes, while the remaining scaffolds were considered as the accessory genome (Williams et al., 2016). Combining the antiSMASH results with the location on core chromosomes, the SM clusters in the core regions were arranged following the order of the core chromosomes of *Fol* 4287 except the LS chromosomes (3, 6, 14, and 15). The types of SM clusters which existed throughout all 21 *F. oxysporum* genomes in the core regions were identified. In particular, the common clusters in *Fol* 4287 and the *Fov* isolates were extracted to visualize the cluster distribution on the chromosomes by the R package, chromoMAP (Anand & Rodriguez Lopez, 2019). In addition, the rest of SM clusters in the *Fov* genomes that were not included in common clusters were visualized on the core regions.

When hybrid clusters were found on the same chromosomal location such as NRPS-NRPSlike and NRPSlike-NRPS, NRPS-T1PKS and T1PKS-NRPS, and NRPSlike-T1PKS and T1PKS-NRPSlike, manual inspection of these clusters throughout all FOSC genomes were performed. The scaffolds (or contigs) have opposite directions between FOSC isolates but the entire components and sequences are almost identical between the hybrid clusters. Thus, these clusters were considered as the same and merged.

4.2.3 *Fov* specific bioinformatics analysis

In order to further characterize the putative SM clusters in the *Fov* genomes, the additional genes from the secondary metabolism Clusters of Orthologous Groups (smCOGs) analysis predicted by antiSMASH were searched manually. If necessary, the amino acid sequences of the core biosynthetic gene in the cluster were extracted and searched by BLASTP. In addition, to

understand the cluster location near telomeres, the SM clusters within 300 kb of the end of the chromosome were investigated throughout *Fov* isolates and *Fol* 4287 for the comparison.

Specifically, three genes, *NRPS1, 2,* and *6*, responsible in the production of siderophores and involved in virulence in *F. graminearum*, were obtained from GenBank (Oide et al., 2006, 2007, 2015). Putative homologs related to siderophore biosynthesis in *Fol* 4287 and the six *Fov* were identified using BLASTP searches with a cutoff e-value of 0.001. Additionally, the cluster containing siderophores potentially responsible for the *NRPS6* biosynthesis in the same chromosomal location (CHR11) among each *Fov* strain as well as *Fol* 4287 was extracted to compare the sequence similarity. The alignments by built-in BLASTN were visualized using EASYFIG v.2.2.5 (Sullivan et al., 2011).

4.3 Results

4.3.1 Inventory of putative secondary metabolite cluster types in 21 FOSC genomes

To assess the potential secondary metabolic profiles that exist within the genomes of the FOSC isolates, the genome sequences were searched using antiSMASH v.4 (Blin et al., 2017). Genome sizes ranged from 46-63 Mb, where the smallest was *F. oxysporum* f. sp. *cubense* TR4 and the largest was *F. oxysporum* f. sp. *vasinfectum* race 4 (Table 1). Among the 21 *F. oxysporum* genomes the antiSMASH program predicted, the number of SM clusters ranged from 38 (*F. oxysporum* f. sp. *cubense* race 4) to 50 (*F. oxysporum* f. sp. *conglutinans* 54008) (Table 2). The putative SMs were classified as 14 different cluster types, but NRPS, PKS, the hybrid NRPS/PKS, and terpene synthase backbone enzymes were the predominant core biosynthetic enzymes in the clusters. The NRPS-like was categorized different from NRPS since it was fragmented. Frequently, the core biosynthetic genes were in combination with another type of core SM gene (e.g., NRPS-

T1PKS) or others such as indole diterpene (e.g., PKS-indole). Bacteriocin, beta-lactone (related to antibiotic compound biosynthesis), and CDPS (tRNA-dependent cyclodipeptide synthases) were also observed. From the module-specific information of antiSMASH, bacteriocin was classified as an unspecified ribosomally synthesized and post-translationally modified peptide product (RiPP) cluster and renamed to RiPP-like in the new version antiSMASH v.5. Only one or two of these RiPP clusters were detected in each *F. oxysporum* genome.

4.3.2 Common clusters in core regions

F. oxysporum f. sp. *lycopersici* (*Fol*) 4287 was used as a reference genome to map the whole genome of individual isolates to distinguish "core" and "accessory" chromosomes/scaffolds. For consistency among the core and accessory regions in the 21 FOSC genomes, the scaffolds of each *F. oxysporum* genome exceeding its coverage over 30% sequence identity were adopted to select and assign to the core regions. Once the FOSC genomes were separated into the core and accessory genomes, the SM gene clusters were assigned as being either part of the core or accessory. The number of SM gene clusters in the core chromosomes/scaffolds ranged from 35 (*F. oxysporum* f. sp. *raphani* 54005) to 46 (*F. oxysporum* f. sp. *melonis* 26406 and f. sp. *vasinfectum* 14-004) (Table 3).

Following the core chromosomes of *Fol* 4287, the SM clusters were arranged on one of the 11 core chromosomes (Table 3). Clusters that were shared between at least 19 of the 21 FOSC genomes and resided on a core chromosome were considered to be a common SM cluster. A total of 28 clusters were common to all 21 FOSC genomes (Figure 1 and Table 3). The distribution of the clusters was visualized for the common clusters as well as specific clusters only in *Fov* genomes (Figure 1). CHR01 encoded three common clusters, one NRPS and two NRPS-like, while CHR02 carried three clusters, an NRPS cluster, a type 1 PKS cluster known to produce

oxyjavanicin, and a terpene synthase cluster responsible for synthesis of squalestatin S1 (Tables 3 and 4). CHR04 had two common clusters, a NRPS-like and a type 1 PKS-NRPS hybrid. The largest number were present on CHR05 with five clusters, a type 1 PKS-NRPS-like hybrid responsible for fusaric acid biosynthesis, a type 1 PKS-NRPS hybrid that potentially produces equisetin, a beta-lactone, as well as two clusters with NRPS-like enzymes as the core biosynthetic unit. CHR07 had three clusters with a NRPS, a NRPS-like, and a terpene synthase. Two type 1 PKS and one type 3 PKS was on CHR08. CHR09 had only one NRPS, which was not found in the Fo5176 genome. CHR10 had two clusters, a NRPS and an indole, and CHR11 had only a single NRPS-like containing cluster. CHR12 had two clusters, an indole and a terpene synthase. Lastly, CHR13 had three clusters, two terpene synthase containing clusters and one type 1 PKS-NRPS or type 1 PKS missing in *Fom* 26406. The reason why the type 1 PKS-NRPS cluster on CHR13 was considered as one of the common clusters was that although Fo5176 and Foco54008 had the type 1 PKS type instead of the hybrid of PKS and NRPS found in the other 18 FOSC isolates, this cluster is similar to the known SM cluster responsible for the synthesis of gibepyrone-A, with more than 40% similarity (Tables 3 and 4).

4.3.3 SM clusters in accessory regions

After assigning secondary metabolite biosynthetic genes to the core genome, by default the remaining SM gene clusters were placed in the accessory genome for their respective isolates. There was high variability in the number and diversity of the SM biosynthetic genes in the accessory genomes of members of the FOSC. Consequently, in *Fol* 4287 used as the reference for the genome structure, five clusters were present in the accessory genome, two NRPS encoding genes, and single clusters containing a type 1 PKS, a terpene synthase, and a beta-lactone (Table 5). The number of SM biosynthetic gene clusters in the accessory genomes, ranged from 1 (*F.*

oxysporum f. sp. *radices-cucumerinum* 016 and f. sp. *vasinfectum* LA3B) to 9 (*F. oxysporum* f. sp. *conglutinans* 54008). No clusters were detected in the accessory genomes for three isolates, the clinical *F. oxysporum* NRRL 32931 isolate and *F. oxysporum* f. sp. *vasinfectum* isolates 14-004 and MDS-12. Only clusters in *F. oxysporum* f. sp. *pisi* HDV247 and f. sp. *raphani* 54005 were a known biosynthetic function able to be postulated for an SM in the accessory genome, enniatin (100%) and gibberellin (71%), respectively.

4.3.4 Identification of *Fov* specific clusters on core regions

The prediction of the SM clusters by antiSMASH in six *Fov* isolates was not distinctive when compared to other FOSC isolates in terms of either the overall number of clusters nor the number found in the core and accessory genomes (Tables 2, 3, and 5). To understand whether specific clusters exist in the six *Fov* isolates including race 7 (25433), the type of cluster, uniqueness, and location in sub-telomeric regions were manually searched (Figure 2). The common clusters in the *Fov* genomes were identified using the existence in at least two *Fov* isolates as the cutoff. Only five *Fov* isolates, the exception being race 7, were visualized in the bar plots due to the fragmentation of the race 7 genome (Figure 2). The clusters in the sub-telomeric regions were defined as being within 300 kb of each chromosome end (5′ and 3′).

The *Fov* 14-004 genome had a distinctly different arsenal of SM clusters when compared to genomes of the other *Fov* isolates, where eight clusters were unique and not found in the other *Fov* isolates. Four clusters were distributed in the middle of chromosomes 11-13 and another four clusters were located in the sub-telomeric regions of CHR04 and CHR11 and there were two in CHR09. Among the four unique clusters in *Fov* 14-004, there were three clusters with a predicted function with over 40% similarity to SM clusters in fungal species: biosynthesis of hexadehydroastechrome/ terezine-D/ astechrome (62%) in the 5′ subtelomeric region of CHR09,

α-acorenol in CHR11 (100%), and fujikurin A/ B/ C/ D in CHR13 (50%). Another isolate which had a unique cluster was *Fov* LA3B, with a hybrid cluster of a NRPS and type 1 PKS on CHR12 with an unknown function. Interestingly, a shared NRPS-type 1 PKS type in five *Fov* and several other FOSC genomes was located on the same chromosome, CHR12, and were postulated to be responsible for ACT-toxin II biosynthesis with 100% similarity (Table 4).

Several SM clusters specific for *Fov* were identified in multiple isolates: one type 1 PKS was in three isolates (TF1, LA3B, and 14-004) and resided on CHR01; a terpene synthase was in all isolates except *Fov* LA3B on CHR02; one cluster containing a NRPS-like core enzyme was located on CHR05 in four isolates with the exceptions being *Fov* 89-1A and MDS-12; a NRPS cluster on CHR08 was present in all the *Fov* isolates; one beta-lactone in five *Fov* isolates except *Fov* race 7 on CHR10 and one of *Fov* TF1 located in the sub-telomeric regions, respectively; a NRPS, NRPS-like cluster was present in four isolates (TF1, LA3B, 89-1A, and MDS-12) and a NRPS cluster was shared between the *Fov* race 7 isolate 25433 and 14-004 on CHR11; and a hybrid of a NRPS-type 1 PKS cluster (related to the ACT-toxin II biosynthetic cluster mentioned above) on CHR12 in five *Fov* isolates, with the exception being LA3B (Figure 2).

Three *Fov* genomes, the race 4, 7, and MDS-12 isolates, had a type 1 PKS cluster on CHR08 which was potentially responsible for depudecin synthesis (83%) (Figure 2 and Table 4). At the sub-telomeric region on CHR12, a NRPS cluster was present in five *Fov* isolates and absent in race 7 (25433). An additional NRPS cluster in *Fov* 89-1A, 14-004, and MDS-12 in a sub-telomeric region of CHR12. In the middle of chromosome 12, a terpene synthase cluster with 100% similarity to proteins involved in koraiol synthesis are found in the *Fov* race 7 and MDS-12 genomes (Table 4). Lastly, in the sub-telomeric region of CHR13, a NRPS cluster present in *Fov* TF1 and LA3B is similar to AbT1 (100%), a precursor of the cyclic peptide aureobasidin A (AbA)

produced by *Aureobasidium pullulans* which has toxicity against several other fungal species such as *Aspergillus* (Slightom et al., 2009). An additional NRPS cluster is present in close proximity in *Fov* MDS-12.

Among the 28 common clusters in all the FOSC genomes, the clusters which were located in the sub-telomeric regions in the *Fov* genomes was further investigated. Two clusters were present in opposite sub-telomeric ends of chromosome 5 in *Fov* race 4 (89-1A), were one was predicted to produce fusaric acid (Fub1) (59%). The cluster to produce fusaric acid was common throughout the FOSC but was not predicted to reside within the sub-telomeric regions in other FOSC genomes. Three isolates (TF1, LA3B, and 14-004) had a cluster containing a NRPS as the core enzyme at the end of CHR09, and two clusters common between the *Fov* TF1 and LA3B genomes were adjacent to each other in the sub-telomeric region on CHR10 and contained a NRPS and indole.

Two SM clusters with a type 1 PKS and terpene synthase as the main core enzymes resided in close proximity in the sub-telomeric region of CHR07 in all the *Fov* genomes**.** The terpene synthase encoded enzymes responsible for gibberellin biosynthesis with 57% similarity in three *Fov* isolates (TF1, LA3B, and MDS-12), and had lower similarity (28%) in the *Fov* race 4 and 7 genomes but was absent in the *Fov* 14-004 genome. Additionally, a type 1 PKS cluster responsible for the synthesis of bikaverin with 71-100% similarity was in the subtelomeric regions (~300 kb) in *Fov* 89-1A and 14-004. Furthermore, the entire bikaverin cluster in TF1, LA3B, and MDS-12 was less than 400 kb from the chromosomal end. Interestingly, the SM clusters of the *Fov* race 4 (89-1A) on CHR07 had a rearrangement event between a type 1 PKS and terpene synthase, which was evident when compared to other *Fov* isolates (Figure 2).

4.3.5 Additional studies of the SM clusters among Fov isolates

The accessory regions of the *Fov* genomes had unique SM clusters, in particular PKS or NRPS clusters depending on the isolates (Table 6). *Fov* TF1 has two PKS containing clusters with an unknown function. In the *Fov* race 4 (89-1A) and 7 (25433) isolates, two additional SM synthetic clusters were found in the accessory regions, containing NRPS as the main enzymes. *Fov* LA3B has an additional NRPS cluster in ACC01 (Table 6). However, the BLASTP results between the core biosynthetic genes revealed no sequence similarity at all. However, BLAST analysis of the core synthetic genes against non-redundant (nr) protein sequences showed more functional annotation, providing more specific cluster functions (Table 6).

4.3.6 Siderophore biosynthesis in *Fov*

To identify putative homologs of the core biosynthetic genes related to siderophores, the amino acid sequences of *NRPS1, 2,* and *6* of *F. graminearum,* were identified in the *Fov* and *Fol* 4287 proteomes. A single conserved siderophore NRPS protein was identified in each proteome except for *Fov* 14-004 (Table 7). In the *Fov* 14-004 isolate, two genes covered the entire length of the *NRPS6* gene with high sequence similarity over 85%. Of the three NRPS proteins involved in siderophore biosynthesis, *NRPS6* was the most conserved among the *Fusarium* spp. (Table 7).

From antiSMASH, the entire cluster responsible to encode *NRSP6* was predicted. Initially, one NRPS cluster in *Fov* 14-004 and MDS-12 (2) and one NRPS,NRPS-like cluster in *Fol* 4287 and *Fov* TF1, 89-1A, LA3B, and MDS-12 (5) on the eleventh chromosome (CHR11) contained the core biosynthetic gene encoding siderophores (Figure 2). The five *Fo* isolates having the NRPS/NRPS-like cluster, has two core genes in each NRPS and NRPS-like region. The core gene in the NRPS-like region has no match with a known function including siderophores. On the other hand, the core biosynthetic gene in the NRPS region of five *Fo* isolates had high sequence similarity with the *NRPS6* (94-100%) than two *NRPS1,* and *2*. The *Fov* MDS-12 had the same

results with other *Fo* isolates (99% similarity). Although two core genes in the NRPS cluster of the *Fov* 14-004 had lower similarity 77 and 60 % with *NRPS6*, the similarity with *NRPS6* was higher than other two genes. Additionally, the highly conserved sequences throughout the entire putative siderophore synthesis clusters of seven *Fo* isolates not only the core biosynthetic gene were observed and visualized vie BLASTN and EASYFIG (Figure 3). Thus, it suggests the *Fov* isolates produce the highly conserved extracellular siderophore, either fusarinin or triacetylfusarinine (Oide et al., 2006).

Furthermore, in two *Fov* isolates, LA3B and 14-004, the additional NRPS cluster related to the siderophore biosynthesis is present on CHR12. The core biosynthetic gene in each NRPS cluster has high sequence similarity with *NRPS1* in *F. graminearum*, which is potentially responsible in a second extracellular siderophore termed malonichrome (Oide et al., 2015).

4.4 Discussion

The repertoire of secondary metabolite gene clusters in *Fusarium oxysporum* species is primarily unique to each isolate. Genomic data of a total of 21 FOSC isolates were used to identify putative SM clusters via the antiSMASH prediction program. The number and type of SM clusters varied depending on the isolates and the total number of SM clusters ranged from 38 to 50. The SM clusters were separated into core and accessory (or lineage-specific) regions, where common clusters in the core regions had high sequence similarity within the FOSC genomes. Twenty-eight clusters were common in the core regions throughout all 21 FOSC genomes. The SM product of some of the clusters was predicted by antiSMASH, and five clusters among the 28 common ones are hypothesized to produce oxyjavanicin and squalestatin S1 (CHR02), fusaric acid and equisetin (CHR05), and gibepyrone-A (CHR13). Some of the known clusters are considered mycotoxins and have been shown to be involved in virulence, depudecin as well as gibepyrone-A and fusaric

acid (Hoogendoorn et al., 2018; Janevska et al., 2016; Studt et al., 2016; Wight et al., 2009). Most of the FOSC genomes harbor a PKS cluster on CHR07 responsible for bikaverin, the SM responsible for the red pigmentation (Wang et al., 2018). However, the SM produced by most of the SM clusters were unable to be predicted. The role of the SM clusters may be particularly significant for the success of members of the FOSC to adapt to the specific ecological niche or lifestyle.

The evolution of biosynthetic diversity in the *Fusarium* species was revealed via searching for the absence/presence matrix and ancestral state reconstruction (Hoogendoorn et al., 2018). Among the eight *Fusarium* species in the study, two *F. oxysporum* isolates, the biocontrol strain Fo47 and the tomato pathogen *Fol* 4287 were included (Reynolds et al., 2017). As with this study, the PKS cluster for the depudecin synthesis was absent in Fo47, but present in *Fol* 4287. The authors hypothesized that the depudecin biosynthetic cluster was recently acquired during evolution and partially lost in Fo47 resulting in the truncation of the genes and a non-functional cluster (Reynolds et al., 2017). Based on the location of the cluster in core regions (CHR08), the degeneration of the depudecin cluster is more likely to explain the variation of the clusters rather than horizontal chromosome transfer. In the same manner, this correlating acquisition or degeneration event of secondary metabolites during evolution increases probability that each *F. oxysporum* isolate has a variety of SM candidates even between closely related isolates.

With regards to the distribution of the SM clusters on the core chromosomes of the *Fov* genomes, the majority of the common 28 clusters were spread throughout the length of the chromosomes, although some of the *Fov* genomes had up to three common clusters in sub-telomeric regions. After excluding the 28 common clusters in the 21 FOSC genomes and analyzing the remaining clusters in the *Fov* isolates, several unique clusters in a single *Fov* isolate were found.

Generally, common clusters which were shared between at least two *Fov* isolates resided at subtelomeric or telomere proximal locations. Intriguingly, the *Fov* race 4 (89-1A) genome with only few unique clusters in its genome had a potential chromosomal rearrangement on chromosome 7. A NRPS-like and a Type 1 PKS cluster were located near the telomere of CHR07, with approximately a 100 kb gap between them. The NRPS-like cluster is one of the 28 common clusters, while the type 1 PKS cluster responsible for bikaverin biosynthesis was conserved in most isolates except three. The distribution of these two clusters existing in four other *Fov* isolates indicates that the NRPS-like cluster was closest to the telomeres, and the T1PKS cluster follows closer to the centromere (Figure 7). However, the order differs in *Fov* 89-1A, where the T1PKS resides proximal to the subtelomeric region.

The intra- or inter- chromosome homologous recombination of repetitive regions can cause deletions and translocations, and small deletions and duplications in core chromosomes have been frequently observed in *Fol* as well as *Zymoseptoria tritici*, especially at the end of chromosomes or in close proximity to repetitive elements (Li et al., 2020; Møller et al., 2018). Repeat-rich sub-telomeric regions consequently undergo chromosomal rearrangements, resulting in rapid evolution, gene expansion, and niche adaptation (Farman, 2007). The *Fov* 89-1A genome has a high proportion of repeat sequences (~20%) in the genome compared to the other *Fov* genomes (Chapter 2), and therefore many hotspots could exist as a source of genetic variation in *Fov* 89-1A (Fierro & Martín, 1999). This potential for increased genetic variability could potentially explain why *Fov* race 4 is the most virulent on cotton, and several evolutionary mechanisms have occurred in repetitive elements, as observed for chromosome 7.

The core genome is generally considered stable and conserved within the same fungal species. However, the accessory regions are dynamic consisting of many evolutionary hotspots

e.g., repetitive elements, and spontaneous accessory chromosome loss can occur (Fokkens et al., 2018; Li et al., 2020; Vlaardingerbroek, Beerens, Schmidt, et al., 2016; Zhang et al., 2020). It is postulated that the poorly conserved, largely uncharacterized SM clusters in the accessory regions can be involved in host-specific virulence. Four *Fov* isolates were the focus of this study, as both *Fov* 14-004 and MDS-12 had no predicted clusters in the accessory genome. *Fov* TF1 has two type 1 PKS clusters in ACC01 that are not found in other *Fov* isolates. The first PKS has a core gene which, according to aniSMASH, possibly encodes a beta-ketoacyl synthase domain-containing protein although BLASTP indicates similarity with fatty acid synthases. The PKS cluster consequently is probably involved in fatty acid synthesis rather than SM production (Herbst et al., 2016). The other PKS is predicted to have a role as a CoA reductase/alcohol dehydrogenase (Adh). The role of Adh proteins is important for carbon and energy metabolism under various environmental conditions,as *adh1* has a role as a fermentative or oxidative enzyme, producing ethanol (de Smidt et al., 2008). In *Fol* 4287 *adh1* was expressed during the initial stage of colonization of tomato, presumably resulting from induction under the oxygen-limited environment. A delay of disease symptoms was observed in *adh1* mutant strains when compared to the wild-type isolate and is necessary for full fungal virulence during infection on the host root (Corrales Escobosa et al., 2011). The *adh1* gene resides on chromosomes 9, but at least four *adh* homologs are additionally present on accessory chromosomes 3 (2 genes) and 5 (2 genes) in the *Fol* 4287 genomes, suggesting alternative Adh proteins may have functional redundancy with *adh1* (Corrales Escobosa et al., 2011). Interestingly, a PKS cluster is present in the supernumerary chromosome 3 of *Fol* 4287, and the core biosynthetic gene has high conservation (94.52% identity) with the core gene in the PKS of *Fov* TF1 related to the *adh1* function. Validation is still necessary to confirm the function of the PKS cluster of *Fov* TF1. The similar proposed function of core

biosynthetic genes and location within the accessory regions suggests that the putative SM cluster related to alcohol dehydrogenase on ACC01 in *Fov* TF1 could have importance in host colonization as observed with the cluster in chromosome 3 in *Fol* 4287 (Corrales Escobosa et al., 2011).

Another three *Fov* isolates, 89-1A, 25433, and LA3B have putative NRPS clusters in accessory regions in each genome, and most of the domains in the NRPS are identified as AMP-dependent synthetases and ligases. Additional BLAST analysis suggests more detailed potential functions of each NRPS cluster. First, the core genes in both NRPSs of the *Fov* 89-1A isolate are commonly predicted as a potential HC-toxin, known as a cyclic tetrapeptide host-specific toxin (HST). Phytotoxins produced by fungi are usually specific to certain plant hosts and are important as virulence factors (Coleman & Mylonakis, 2009). The HC-toxin produced by *Cochliobolus carbonum*, functions as an inhibitor of histone deacetylases and affects to maize infection (Pitkin et al., 1996). The synthetic locus of this compound contains two MFS (major facilitator superfamily) transporters, termed *TOXA* and *TOXB*. The transporters are responsible for efflux of the toxin into the host cells, in this case, maize (Pitkin et al., 1996). The NRPS cluster in ACC45 includes a predicted ATP-binding cassette (ABC) transporter, one of the main classes of transporters responsible for secondary metabolite efflux from fungal cells (Coleman & Mylonakis, 2009). Therefore, the presence of these efflux pumps suggests that the product of the NRPS cluster in the accessory region is toxic and is extracellular and may mediate host-pathogen interactions during infection.

A single NRPS cluster in ACC01 of the *Fov* LA3B genome includes a core gene related to beauvericin/enniatin mycotoxin biosynthesis and a predicted ABC transporter. These mycotoxins are generally synthesized by various *Fusarium* species and are food contaminants, particularly in

cereal grains and are concern due to the effects on human and animal health (Prosperini et al., 2017; Wu et al., 2018). The close phylogenetic relationship between *Fov* LA3B and *F. oxysporum* f. sp. *pisi* (*Fop*) HDV247 explains the shared placement of the beauvericin/enniatin cluster in putative accessory regions. The core genes between *Fov* LA3B and FopHDV247 are over 99% in sequence similarity, and the sequences of the entire NRPS cluster between both isolates has high conservation (not shown). It implies that some cotton *Fov* isolates, as well as other FOSC isolates, harbor potential toxic compounds although the presence of phytotoxins are rarely described in *F. oxysporum* isolates. In addition, putative SM clusters can potentially be laterally acquired from donor fungal isolates outside the FOSC.

Lastly, *Fov* 25433 has two siderophore related clusters in the accessory regions (JH657951 and JH657976). As mentioned previously, siderophores are low molecular weight, iron-binding molecules that enable iron uptake or storage in microorganisms (Lee et al., 2005; Oide et al., 2006, 2007, 2015). The NRPS encoding core gene from the cluster in the JH657951 contig was searched against the *Fov* 25433 proteome to identify the corresponding gene, where both FOTG_11400 and FOTG_11399 (99%) covered the entire coding sequence. The BLAST results against non-redundant protein sequences (nr) showed the core gene in the JH657951 contig was responsible for the production of the ferrichrome-type siderophore, malonichrome (*NRPS1*) in *Fusarium* species (Oide et al., 2015). Two core genes exist in the JH657976 contig and one gene (FOTG_13635) are related to ferrirhodin synthetase while another gene partially matched to proteomes in the *Fov* race 7. In *F. sacchari*, ferrirhodin is a secreted siderophore, but the core gene has over 87% sequence similarity with *NRPS2* in *F. gramenearium*, involved in intracellular iron storage and production of ascospores (Munawar et al., 2013; Oide et al., 2007). The lack of any

type of transporter within the NRPS clusters and their close evolutionary relationships indicate that *Fov* race 7 may have additional siderophore functions in the accessory regions.

In *Fov* 14-004 no clusters in the accessory genome were identified, although the SM cluster repertoire is unique in *Fov* 14-004 when compared to the other five *Fov* genomes. Our genome-wide study revealed that this isolate has many common SM clusters with the closely related clinical isolate *F. oxysporum* NRRL 32931. *Fov* 14-004 and NRRL 32931 are predicted to have 46 and 44 SM clusters and they all reside in the core genome. *Fov* 14-004 has three additional clusters on chromosomes 11 (a CDPS), 12 (a NRPS), and 13 (a bacteriocin) while NRRL 32931 has a NRPS cluster on chromosome 8. In addition to the common 28 SM clusters in the core genomes of all of the FOSC, the *Fov* 14-004 and NRRL 32931 genomes had six SM clusters that were rarely found in the other FOSC genomes which were located on chromosomes 1, 4, 9, 11 (2), and 13. Comparison of the amino acid sequences of the shared core biosynthetic enzyme between these two genomes had 89-99% identity. The phylogenetic placement (Chapter 2) showed that *Fov* 14-004 is phylogenetically close with NRRL 32931 and other FOSC isolates with host specificity other than cotton. In addition, most of the accessory regions (ACC01-04) of *Fov* 14-004 are syntenic with loci in other *Fov* isolates, (TF1 and MDS-12). Altogether, it indicates that this isolate (*Fov* 14-004) may have acquired the accessory chromosomes from other *Fov* isolates during co-existence in the same cotton plant, and obtained the genes involved in pathogenicity on cotton by horizontal chromosome transfer (Chen et al., 2018; Vlaardingerbroek, Beerens, Rose, et al., 2016).

Members of the FOSC are prolific producers of secondary metabolites, which may be a contributing factor in their ubiquitous nature and broad host range. Within those infecting cotton, PKS and NRPS clusters were distributed and expanded across the *Fov* genomes and the SM biosynthetic clusters contributed to the uniqueness and diversity of each isolate. Although the

analysis of the genome sequences uncovered a multitude of SM candidates, the biosynthetic product of many of the SM clusters is largely unknown due to the involvement of many genes and tight control for their synthesis. Further functional analysis to validate individual gene clusters is necessary to further explore the arsenal of SM produced by members of the FOSC.

Reference

Anand, L., & Rodriguez Lopez, C. M. (2019). ChromoMap: An R package for Interactive Visualization and Annotation of Chromosomes. *BioRxiv*, 605600. https://doi.org/10.1101/605600

Blin, K., Wolf, T., Chevrette, M. G., Lu, X., Schwalen, C. J., Kautsar, S. A., Suarez Duran, H. G., de Los Santos, E. L. C., Kim, H. U., Nave, M., Dickschat, J. S., Mitchell, D. A., Shelest, E., Breitling, R., Takano, E., Lee, S. Y., Weber, T., & Medema, M. H. (2017). AntiSMASH 4.0 - improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Research*, *45*(W1), W36–W41. https://doi.org/10.1093/nar/gkx319

Cairns, T., & Meyer, V. (2017). *In silico* prediction and characterization of secondary metabolite biosynthetic gene clusters in the wheat pathogen *Zymoseptoria tritici*. *BMC Genomics*, *18*(1), 631. https://doi.org/10.1186/s12864-017-3969-y

Chen, J. Y., Liu, C., Gui, Y. J., Si, K. W., Zhang, D. D., Wang, J., Short, D. P. G., Huang, J. Q., Li, N. Y., Liang, Y., Zhang, W. Q., Yang, L., Ma, X. F., Li, T. G., Zhou, L., Wang, B. L., Bao, Y. M., Subbarao, K. v., Zhang, G. Y., & Dai, X. F. (2018). Comparative genomics reveals cotton-specific virulence factors in flexible genomic regions in *Verticillium dahliae* and evidence of horizontal gene transfer from *Fusarium*. *New Phytologist*, *217*(2), 756–770. https://doi.org/10.1111/nph.14861

Coleman, J. J., & Mylonakis, E. (2009). Efflux in Fungi: La Pièce de Résistance. *PLoS Pathogens*, *5*(6), e1000486. https://doi.org/10.1371/journal.ppat.1000486

Corrales Escobosa, A. R., Rangel Porras, R. A., Meza Carmen, V., Gonzalez Hernandez, G. A., Torres Guzman, J. C., Wrobel, K., Wrobel, K., Roncero, M. I. G., & Gutierrez Corona, J. F. (2011). *Fusarium oxysporum Adh1* has dual fermentative and oxidative functions and is involved in fungal virulence in tomato plants. *Fungal Genetics and Biology*, *48*(9), 886–895. https://doi.org/10.1016/j.fgb.2011.06.004

de Smidt, O., du Preez, J. C., & Albertyn, J. (2008). The alcohol dehydrogenases of *Saccharomyces cerevisiae*: a comprehensive review. *FEMS Yeast Research*, *8*(7), 967–978. https://doi.org/10.1111/j.1567-1364.2008.00387.x

Desjardins, A. E., & Proctor, R. H. (2007). Molecular biology of *Fusarium* mycotoxins. *International Journal of Food Microbiology*, *119*(1–2), 47–50. https://doi.org/10.1016/j.ijfoodmicro.2007.07.024

Farman, M. L. (2007). Telomeres in the rice blast fungus *Magnaporthe oryzae*: the world of the end as we know it. *FEMS Microbiology Letters*, *273*(2), 125–132. https://doi.org/10.1111/j.1574-6968.2007.00812.x

Fierro, F., & Martín, J. F. (1999). Molecular mechanisms of chromosomal rearrangement in fungi. *Critical Reviews in Microbiology*, *25*(1), 1–17. https://doi.org/10.1080/10408419991299185

Fokkens, L., Shahi, S., Connolly, L. R., Stam, R., Schmidt, S. M., Smith, K. M., Freitag, M., & Rep, M. (2018). The multi-speed genome of *Fusarium oxysporum* reveals association of histone modifications with sequence divergence and footprints of past horizontal chromosome transfer events. *BioRxiv*, 465070. https://doi.org/10.1101/465070

Graham-Taylor, C., Kamphuis, L. G., & Derbyshire, M. C. (2020). A detailed in silico analysis of secondary metabolite biosynthesis clusters in the genome of the broad host range plant pathogenic fungus *Sclerotinia sclerotiorum*. *BMC Genomics*, *21*(1), 7. https://doi.org/10.1186/s12864-019-6424-4

Herbst, D. A., Jakob, R. P., Zähringer, F., & Maier, T. (2016). Mycocerosic acid synthase exemplifies the architecture of reducing polyketide synthases. *Nature*, *531*(7595), 533–537. https://doi.org/10.1038/nature16993

Hoogendoorn, K., Barra, L., Waalwijk, C., Dickschat, J. S., van der Lee, T. A. J., & Medema, M. H. (2018). Evolution and Diversity of Biosynthetic Gene Clusters in Fusarium. *Frontiers in Microbiology*, *9*(JUN), 1158. https://doi.org/10.3389/fmicb.2018.01158

Janevska, S., Arndt, B., Niehaus, E. M., Burkhardt, I., Rösler, S. M., Brock, N. L., Humpf, H. U., Dickschat, J. S., & Tudzynski, B. (2016). Gibepyrone biosynthesis in the rice pathogen *Fusarium fujikuroi* is facilitated by a small polyketide synthase gene cluster. *Journal of Biological Chemistry*, *291*(53), 27403–27420. https://doi.org/10.1074/jbc.M116.753053

Keller, N. P. (2019). Fungal secondary metabolism: regulation, function and drug discovery. In *Nature Reviews Microbiology* (Vol. 17, Issue 3, pp. 167–180). Nature Publishing Group. https://doi.org/10.1038/s41579-018-0121-1

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., & Salzberg, S. L. (2004). Open Access Versatile and open software for comparing large genomes. *Genome Biology*, *5*(2), 12. https://doi.org/https://doi.org/10.1186/gb-2004-5-2-r12

Lee, B. N., Kroken, S., Chou, D. Y. T., Robbertse, B., Yoder, O. C., & Turgeon, B. G. (2005). Functional analysis of all nonribosomal peptide synthetases in *Cochliobolus heterostrophus* reveals a factor, *NPS6*, involved in virulence and resistance to oxidative stress. *Eukaryotic Cell*, *4*(3), 545–555. https://doi.org/10.1128/EC.4.3.545-555.2005

Li, J., Fokkens, L., Conneely, L. J., & Rep, M. (2020). Partial pathogenicity chromosomes in *Fusarium oxysporum* are sufficient to cause disease and can be horizontally transferred. *Environmental Microbiology*, *22*(12), 4985–5004. https://doi.org/10.1111/1462-2920.15095

Møller, P., Wils, R. S., Jensen, D. M., Andersen, M. H. G., & Roursgaard, M. (2018). Telomere dynamics and cellular senescence: an emerging field in environmental and occupational toxicology. *Critical Reviews in Toxicology*, *48*(9), 761–788. https://doi.org/10.1080/10408444.2018.1538201

Munawar, A., Marshall, J. W., Cox, R. J., Bailey, A. M., & Lazarus, C. M. (2013). Isolation and Characterisation of a Ferrirhodin Synthetase Gene from the Sugarcane Pathogen *Fusarium sacchari*. *ChemBioChem*, *14*(3), 388–394. https://doi.org/10.1002/cbic.201200587

Munkvold, G. P. (2017). *Fusarium* species and their associated mycotoxins. In *Methods in Molecular Biology* (Vol. 1542, pp. 51–106). Humana Press Inc. https://doi.org/10.1007/978-1-4939-6707-0_4

Oide, S., Berthiller, F., Wiesenberger, G., Adam, G., & Turgeon, B. G. (2015). Individual and combined roles of malonichrome, ferricrocin, and TAFC siderophores in *Fusarium graminearum* pathogenic and sexual development. *Frontiers in Microbiology*, *5*(DEC), 759. https://doi.org/10.3389/fmicb.2014.00759

Oide, S., Krasnoff, S. B., Gibson, D. M., & Turgeon, B. G. (2007). Intracellular siderophores are essential for ascomycete sexual development in heterothallic *Cochliobolus heterostrophus* and homothallic Gibb*erella zeae*. *Eukaryotic Cell*, *6*(8), 1339–1353. https://doi.org/10.1128/EC.00111-07

Oide, S., Moeder, W., Krasnoff, S., Gibson, D., Haas, H., Yoshioka, K., & Turgeon, B. G. (2006). *NPS6*, Encoding a Nonribosomal Peptide Synthetase Involved in Siderophore-Mediated Iron Metabolism, Is a

Conserved Virulence Determinant of Plant Pathogenic Ascomycetes. *The Plant Cell*, *18*(10), 2836–2853. https://doi.org/10.1105/tpc.106.045633

Pitkin, J. W., Panaccione, D. G., & Walton, J. D. (1996). A putative cyclic peptide efflux pump encoded by the *TOXA* gene of the plant-pathogenic fungus *Cochliobolus carbonum*. *Microbiology*, *142*(6), 1557–1565. https://doi.org/10.1099/13500872-142-6-1557

Prosperini, A., Berrada, H., Ruiz, M. J., Caloni, F., Coccini, T., Spicer, L. J., Perego, M. C., & Lafranconi, A. (2017). A review of the mycotoxin enniatin B. *Frontiers in Public Health*, *5*, 1. https://doi.org/10.3389/fpubh.2017.00304

Pusztahelyi, T., Holb, I. J., & Pócsi, I. (2015). Secondary metabolites in fungus-plant interactions. *Frontiers in Plant Science*, *6*(AUG), 573. https://doi.org/10.3389/fpls.2015.00573

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Reynolds, H. T., Slot, J. C., Divon, H. H., Lysøe, E., Proctor, R. H., & Brown, D. W. (2017). Differential Retention of Gene Functions in a Secondary Metabolite Cluster. *Molecular Biology and Evolution*, *34*(8), 2002–2015. https://doi.org/10.1093/molbev/msx145

Seo, S., Pokhrel, A., & Coleman, J. J. (2020). The genome sequence of five genotypes of *Fusarium oxysporum* f. sp. *vasinfectum*: A resource for studies on Fusarium wilt of cotton. *Molecular Plant-Microbe Interactions*, *33*(2), 138–140. https://doi.org/10.1094/MPMI-07-19-0197-A

Slightom, J. L., Metzger, B. P., Luu, H. T., & Elhammer, A. P. (2009). Cloning and molecular characterization of the gene encoding the Aureobasidin A biosynthesis complex in *Aureobasidium pullulans* BP-1938. *Gene*, *431*(1–2), 67–79. https://doi.org/10.1016/j.gene.2008.11.011

Studt, L., Janevska, S., Niehaus, E. M., Burkhardt, I., Arndt, B., Sieber, C. M. K., Humpf, H. U., Dickschat, J. S., & Tudzynski, B. (2016). Two separate key enzymes and two pathway-specific transcription factors are involved in fusaric acid biosynthesis in *Fusarium fujikuroi*. *Environmental Microbiology*, *18*(3), 936–956. https://doi.org/10.1111/1462-2920.13150

Sullivan, M. J., Petty, N. K., & Beatson, S. A. (2011). Easyfig: a genome comparison visualizer. *Bioinformatics*, *27*(7), 1009–1010. https://doi.org/10.1093/bioinformatics/btr039

Vlaardingerbroek, I., Beerens, B., Rose, L., Fokkens, L., Cornelissen, B. J. C., & Rep, M. (2016). Exchange of core chromosomes and horizontal transfer of lineage-specific chromosomes in *Fusarium oxysporum*. *Environmental Microbiology*, *18*(11), 3702–3713. https://doi.org/10.1111/1462-2920.13281

Vlaardingerbroek, I., Beerens, B., Schmidt, S. M., Cornelissen, B. J. C., & Rep, M. (2016). Dispensable chromosomes in *Fusarium oxysporum* f. sp. *lycopersici*. *Molecular Plant Pathology*, *17*(9), 1455–1466. https://doi.org/10.1111/mpp.12440

Wang, Q., Cobine, P. A., & Coleman, J. J. (2018). Efficient genome editing in *Fusarium oxysporum* based on CRISPR/Cas9 ribonucleoprotein complexes. *Fungal Genetics and Biology*, *117*, 21–29. https://doi.org/10.1016/j.fgb.2018.05.003

Wight, W. D., Kim, K. H., Lawrence, C. B., & Walton, J. D. (2009). Biosynthesis and role in virulence of the histone deacetylase inhibitor depudecin from *Alternaria brassicicola*. *Molecular Plant-Microbe Interactions*, *22*(10), 1258–1267. https://doi.org/10.1094/MPMI-22-10-1258

Williams, A. H., Sharma, M., Thatcher, L. F., Azam, S., Hane, J. K., Sperschneider, J., Kidd, B. N., Anderson, J. P., Ghosh, R., Garg, G., Lichtenzveig, J., Corby Kistler, H., Shea, T., Young, S., Buck, S.-A. G., Kamphuis, L. G., Saxena, R., Pande, S., Ma, L.-J., … Singh, K. B. (2016). Comparative genomics and prediction of conditionally dispensable sequences in legume-infecting *Fusarium oxysporum formae speciales* facilitates identification of candidate effectors. *BMC Genomics*, *17*(191). https://doi.org/10.1186/s12864-016-2486-8

Wu, Q., Patocka, J., Nepovimova, E., & Kuca, K. (2018). A review on the synthesis and a review on the synthesis and bioactivity aspects of beauvericin, a *Fusarium* mycotoxin. *Frontiers in Pharmacology*, *9*(NOV), 1338. https://doi.org/10.3389/fphar.2018.01338

Zhang, Y., Yang, H., Turra, D., Zhou, S., Ayhan, D. H., DeIulio, G. A., Guo, L., Broz, K., Wiederhold, N., Coleman, J. J., Donnell, K. O., Youngster, I., McAdam, A. J., Savinov, S., Shea, T., Young, S., Zeng, Q., Rep, M., Pearlman, E., … Ma, L. J. (2020). The genome of opportunistic fungal pathogen *Fusarium oxysporum* carries a unique set of lineage-specific chromosomes. *Communications Biology*, *3*(1), 1–12. https://doi.org/10.1038/s42003-020-0770-2

Table 1. Details of the FOSC genomes used in this study.

f. sp.	Strain or NRRL #	Race	Size (Mb)	GC%	Scaffolds	Accession	Name	AntiSMASH
Fo47	Fo47	-	49.6646	47.7	124	GCA_000271705.2	Fo47	43
Fo5176	Fo5176	-	54.7676	47.8	7858	GCA_000222805.1	Fo5176	45
cepae	Fus2	-	53.4258	47.72	34	GCA_003615085.1	FoceFus2	46
conglutinans	54008	2	53.5754	47.7	2552	GCA_000260215.2	Foco54008	50
cubense	N2	1	47.6574	48	1341	GCA_000350345.1	Focu1	42
	B2	4	52.9562	48.2	840	GCA_000350365.1	Focu4	38
	54006	TR4	46.5538	47.5	418	GCA_000260195.2	FocuTR4	42
lycopersici	4287	-	61.3869	48.43	114	GCA_000149955.2	Fol4287	43
	MN25	2	48.6374	47.7	388	GCA_000259975.2	FolMN25	46
melonis	26406	-	54.0343	47.5	1146	GCA_000260495.2	Fom26406	48
FoNRRL32931	FoNRRL32931	-	47.9063	47.6	168	GCA_000271745.2	FoNRRL32931	44
pisi	HDV247	-	55.1882	47.6	472	GCA_000260075.2	FopHDV247	42
raphani	54005	3	53.4994	47.8	1218	GCA_000260235.2	For54005	40
radicis-cucumerinum	Forc016	-	52.9083	47.68	33	GCA_001702695.2	Forc016	45
radicis-lycopersici	26381	-	49.3593	47.6	418	GCA_000260155.3	Forl26381	43
vasinfectum	TF1	1	50.0435	47.4	17	GCA_009602505.1	FovTF1	41
	89-1A	4	63.3239	47.65	94	GCA_009602635.1	Fov891A	40
	25433	7	52.9144	47.7	985	GCA_000260175.2	Fov25433	41
	LA3B	LA108	50.4378	47.48	17	GCA_009602545.1	FovLA3B	39
	14-004	LA127/140	51.6366	47.35	18	GCA_009602625.1	Fov14004	46
	LA 1E	MDS12	53.5323	47.44	26	GCA_009602615.1	FovMDS12	40

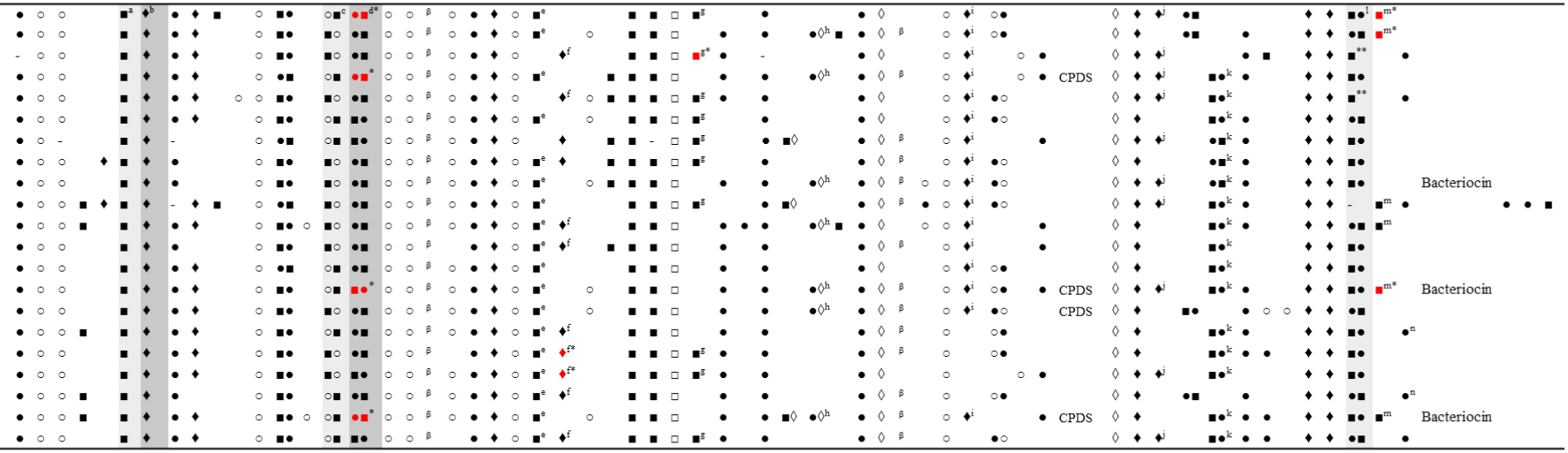
Table 2. The number of secondary metabolite cluster types from antiSMASH in 21 FOSC genomes.

Name	# of clusters	NRPS	NRPSlike	NRPS-NRPSlike[a]	T1PKS	T3PKS	NRPS-T1PKS[b]	NRPSlike-T1PKS[c]	NRPS-indole	T1PKS-indole		Bacteriocin	Betalactone	CDPS	Indole	Terpene
Fo47	43	8	10	1	6	1	4	1	1			2		2	7	
Fo5176	45	9	10		7	1	3	1				2	1	2	9	
FoceFus2	46	9	10		5	1	5	1	1			2	1	2	9	
Foco54008	50	10	11	1	9	1	3	1				2	1	2	9	
Focu1	42	8	10	1	6	1	4	1				2		2	7	
Focu4	38	7	7		5	1	4	1		1		2		2	8	
FocuTR4	42	6	8	1	7	1	4	1		1		2		2	9	
Fol4287	43	7	8	1	8	1	4	1				2		2	9	
FolMN25	46	8	12	1	5	1	4	1	1	1	1	2		2	7	
Fom26406	48	10	9	1	9	1	3	1		1		2		2	9	
FoNRRL32931	44	9	10		7	1	4	1	1			1		2	8	
FopHDV247	42	9	8		6	1	6	1				2		2	7	
For54005	40	7	8	1	5	1	5	1				2		2	8	
Forc016	45	8	10	1	5	1	4	1	1		1	2	1	2	8	
Forl26381	43	8	10	1	5	1	4	1	1			2	1	2	7	
FovTF1	41	8	8	1	7	1	4	1				2		2	7	
Fov891A	40	10	7	1	5	1	4	1				2		2	7	
Fov25433	41	9	9		5	1	4	1				2		2	8	
FovLA3B	39	9	8	1	5	1	4	1				2		2	6	
Fov14004	46	9	10		6	1	4	1	1	1	1	2	1	2	7	
FovMDS12	40	9	7	1	5	1	4	1				2		2	8	
Variation	38-50	6-9	7-12	-	5-9	1	3-6	2	-	-	-	1-2	-	2	6-9	

NRPS: non-ribosomal peptide synthetases, T1PKS: type I polyketide synthases, T3PKS: type III polyketide synthases, and CDPS: tRNA-dependent cyclodipeptide synthases

Based on results from classification of individual chromosomes and gene location on them throughout all Fo strains, such clusters are considered as the same cluster: [a]: NRPS-NRPS-like and NRPS-like-NRPS, [b]: NRPS-T1PKS and T1PKS-NRPS and [c]: NRPS-like-T1PKS and T1PKS-NRPS-like.

Table 3. Profile of secondary metabolite clusters in each core chromosomes



CCs: Core chromosomes

●: NRPS, ○: NRPSlike, ■: T1PKS, □: T3PKS, ◊: indole, ♦: terpene, ᵝ: betalactone

Known and common clusters in all 21 FOSC (or missing up to two strains allowed), are shaded with light or dark gray. The clusters having the similarity less than 40% to known functions are excluded but if other strains have the same known cluster with over 40% similarity, are marked with the superscripted symbol (*) instead of the elimination. The double superscripted symbol (**) indicates two isolates have a T1PKS instead of the hybrid of T1PKS and NRPS but are considered as the same cluster based on the same cluster function with other 18 Fo strains. The details are below with the matching alphabet:

[a]: oxyjavanicin

[b]: squalestatin S1

[c]: fusaric acid

[d]: equisetin (the cluster having the similarity less than 40% marked)

[e]: bikaverin (all isolates have except three)

[f]: gibberellin (9 isolates have, and the cluster having the similarity less than 40% marked)

[g]: depudecin (10 isolates have, and the cluster having the similarity less than 40% marked)

[h]: hexadehydroastechrome / terezine-D / astechrome (7 isolates have)

[i]: α-acorenol (16 isolates)

[j]: koraiol (10 isolates)

[k]: ACT-Toxin II (16 isolates have)

[l]: gibepyrone-A (one isolate misses)

[m]: fujikurin A/ B/ C/ D (6 isolates have, and the cluster having the similarity less than 40% marked)

[n]: AbT1

Table 4. Information about the known clusters from antiSMASH.

Location	Most similar known cluster	Cluster type	Strains #	Similarity (%)	Existence in *Fov* strains	Notes
CHR02	Oxyjavanicin	T1PKS	All	40	All	
	Squalestatin S1	Terpene	All	40	All	
CHR05	Fusaric acid	T1PKS,NRPSlike	All	59-63	All	
	Equisetin	NRPS,T1PKS	All	36-45	All	Clusters which have similarity less than 40% exist in 4 strains with 36%.
CHR07	Bikaverin	T1PKS	18	57-100	All	
CHR11	α-acorenol	Terpene	16	100	Fov14004 only	
CHR12	ACT-Toxin II	T1PKS,NRPS	16	100	Five *Fov* strains, except FovLA3B	
CHR13	Gibepyrone-A	NRPS,T1PKS	20	40-80	All	
CHR07	Gibberellin	Terpene	9	57-71	Five *Fov* strains except Fov14004	Two strains including *Fov* race 4 with 28%.
CHR08	Depudecin	T1PKS	10	50-100	Three *Fov* race 4, 7 and MDS-12 isolates	One strain with 33%.
CHR09	Hexadehydroastechrome / terezine-D / astechrome	NRPS,T1PKS	7	62-75	Only Fov14004	
CHR12	Koraiol	Terpene	10	100	Two *Fov* race 7 and MDS-12 isolates	
CHR13	Fujikurin A/ B/ C/ D	T1PKS	6	50	Only Fov14004	Three strains with 33%.
CHR13	AbT1	NRPS	2	100	Two FovTF1 and LA3B	
Accessory regions	Enniatin	NRPS	FopHDV247	100	None	
	Gibberellin	Terpene	For54005	71	None	

Table 5. Secondary metabolite cluster in potential accessory scaffolds in FOSC.

FOSC	Total	# in CCs	# in ACCs	CHR03	CHR14		78*	79*										
Fol4287	43	38	5	■	●	♦	●	β										
Fo47	43	41	2		●					○								
Fo5176	45	38	7	■	●		●	β		○				●■				CDPS
FoceFus2	46	43	3			♦				○					■●			
Foco54008	50	41	9	■	●		●	β	●	○		■	■					CDPS
Focu1	42	37	5	■	●		●	β		○								
Focu4	38	36	2	■	●													
FocuTR4	42	39	3	■		♦										■◊		
FolMN25	46	42	4		●					○	○					■◊		
Fom26406	48	46	2		●					○								
FoNRRL32931	44	44	-															
FopHDV247	42	36	6	■	●		●[a]			○					■●	■●		
For54005	40	35	5	■	●	♦[b]		β							■●			
Forc016	45	44	1							○								
Forl26381	43	41	2	■	●													
FovTF1	41	39	2	■								■						
Fov891A	40	38	2		●		●											
Fov25433	41	38	3		●		●	β										
FovLA3B	39	38	1		●													
Fov14004	46	46	-															
FovMDS12	40	40	-															

CCs means core chromosomes and ACCs is accessory chromosomes.

●: NRPS, ○: NRPSlike, ■: T1PKS, □: T3PKS, ◊: indole, ♦: terpene, β: betalactone

*: the number 78 and 79 means supercontigs which were fragments during Fol genome assembly. However, the information of *Fol* accessory chromosomes is rarely associated with clusters found in other strains in FOSC.

Known clusters are stated below:

[a]: enniatin

[b]: gibberellin

Table 6. Analyses of predicted secondary metabolites in the accessory regions of the *Fov* genomes.

Fov isolate	Cluster type	Location	# core gene	Annotation in each genome	Annotation in antiSMASH (SMCOG) *	Hit cluster among the top 25 (BLASTP) **				
						Annotation	Species name	% similarity	% sequence coverage	E value
TF1	T1PKS	ACC01	1	*F02700*	Beta-ketoacyl synthase	Mycocerosic acid synthase	*F. odoratissimum*	94	90	0
						Phthiocerol synthesis polyketide synthase type I PpsD	*F. oxysporum* f. sp. *cubense*	96	44	0
						Fatty acid synthase S-acetyltransferase	*Aspergillus candidus*	45	99	0
						Acyl transferase/acyl hydrolase/lysophospholipase	*Penicillium griseofulvum*	58	97	0
	T1PKS	ACC01	1	*F06929*	Crotonyl-CoA reductase / alcohol dehydrogenase	Compactin diketide synthase mokB	*F. oxysporum*	94	100	0
89-1A	NRPS	ACC06	1	*F04659*	Condensation domain-containing protein	Acetyl-CoA synthetase-like protein	*Lindgomyces ingoldianus*	56	99	6.00E-119
						HC-toxin synthetase	*Pyrenophora teres* f. *teres*	55	91	3.00E-112
						AMP-binding enzyme	*P. teres* f. *maculata*	54	92	4.00E-111
						Destruxin synthetase	*P. teres* f. *teres*	54	92	5.00E-111
	NRPS	ACC45	1	*F00104*[a]	Condensation domain-containing protein	Acetyl-CoA synthetase-like protein	*Lindgomyces ingoldianus*	56	99	6.00E-119
						HC-toxin synthetase	*P. teres* f. *teres*	55	91	3.00E-112
						AMP-binding enzyme	*L. ingoldianus*	54	92	4.00E-111
						Destruxin synthetase	*P. teres* f. *teres*	54	92	5.00E-111
			2	*F00104*[a]	AMP-dependent synthetase and ligase	AMP-binding enzyme	*P. teres* f. *maculate*	58	98	0
						Acetyl-CoA synthetase-like protein	*L. ingoldianus*	58	97	0
						Destruxin synthetase	*P. teres* f. *teres*	61	98	0
						EntF Non-ribosomal peptide synthetase module protein	*P. teres* f. *maculate*	57	98	0
						HC-toxin synthetase	*P. tritici-repentis* Pt-1C-BFP	56	99	0
			3	- [b]	AMP-dependent synthetase and ligase	AMP-binding enzyme	*P. teres* f. *maculate*	57	79	0
						Destruxin synthetase	*P. teres* f. *teres*	56	79	0
						EntF Non-ribosomal peptide synthetase module protein	*P. teres* f. *maculate*	56	79	0
						Cyclosporine synthetase	*P. teres* f. *maculata*	56	96	0
LA3B	NRPS	ACC01	1	*F00155*	AMP-dependent synthetase and ligase	Cyclic peptide synthetase	*F. oxysporum*	98	99	0
						Enniatin synthase	*F. austroafricanum*	82	99	0
						Beauvericin nonribosomal cyclodepsipeptide synthetase bea1	*F.* sp. DS 682	70	99	0
25433	NRPS	JH657951.1[c]	1	*FOTG_11399*	AMP-dependent synthetase and ligase	Ferrichrome siderophore peptide synthetase	*F. oxysporum* f. sp. *cubense* race 1	96	91	0
	NRPS	JH657976.1[c]	1	*FOTG_13635*	AMP-dependent synthetase and ligase	Ferrirhodin synthetase	*F. sacchari*	73	100	0

	2	-[d]	AMP-dependent synthetase and ligase	Ferrirhodin synthetase	*F. sacchari*	80	96	1.00E-165

[*]: SMCOG is the secondary metabolism Clusters of Orthologous Groups predicted by antiSMASH program.

[**]: Except hypothetical proteins, all the known functions of each core synthetic gene were described within the top 25.

[a]: In case of the NRPS cluster in ACC45, two genes among three are originally predicted as a single gene, *F00104*.

[b]: Although antiSMASH results show the third core gene in the NRPS cluster on ACC45, the gene has no match with the annotation.

[c]: Due to the difference of the *Fov* race 7 genome with other Fov genomes, the scaffold/contig names are stated instead of ACC.

[d]: The corresponding core gene is hardly found in the annotation files because of many partial genes.

Table 7. The core genes responsible for siderophore biosynthesis in *Fol* 4287 and *Fov* genomes.

Isolate	*NRPS1*	Protein length	% identity	*NRPS2*	Protein length	% identity	*NRPS6*	Protein length	% identity
F. graminearum	*FGSG_11026*	4747		*FGSG_05372*	4841		*FGSG_03747*	2071	
Fol4287	*FOXG_17422*	4759	69.526	*FOXG_06448*	3914	72.021	*FOXG_09785*	1923	86.184
FovTF1	*F01079*	4733	68.621	*F10302*	4848	70.196	*F07873*	2067	87.421
Fov89-1A	*F01880*	4759	69.42	*F14468*	4848	70.052	*F11256*	2067	87.178
Fov25433	*FOTG_11399*	4759	69.42	*FOTG_03319*	4869	70.052	*FOTG_12753*	2067	87.178
FovLA3B	*F00969*	4726	68.753	*F09756*	3277	68.49	*F05939*	2067	87.518
Fov14-004	*F11843*	4726	68.733	*F12054*	3277	68.399	*F07475*[*]	1279	88.842
							F07486[*]	731	85.109
FovMDS-12	*F15236*	4727	68.878	*F13059*	4848	70.052	*F08266*	2067	87.518

[*]: Two genes in the *Fov* 14-004 cover the entire *NRPS6* gene.

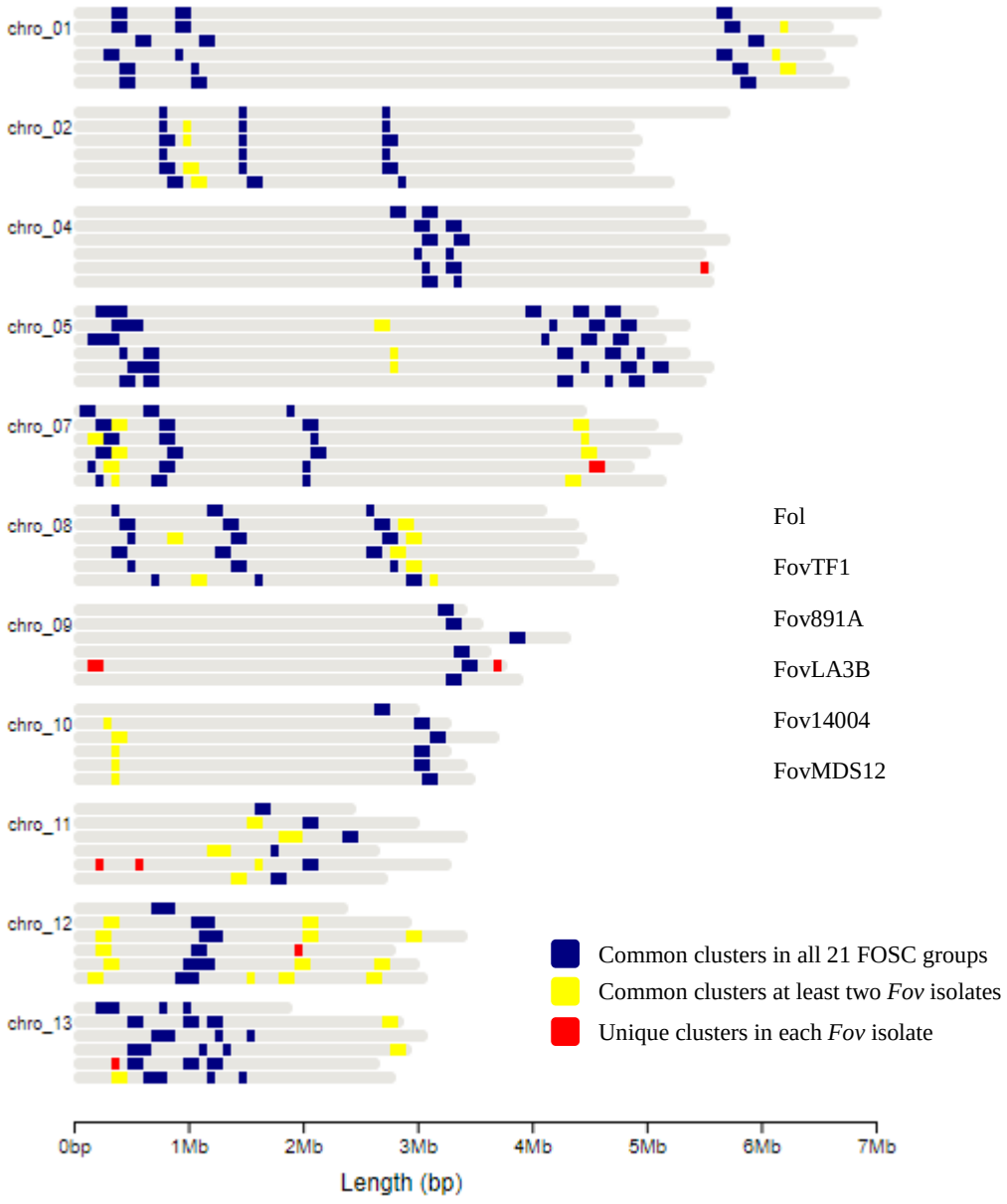


Figure 1. Physical distribution of common secondary metabolite clusters on core regions of *Fol* 4287 and five *Fov* genomes. Twenty-eight common SM clusters existing in the FOSC were placed on the eleven core chromosomes (blue). In addition to the common clusters, the remaining predicted SM clusters in the five *Fov* genomes are indicated in the bar graph. The figure was generated by the R package, chromoMAP.

Figure 2. Distribution of gene clusters common, unique, and located in sub-telomeric regions on the *Fov* core chromosomes. The SM clusters within 300 kb of each core scaffold end are shown.

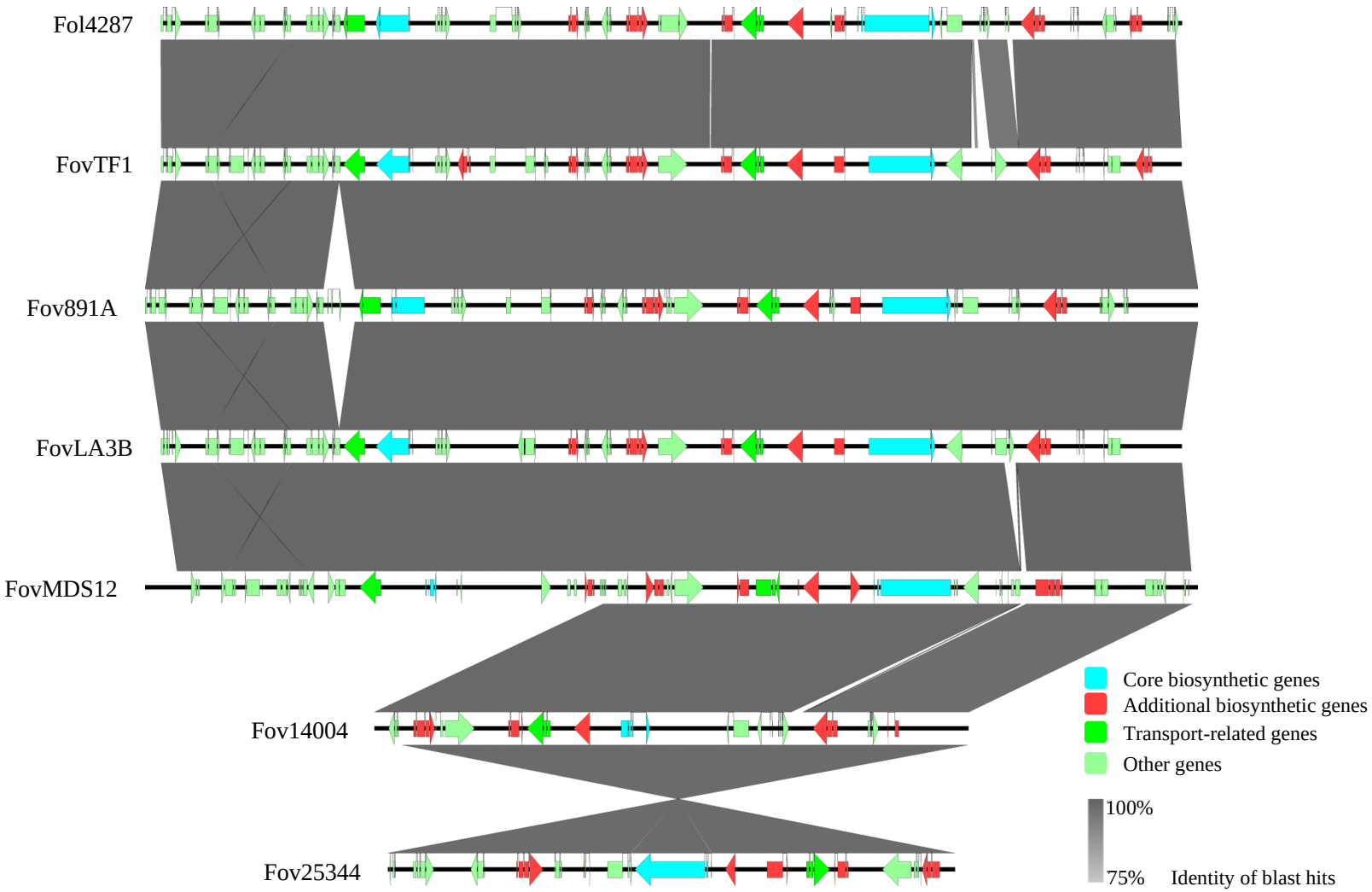


Figure 3. Sequence comparison of a putative siderophore cluster of *Fol* 4287 and six *Fov* genomes including Fov25433. One NRPS-like,NRPS clusters related to the siderophore biosynthesis are present in *Fol* 4287, and *Fov* race 1, 4, and LA108, LA 1E (MDS-12) isolates. Other two isolates,

Fov race LA127/140 and race 7 potentially had the only NRPS type cluster predicted from the antiSMASH results. Arrows indicate the gene annotation. Connecting lines represent sequence similarities between isolates. The figure was generated by EASYFIG.